Topics in Applied Physics
Volume 107

Topics in Applied Physics

Topics in Applied Physics is a well-established series of review books, each of which presents a comprehensive survey of a selected topic within the broad area of applied physics. Edited and written by leading research scientists in the field concerned, each volume contains review contributions covering the various aspects of the topic. Together these provide an overview of the state of the art in the respective field, extending from an introduction to the subject right up to the frontiers of contemporary research.
Topics in Applied Physics is addressed to all scientists at universities and in industry who wish to obtain an overview and to keep abreast of advances in applied physics. The series also provides easy but comprehensive access to the fields for newcomers starting research.
Contributions are specially commissioned. The Managing Editors are open to any suggestions for topics coming from the community of applied physicists no matter what the field and encourage prospective editors to approach them with ideas.

Sebastian Volz (Ed.)

Microscale and Nanoscale Heat Transfer

With 144 Figures and 7 Tables

In Collaboration with Rémi Carminati, Patrice Chantrenne, Stefan Dilhaire, Séverine Gomez, Nathalie Trannoy, and Gilles Tessier

Springer

Dr. Sebastian Volz
Labortoire d'Enerégtique Moléculaire et Macroscopique, Combustion
Ecole Central Paris
Grande Voie des Vignes
92295 Châtenay Malabry, France
volz@em2c.ecp.fr

Library of Congress Control Number: 2006934584

Physics and Astronomy Classification Scheme (PACS):
65.80.+n, 82.53.Mj, 81.16.-c, 44.10.+i, 44.40.+a, 82.80.Kq

ISSN print edition: 0303-4216
ISSN electronic edition: 1437-0859
ISBN-10 3-540-36056-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-36056-8 Springer Berlin Heidelberg New York

Typesetting: DA-TEX · Gerd Blumenstein · www.da-tex.de
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper 57/3100/YL 5 4 3 2 1 0

Preface

The development of micro- and nanoscale fabrication techniques has triggered a broad scientific and technical revolution. A prime example is provided by microelectronics, which has now become nanoelectronics. Other evolutionary breakthroughs are now clearly established in the fields of optoelectronics, materials, the production and conversion of energy, and techniques for data processing and communications.

A remarkable feature of this trend is the way it has brought together physicists and engineers. On the one hand, the classical laws used to model macroscopic systems are generally unsuitable when system sizes approach characteristic microscopic scales, such as the mean free path or the length of carriers. The physical description of the individual or collective behaviour of the basic elements must then be reassessed. On the other hand, the development and integration of physical ideas exploiting very small structures, such as ultrathin films, superlattices, nanowires, and nanoparticles, in order to improve an industrial system, requires the physicist to understand some of the more technical aspects of engineering.

The international community of thermal scientists, whether in research or engineering, base their approach on the mass, momentum and energy conservation equations associated with the laws of diffusion for conduction (Fourier) and for mass transfer (Fick), and Newton's law for conduction–convection. For radiation, the radiative transfer equation is widely used to treat semi-transparent media, grey or otherwise.

But this theoretical framework can no longer describe the conductive and conductive–convective transfer regimes on very small space and time scales, simply because the carriers undergo too few collisions. As the radiated thermal wavelengths are of the order of a few microns, the radiative transfer equation, and even the whole notion of luminance, become quite inappropriate on submicron scales.

One does not even need to approach the limits of macroscopic models to observe that the phenomenology of heat transfer is quite different on the micron and centimeter length scales. Whilst heat transfer is generally felt to be a slow process – the time scale for heat conduction in macroscopic systems ($\sim$ 50 cm) is a few minutes – the propagation of heat is an extremely efficient process on the microscale ($\sim$ 10 ns). Indeed, the diffusion time is proportional to the square of the length. Moreover the thermal resistances of

microscale structures are so small that they become of the same order as the interface resistances between such structures. Microscale heat transfer thus occurs practically without inertia, and is essentially equivalent to interface heat transfer. Naturally, this is even more true for nanoscale heat transfer.

From the experimental standpoint, very weak and highly localised contributions must be detected in order to measure the conductive flux in nanostructures. For example, the methods used must not introduce high contact thermal resistances. Ultrafast optical methods (nano- to picosecond) and near-field microscopy are best suited to satisfy these criteria.

It is therefore clear that the study of heat transfer on micro- and nanoscales requires a quite new approach on the part of the thermal science community. The task here is to integrate the new physical models and also the novel experimental devices now available to treat energy exchanges in micro- and nanostructures.

There are many consequences for industry:

- In housing, superinsulating nanoporous materials can limit heat losses whilst increasing the ground surface, and their conductivity in vacuum is smaller than that of air.
- Nanofluids, i.e., heat-carrying liquids transporting nanoparticles, have conductivities 10–40% higher than those of the base fluid and hence a greatly enhanced transfer efficiency.
- In the nanoelectronics of processors, heating problems have led manufacturers to slow down the miniaturisation trend by switching to multi-unit structures in which several computing units are integrated into the same chip.
- Data storage will for its part be heat-assisted. Heating can activate or inhibit magnetisation reversal. It can also change the phase or the geometry of a storage medium, and this over nanoscale areas.
- Thermoelectric energy conversion is currently undergoing a revolution through manipulation of the thermophysical properties of nanostructured materials. In 2002, certain superlattice alloys were able to produce an intrinsic performance coefficient twice as high as had ever been measured for a bulk solid material. This breakthrough was achieved by improving thermal properties.

In all these fields of application, our understanding of the relevant heat mechanisms and the associated modelling tools remains poor or at best imperfect.

The present book brings together for the first time the physical ideas and formalism as well as the experimental tools making up this new field of thermal science. Although these are usually considered to be the jurisdiction of the physicist, the aim of the book remains quite concrete, since it seeks to solve the problems of heat transfer in micro- and nanostructured materials. The book itself results from a collaborative network in France known as the *Groupement de Recherche Micro et Nanothermique* (GDR), bringing

together teams organised by a unit of the *Centre national de la recherche scientifique* (CNRS)[1] and a unit of the department[2] of *Sciences pour l'Ingénieur*. This group combines research centres involved in thermal science, solid state physics, optics and microsystems. Each chapter has been written by one or several authors – sometimes belonging to different research teams – and then edited by experts and non-experts in the GDR.

The first part of the book is theoretical, making the connection between the fundamental approaches to energy transfer and the quantities describing heat transfer. Chapter 1 considers the limits of classical models on small scales. Chapters 2, 3 and 4 then treat the physical models describing heat transfer in gases, conduction, and radiation, respectively, all on these small scales.

The second part of the book covers the numerical tools that can be implemented to solve the previously formulated equations in concrete situations. Chapters 5 and 6 examine solutions of the Boltzmann and Maxwell equations, respectively. Having discussed continuum models, microscopic simulations are tackled in Chap. 7 via the Monte Carlo method and in Chap. 8 via the technique of molecular dynamics simulations. In each chapter, it is shown how to calculate a heat flux or conductivity explicitly through various examples.

The last part of the book deals with experimental approaches. Chapter 9 introduces different forms of near-field microscopy and discusses their applications in thermal science. A thermal microscope is presented in some detail with example applications. Chapter 10 discusses optical techniques as provided by the photothermal microscope and reflectometry, whilst Chap. 11 brings together optical and near-field microscopy in a single hybrid system. This series of chapters on microscopy is followed by two chapters presenting the thermal applications of femtosecond lasers in pump–probe configurations. Chapter 12 deals with the electron–photon interaction on ultrashort time scales and Chap. 13 treats of thermal–acoustic coupling in various types of structure.

The book thus constitutes a particularly complete and original collection of ideas, models, numerical methods and experimental tools that will prove invaluable in the study of micro- and nano-heat transfer. It should be of interest to research scientists and thermal engineers who wish to carry out theoretical research or metrology in this field, but also to physicists concerned with the problems of heat transfer, or teachers requiring a solid foundation for an undergraduate university course in this area.

[1] The French National Research Institute.
[2] Science for Engineering.

Acknowledgements

The strength of this book mostly relies on the collaborative effort of my dear colleagues. I am glad to express my deep thanks to Rémi Carminati, Patrice Chantrenne, Bernard Cretin, Stefan Dilhaire, Danièle Fournier, Séverine Gomez, Jean-Jacques Greffet, Karl Joulain, Denis Lemonnier, Bernard Perrin, Nathalie Trannoy, Gilles Tessier, Fabrice Vallée and Pascal Vairac for providing a work of highest quality in their field of expertise.

Paris, *Sebastian Volz*
November, 2005

Contents

List of Contributors

Rémi Carminati
Professeur à l'Ecole Centrale Paris

Patrice Chantrenne
Maître de Conférence à l'INSA de Lyon

Bernard Cretin
Professeur à l'Ecole Nationale Supérieure de Mécanique
et Microtechniques de Besançon

Stefan Dilhaire
Maître de Conférence à l'Université de Bordeaux I

Danièle Fournier
Professeur à l'Université Pierre et Marie Curie (Paris VI)

Séverine Gomes
Chargé de Recherche CNRS au Centre de Thermique de Lyon
de l'INSA de Lyon

Jean-Jacques Greffet
Professeur à l'Ecole Centrale Paris

Karl Joulain
Maître de Conférence à l'Ecole Nationale Supérieure de Mécanique
et d'Aérotechnique

Denis Lemonnier
Chargé de Recherche CNRS au Laboratoire d'Etudes Thermiques
de l'Ecole Nationale Supérieure de Mécanique et d'Aérotechnique

Bernard Perrin
Directeur de Recherche CNRS à l'Institut des Nanosciences de Paris

Gilles Tessier
Maître de Conférence à l'Ecole Supérieure de Physique
et de Chimie Industrielles de Paris

Nathalie Trannoy
Maître de Conférence à l'Université de Reims

Pascal Vairac
Maître de Conférence à l'Ecole Nationale Supérieure
de Mécanique et Microtechniques de Besançon

Fabrice Vallée
Directeur de Recherche CNRS au Centre
de Physique Moléculaire Optique et Hertzienne, Université Bordeaux I

Sebastian Volz
Chargé de Recherche CNRS au Laboratoire d'Energétique Moléculaire
et Macroscopique, Combustion de l'Ecole Centrale Paris

Topics in Applied Physics
Volume 107

Topics in Applied Physics

Topics in Applied Physics is a well-established series of review books, each of which presents a comprehensive survey of a selected topic within the broad area of applied physics. Edited and written by leading research scientists in the field concerned, each volume contains review contributions covering the various aspects of the topic. Together these provide an overview of the state of the art in the respective field, extending from an introduction to the subject right up to the frontiers of contemporary research.
Topics in Applied Physics is addressed to all scientists at universities and in industry who wish to obtain an overview and to keep abreast of advances in applied physics. The series also provides easy but comprehensive access to the fields for newcomers starting research.
Contributions are specially commissioned. The Managing Editors are open to any suggestions for topics coming from the community of applied physicists no matter what the field and encourage prospective editors to approach them with ideas.

Sebastian Volz (Ed.)

Microscale and Nanoscale Heat Transfer

With 144 Figures and 7 Tables

In Collaboration with Rémi Carminati, Patrice Chantrenne, Stefan Dilhaire, Séverine Gomez, Nathalie Trannoy, and Gilles Tessier

Springer

Dr. Sebastian Volz
Labortoire d'Enerégtique Moléculaire et Macroscopique, Combustion
Ecole Central Paris
Grande Voie des Vignes
92295 Châtenay Malabry, France
volz@em2c.ecp.fr

Library of Congress Control Number: 2006934584

Physics and Astronomy Classification Scheme (PACS):
65.80.+n, 82.53.Mj, 81.16.-c, 44.10.+i, 44.40.+a, 82.80.Kq

ISSN print edition: 0303-4216
ISSN electronic edition: 1437-0859
ISBN-10 3-540-36056-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-36056-8 Springer Berlin Heidelberg New York

Typesetting: DA-TEX · Gerd Blumenstein · www.da-tex.de
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper 57/3100/YL 5 4 3 2 1 0

Preface

The development of micro- and nanoscale fabrication techniques has triggered a broad scientific and technical revolution. A prime example is provided by microelectronics, which has now become nanoelectronics. Other evolutionary breakthroughs are now clearly established in the fields of optoelectronics, materials, the production and conversion of energy, and techniques for data processing and communications.

A remarkable feature of this trend is the way it has brought together physicists and engineers. On the one hand, the classical laws used to model macroscopic systems are generally unsuitable when system sizes approach characteristic microscopic scales, such as the mean free path or the length of carriers. The physical description of the individual or collective behaviour of the basic elements must then be reassessed. On the other hand, the development and integration of physical ideas exploiting very small structures, such as ultrathin films, superlattices, nanowires, and nanoparticles, in order to improve an industrial system, requires the physicist to understand some of the more technical aspects of engineering.

The international community of thermal scientists, whether in research or engineering, base their approach on the mass, momentum and energy conservation equations associated with the laws of diffusion for conduction (Fourier) and for mass transfer (Fick), and Newton's law for conduction–convection. For radiation, the radiative transfer equation is widely used to treat semi-transparent media, grey or otherwise.

But this theoretical framework can no longer describe the conductive and conductive–convective transfer regimes on very small space and time scales, simply because the carriers undergo too few collisions. As the radiated thermal wavelengths are of the order of a few microns, the radiative transfer equation, and even the whole notion of luminance, become quite inappropriate on submicron scales.

One does not even need to approach the limits of macroscopic models to observe that the phenomenology of heat transfer is quite different on the micron and centimeter length scales. Whilst heat transfer is generally felt to be a slow process – the time scale for heat conduction in macroscopic systems ($\sim$ 50 cm) is a few minutes – the propagation of heat is an extremely efficient process on the microscale ($\sim$ 10 ns). Indeed, the diffusion time is proportional to the square of the length. Moreover the thermal resistances of

microscale structures are so small that they become of the same order as the interface resistances between such structures. Microscale heat transfer thus occurs practically without inertia, and is essentially equivalent to interface heat transfer. Naturally, this is even more true for nanoscale heat transfer.

From the experimental standpoint, very weak and highly localised contributions must be detected in order to measure the conductive flux in nanostructures. For example, the methods used must not introduce high contact thermal resistances. Ultrafast optical methods (nano- to picosecond) and near-field microscopy are best suited to satisfy these criteria.

It is therefore clear that the study of heat transfer on micro- and nanoscales requires a quite new approach on the part of the thermal science community. The task here is to integrate the new physical models and also the novel experimental devices now available to treat energy exchanges in micro- and nanostructures.

There are many consequences for industry:

- In housing, superinsulating nanoporous materials can limit heat losses whilst increasing the ground surface, and their conductivity in vacuum is smaller than that of air.
- Nanofluids, i.e., heat-carrying liquids transporting nanoparticles, have conductivities 10–40% higher than those of the base fluid and hence a greatly enhanced transfer efficiency.
- In the nanoelectronics of processors, heating problems have led manufacturers to slow down the miniaturisation trend by switching to multi-unit structures in which several computing units are integrated into the same chip.
- Data storage will for its part be heat-assisted. Heating can activate or inhibit magnetisation reversal. It can also change the phase or the geometry of a storage medium, and this over nanoscale areas.
- Thermoelectric energy conversion is currently undergoing a revolution through manipulation of the thermophysical properties of nanostructured materials. In 2002, certain superlattice alloys were able to produce an intrinsic performance coefficient twice as high as had ever been measured for a bulk solid material. This breakthrough was achieved by improving thermal properties.

In all these fields of application, our understanding of the relevant heat mechanisms and the associated modelling tools remains poor or at best imperfect.

The present book brings together for the first time the physical ideas and formalism as well as the experimental tools making up this new field of thermal science. Although these are usually considered to be the jurisdiction of the physicist, the aim of the book remains quite concrete, since it seeks to solve the problems of heat transfer in micro- and nanostructured materials. The book itself results from a collaborative network in France known as the *Groupement de Recherche Micro et Nanothermique* (GDR), bringing

together teams organised by a unit of the *Centre national de la recherche scientifique* (CNRS)[1] and a unit of the department[2] of *Sciences pour l'Ingénieur*. This group combines research centres involved in thermal science, solid state physics, optics and microsystems. Each chapter has been written by one or several authors – sometimes belonging to different research teams – and then edited by experts and non-experts in the GDR.

The first part of the book is theoretical, making the connection between the fundamental approaches to energy transfer and the quantities describing heat transfer. Chapter 1 considers the limits of classical models on small scales. Chapters 2, 3 and 4 then treat the physical models describing heat transfer in gases, conduction, and radiation, respectively, all on these small scales.

The second part of the book covers the numerical tools that can be implemented to solve the previously formulated equations in concrete situations. Chapters 5 and 6 examine solutions of the Boltzmann and Maxwell equations, respectively. Having discussed continuum models, microscopic simulations are tackled in Chap. 7 via the Monte Carlo method and in Chap. 8 via the technique of molecular dynamics simulations. In each chapter, it is shown how to calculate a heat flux or conductivity explicitly through various examples.

The last part of the book deals with experimental approaches. Chapter 9 introduces different forms of near-field microscopy and discusses their applications in thermal science. A thermal microscope is presented in some detail with example applications. Chapter 10 discusses optical techniques as provided by the photothermal microscope and reflectometry, whilst Chap. 11 brings together optical and near-field microscopy in a single hybrid system. This series of chapters on microscopy is followed by two chapters presenting the thermal applications of femtosecond lasers in pump–probe configurations. Chapter 12 deals with the electron–photon interaction on ultrashort time scales and Chap. 13 treats of thermal–acoustic coupling in various types of structure.

The book thus constitutes a particularly complete and original collection of ideas, models, numerical methods and experimental tools that will prove invaluable in the study of micro- and nano-heat transfer. It should be of interest to research scientists and thermal engineers who wish to carry out theoretical research or metrology in this field, but also to physicists concerned with the problems of heat transfer, or teachers requiring a solid foundation for an undergraduate university course in this area.

[1] The French National Research Institute.
[2] Science for Engineering.

Acknowledgements

The strength of this book mostly relies on the collaborative effort of my dear colleagues. I am glad to express my deep thanks to Rémi Carminati, Patrice Chantrenne, Bernard Cretin, Stefan Dilhaire, Danièle Fournier, Séverine Gomez, Jean-Jacques Greffet, Karl Joulain, Denis Lemonnier, Bernard Perrin, Nathalie Trannoy, Gilles Tessier, Fabrice Vallée and Pascal Vairac for providing a work of highest quality in their field of expertise.

Paris, *Sebastian Volz*
November, 2005

Laws of Macroscopic Heat Transfer and Their Limits

Jean-Jacques Greffet

Ecole Centrale Paris, Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion (EM2C), Centre National de la Recherche Scientifique, 92295 Châtenay-Malabry Cedex
greffet@em2c.ecp.fr

Abstract. In this introductory text, we examine the three mechanisms of heat transfer. For each one, we review the main ideas used in the traditional macroscopic description of heat transfer. This is followed by a discussion of the length and time scales characterising these transfer mechanisms. We then study the hypotheses underlying these models in order to determine their field of validity. We outline transfer mechanisms beyond the validity of macroscopic laws. The latter will be discussed in more depth throughout the book.

1 Heat Conduction in Solids

1.1 Macroscopic Approach

Fourier Law

Heat conduction in a homogeneous medium is described by Fourier's law, which relates the flux to the temperature gradient by

$$\boldsymbol{\phi} = -k\boldsymbol{\nabla}T\,, \tag{1}$$

where k is the thermal conductivity.

Heat Equation

Energy conservation is expressed locally by

$$\rho c_{\mathrm{p}}\frac{\partial T}{\partial t} = -\boldsymbol{\nabla}\cdot\boldsymbol{\phi}\,, \tag{2}$$

where c_{p} is the specific heat capacity at constant pressure. Inserting the Fourier law into this expression and assuming that the thermal conductivity is homogeneous, we obtain a diffusion equation for the temperature field, viz.,

$$\rho c_{\mathrm{p}}\frac{\partial T}{\partial t} = k\nabla^2 T\,. \tag{3}$$

Defining the thermal diffusivity by $a = k/\rho c_{\mathrm{p}}$, (3) becomes

$$\nabla^2 T = \frac{1}{a}\frac{\partial T}{\partial t}\,. \tag{4}$$

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 1–13 (2007)

1.2 Characteristic Length and Time Scales

Characteristic Length and Time Scales for Conduction

The heat equation can be written in dimensionless form by setting $t = t^+ t_0$ and $x = x^+ L$, where t_0 and L are a characteristic time and length for the problem. The equation then becomes

$$\nabla_+^2 T = \frac{L^2}{a t_0} \frac{\partial T}{\partial t^+} \,. \tag{5}$$

The problem now depends only on the dimensionless quantity at_0/L^2, known as the Fourier number. In order to visualise the typical behaviour of a diffusive phenomenon, it is useful to study the solution to the problem described by

$$\nabla^2 T - \frac{1}{a} \frac{\partial T}{\partial t} = S\delta(t)\delta(\boldsymbol{r}) \,, \tag{6}$$

which corresponds to the temperature field created by a heat pulse at $t = 0$ and $\boldsymbol{r} = 0$. Choosing boundary conditions for an infinite medium in which the temperature decreases to zero at infinity, the solution is

$$T(\boldsymbol{r}, t) = \frac{S}{(4\pi a t)^{3/2}} \exp\Big(-\frac{r^2}{4at}\Big) \,. \tag{7}$$

This solution describes a Gaussian spread of energy in space, with characteristic width $2\sqrt{at}$. We observe that, after a time t has elapsed, the heat is spread out over a sphere of radius $2\sqrt{at}$.

Applications

A practical application of this notion can be formulated in the following question: when a bead of radius R is heated by absorption of heat at the surface, how long will it take for the object to thermalise? The answer is simply the time required for the heat to diffuse throughout its volume, i.e., a time of order $t_{\mathrm{cd}} = R^2/a$.

If the bead is now heated in a way that varies sinusoidally in time, i.e.,

$$Q\left[1 + \sin\frac{2\pi t}{T}\right] \,,$$

the behaviour will be quite different depending on whether T is large or small compared with the conduction time. If $T \gg t_{\mathrm{cd}}$, the system gradually thermalises, the bead becoming isothermal and following the temporal variations imposed upon it. But if $T \ll t_{\mathrm{cd}}$, the system cannot follow the imposed variations. It behaves like a low-pass filter and only responds to the mean value of the imposed flux. Moreover, the temperature field is not uniform.

Let us now examine several orders of magnitude. For most condensed materials, the order of magnitude of the thermal diffusivity is $10^{-6}\,\mathrm{m}^2 \cdot \mathrm{s}^{-1}$.

– For a sphere of radius 6400 km, the diffusion time is about

$$40 \times \frac{10^{12}}{10^{-6}} = 4 \times 10^{19}\,\mathrm{s}\,.$$

Since one year is approximately $\pi \times 10^7$ s, this gives a value of around 10^{12} yr. The Earth is thus unlikely to be isothermal, something we already knew!
– For a distance of 0.3 m, we obtain 10^5 s or one day. We deduce that a water pipe buried at a depth of 30 cm can get through the night without freezing.
– For a sphere of diameter 1 cm, we obtain 100 s.
– For a sphere of diameter 100 nm, the homogenisation time drops to 10^{-8} s = 10 ns.
– What about a sphere of diameter 1 nm? We find a time of 1 ps, which is correct to within an order of magnitude.

However, the above arguments are no longer applicable. We have reached the limits of validity of our macroscopic heat transfer models, for we now encounter another characteristic time of the physical problem. A difficulty arises here because the picosecond is the typical phonon relaxation time.

Short-Time Limit of the Diffusion Equation

The problems arising from the relaxation time of the system appear very clearly in the diffusion equation at short time scales. Indeed, looking at the solution for the response to a localised heat pulse, we find that the temperature field is modified instantaneously at every point of space from the moment the pulse comes into effect. Now at a time t, the temperature cannot have been modified beyond a distance ct, otherwise the effect of the pulse would have propagated faster than the speed of light. The diffusion equation must therefore be modified.

The physical origin of the problem lies in the fact that the Fourier law describes an instantaneous response to the excitation. This basically amounts to saying that we have neglected the response time of the system. Of course, the response to a temperature gradient imposed suddenly in this way cannot be instantaneous. The appropriate time scale can be ascertained for a linear system in the following way. We know the time scale of the response to a step-shaped excitation: it is the time required for equilibrium to be reestablished, i.e., the relaxation time. This is determined by microscopic collision processes. Now for any linear system, the response to a step function and the response to a Dirac function are related by differentiation. The natural time scale of the physical problem which marks out the domain of validity of the Fourier law is thus the relaxation time.

From the Boltzmann equation, the Fourier law can be modified to [1]

$$\tau \frac{\partial \boldsymbol{\phi}}{\partial t} + \boldsymbol{\phi} = -k \boldsymbol{\nabla} T\,, \tag{8}$$

where τ is the mean relaxation time. The resulting diffusion equation contains an extra term corresponding to propagation at speed $v = \sqrt{a/\tau}$. It is clear from this equation that, if the characteristic time scale t_0 is much longer than the relaxation time τ, the Fourier law used earlier is valid. However, if the time t_0 characterising the evolution of the system is shorter than the relaxation time, the traditional model is no longer valid. The above correction must then be taken into account, which amounts to saying that the Fourier law is dispersive.

Equivalently, we may compare a characteristic length scale for evolution of the system with the mean free path. If the latter is much shorter than the characteristic length, the local Fourier law is valid. If not, non-local effects must be taken into account. In other words, the energy flux at a point depends on the temperature gradient in a neighbourhood of this point.

Harmonic Conduction Regime

Let us consider the shape of a temperature field when we apply a modulated heat source with angular frequency ω. Seeking a solution of form

$$T(\boldsymbol{r}, t) = T_0(\boldsymbol{r}) + \theta(\boldsymbol{r}) \exp(-\mathrm{i}\omega t)\,,$$

we obtain for the time-dependent part

$$\nabla^2\theta + \frac{\mathrm{i}\omega}{a}\theta = 0\,. \tag{9}$$

This equation has an analogous structure to the Helmholtz equation, i.e., a harmonic propagation equation. One may then speak of thermal waves. If we now seek a solution of the form $\exp(\mathrm{i}Kx)$, we find a complex wave vector

$$K = \frac{1+\mathrm{i}}{\sqrt{2}}\sqrt{\frac{\omega}{a}}\,. \tag{10}$$

The imaginary part of K describes attenuation of the wave during its propagation. As discussed above, the system cannot respond to variations that occur too quickly. This behaviour shows up here when we observe that high frequency waves are damped over very short distances. It is easy to establish the order of magnitude of the wavelength or the attenuation distance. One only has to consider the Fourier number, using the period T as the characteristic time, which yields $l = \sqrt{aT}$. Note that the amplitude is usually written in the form

$$\exp\frac{\mathrm{i}2\pi x}{\lambda}\exp\Big(-\frac{x}{\delta}\Big)\,.$$

The decay length is thus the wavelength divided by 2π. Thermal waves are therefore highly damped waves. Put another way, we are always in the near field when we observe thermal waves.

Looking again at the modified Fourier law (8), we observe that the conductivity depends on the frequency. In the monochromatic regime, we have

$$\boldsymbol{\phi} = -\frac{k}{1 - \mathrm{i}\omega\tau}\boldsymbol{\nabla}\theta\,. \tag{11}$$

The time constant τ here is related to the phonon relaxation time, which is in the picosecond range. There is no point in taking this effect into account for frequencies less than $1/\tau$, i.e., a few THz. We may thus consider that thermal conductivity is non-dispersive at low frequencies. This is analogous to what happens with electrical conductivity, where dispersion effects can be ignored up to a few GHz.

1.3 Short-Scale Transfer

To summarise, the main limitations of the macroscopic approach to conduction correspond to length and time scales comparable with the phonon mean free path and the phonon relaxation time, respectively. In the Chapter on Electrons and Phonons by *Greffet* in this volume, which deals with phonons, we shall see that conductive transport can be described using a Boltzmann equation. This type of equation will be investigated further in the Chapters by *Carminati* in this volume, which treat the problems of transport in dilute media and radiative transfer, respectively. The transition between a ballistic transport regime and a diffusive regime will be discussed.

Another type of limitation arises when we consider geometrical dimensions comparable with the phonon wavelength. In this situation, mode quantisation effects become relevant. Consider the case of a wire a few nanometers in diameter. The number of modes that can propagate is then discrete. This can lead to quantisation of the conductance, a phenomenon that is well known in electrical conduction [2] and which has recently been discussed in the context of heat conduction [3].

2 Conduction in Fluids. Convection

2.1 Macroscopic Approach

Introduction

When we consider heat transfer in a fluid at rest, the heat flux is due to conduction and can be described by the Fourier law. In the presence of gravity, the density variation induced by temperature differences leads to forces which can produce movement within the fluid. The phenomenon is called natural convection. When it occurs, this transfer mechanism is much more efficient than the conduction phenomenon. One also speaks of convection to describe

energy transfer between a surface and a fluid. In the following, we shall not discuss convection phenomena with mass transfer which are truly macroscopic effects. We shall only address conduction in gases in the vicinity of a surface.

Transfer Near a Surface

We shall consider the problem of energy transfer between a surface separating a solid ($x < 0$) and a fluid ($x > 0$). It can be shown classically that the heat flux per unit area ϕ can be expressed in the form

$$\phi = h(T_\mathrm{s} - T_\mathrm{f})\,, \tag{12}$$

where T_s and T_f are the temperatures of the surface and the fluid some distance from the surface. The origin of this transfer is heat conduction in the boundary layer. Owing to the condition of zero velocity at the surface itself, the fluid has no overall motion along the surface and there is no convection, i.e., no energy transfer associated with a mass transfer. The only energy transport phenomenon is conduction in the fluid within the boundary layer. We may thus express the fact that the flux at the surface is oriented normally to the surface and has the form

$$\phi = -k_\mathrm{f}\frac{\mathrm{d}T}{\mathrm{d}x}\,, \tag{13}$$

where k_f is the thermal conductivity of the fluid. This expression can be written in a dimensionless way by introducing the dimensionless temperature

$$T(x) - T_\mathrm{s} = (T_\mathrm{f} - T_\mathrm{s})T^+(x)$$

and a dimensionless variable $x = x^+L$, where L is a characteristic length for the system. We then have

$$\phi = -\frac{k_\mathrm{f}(T_\mathrm{f} - T_\mathrm{s})}{L}\left(-\frac{\mathrm{d}T^+}{\mathrm{d}x^+}\right) = h(T_\mathrm{s} - T_\mathrm{f})\,, \tag{14}$$

where $h = k_\mathrm{f}Nu/L$ is the exchange coefficient and the dimensionless number Nu is the Nusselt number. The latter represents the dimensionless temperature gradient in the boundary layer, i.e., $Nu = \mathrm{d}T^+/\,\mathrm{d}x^+$.

Characteristic Convection Time

It is useful to introduce a characteristic time scale associated with the convective flow. This time scale arises when we consider a sphere of matter with volume V and area S, cooling in a fluid. The energy balance equation is

$$\rho c_\mathrm{p} V\frac{\partial T}{\partial t} = hS(T - T_\mathrm{f})\,. \tag{15}$$

We have assumed here that the sphere is isothermal which is justified if the system evolves on a longer time scale than the conduction time. When we solve this equation, we obtain an exponential temperature decrease with a time constant τ_{cv} known as the convection time:

$$\tau_{\mathrm{cv}} = \frac{\rho c_{\mathrm{p}} V}{hS} \,. \tag{16}$$

2.2 Short-Scale Transfer. Ballistic Transport

Let us now consider the transfer between two parallel planes separated by a distance d and with different temperatures. Molecules leaving one surface carry out a random walk, undergoing collisions with other molecules. The key parameter in this process is the mean free path, i.e., the average distance travelled between two successive collisions. Clearly, if the distance d between the two planes is small compared with the mean free path, the phenomenon here will be profoundly different. The transfer is then said to be ballistic, since molecules leaving one surface will go directly to the other without collision. This is called the ballistic or rarefied gas regime.

To get a more quantitative idea, the order of magnitude of the mean free path in air under normal conditions is around 40 nm. Hence, the energy transfer between the apex of an AFM tip and a sample surface is ballistic. It should be noted that ballistic transfer is considerably more efficient than diffusive transfer. This is easy to understand. In the ballistic regime, every molecule leaving a surface will reach the opposite surface and will thus play an efficient role in the transfer of energy. In the diffusive regime, some molecules undergo collisions in which they are scattered forward, in which case the situation is much the same as before. However, some molecules are scattered backwards and this reduces the efficiency of the energy transfer.

The relevant regime can be characterised by two characteristic time scales. The first is the collision time and the second is the time-of-flight between the two surfaces. As long as the time-of-flight is shorter than the collision time, the regime is ballistic. To summarise, the usual approach to convective heat transfer based on the use of an exchange coefficient is valid for distances greater than the mean free path and time scales longer than the collision time.

3 Radiation

3.1 Macroscopic Approach

Assumptions

The phenomenological description of radiative transfer is based on the notion of specific intensity and the theory of geometric optics.

Specific Intensity

A radiative energy flux crossing a surface of area $\mathrm{d}S$ in the direction $\boldsymbol{u}$, in a solid angle $\mathrm{d}\Omega$ and a frequency band $[\nu, \nu + \mathrm{d}\nu]$ is expressed in the form

$$\mathrm{d}\phi_\nu = L_\nu(\boldsymbol{u}, \boldsymbol{r})\, \mathrm{d}S \cos\theta\, \mathrm{d}\Omega\, \mathrm{d}\nu\,. \tag{17}$$

The quantity L is the specific intensity. It depends on the frequency, the direction and the point considered. It can be interpreted as the product of the number $n_\nu(\boldsymbol{u}, \boldsymbol{r})$ of photons per unit volume propagating in the direction $\boldsymbol{u}$ with the energy $h\nu$ per photon and the speed of propagation c.

Equilibrium Specific Intensity

In thermodynamic equilibrium, each mode is occupied by a number of photons given by the Bose–Einstein distribution. The number n_ν introduced above is then given by

$$n_\nu(\boldsymbol{u}, \boldsymbol{r}) = \frac{1}{\exp(h\nu/kT) - 1}\,. \tag{18}$$

It remains to determine the number of modes per unit volume in the solid angle $\mathrm{d}\Omega$. This can be found by counting the number of electromagnetic modes in a finite cavity of side L and then letting L tend to infinity. The result is $2\nu^2/c^3$. From there, we retrieve the expression for the specific intensity of a black body, which is the specific intensity of equilibrium radiation, i.e., the radiation when a system is in thermodynamic equilibrium:

$$L^0_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT) - 1}\,. \tag{19}$$

It should be noted that the only solutions of Maxwell's equations taken into account in the above reasoning are plane waves propagating in a vacuum. Near an interface, inside a cavity of size comparable with the wavelength, or in a wave guide, the solutions to Maxwell's equations are different. The modes change from one physical system to another. At equilibrium, each mode remains occupied with a mean occupation number given by Bose–Einstein statistics. However, the number of modes per unit volume is no longer $2\nu^3/c^3$ and can vary significantly from one system to another.

Emitted and Absorbed Flux

At equilibrium, the specific intensity is denoted by L^0_ν. The specific intensity leaving a surface is the sum of the emitted and reflected specific intensities. The emitted specific intensity is thus a fraction of the equilibrium specific intensity. This fraction is given by the emissivity ϵ_ν. Hence,

$$L^{\mathrm{e}}_\nu(T) = \epsilon_\nu L^0_\nu(T)\,. \tag{20}$$

The absorptance (or absorptivity) is defined as the ratio of the power absorbed per unit area and the power incident per unit area. It is denoted by α_ν. From the reciprocity theorem, it can be shown that the polarised directional monochromatic absorptance is equal to the polarised directional monochromatic emissivity [4]:

$$\alpha_\nu(\boldsymbol{u}) = \epsilon_\nu(\boldsymbol{u})\,. \tag{21}$$

The physical origin of this equality can be understood by noting that the absorptance of a surface is nothing other than the transmission factor of the interface. Indeed, all the energy passing through the surface is subsequently absorbed in the semi-infinite medium. The emissivity can also be interpreted as a transmission factor. Inside the opaque medium, there is radiation which is transmitted across the interface before propagating in the vacuum. Looking at the situation in this way, the emissivity is once again a transmission factor of the interface. Kirchhoff's law then arises as a consequence of the equality between the vacuum–medium and medium–vacuum transmission factors. This equality in turn follows from the reciprocity principle as demonstrated in the theory of electromagnetism. It expresses the fact that the signal delivered by a detector does not change when source and detector swap positions. A more detailed discussion can be found in [5].

3.2 Characteristic Length and Time Scales

Several length scales arise in radiative problems: the wavelength, the skin depth, and the coherence lengths.

Wavelength

An obvious characteristic length is the wavelength. It determines several types of behaviour. On the one hand, if the relevant distances are smaller than the wavelength, retardation effects can be neglected. The problem can then be treated as a problem of electrostatics as far as the spatial variation of the field is concerned. One should be careful, however, because certain properties of the medium such as the dielectric constant still depend on the frequency.

The wavelength is also the length scale that determines the range of validity of geometric optics. When structures become comparable in size with the wavelength, it becomes difficult to justify using light rays represented by straight lines.

Finally, let us note the importance of the wavelength in transfer modes between objects that are smaller than the wavelength. When an object is smaller than the wavelength, it can be assimilated from the radiative point of view with a dipole. For concreteness, consider two particles measuring a few tens of nanometers and placed 100 nm apart. This situation can be encountered for example in soot, a substance made up of clusters of carbon

nanoparticles with sizes of the order of about ten nanometers. Each nanometric particle can be described by a random electric dipole due to thermally induced charge fluctuations. It thus creates a dipole field. At a distance of around 100 nm, the particle creates a field dominated by the $1/r^3$ terms of the dipole radiation. The $1/r$ terms normally kept in the far field are negligible in this case! It is clear that the notions of emissivity and absorptance are inappropriate here. What we have just described is in fact the counterpart of the van der Waals forces for energy transfer. One refers to energy transfer, while the other refers to momentum transfer. Note finally that this type of transfer which takes place via terms of electrostatic type in the electromagnetic field, i.e., dipole–dipole coupling, can be described as near-field radiative transfer. In the context of energy exchange between molecules and their surroundings, this mode of transfer is often called non-radiative transfer. This term is used because, in the language of quantum physics, this process does not correspond to a photon emission and absorption process. However, it is nevertheless a transfer of energy due to electromagnetic interactions.

Note also that the Fresnel reflection factor is defined for a plane wave [6]. In practice, a real beam can be assimilated with a plane wave if it is highly directional, which amounts to saying that, at its narrowest point, the beam radius is much greater than the wavelength. It follows that the concept of reflection factor is meaningless for length scales shorter than the wavelength.

Skin Depth

When a medium is opaque, the field penetrates to a depth known as the skin depth. It is denoted by $\lambda/2\pi\Im(n)$, where $\Im(n)$ is the imaginary part of the complex refractive index n. The power is thus absorbed at the surface throughout a layer of this thickness. Likewise, the emitted field is produced in this layer. One may thus apply the idea of emissivity, provided that the temperature field is uniform over the skin depth. If this is not the case, the problem has to be treated as a semi-transparent medium.

Coherence Length

Another idea plays an important part in thermal emission phenomena. This is the idea of spatial coherence length. When the field is incoherent, we add together the intensities of the contributions from the different points on the surface. In the case of a coherent field, the contributions from different points of the surface can interfere with one another. This leads to a quite different result. It can be shown that the coherence length is much shorter than the wavelength when the media do not carry surface waves [5]. Otherwise, waves propagating across the surface create a highly coherent field at the interface. This leads to large anomalies in the emissivity and absorptance which can be exploited to produce a profound modification in the radiative properties of surfaces.

Attenuation Constant in a Diffusive Medium

In a diffusive medium, there are two characteristic lengths associated with the electromagnetic field. The first corresponds to what is called the visibility in our weather reports. This is the distance characterising the attenuation of a collimated beam, as given by the Beer–Lambert law. It is called the extinction length. Extinction is due to absorption and scattering. The second characteristic length corresponds to the distance beyond which the radiation becomes isotropic so that the radiative flux tends to zero. This is called the transport length.

Applying these definitions to the case of a cloud layer, the first distance is that at which we may still glimpse several rays of sunshine, so that solid objects still cast a faint shadow. The second definition corresponds to the thickness at which the cloud layer becomes more or less opaque. This idea will be further discussed in the Chapter on transport in dilute media by *Carminati* in this volume.

Electromagnetic Origins of Thermal Radiation

The radiation from a heat source can be calculated within the framework of electromagnetic theory. The method consists in applying the usual theory of antennas. With this approach, we can relate the radiated fields to currents. For thermal radiation, the currents are due to random motions of charges in the matter, i.e., electrons in metals and ions in polar crystals. Each volume element then behaves as a radiating random dipole moment. The problem of thermal radiation from a surface reduces to that of a dipole behind an interface. Since the average motion of the charges is zero, the mean radiated field is zero. However, the mean squared value of the field is not zero. Carrying out the calculation, one directly obtains the expression for the emissivity as a function of the refractive index, as will be discussed in detail in the Chapter by *Joulain* in this volume (see also [5]).

Flux Between Two Planes as a Function of Their Separation

At very short distances, radiative transfer is dominated by terms of electrostatic type, i.e., the components of the dipole field that decay as $1/r^3$ with distance r. When there are surface waves, they are responsible for the main part of the energy transfer. This contribution may be viewed as a heat flow due to the tunnel effect. It can also be interpreted as a collision between phonons in the upper interface and phonons in the lower interface in the case of polar crystals. For metallic surfaces, one has collisions between surface plasmons.

4 Conclusion

To conclude, the main message to be understood from this brief overview is that what we call nanoheat transfer in a rather convenient abbreviation refers to heat transfer phenomena in contexts where we go beyond the usually applied macroscopic laws. The frontier here is neither the micron nor the nanometer. Rather, it is fixed by the length scales (mean free path) and time scales (relaxation time) of the relevant physical systems. New phenomena appear when we consider systems on length and time scales comparable with or shorter than these intrinsic length and time scales.

References

[1] S. Volz: Phys. Rev. Lett. **87**, 074301 (2001) 3
[2] N. Agraït, A. Levy Yeyati, J. M. van Ruitenbeek: Phys. Rep. **337**, 81 (2003) 5
[3] K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes: Nature (London) **404**, 974 (2000) 5
[4] J. J. Greffet, M. Nieto-Vesperinas: J. Opt. Soc. Am. A **10**, 2735 (1998) 9
[5] K. Joulain, J. P. Mulet, F. Marquier, R. Carminati, J. J. Greffet: Surf. Sci. Rep. **57**, 59–112 (2005) 9, 10, 11
[6] J. J. Greffet, R. Carminati: Intl. J. Heat and Tech. **18**, 81 (2000) 10

Index

Transport in Dilute Media

Rémi Carminati

Ecole Centrale Paris, Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion (EM2C), Centre National de la Recherche Scientifique, 92295 Châtenay-Malabry Cedex
remi.carminati@em2c.ecp.fr

Abstract. This Chapter is an introduction to the kinetic theory of gases. As part of a book on micro and nanoscale heat transfer, the aims are twofold:

- To introduce the necessary concepts and tools, and in particular, the idea of a distribution function and the Boltzmann equation, to describe heat transfer in dilute gases on short length and time scales.
- To introduce general notions in the theory of transport, based on the kinetic approach, which will prove useful in later Chapters of the book, especially for describing the transport of electrons and phonons in solids.

The Chapter is organised as follows. We begin by introducing the ideas of distribution function, average and flux. We then discuss the particular context of thermodynamic equilibrium and show that, to describe systems that are out of equilibrium, which provide the conditions for macroscopic transfer, one must be able to calculate the distribution function in the most general situation. We introduce the underlying formalism of the Boltzmann equation and a highly simplified model based on the relaxation time. We can then discuss the idea of local thermodynamic equilibrium (LTE), and also situations that are close enough to LTE to be treated by perturbation methods. We shall show in particular how to demonstrate the Fourier law in this regime and obtain an expression for the thermal conductivity of a gas. We then turn to non-LTE regimes and in particular the ballistic transport regime which arises when the characteristic size of the system is smaller than the mean free path (or the observation time is shorter than the average time elapsed between two collisions). We end with a concrete example in which we compare and comment upon the orders of magnitude of exchanged fluxes in different regimes (convection, Fourier-type conduction, ballistic transport).

1 Distribution Function and Flux

1.1 Distribution Function

We shall be concerned here with a monatomic gas (no internal degrees of freedom), made up of N identical atoms of mass m contained within a volume V. The gas is assumed to be dilute, i.e., the average distance $d \sim (V/N)^{1/3}$ between molecules is much greater than the range a of the interaction potential. It is also assumed to be a classical gas in the sense that the de Broglie wavelength $\lambda = h/\sqrt{3mk_{\mathrm{B}}T} \ll d$. Each atom is therefore considered to be a classical particle with well defined position and velocity.

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 15–35 (2007)

In order to describe the properties of the gas, either in or out of equilibrium, we introduce the one-particle distribution function or velocity distribution function $f(\boldsymbol{r}, \boldsymbol{v}, t)$ defined by

$$\mathrm{d}n(\boldsymbol{r}, \boldsymbol{v}, t) = f(\boldsymbol{r}, \boldsymbol{v}, t)\, \mathrm{d}^3 r\, \mathrm{d}^3 v\,, \tag{1}$$

where $\mathrm{d}n$ is the number of particles at time t with position inside the volume element $\mathrm{d}^3 r$ centered on the point $\boldsymbol{r}$ and velocity within the element $\mathrm{d}^3 v$ of the velocity space centered on $\boldsymbol{v}$. The distribution function is the fundamental quantity in the kinetic theory of transport for gases and solids (see the Chapter by *Greffet* on Electrons and Phonons in this volume).

The particle number appearing in (1) should be understood as an average over a volume element $\mathrm{d}^3 r$ that is infinitely small on the macroscopic scale and over a time scale of the order of the time required by the particles to cross this element. Two comments are in order here:

- The infinitesimal volume element must be large compared with the size of the molecules. An order of magnitude for the molecular size is provided by the range a of the interaction potential. This implies that the distances L over which spatial variations of the distribution function become significant must be larger than a.
- The size of the infinitesimal volume element can be chosen arbitrarily with respect to the average intermolecular distance d. This choice affects the meaning attributed to value of the particle number $\mathrm{d}n$ in (1). If $\mathrm{d}^3 r$ is large compared with d, fluctuations in the number of particles in $\mathrm{d}^3 r$ will be small and $\mathrm{d}n$ will be a macroscopic quantity. Otherwise the volume $\mathrm{d}^3 r$ contains on average a small number of particles and the fluctuations in this number are of the order of its average value (a more detailed discussion can be found in [1]). In this case the distribution function fluctuates over a spatial length scale L such that $a \ll L \ll d$ and on a time scale θ such that $\tau_0 \ll \theta \ll \tau$, where τ_0 is the duration of a collision and τ is the average time between consecutive collisions. This situation is encountered in the context of short-scale transport, of the kind we shall be concerned with here.

1.2 Averages

Let G be a microscopic quantity associated with the gas particles. The macroscopic average of G at position $\boldsymbol{r}$ and time t (in the sense defined by the volume element $\mathrm{d}^3 r$) is

$$\langle G(\boldsymbol{r}, t)\rangle = \frac{1}{n(\boldsymbol{r}, t)} \int f(\boldsymbol{r}, \boldsymbol{v}, t) G(\boldsymbol{r}, \boldsymbol{v}, t)\, \mathrm{d}^3 v\,, \tag{2}$$

where $n(\boldsymbol{r}, t)$ is the number of particles per unit volume at point $\boldsymbol{r}$ and time t, given by

$$n(\boldsymbol{r}, t) = \int f(\boldsymbol{r}, \boldsymbol{v}, t)\, \mathrm{d}^3 v\,. \tag{3}$$

For example, the instantaneous local average velocity $\langle \boldsymbol{v} \rangle$ (also called the hydrodynamic velocity) is given by

$$\langle \boldsymbol{v}(\boldsymbol{r}, t) \rangle = \frac{1}{n(\boldsymbol{r}, t)} \int f(\boldsymbol{r}, \boldsymbol{v}, t) \boldsymbol{v}\, \mathrm{d}^3 v\,. \tag{4}$$

This is the velocity used in fluid mechanics. Likewise, the internal energy per unit volume is given by

$$\langle E(\boldsymbol{r}, t) \rangle = \frac{1}{n(\boldsymbol{r}, t)} \int f(\boldsymbol{r}, \boldsymbol{v}, t) \frac{1}{2} m (\boldsymbol{v} - \langle \boldsymbol{v} \rangle)^2\, \mathrm{d}^3 v\,, \tag{5}$$

where $m(\boldsymbol{v} - \langle \boldsymbol{v} \rangle)^2/2$ is the kinetic energy of a particle in a reference frame moving with the fluid.

1.3 Conductive Flux

The distribution function can be used to calculate fluxes associated with the motion of the carriers, in this case the gas particles. Let G be an arbitrary microscopic quantity and $\mathrm{d}S$ a surface element with unit normal $\boldsymbol{n}$. The flux density (or flux per unit area) of the quantity G through the surface element $\mathrm{d}S$ is defined as

$$\phi_G = \int G(\boldsymbol{r}, \boldsymbol{v}, t) f(\boldsymbol{r}, \boldsymbol{v}, t) (\boldsymbol{v} - \langle \boldsymbol{v} \rangle) \cdot \boldsymbol{n}\, \mathrm{d}^3 v\,. \tag{6}$$

This corresponds to a diffusive flux, as can be seen from the presence of the relative velocity $\boldsymbol{v} - \langle \boldsymbol{v} \rangle$ with respect to the ensemble motion of an elementary volume of gas.

The conductive flux which is of particular importance in the study of heat transfer corresponds to the transport of kinetic energy:

$$\phi_{\mathrm{cd}} = \int \frac{1}{2} m (\boldsymbol{v} - \langle \boldsymbol{v} \rangle)^2 f(\boldsymbol{r}, \boldsymbol{v}, t) (\boldsymbol{v} - \langle \boldsymbol{v} \rangle) \cdot \boldsymbol{n}\, \mathrm{d}^3 v\,. \tag{7}$$

2 Thermodynamic Equilibrium

2.1 Definition

Thermodynamic equilibrium is a fundamental concept which is often used intuitively. A formal definition can nevertheless be given [2]:

> If a closed system is in a state for which, in every macroscopic subsystem and at every moment of time, the macroscopic physical quantities are equal to their average value to very high accuracy, then the system is said to be in statistical equilibrium (or thermodynamic equilibrium, or again, thermal equilibrium).

Another definition in the form of a postulate upon which equilibrium statistical physics can be constructed has been given by *Callen* [3]:

> There exist particular states (called equilibrium states) of simple systems that, macroscopically, are characterized completely by the internal energy U, the volume V, and the mole numbers $N_1, N_2, \ldots, N_r$ of the chemical components.

A simple system here would be an uncharged gas, for example. It is important to note that, at equilibrium, the macroscopic quantities are homogeneous and static. There is no macroscopic flux in this context.

2.2 Equilibrium Distribution Function

The distribution function at thermodynamic equilibrium at temperature T can be calculated using standard techniques of equilibrium statistical physics (see Appendix A). We obtain the Maxwell–Boltzmann distribution

$$f^0 = n \left(\frac{m}{2\pi k_\mathrm{B} T} \right)^{3/2} \exp\left(-\frac{m v^2}{2 k_\mathrm{B} T} \right) , \tag{8}$$

where $n = N/V$ is the number of particles per unit volume and k_B is the Boltzmann constant. This distribution function is homogeneous and stationary and depends only on the speed $v = |\boldsymbol{v}|$ of the particles.

The function f^0 is even with respect to the three components of the velocity $\boldsymbol{v}$. Inserting it into (4) and (6), we thus find that, at equilibrium, the instantaneous average local velocity is zero and all macroscopic fluxes are zero. In order to describe transport situations, we must find a way to calculate the distribution function out of equilibrium. To do so, we use a dynamical equation for the distribution function, first introduced by Boltzmann at the end of the nineteenth century. This nonlinear equation is extremely difficult to solve in the general case. We therefore introduce a simplification by considering situations out of equilibrium but nevertheless close to a state known as local thermodynamic equilibrium. A perturbative approach can then be used to linearise the problem. This is the subject of the rest of the Chapter.

3 Boltzmann Equation

3.1 Dynamical Equation for the Distribution Function

Free Particles

The case of non-interacting particles is an ideal one. It applies to a good approximation in the case of rarefied media to be discussed later, e.g., in cases where the characteristic length scale of the system is shorter than the mean free path.

In this case, it can be shown that (see Appendix B) the distribution function satisfies

$$\frac{\mathrm{d}f}{\mathrm{d}t}(\boldsymbol{r},\boldsymbol{v},t) = \frac{\partial f}{\partial t} + \boldsymbol{v}\cdot\frac{\partial f}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m}\cdot\frac{\partial f}{\partial \boldsymbol{v}} = 0\,, \tag{9}$$

where $\boldsymbol{F}$ is the external force applied to each particle, e.g., an electrical force if the gas is made up of charged particles and placed in an external electric field. The notation $\partial/\partial\boldsymbol{r}$ ($\partial/\partial\boldsymbol{v}$) denotes the gradient with respect to the variable $\boldsymbol{r}$ ($\boldsymbol{v}$).

Collisions and Their Role

Collisions between particles (and collisions with the walls of a container, where appropriate) have the effect of suddenly changing the state of a given particle, in a time that is assumed to be infinitely short and over a length scale that is considered to be infinitely small, in conformity with the assumption of a dilute medium. In the presence of collisions, we then have, quite generally,

$$\frac{\partial f}{\partial t} + \boldsymbol{v}\cdot\frac{\partial f}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m}\cdot\frac{\partial f}{\partial \boldsymbol{v}} = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}}\,. \tag{10}$$

The right-hand side expresses the variation per unit time of the distribution function $f(\boldsymbol{r},\boldsymbol{v},t)$ due to collisions. It can be written formally as

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}} = \Gamma^{+} - \Gamma^{-}\,,$$

where Γ^{+} expresses the effect of all collisions producing a particle in the state $(\boldsymbol{r},\boldsymbol{v},\mathrm{d}^3r,\mathrm{d}^3v)$ from some different state, and Γ^{-} expresses the effect of all collisions which, starting from a particle in the state $(\boldsymbol{r},\boldsymbol{v},\mathrm{d}^3r,\mathrm{d}^3v)$, produce a particle in some different state.

Boltzmann was the first to give an explicit expression for the collision term, for the case where only binary collisions (involving just two particles at a time) were taken into account. This hypothesis is natural in the context of dilute media, since the probability of three or more bodies colliding at the same instant is then extremely small. The demonstration is also based

on the assumption of molecular chaos, the states of colliding particles being statistically uncorrelated, and the hypothesis of elastic collisions. The proof itself is rather tedious [4, 5] and it will be sufficient here merely to quote the result, since it will not be used later:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int_{\boldsymbol{v}_1}\int_{4\pi}\Big[f(\boldsymbol{r},\boldsymbol{v}_1',t)f(\boldsymbol{r},\boldsymbol{v}',t)-f(\boldsymbol{r},\boldsymbol{v}_1,t)f(\boldsymbol{r},\boldsymbol{v},t)\Big]|\boldsymbol{v}-\boldsymbol{v}_1|\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\,\mathrm{d}^3v_1\,\mathrm{d}\Omega\,. \tag{11}$$

In this equation, we consider a binary collision involving just two particles with velocities $\boldsymbol{v}$ and $\boldsymbol{v}_1$ before the collision and $\boldsymbol{v}'$ and $\boldsymbol{v}_1'$ after the collision. The latter two outgoing velocities can be expressed in terms of the incoming velocities in the case of elastic collisions. The solid angle $\mathrm{d}\Omega$ contains the direction of the relative velocity $\boldsymbol{v}_1'-\boldsymbol{v}'$ after the collision and is called the scattering solid angle, whilst $\mathrm{d}\sigma/\,\mathrm{d}\Omega$ is the differential scattering cross-section.

The original Boltzmann equation as formulated by Boltzmann himself at the end of the nineteenth century is the dynamical equation (10) in which the right-hand side has been replaced by the expression for the collision term in (11). By abuse of language, (10) is often referred to as the Boltzmann equation, whatever model is used to express the right-hand side.

3.2 The Relaxation Time Model

The collision term (11) is rather difficult to handle and the full Boltzmann equation is nonlinear. In many situations, one can justify a rather rough approximation for the collision integral based on the following observation. When the gas evolves freely from a non-equilibrium state, it will tend asymptotically to an equilibrium state. It is the collisions that are responsible for this return to equilibrium. We then have

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f-f^{(0)}}{\tau(\boldsymbol{v})}\,, \tag{12}$$

where $\tau(\boldsymbol{v})$ is a relaxation time describing the return to equilibrium within the volume element under consideration. The function $f^{(0)}$ is an equilibrium distribution function. It is called the local equilibrium distribution because it generally depends on the volume element. This idea will be made more precise in the next section. The relaxation time τ is often identified with the collision time, i.e., the average time elapsed between consecutive collisions. It depends a priori on the particle velocities and the interaction potential. When it cannot be calculated, this parameter is fitted to experimental data in the final formulas.

4 Local Thermodynamic Equilibrium. Perturbation Method

4.1 Dimensionless Boltzmann Equation

The idea underlying any perturbative approach is to seek a solution of an equation in the form of an expansion in powers of some dimensionless parameter that is much smaller than unity. In order to determine such a parameter in the present case, let us consider the characteristic length and time scales associated with the distribution function:

- A characteristic length L for the spatial variation determined by the geometry of the system.
- A characteristic time θ which is the macroscopic time associated with the dynamical evolution of the system. This quantity has a clear meaning when the system is not static.
- A characteristic scale U for variations in the speed of a molecule on the spatial scale L (or over the time θ) under the action of an external force $\boldsymbol{F}$. We may write directly $U = FL/mv$ ($U = F\theta/m$), where F is the order of magnitude of the force $\boldsymbol{F}$ and v is the order of magnitude of the particle speed (which may be identified with the root mean square speed, for example).

Introducing these quantities into the Boltzmann equation in the relaxation time model, we obtain

$$\frac{\tau}{\theta}\frac{\partial f}{\partial t^*} + \frac{v\tau}{L}\boldsymbol{v}^* \cdot \frac{\partial f}{\partial \boldsymbol{r}^*} + \frac{F\tau}{mU}\boldsymbol{F}^* \cdot \frac{\partial f}{\partial \boldsymbol{v}^*} = -\left[f - f^{(0)}\right] . \tag{13}$$

The quantities marked with an asterisk are all dimensionless. The derivatives are carried out with respect to dimensionless variables and all terms involving the distribution function f have the same order of magnitude.

4.2 Mean Free Path. Collision Time. Knudsen Number

We now consider a stationary system. Equation (13) becomes

$$\frac{v\tau}{L}\left(\boldsymbol{v}^* \cdot \frac{\partial f}{\partial \boldsymbol{r}^*} + \boldsymbol{F}^* \cdot \frac{\partial f}{\partial \boldsymbol{v}^*}\right) = -\left[f - f^{(0)}\right] , \tag{14}$$

where U has been replaced by its expression $U = FL/mv$. We see that the variations in f are driven by the dimensionless number $K_n = v\tau/L$, known as the Knudsen number. It can also be written as

$$K_n = \frac{l}{L} , \tag{15}$$

where $l = v\tau$ is the mean free path of the particles in the gas. The Knudsen number is therefore the ratio of the particle mean free path and the characteristic length scale of the system.

If we now consider a homogeneous but non-stationary system, it is the second term on the left-hand side of (13) which vanishes and we obtain

$$\frac{\tau}{\theta}\left(\frac{\partial f}{\partial t^*} + \boldsymbol{F}^* \cdot \frac{\partial f}{\partial \boldsymbol{v}^*}\right) = -\left[f - f^{(0)}\right], \tag{16}$$

where U has been replaced this time by its expression $U = F\theta/m$. The dimensionless number driving variations in f this time is still the Knudsen number, but now expressed in the form

$$K_n = \frac{\tau}{\theta}. \tag{17}$$

The Knudsen number is thus the ratio of the collision time and the macroscopic time scale characterising the dynamical evolution of the system. Note that, given the equality $L = v\theta$ which is always valid up to an order of magnitude, the two definitions of the Knudsen number are equivalent.

To exemplify this, let us evaluate the order of magnitude of the mean free path l and the collision time τ for a gas in standard conditions. The mean free path is given by

$$l = \frac{1}{n\sigma}, \tag{18}$$

where n is the density and σ the scattering cross-section. The collision time is given by

$$\tau = \frac{l}{v}, \tag{19}$$

where v is a typical particle speed, often taken as their average speed. For a gas at room temperature and atmospheric pressure, we have $n \approx 2.5 \times 10^{25}\,\mathrm{m}^{-3}$. The cross-section is of the order of πa^2, where a is the range of the interaction potential, which gives $\sigma \approx 10^{-18}\,\mathrm{m}^2$. We thus obtain $l \approx 40\,\mathrm{nm}$. For nitrogen, the average molecular speed is $7 \times 10^4\,\mathrm{cm \cdot s^{-1}}$. The collision time is then of the order of $\tau \approx 10^{-9}\,\mathrm{s}$.

Depending on the value of the Knudsen number, we may specify three regimes:

- If $K_n \ll 1$, the regime is said to be collisional. This is the regime considered in fluid mechanics and classical heat transfer, where the macroscopic laws such as the Navier–Stokes equations, the Fourier law, and so on, are valid. In this regime, the system is close to local thermodynamic equilibrium, a notion we shall make more precise shortly.
- If $K_n \gg 1$, the regime is described as rarefied. On spatial scales of the order of the system size, or over times of the order of the observation time, collisions between molecules play almost no role.

– If $K_n \sim 1$, we are in an intermediate regime for which it is very difficult to simplify the Boltzmann equation and a solution of the full equation is unavoidable.

The last two regimes are particularly important in micro and nanoscale heat transfer.

4.3 Local Thermodynamic Equilibrium

The distribution function $f^{(0)}$ appearing on the right-hand side of (12) is the one making the collision term equal to zero. To determine it, we seek the function f which makes the collision integral (11) equal to zero. We obtain the following expression (see [5] for details of the calculation):

$$f^{(0)}(\boldsymbol{r},\boldsymbol{v},t) = n(\boldsymbol{r},t)\left[\frac{m}{2\pi k_{\mathrm{B}} T(\boldsymbol{r},t)}\right]^{3/2} \exp\left\{-\frac{m\left[\boldsymbol{v}-\langle \boldsymbol{v}(\boldsymbol{r},t)\rangle\right]^2}{2k_{\mathrm{B}}T(\boldsymbol{r},t)}\right\}. \tag{20}$$

This distribution is formally the same as the full thermodynamic equilibrium distribution (8), except that it involves the density $n(\boldsymbol{r},t)$, the temperature $T(\boldsymbol{r},t)$, and the average speed $\langle \boldsymbol{v}(\boldsymbol{r},t)\rangle$, which are local and instantaneous quantities. They are in fact slowly varying functions of $\boldsymbol{r}$ and t, varying on the macroscopic scale defined by the volume element $\mathrm{d}^3 r$. The distribution (20) is called the local thermodynamic equilibrium function (LTE).

At this point we observe that:

– The LTE corresponds to a non-equilibrium situation imposed by external constraints, but in which each volume element $\mathrm{d}^3 r$ is at every moment of time in an equilibrium characterised by the temperature $T(\boldsymbol{r},t)$ and the density $n(\boldsymbol{r},t)$ that are generally different from the values in a neighbouring volume element.
– The LTE distribution $f^{(0)}$ makes the collision term equal to zero, but it is not a solution of the Boltzmann equation. Indeed, we find that

$$\frac{\partial f^{(0)}}{\partial t} + \boldsymbol{v}\cdot\frac{\partial f^{(0)}}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m}\cdot\frac{\partial f^{(0)}}{\partial \boldsymbol{v}} \neq 0\,.$$

For systems in the collisional regime ($K_n \ll 1$), we may seek a solution of the Boltzmann equation which represents a small correction with respect to the LTE distribution. To do this, we introduce an expansion of the solution in powers of K_n. For systems close to LTE such as those encountered in fluid mechanics or classical heat transfer problems, the first order correction is generally good enough to calculate the fluxes.

4.4 Perturbation Method. Linear Response

The perturbation method consists in seeking a solution f of the Boltzmann equation in the form

$$f(\boldsymbol{r},\boldsymbol{v},t) = f^{(0)}(\boldsymbol{r},\boldsymbol{v},t) + f^{(1)}(\boldsymbol{r},\boldsymbol{v},t) + f^{(2)}(\boldsymbol{r},\boldsymbol{v},t) + \cdots , \tag{21}$$

where $f^{(i)}$ is order i in K_n. Keeping only terms to first order, we have

$$\begin{aligned} &f(\boldsymbol{r},\boldsymbol{v},t) = f^{(0)}(\boldsymbol{r},\boldsymbol{v},t) + f^{(1)}(\boldsymbol{r},\boldsymbol{v},t)\,, \quad \text{where} \\ &f^{(1)}(\boldsymbol{r},\boldsymbol{v},t) \ll f^{(0)}(\boldsymbol{r},\boldsymbol{v},t)\,. \end{aligned} \tag{22}$$

Substituting this expansion into the Boltzmann equation (12) and keeping only first order terms in K_n, we obtain

$$f^{(1)}(\boldsymbol{r},\boldsymbol{v},t) = -\tau(\boldsymbol{v})\left[\frac{\partial f^{(0)}}{\partial t} + \boldsymbol{v}\cdot\frac{\partial f^{(0)}}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m}\cdot\frac{\partial f^{(0)}}{\partial \boldsymbol{v}}\right]. \tag{23}$$

This first order solution can now be used to calculate the diffusion fluxes by means of (6). We shall illustrate the method by working out the conductive flux in a gas, which will lead to the Fourier law.

4.5 Fourier Law and Thermal Conductivity

Consider a gas in the stationary regime, with uniform pressure and zero hydrodynamic velocity, so that the gas is at rest on the macroscopic scale. Suppose also that there is no external force field.

To simplify the notation, we write $f^{(1)} = f^{(0)}\eta$, where $\eta \ll 1$. Then by (23), we have

$$\eta = -\tau\frac{1}{f^{(0)}}\boldsymbol{v}\cdot\frac{\partial f^{(0)}}{\partial \boldsymbol{r}} = -\tau\boldsymbol{v}\cdot\frac{\partial \ln f^{(0)}}{\partial \boldsymbol{r}}\,. \tag{24}$$

Since the function $f^{(0)}$ only depends on $\boldsymbol{r}$ through $n(\boldsymbol{r})$ and $T(\boldsymbol{r})$, we have from (20),

$$\eta = -\tau\boldsymbol{v}\cdot\left(\frac{1}{n}\frac{\partial n}{\partial \boldsymbol{r}} - \frac{3}{2T}\frac{\partial T}{\partial \boldsymbol{r}} + \frac{mv^2}{2k_{\mathrm{B}}T^2}\frac{\partial T}{\partial \boldsymbol{r}}\right). \tag{25}$$

The pressure is uniform and in each volume element in LTE we have $P = n(\boldsymbol{r})k_{\mathrm{B}}T(\boldsymbol{r})$. This implies that

$$\frac{1}{n}\frac{\partial n}{\partial \boldsymbol{r}} = -\frac{1}{T}\frac{\partial T}{\partial \boldsymbol{r}}\,.$$

Hence, (25) simplifies to

$$\eta = -\tau\left(\frac{5}{2} - \frac{mv^2}{2k_{\mathrm{B}}T}\right)\boldsymbol{v}\cdot\frac{1}{T}\frac{\partial T}{\partial \boldsymbol{r}}\,. \tag{26}$$

Having established an expression for η, we may now calculate the conductive flux. Its component in direction j is

$$\phi_j = \int \frac{1}{2} m v^2 f^{(0)} (1+\eta) v_j \, \mathrm{d}^3 v = \int \frac{1}{2} m v^2 f^{(0)} \eta v_j \, \mathrm{d}^3 v \,, \tag{27}$$

because the contribution to the integral from the term containing just $f^{(0)}$ is zero, this function having the same form as the perfect equilibrium distribution function f^0 (and the flux is zero in perfect equilibrium). Substituting (26) in (27), we obtain

$$\phi_j = \frac{nm\tau}{2} \left(\frac{n}{2\pi k_\mathrm{B} T} \right)^{3/2} \int v^2 v_j \left(\frac{5}{2} - \frac{m v^2}{2 k_\mathrm{B} T} \right) \times \exp\left(- \frac{m v^2}{2 k_\mathrm{B} T} \right) v_i \frac{1}{T} \frac{\partial T}{\partial x_i} \, \mathrm{d}^3 v \,, \tag{28}$$

where we have used the Einstein summation convention (summing over any repeated indices). Only the term $i = j$ gives a nonzero contribution. Finally, we obtain

$$\phi_j = -\frac{5}{2} \frac{n k_\mathrm{B}^2 T \tau}{m} \frac{\partial T}{\partial x_j} \,. \tag{29}$$

This expression shows that the conductive flux is proportional to the temperature gradient. We thus retrieve the Fourier law, demonstrated here for a system close to LTE (collisional regime, $K_n \ll 1$). We may also identify the thermal conductivity

$$\kappa = \frac{5}{2} \frac{n k_\mathrm{B}^2 T \tau}{m} = \frac{5}{2} \frac{k_\mathrm{B} P \tau}{m} \,. \tag{30}$$

Note that, in this expression, there is an implicit temperature dependence in the collision time τ, depending on the interaction potential. For example, for a hard sphere potential, we have $\tau = 1/n\pi a^2 v$, where $v \propto \sqrt{T}$, which gives $\kappa \propto \sqrt{T}$.

5 Example of a Non-LTE System. Short-Scale Conduction in a Gas

One of the difficulties in modelling micro and nanoheat transfer lies in the fact that the Knudsen number satisfies $K_n \sim 1$, or even $K_n \gg 1$. One cannot therefore apply the above perturbation method. In a system of size L shorter than the mean free path, e.g., a 10 nm cavity containing a gas in normal conditions, particles in ballistic flight from one wall to another play a key role. For these particles, the particle–particle collision term has little effect,

and the distribution function evolves within the system according to (9). It is therefore essential to take into account the boundary conditions at the walls, i.e., collisions with the walls.

It is also interesting to note that *Boltzmann* wrote as early as 1895 [4]:

> At constant temperature, the theory shows that the mean free path is inversely proportional to the density, and experiment rigorously confirms this. Peculiar phenomena should thus arise when the size of the container is comparable with the mean free path [...]. At low pressures and in rather small containers [...] other properties are involved, precisely when the distance between unequally heated regions has become a small multiple of the mean free path, superposing or partially substituting a direct form of transport by bombardment from one wall to the other on the energy transport by diffusion or circulation between neighbouring regions.

Before presenting an example of transfer in rarefied media, we shall first discuss the idea of temperature on short length and time scales.

5.1 Can One Speak of Temperature on Short Scales?

When we wish to speak of micro and nanoscale heat transfer, the question inevitably arises as to whether one can give meaning to the idea of temperature on short space and time scales. This question can be tackled on the basis of the discussion given at the end of Sect. 1.1.

The distribution function f can be defined on any length scale that is large compared with the molecular size a, i.e., the range of the interaction potential, and on any time scale that is long compared with the duration τ_0 of a collision. Once f has been defined, it is always possible to define and calculate the average value of a quantity using (2). For example, we may define the average kinetic energy per unit volume (internal energy) of the gas at a point $\boldsymbol{r}$ and time t by

$$\langle E(\boldsymbol{r},t)\rangle = \int f(\boldsymbol{r},\boldsymbol{v},t)\frac{1}{2}mv^2\,\mathrm{d}^3v\,, \tag{31}$$

assuming for simplicity that the hydrodynamic velocity is zero. If the system has characteristic size $L < l$, where l is the mean free path, the distribution function is defined on a volume element d^3r whose characteristic size, which we shall write as $(\mathrm{d}^3r)^{1/3}$, satisfies $a \ll (\mathrm{d}^3r)^{1/3} \ll L < l$. According to the discussion in Sect. 1.1, the distribution function f then fluctuates over a length scale that is small compared with the mean free path l, and over a time scale that is small compared with the collision time τ. The same is then true of the average kinetic energy $\langle E(\boldsymbol{r},t)\rangle$.

For a dilute gas, we can always give a *formal* definition of the local, instantaneous temperature by the relation

$$\langle E(\boldsymbol{r},t)\rangle = \tfrac{3}{2}n(\boldsymbol{r},t)k_{\mathrm{B}}T(\boldsymbol{r},t)\,. \tag{32}$$

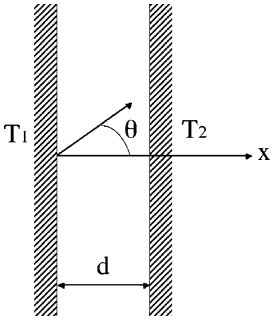

Fig. 1. Calculating the conductive flux in the ballistic regime. The two flat walls are assumed infinite and their separation d is much smaller than the mean free path of the gas molecules

The quantity $T(\boldsymbol{r}, t)$ defined in this way is an average quantity in the sense of (31), but one which fluctuates spatially on a length scale that is small compared with the mean free path l, i.e., on a length scale that is small compared with 10–100 nm for a gas in standard conditions, and on a time scale that is small compared with the collision time τ, which is of the order of 10^{-9} s. It is then difficult to attribute any real practical interest to this quantity, although it can always be defined formally in this way.

5.2 Calculating the Conductive Flux in the Ballistic Regime

To illustrate the calculation of the conductive flux in a gas in the rarefied regime, we consider the arrangement shown in Fig. 1. Two flat walls are separated by a distince d and held at slightly different temperatures T_1 and T_2. The gas is taken to be nitrogen (mass of one molecule 10^{-26} kg), with density $n = 2.5 \times 10^{25}\,\mathrm{m}^{-3}$, which is correct for a gas at equilibrium at atmospheric pressure and temperature $T = 300\,\mathrm{K}$. We assume that $T_1 = 320\,\mathrm{K}$ and $T_2 = 300\,\mathrm{K}$, and also that the cavity has nanometric dimensions $d = 5\,\mathrm{nm}$. Under these conditions, the mean free path is $l = 40\,\mathrm{nm}$, whence the Knudsen number is $K_n = l/d \gg 1$. We are thus in the rarefied gas regime and energy transfer from one wall to the other occurs mainly by ballistic flight of the molecules. In the gas, the distribution function thus evolves according to (9).

The main part of the problem is to establish the boundary conditions on the walls, which alone determine the value of the distribution function at every point. One thus makes the following hypotheses:

– Particles moving towards $x > 0$ had their last collision with the wall at temperature T_1. Assuming that one collision is enough to get the molecule into equilibrium with the wall,[1] these molecules are therefore characterised by a distribution function f^+ which is an equilibrium distribution having the form of (8) for the temperature T_1. Hence, $f^+ = A \exp(-mv^2/2k_\mathrm{B}T_1)$, where A is a constant to be determined (in particular, because we do not know the density of these molecules).

[1] This hypothesis can be improved by introducing an accommodation factor at the wall. See, for example [6].

– Likewise, particles moving twoards $x < 0$ had their last collision with the wall at temperature T_2 and are characterised by an equilibrium distribution function $f^- = B \exp(-mv^2/2k_\mathrm{B}T_2)$, where B is another constant to be determined.

The two constants A and B are found by expressing the fact that the total particle density is n, and also the fact that the particle flux is zero at every point, i.e., there is no average mass transfer. The calculation is detailed in Appendix C. We find

$$A = 2n\frac{\sqrt{T_2}}{\sqrt{T_1}+\sqrt{T_2}}\left(\frac{m}{2\pi k_\mathrm{B}T_1}\right)^{3/2}, \quad B = 2n\frac{\sqrt{T_1}}{\sqrt{T_1}+\sqrt{T_2}}\left(\frac{m}{2\pi k_\mathrm{B}T_2}\right)^{3/2}. \tag{33}$$

The conductive flux in direction Ox is thus

$$\phi_\mathrm{cd} = \int \frac{1}{2} m v^2 f(\boldsymbol{v}) v_x \,\mathrm{d}^3 v\,, \tag{34}$$

where $f = f^+$ if $v_x > 0$ and $f = f^-$ if $v_x < 0$. Substituting the expressions for f^+ and f^- into (34), we obtain

$$\phi_\mathrm{cd} = \frac{m}{2}\int_0^{+\infty} v^5 f^+ \,\mathrm{d}v \int_0^{\pi/2} 2\pi \sin\theta\cos\theta\,\mathrm{d}\theta + \frac{m}{2}\int_0^{+\infty} v^5 f^- \,\mathrm{d}v \int_{\pi/2}^{\pi} 2\pi \sin\theta\cos\theta\,\mathrm{d}\theta\,. \tag{35}$$

Finally, we obtain

$$\phi_\mathrm{cd} = \frac{n\sqrt{T_1T_2}(2k_\mathrm{B})^{3/2}}{\sqrt{\pi m}(\sqrt{T_1}+\sqrt{T_2})}(T_1 - T_2)\,. \tag{36}$$

We observe that the flux is independent of the distance between the walls, provided that the ballistic hypothesis remains valid. Moreover, when $T_1 \sim T_2 = T$, we may define an equivalent exchange coefficient h_eq in the way usually done in convection studies. We obtain

$$h_\mathrm{eq} = \frac{nT(2k_\mathrm{B})^{3/2}}{2\sqrt{\pi m T}}\,. \tag{37}$$

For nitrogen, with the numerical values specified above, we obtain $h_\mathrm{eq} \approx 1.3 \times 10^5\,\mathrm{W}\cdot\mathrm{m}^{-2}\cdot\mathrm{K}^{-1}$.

5.3 Transitions Between Regimes

To end this discussion, it is interesting to determine some orders of magnitude of typical dimensions and exchange coefficients characterising the various

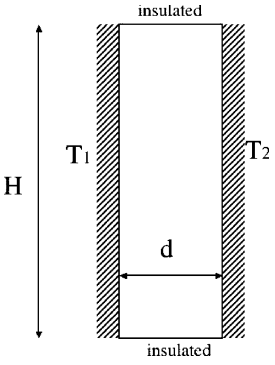

Fig. 2. Geometry of the cavity used to examine transitions between regimes

transfer regimes in a flat gas film: convective transfer, diffusive conductive transfer (Fourier law), semi-ballistic transfer, and pure ballistic transfer. To do this, we shall examine the flux transferred between the two vertical walls of a long cavity, like the one shown in Fig. 2. We assume that the cavity is filled with air at atmospheric pressure. The left-hand wall is held at temperature $T_1 = 320\,\mathrm{K}$ and the right-hand wall at temperature $T_2 = 300\,\mathrm{K}$.

Convective Regime

For concreteness, let us take the dimensions of the cavity to be $H = 1\,\mathrm{m}$ and $d = 10\,\mathrm{cm}$. The air contained within the cavity undergoes natural convection. The Rayleigh number associated with the size d is

$$Ra_d = \frac{g\beta(T_1 - T_2)d^3}{\alpha\nu}\,, \tag{38}$$

where g is the acceleration due to gravity, β is the thermal expansion coefficient ($\beta = 1/T$ for a perfect gas), α is the thermal diffusivity of air, and ν is the viscosity, all evaluated at temperature $T_\mathrm{m} = (T_1 + T_2)/2$. This gives $Ra_d = 1.75 \times 10^6$. The flux transferred through the fluid is then [7]

$$\phi_\mathrm{cv} = \frac{\lambda_\mathrm{eq}}{d}(T_1 - T_2)\,. \tag{39}$$

The apparent conductivity λ_eq is given in terms of the Rayleigh number and the aspect ratio $A = H/d$ of the cavity by [7]

$$\frac{\lambda_\mathrm{eq}}{\lambda} = 0.073 Ra_d^{1/3} A^{-1/9}\,, \tag{40}$$

where λ is the conductivity of the fluid at rest at temperature T_m. We obtain here $\lambda_\mathrm{eq} = 0.2\,\mathrm{W\cdot m^{-1}\cdot K^{-1}}$. One can introduce a convection transfer coefficient $h = \lambda_\mathrm{eq}/d$, which is equal to $2\,\mathrm{W\cdot m^{-2}\cdot K^{-1}}$ in the present case. This value of h, of the order of a few $\mathrm{W\cdot m^{-2}\cdot K^{-1}}$, is typical for natural convection transfer in gases under standard conditions.

Diffusive Conductive Regime

If we reduce the cavity size so that the dimensions are now $H = 1\,\mathrm{cm}$ and $d = 1\,\mathrm{mm}$ (the aspect ratio therefore remaining the same), the Rayleigh number becomes $Ra_d = 1.75$. We are now in the pure conduction regime, i.e., there is no macroscopic fluid motion and natural convection cannot develop [7]. This is the regime one finds between the two panes of a double-glazed window panel. The equivalent transfer coefficient is then $h = \lambda/d = 30\,\mathrm{W}\cdot\mathrm{m}^{-2}\cdot\mathrm{K}^{-1}$.

The regime considered here is collisional ($K_n \ll 1$). Transport is diffusive and obeys the Fourier law. It is worth asking how big d can be for such a regime, and for the above expression for h to remain valid. We know that the Fourier law can be demonstrated for $K_n \ll 1$, i.e., provided that d is large compared with the mean free path l of the molecules in the gas. Since $l = 40\,\mathrm{nm}$ for a gas at atmospheric pressure and standard temperatures, the diffusive regime exists at least up to sizes $d \sim 400\,\mathrm{nm}$. For this size limit, the coefficient h already assumes a very high value, viz., $h = 7.5 \times 10^4\,\mathrm{W}\cdot\mathrm{m}^{-2}\cdot\mathrm{K}^{-1}$.

Semi-Ballistic and Ballistic Regimes

When the system size is further reduced, e.g., for a cavity with H a few mm and $d = 5\,\mathrm{nm}$, we enter a regime where ballistic transport begins to play a significant role.

When $K_n \sim 1$, the transport is semi-ballistic. When the molecules go from one wall to another, they are involved in at most a small number of collisions. A precise assessment of the flux transferred requires full solution of the Boltzmann equation.

When $K_n > 1$, the transfer becomes largely ballistic and the flux is given to a good approximation by (36). In particular, it becomes independent of the distance d and the order of magnitude of the transfer coefficient is $h \sim 10^5\,\mathrm{W}\cdot\mathrm{m}^{-2}\cdot\mathrm{K}^{-1}$.

Finally, it is important to note that:

- The mean free path is inversely proportional to the density. Depending on the density of the gas, the ballistic regime is reached for different values of the system size.
- Equation (37) shows that the transfer coefficient in the ballistic regime varies linearly with the gas density. At very short distances, a modification in the gas density (whilst remaining within the ballistic regime) can significantly change the level of flux transferred by conduction.

6 Conclusion

In this Chapter we have introduced the basic formalism of the Boltzmann equation. We have discussed situations close to LTE, where it is possible to

carry out perturbative calculations of fluxes and transport coefficients. We have also examined non-LTE situations such as are frequently encountered in micro and nanoscale heat transfer. We have gone into some detail concerning the example of ballistic conduction in a gas and given orders of magnitude of the equivalent conductivities (or equivalent exchange coefficients) in the various transport regimes (ballistic, semi-ballistic and diffusive). All these methods and all these ideas extend to the description of electron and phonon transport in solids.

A Equilibrium Distribution Function

Consider a monatomic gas consisting of N particles in a volume V, in equilibrium with a thermostat at temperature T. The average number of particles with positions in a volume element d^3r centered on the point $\boldsymbol{r}$ and with velocities in an element d^3v of the velocity space centered on $\boldsymbol{v}$ is

$$\mathrm{d}N = f^0\,\mathrm{d}^3r\,\mathrm{d}^3v\,, \tag{41}$$

where f^0 is the equilibrium distribution function.

The usual approach in equilibrium statistical physics is to express $\mathrm{d}N$ as the product of the number of quantum states (translational states here) accessible to a particle, written $\mathrm{d}N_{\mathrm{acc}}$, and the average number of particles $\bar{N}$ per state. For a classical gas, the average number of particles is given by the Maxwell–Boltzmann statistic

$$\bar{N} = N\frac{\exp(-mv^2/2k_{\mathrm{B}}T)}{\zeta}\,, \tag{42}$$

where ζ is the one-particle partition function calculated in the classical limit (of large quantum numbers):

$$\begin{aligned}\zeta &= \sum_{\text{states } p}\exp\left(-\frac{E_p}{k_{\mathrm{B}}T}\right) = \int_0^{+\infty}\exp\left(-\frac{p^2}{2mk_{\mathrm{B}}T}\right)\frac{4\pi p^2V}{h^3}\,\mathrm{d}p\\ &= \frac{V}{h^3}(2\pi mk_{\mathrm{B}}T)^{3/2}\,,\end{aligned} \tag{43}$$

where h is Planck's constant. In the classical approximation, the number of accessible states can be written

$$\mathrm{d}N_{\mathrm{acc}} = \frac{\mathrm{d}^3r m^3\,\mathrm{d}^3v}{h^3}\,. \tag{44}$$

Hence, finally,

$$\mathrm{d}N = m^3\frac{N}{V}\exp\left(-\frac{mv^2}{2k_{\mathrm{B}}T}\right)\frac{1}{(2\pi mk_{\mathrm{B}}T)^{3/2}}\,\mathrm{d}^3r\,\mathrm{d}^3v = f^0\,\mathrm{d}^3r\,\mathrm{d}^3v\,, \tag{45}$$

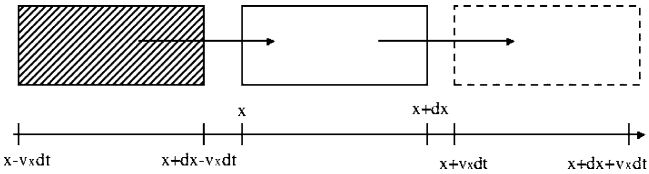

Fig. 3. Schematic representation of the volume over which we calculate the number of particles with given velocity $\boldsymbol{v}$. Due to the velocity component in the x direction, the particles which occupied the central volume element at time t will occupy the *dotted volume element* at time $t+\mathrm{d}t$. The particles which occupy the *central volume element* at time $t+\mathrm{d}t$ are those which occupied the *shaded element* at time t

and we may identify

$$f^0 = n\left(\frac{m}{2\pi k_\mathrm{B}T}\right)^{3/2}\exp\left(-\frac{mv^2}{2k_\mathrm{B}T}\right), \tag{46}$$

where $n = N/V$ is the number of particles per unit volume.

B Dynamical Evolution of the Distribution Function for Free Particles

In this appendix, we shall demonstrate (9). Demonstrations based on other approaches can be found in [1, 4, 5]. The proof given here is inspired by the discussion in [8].

For a gas of non-interacting particles, consider those particles with well-defined velocity lying in the range from $\boldsymbol{v}$ to $\boldsymbol{v}+\mathrm{d}^3v$. Fix a spatial volume element d^3r and consider among these particles all those with positions between $\boldsymbol{r}$ and $\boldsymbol{r}+\mathrm{d}^3r$. At time t, we denote the number of such particles by $\mathrm{d}n(\boldsymbol{r},\boldsymbol{v},t)$. We shall now ascertain how this number varies between times t and $t+\mathrm{d}t$, keeping only first order contributions.

Consider first the Ox direction. At time t, the $\mathrm{d}n$ particles occupy the parallelepiped defined by $(x, x+\mathrm{d}x, y, y+\mathrm{d}y, z, z+\mathrm{d}z)$ (see Fig. 3). At time $t+\mathrm{d}t$:

– these same molecules occupy the volume element defined by

$$(x + v_x\,\mathrm{d}t, x + \mathrm{d}x + v_x\,\mathrm{d}t, y, y+\mathrm{d}y, z, z+\mathrm{d}z)\,,$$

– the molecules occupying the volume element $(x, x+\mathrm{d}x, y, y+\mathrm{d}y, z, z+\mathrm{d}z)$ are those which occupied the volume element

$$(x - v_x\,\mathrm{d}t, x + \mathrm{d}x - v_x\,\mathrm{d}t, y, y+\mathrm{d}y, z, z+\mathrm{d}z)\,,$$

at time t. The number of such molecules is $\mathrm{d}n(x - v_x\,\mathrm{d}t, y, z, \boldsymbol{v}, t)$.

The change in the number of molecules with velocity in the range from $\boldsymbol{v}$ to $\boldsymbol{v} + \mathrm{d}^3 v$ and with position between $\boldsymbol{r}$ and $\boldsymbol{r} + \mathrm{d}^3 r$, due to the velocity component in the x direction, is therefore

$$\begin{aligned} &\mathrm{d}n(x - v_x\,\mathrm{d}t, y, z, \boldsymbol{v}, t) - \mathrm{d}n(x, y, z, \boldsymbol{v}, t) \\ &\qquad = \Big[f(x - v_x\,\mathrm{d}t, y, z, \boldsymbol{v}, t) - f(x, y, z, \boldsymbol{v}, t)\Big]\,\mathrm{d}^3 r\,\mathrm{d}^3 v \\ &\qquad = -v_x\,\mathrm{d}t \frac{\partial f}{\partial x}\,\mathrm{d}^3 r\,\mathrm{d}^3 v\,. \end{aligned} \tag{47}$$

This result can be generalised, first by considering all three directions Ox, Oy, and Oz, and secondly, by observing that in the presence of an external force field, $\boldsymbol{F}\,\mathrm{d}t/m$ plays the same role for the velocity $\boldsymbol{v}$ as $\boldsymbol{v}\,\mathrm{d}t$ for the position $\boldsymbol{r}$. In one case we examine molecules with a given velocity and seek the variation in their number within a volume element $\mathrm{d}^3 r$, whilst in the other case we are concerned with molecules having a given position and we seek the variation in their number in a volume element $\mathrm{d}^3 v$ of the velocity space, under the action of the force $\boldsymbol{F}$. Finally, we obtain

$$\begin{aligned} \mathrm{d}n(\boldsymbol{r}, \boldsymbol{v}, t + \mathrm{d}t) - \mathrm{d}n(\boldsymbol{r}, \boldsymbol{v}, t) &= \frac{\partial f}{\partial t}\,\mathrm{d}^3 r\,\mathrm{d}^3 v\,\mathrm{d}t \\ &= -\boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}}\,\mathrm{d}^3 r\,\mathrm{d}^3 v\,\mathrm{d}t - \frac{\boldsymbol{F}}{m} \cdot \frac{\partial f}{\partial \boldsymbol{v}}\,\mathrm{d}^3 r\,\mathrm{d}^3 v\,\mathrm{d}t\,, \end{aligned}$$

whence we deduce the required result

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}}{m} \cdot \frac{\partial f}{\partial \boldsymbol{v}} = 0\,. \tag{48}$$

C Calculating the Constants A and B for the Flux in the Ballistic Regime

To find the constants A and B in (33), we first express the fact that the particle density is known and equal to n:

$$n = \int f(\boldsymbol{v})\,\mathrm{d}^3 v\,, \tag{49}$$

where $f = f^+$ if $v_x > 0$ and $f = f^-$ if $v_x < 0$. In spherical coordinates, we have $\mathrm{d}^3 v = 2\pi \sin\theta v^2\,\mathrm{d}\theta\,\mathrm{d}v$, whence

$$\begin{aligned} n = A \int_0^{+\infty} v^2 \exp\left(-\frac{mv^2}{2k_\mathrm{B}T_1}\right)\mathrm{d}v \int_0^{\pi/2} 2\pi \sin\theta\;\mathrm{d}\theta \\ + B \int_0^{+\infty} v^2 \exp\left(-\frac{mv^2}{2k_\mathrm{B}T_2}\right)\mathrm{d}v \int_{\pi/2}^{\pi} 2\pi \sin\theta\,\mathrm{d}\theta\,, \end{aligned} \tag{50}$$

which eventually yields

$$n = \frac{A}{2}\left(\frac{2\pi k_{\mathrm{B}} T_1}{m}\right)^{3/2} + \frac{B}{2}\left(\frac{2\pi k_{\mathrm{B}} T_2}{m}\right)^{3/2} . \tag{51}$$

To obtain a second equation relating A and B, we express the fact that the particle flux in the direction Ox is zero at every point (no macroscopic matter transfer):

$$q_m = \int f(\boldsymbol{v}) v_x \, \mathrm{d}^3 v = 0 \, , \tag{52}$$

where $f = f^+$ if $v_x > 0$ and $f = f^-$ if $v_x < 0$. We thus obtain

$$\begin{aligned} 0 = A \int_0^{+\infty} v^3 \exp\left(-\frac{mv^2}{2k_{\mathrm{B}} T_1}\right) \mathrm{d}v \int_0^{\pi/2} 2\pi \sin\theta \cos\theta \, \mathrm{d}\theta \\ + B \int_0^{+\infty} v^3 \exp\left(-\frac{mv^2}{2k_{\mathrm{B}} T_2}\right) \mathrm{d}v \int_{\pi/2}^{\pi} 2\pi \sin\theta \cos\theta \, \mathrm{d}\theta \, , \end{aligned} \tag{53}$$

which yields finally,

$$A T_1^2 = B T_2^2 \, . \tag{54}$$

The two equations (51) and (54) completely determine A and B, leading to the expressions in (33).

References

[1] L. D. Landau, E. M. Lifshitz: *Cinétique Physique*, vol. 10, Cours de Physique Théorique (Mir, Moscow 1990) 16, 32

[2] L. D. Landau, E. M. Lifshitz: *Statistical Physics*, vol. 5, part I, Course of Theoretical Physics (Pergamon Press, Oxford 1980) 17

[3] H. B. Callen: *Thermodynamics* (Wiley, New York 1985) 18

[4] L. Boltzmann: *Leçons sur la Théorie des Gaz* (Editions J. Gabay, Paris 1987) reprint of the French translation (1902) of the original work by Boltzmann (1895) 20, 26, 32

[5] J. H. Ferziger, H. G. Kaper: *Mathematical Theory of Transport Processes in Gases* (North Holland, Amsterdam 1972) 20, 23, 32

[6] R. J. Coruccini: Vacuum **7**–**8**, 19 (1959) 27

[7] J. Taine, J.-P. Petit: *Heat Transfer* (Prentice Hall 1993) section entitled *Basic Data* 29, 30

[8] B. Diu, G. Guthmann, D. Lederer, B. Roulet: *Physique Statistique* (Hermann, Paris 1989) 32

Index

Electrons and Phonons

Jean-Jacques Greffet

Ecole Centrale Paris, Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion (EM2C), Centre National de la Recherche Scientifique, 92295 Châtenay-Malabry Cedex
greffet@em2c.ecp.fr

Abstract. The aim of this Chapter is not to produce a complete course on electrons and phonons, but rather to give a concise outline of a certain number of their basic properties in order to be able to describe transport phenomena in crystalline solids. Indeed, transport phenomena will provide the guiding motivation.

Transport phenomena in dilute media were discussed with the help of the Boltzmann equation in the Chapter on Transport in Dilute Media by *Carminati* in this volume, in the context of the kinetic approach. The idea is to establish an expression for the particle flux through a given area. We then deduce the flux of any quantity transported by each particle. The Boltzmann equation is used to determine the velocity distribution of the particles. It can be solved fairly straightforwardly if the system is close to equilibrium. We thus introduce the idea of local thermodynamic equilibrium (LTE). As we have seen, this notion can only be defined for length and time scales greater than the mean free path and the average time between consecutive collisions, respectively.

In this Chapter, we explain how to transpose this kinetic approach[1] to the case of electrons and phonons. The first step is to define the fluxes. The second is to obtain the counterpart of the velocity distribution function. A difficulty arises because we can no longer apply classical mechanics. The system is described using the wave functions of quantum mechanics. We must first revise the notion of flux using the language of waves. It is no longer useful to introduce the particle aspect when expressing the fluxes in this context. The next step is to find the counterpart of the velocity distribution function. The velocity is not an observable for an electron in a crystal. What plays the role of the velocity distribution function is the average occupation number of a state. At equilibrium, this is given by the Fermi–Dirac distribution. We are still in the framework of the wave description, since the stationary states are described by wave functions. However, it is the Boltzmann equation which provides a way of determining the correction required to take into account an imbalance due to the application of a temperature or potential gradient. This is done by returning to a particle view of electrons or phonons. The problem will thus be to see how to revert to a so-called semi-classical approach in terms of particles in order to describe these objects.

[1] The word 'kinetic' expresses the fact that transport is described by studying the motion of the particles that transport the relevant quantity, e.g., energy, charge, momentum, etc.

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 37–54 (2007)

1 Electrons

1.1 Free Electrons

Introduction

The model of the free electron can be used to describe electrons in the conduction band. It amounts to replacing the interactions between an electron and all other electrons and nuclei in the solid by a uniform potential that can be chosen equal to zero. It is thus assumed that the electrons are non-interacting. The state of the system is described by a wave function satisfying the Schrödinger equation

$$-\sum_{i=1}^{N}\frac{\hbar^2}{2m}\Delta\Phi(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)=E\Phi(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)\,. \tag{1}$$

Rather than studying the N-electron system, we shall limit the discussion here to the one-electron wave function. This simplification is possible because the wave function for an N-electron system can be written as a product of N elementary wave functions:

$$\Phi(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)=\Psi(\boldsymbol{r}_1)\Psi(\boldsymbol{r}_2)\ldots\Psi(\boldsymbol{r}_N)\,, \tag{2}$$

where $\Psi(\boldsymbol{r})$ is the solution of the time-independent one-electron Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\Psi=E\Psi\,. \tag{3}$$

As electrons are fermions, the wave function must change sign when two electrons are permuted. This is the antisymmetrisation principle. The linear combination of the one-electron wave functions with this property is called the Slater determinant. For most of the applications we shall be concerned with, it will not be essential to use the Slater determinant. Note in particular that, when we use Fermi–Dirac statistics, we always work with states describing a single particle. Note also that, in order to define the state of an electron, one must also specify its spin state. We omit the spin index in the following for notational simplicity.

Energy Spectrum and Wave Function

The energy spectrum of free electrons is obtained by seeking the solution of the time-independent Schrödinger equation. The solution has the form $\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})/\sqrt{V}$, where the denominator is a normalisation factor ensuring that the probability of finding the electron somewhere in the volume V of the crystal is unity. Note that each of these stationary states or eigenstates

gives a constant probability density equal to $1/V$. An electron described by a stationary state is thus delocalised. The energy of the electron is given by

$$E = \frac{\hbar^2 k^2}{2m} \,. \tag{4}$$

The dispersion relation between the energy and the wave vector is thus represented by a parabola.

It should be emphasised that the boundary conditions have not been explicitly taken into account in the above. The discussion here is therefore only valid for an infinitely extended system. In practice, the energy levels of a given finite system will be modified. This is the same as the well known effect in acoustics. In empty space, we have a dispersion relation of type $\Omega = vK$, where v is the speed of sound. In a closed container, a number of discrete modes arise, and the smaller the container, the higher the natural frequencies will become. In the same way, electronic states are discretised when the finite size of the crystal is taken into account.

Density of States

To derive the energy or the conductivity of a solid, we must consider the contribution of each electron and sum over all electrons. To do this, we have to know the number of states $\mathrm{d}n(E)$ with energy in the range from E to $E + \mathrm{d}E$. We introduce the density of states $g(E)$ with the relation

$$\mathrm{d}n(E) = g(E)\,\mathrm{d}E \,.$$

In order to count the states, we seek all possible solutions of the problem. The exponential solution that we introduced above is the general solution of the problem. Boundary conditions must be taken into account. In the case of a box-shaped crystal with sides L_1, L_2, and L_3, we require the wave function to be zero on the walls of the box. This leads to stationary waves of type $\sin(k_x x)\sin(k_y y)\sin(k_z z)$, with the condition $E = \hbar^2 k^2/2m$. Solving the problem in this way, we can then count the states. However, this gives stationary states. To study transport phenomena, it is useful to work with plane waves. Indeed, the flux associated with plane waves is nonzero, whereas the particle flux associated with a stationary wave is zero. The difficulty here is that we cannot find a solution of this type which satisfies the boundary conditions. To get round this problem, one usually works with periodic boundary conditions (Born–von Karman boundary conditions). We require the wave function at $x = 0$ to equal the wave function at $x = L_1$, the other end of the crystal. This maneuver allows one to introduce the finite size of the crystal whilst continuing to work with propagative solutions. This then requires $\exp(\mathrm{i}k_x L_1) = 1$, whence $k_x = 2\pi p/L_1$, where p is an integer. The possible values of the wave vector belong to a discrete set. For a one-dimensional

problem, we thus have a number $\mathrm{d}N(k_x)$ of states with wave vector in the range from k_x to $k_x + \mathrm{d}k_x$ given by

$$\mathrm{d}N(k_x) = 2\frac{\mathrm{d}k_x}{2\pi/L_1}\,, \tag{5}$$

where the extra factor of 2 arises due to spin degeneracy. In two dimensions, we have

$$\mathrm{d}N(k_x, k_y) = 2L_1L_2\frac{\mathrm{d}k_x\,\mathrm{d}k_y}{(2\pi)^2}\,, \tag{6}$$

and in three dimensions,

$$\mathrm{d}N(\boldsymbol{k}) = 2L_1L_2L_3\frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\,. \tag{7}$$

Note that the number of states per unit volume is therefore

$$\mathrm{d}n(\boldsymbol{k}) = \frac{\mathrm{d}N(\boldsymbol{k})}{V} = \frac{1}{4\pi^3}\,\mathrm{d}^3\boldsymbol{k}\,. \tag{8}$$

The density of states is thus uniform in the reciprocal space.

It should be noted that the interval between two consecutive states depends sensitively on the system size. As the size of the system decreases, the density of states falls quickly, or put another way, the separation between consecutive states increases. For example, it is now possible to fabricate semiconducting systems with sizes of a few tens of nanometers, where electrons are confined in one, two or three dimensions. The energy levels depend sensitively on the size in such systems.

We can now find an expression for the density of states $g(E)$. In the reciprocal space, consider the surface defined by $E(\boldsymbol{k}) = E_0$. Let $S(E)$ be its area. A neighbouring surface corresponding to the energy $E_0 + \mathrm{d}E_0$ is separated by a distance $\mathrm{d}k$ in the reciprocal space such that $\mathrm{d}E_0 = |\nabla_{\boldsymbol{k}}E(\boldsymbol{k})|\,\mathrm{d}k$. The volume $S(E)\,\mathrm{d}k$ of the reciprocal space contains a number of states per unit volume equal to $\mathrm{d}n = S(E)\,\mathrm{d}k/4\pi^3$. We can now deduce that the density of states is given by

$$\mathrm{d}n(E) = g(E)\,\mathrm{d}E = \frac{S(E)}{4\pi^3}\frac{\mathrm{d}E}{|\nabla_{\boldsymbol{k}}E(\boldsymbol{k})|}\,. \tag{9}$$

Localised Electrons

Although the stationary states are delocalised, it is also possible to describe a localised electron. To do so, we must construct a wave packet. The velocity associated with such a wave packet is the velocity of the envelope of the wave packet, known as the group velocity. It can be shown [1–3] to be equal to

$$\frac{1}{\hbar}\nabla_{\boldsymbol{k}}E(\boldsymbol{k})\,.$$

1.2 Electrons in a Periodic Potential

It can be shown that any solution of the Schrödinger equation with a periodic potential has the form of a Bloch wave [1–3]:

$$\Psi_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}) \,, \tag{10}$$

where $u_{\boldsymbol{k}}(\boldsymbol{r})$ is a periodic function with period $\boldsymbol{R}$ equal to a vector in the Bravais lattice of the crystal, and $\boldsymbol{k}$ belongs to the first Brillouin zone. The Bloch wave is normalized over a cell of the Bravais lattice.

When a periodic potential is taken into account, gaps are introduced into the energy spectrum. The last partially occupied band is called the conduction band. The band immediately below it is called the valence band. This has many important consequences for transport phenomena. The distortion of the dispersion relation near the edges of the Brillouin zone leads to two effects:

- the density of states $g(E)$ has a peak, the so-called van Hove singularity,
- the group velocity is zero at the edge of the Brillouin zone.

These two effects balance one another in the context of transport phenomena. Indeed, we have seen that the density of states goes as $1/|\nabla_{\boldsymbol{k}} E|$, whereas the group velocity goes as $|\nabla_{\boldsymbol{k}} E|$. We shall see later on that it is indeed this velocity that comes into expressions for fluxes.

1.3 Electrical Conduction

Particle Current Density

In this section, we shall establish the form of the electrical current density. We begin by recalling the expression for the particle current density $\boldsymbol{J}$, given by

$$\boldsymbol{J} = \frac{\hbar}{m} \Im \left(\Psi^* \boldsymbol{\nabla} \Psi \right) \,, \tag{11}$$

for a wave function Ψ. This result is deduced from the fact that the electron probability density is given by the squared modulus of the wave function. Since the particle number is conserved, there has to be a probability current density such that the continuity equation

$$\frac{\partial |\Psi|^2}{\partial t} + \operatorname{div} \boldsymbol{J} = 0 \tag{12}$$

is valid. This equation is easy to derive using the Schrödinger equation to calculate $\partial |\Psi|^2 / \partial t$.

Note that the expectation value of the particle current in a Bloch state is equal to the expectation value of the velocity operator. Indeed, the expectation value of the current operator in a Bloch state $\Psi_{\boldsymbol{k}}(\boldsymbol{r})$ is

$$\boldsymbol{J}_{\boldsymbol{k}} = \frac{\hbar}{m}\Im\left[\int \mathrm{d}^3\boldsymbol{r}\Psi^*_{\boldsymbol{k}}(\boldsymbol{r})\boldsymbol{\nabla}\Psi_{\boldsymbol{k}}(\boldsymbol{r})\right], \tag{13}$$

where the integral is taken over a Bravais cell. This result can be related to the expectation value of the velocity operator in this same state:

$$\begin{aligned}\langle\boldsymbol{v}\rangle_{\boldsymbol{k}} &= \left\langle\frac{\boldsymbol{P}}{m}\right\rangle_{\boldsymbol{k}}\\ &= \frac{\hbar}{\mathrm{i}m}\langle\boldsymbol{\nabla}\rangle_{\boldsymbol{k}}\\ &= \frac{\hbar}{\mathrm{i}m}\int \mathrm{d}^3\boldsymbol{r}\Psi^*_{\boldsymbol{k}}(\boldsymbol{r})\boldsymbol{\nabla}\Psi_{\boldsymbol{k}}(\boldsymbol{r})\,.\end{aligned} \tag{14}$$

Electric Current Density

To obtain the contribution to the electric current density from a stationary state $\Psi_{\boldsymbol{k}}$, we simply multiply the particle current density by the electron charge:

$$\boldsymbol{j}^{\mathrm{q}}_{\boldsymbol{k}} = -e\boldsymbol{J}_{\boldsymbol{k}} = -e\langle\boldsymbol{v}\rangle_{\boldsymbol{k}}\,. \tag{15}$$

To calculate the current density, we must now sum over all states. To do this, we use the density of states in the reciprocal space and integrate over the first Brillouin zone to give

$$\boldsymbol{j}^{\mathrm{q}} = \int \frac{\mathrm{d}^3\boldsymbol{k}}{4\pi^3}n^0(\boldsymbol{k})(-e)\langle\boldsymbol{v}\rangle_{\boldsymbol{k}}\,, \tag{16}$$

where $\mathrm{d}^3\boldsymbol{k}/4\pi^3$ is the density of states per unit volume and $n^0(\boldsymbol{k})$ is the average occupation number of each state given by equilibrium Fermi–Dirac statistics:

$$n^0(\boldsymbol{k}) = \frac{1}{\exp\frac{E(\boldsymbol{k})-\mu}{kT}+1}\,. \tag{17}$$

When all the states are occupied, we see that they can be associated in pairs with opposite velocities. This shows that a filled band does not conduct a current. In the same manner, at equilibrium, the states characterised by $\boldsymbol{k}$ and $-\boldsymbol{k}$ balance one another in a partly filled band. We must therefore calculate the perturbation of $n(\boldsymbol{k})$ out of equilibrium. We shall investigate two scenarios: the collisional regime described by Ohm's law and the ballistic regime.

1.4 Semi-Classical Approach

In order to calculate the average occupation number of the quantum states when the system is perturbed from equilibrium, e.g., by an electric field, we shall use the Boltzmann equation (10) in the Chapter on Transport in Dilute Media by *Carminati* in this volume. One of the key features of this approach is that one refers to particles localised in space, so that the argument does not strictly apply to electrons described by stationary states. Of course, the Bloch states are delocalised throughout the crystal. We shall show that the particle behaviour of an electron can be retrieved from a quantum description of the states. Having done so, we will be able to use the Boltzmann equation to calculate the electrical conductivity. The aim of the following discussion will be to summarise the main results from the semi-classical description of electrons. We shall then ask to what extent an electron can be represented by a point particle of mass m with motion described by classical mechanics. There are two issues here: the validity of a description in which a particle has a well defined position, and the validity of the equations of classical mechanics.

Spatial Extent of an Electron

The first point to consider concerns the localisation of an electron. We have seen that the idea of a wave packet can be used to move from a wave description in which the electron is delocalised to a situation in which the electron is localised in space. However, if we wish the energy and momentum of the electron to be well defined, the width of the wave packet in the reciprocal space must be small compared with the size of the first Brillouin zone. This amounts to saying that the width Δk_x of the wave packet must be very small compared with $2\pi/a$, where a is the lattice parameter of the crystal. Moreover, a wave packet of width Δk_x in the reciprocal space has a width $\Delta x \geq 2\pi/\Delta k_x$ in the direct space. This follows from the properties of the Fourier transform. Finally, we see that the spatial extent Δx of the wave packet must be much greater than the size of a unit cell in the crystal lattice. We can only therefore treat the electron as a point particle in systems for which all characteristic length scales are much greater than the lattice parameter.

Electron Dynamics

The second point concerns the validity of the classical equations of motion. To answer this question, one must use the Ehrenfest theorem [3]:

$$\frac{\mathrm{d}\langle \boldsymbol{R} \rangle}{\mathrm{d}t} = \frac{\langle \boldsymbol{P} \rangle}{m}\,, \qquad \frac{\mathrm{d}\langle \boldsymbol{P} \rangle}{\mathrm{d}t} = -\langle \boldsymbol{\nabla} V \rangle\,, \tag{18}$$

where $\langle \boldsymbol{R} \rangle$ and $\langle \boldsymbol{P} \rangle$ are the expectation values of the position and momentum operators when the electron is in a specified stationary state. These

equations have precisely the same structure as the classical equations of motion. However, there is one significant difference: they refer to the average value of the potential gradient. One needs the potential gradient evaluated at $\langle \boldsymbol{R} \rangle$ to retrieve the classical equation. If the extent of the wave packet is very small compared with the length scale over which the potential varies, the classical approximation will be justified. This is generally the case for the type of electrical potential one might apply to a solid, which would vary over length scales of a few microns to a few meters. In contrast, the spatial extent of the wave packet is much greater than the lattice parameter, and this is the length scale associated with the periodic potential due to the crystal lattice. There are therefore internal forces which cannot be taken into account in the context of the classical approximation. We shall see that these forces can be accounted for by introducing the idea of an effective mass.

We shall now give the main results justifying a classical treatment of electron motion. To begin with, it can be shown that the expectation value $\langle \boldsymbol{v} \rangle_{\boldsymbol{k}}$ of the velocity operator in a Bloch state:

$$\langle \boldsymbol{v} \rangle_{\boldsymbol{k}} = \frac{1}{\hbar} \boldsymbol{\nabla}_{\boldsymbol{k}} E(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}} \omega(\boldsymbol{k}) \,. \tag{19}$$

In other words, the expectation value of the velocity in a Bloch state happens to equal the group velocity. Ehrenfest's theorem shows that it is this value that corresponds to the notion of velocity in classical mechanics, i.e., the ratio of momentum to mass.

A second key result is the link between the central value of a wave packet describing an electron and the force applied to this electron. Let $\boldsymbol{k}_0$ be the central position of the wave packet in the reciprocal space. This value is time dependent because the forces applied to the electron modify its energy. Over a time $\mathrm{d}t$, a force $\boldsymbol{F}$ supplies energy $\mathrm{d}E = \boldsymbol{F} \cdot \boldsymbol{v} \, \mathrm{d}t$ to the electron. The electron thus acquires an energy $\mathrm{d}E = \boldsymbol{\nabla}_{\boldsymbol{k}} E(\boldsymbol{k}) \cdot \mathrm{d}\boldsymbol{k}_0$. Noting that the electron velocity is given by $\boldsymbol{\nabla}_{\boldsymbol{k}} E(\boldsymbol{k})/\hbar$, we obtain

$$\hbar \frac{\mathrm{d}\boldsymbol{k}_0}{\mathrm{d}t} = \boldsymbol{F} \,. \tag{20}$$

Finally, we introduce an effective mass in order to write down an equation relating the time derivative of the velocity to the external forces applied to the particle. To do so, we calculate the time derivative of the average velocity of the wave packet:

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{1}{\hbar} \frac{\mathrm{d}(\nabla_i E)_{\boldsymbol{k}_0(t)}}{\mathrm{d}t} \,. \tag{21}$$

In this expression, only the central position $\boldsymbol{k}_0(t)$ of the wave packet is time dependent. Hence,

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{1}{\hbar} \frac{\partial^2 E}{\partial k_i \partial k_j} \frac{\mathrm{d}k_{0j}}{\mathrm{d}t} = \left(\frac{1}{m} \right)_{ij} F_j \,, \tag{22}$$

where we have introduced the effective mass tensor defined by

$$\left(\frac{1}{m}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_i \partial k_j}\,. \tag{23}$$

This relation can be used to relate the velocity to the external forces in a way that maintains the structure of Newton's equation. The forces due to the periodic potential in the crystal are accounted for by including the effective mass term. As can be seen, this effective mass is completely determined once we know the dispersion relation. Note in particular that the effective mass, which is the reciprocal of the radius of curvature of the dispersion relation, can be negative. This means that a force oriented in the positive x direction can produce motion in the negative x direction, for example. This can be understood when we remember that the electrons are described by a wave. At the edge of the Brillouin zone, the wave associated with the electron undergoes Bragg reflection, which corresponds to motion in the opposite direction. Although we have arrived at a classical type of formulation, we stress that wave effects have been taken into account in an effective manner by introducing this effective mass, which itself depends on the dispersion relation of the electron waves.

1.5 Electrical Conductivity in the Collisional Regime

In the collisional regime, the distribution is close to the equilibrium distribution given by the Fermi–Dirac expression. We seek a correction describing the perturbation introduced by the presence of an electric field. We follow the same procedure as we did for conduction in dilute media. We use the Boltzmann equation, which governs the evolution of the average occupation number of each state:

$$\frac{\partial n(\boldsymbol{r},\boldsymbol{k},t)}{\partial t} + \frac{\partial \boldsymbol{r}}{\partial t}\cdot\boldsymbol{\nabla} n(\boldsymbol{r},\boldsymbol{k},t) + \frac{\partial \boldsymbol{k}}{\partial t}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}} n(\boldsymbol{r},\boldsymbol{k},t) = \left(\frac{\partial n}{\partial t}\right)_{\text{coll}}\,. \tag{24}$$

We observe a certain number of differences. To begin with, the appropriate variables here are no longer the position and the velocity but the position and the wave vector. Furthermore, the Maxwell–Boltzmann velocity distribution function is replaced by a Fermi–Dirac distribution. We should stress an important point at this stage. The Boltzmann equation was established using the equations of classical mechanics. It follows directly from the Liouville theorem. In order to use it here, we must therefore consider the quasi-classical approximation:

$$\frac{\partial n(\boldsymbol{r},\boldsymbol{k},t)}{\partial t} + \langle \boldsymbol{v}\rangle_{\boldsymbol{k}}\cdot\boldsymbol{\nabla} n(\boldsymbol{r},\boldsymbol{k},t) + \frac{\mathrm{d}\boldsymbol{k}_0}{\mathrm{d}t}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}} n(\boldsymbol{r},\boldsymbol{k},t) = \left(\frac{\partial n}{\partial t}\right)_{\text{coll}}\,, \tag{25}$$

where $\boldsymbol{r}$ stands for the mean position of a wave packet, $\partial\boldsymbol{r}/\partial t$ is given by the mean value of the velocity operator which is given by the group velocity, and

$\partial \boldsymbol{k}/\partial t$ is $\mathrm{d}\boldsymbol{k}_0/\,\mathrm{d}t = \boldsymbol{F}/\hbar$. It is difficult to determine an explicit form for the collision term. Within the framework of the relaxation time approximation, we obtain

$$n(\boldsymbol{r},\boldsymbol{k},t) = n^0(\boldsymbol{r},\boldsymbol{k},t) - \frac{\tau}{\hbar}\boldsymbol{F}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}} n^0(\boldsymbol{r},\boldsymbol{k},t)\,. \tag{26}$$

Note that the force is related to the electric field by $\boldsymbol{F} = -e\boldsymbol{E}$. Substituting this expression for n into the equation giving the flux, we obtain

$$\boldsymbol{j}^{\mathrm{q}} = \int e^2 \langle \boldsymbol{v}\rangle_{\boldsymbol{k}} \frac{1}{4\pi^3}\frac{\tau}{\hbar}\boldsymbol{E}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}} n^0(\boldsymbol{r},\boldsymbol{k},t)\,\mathrm{d}^3\boldsymbol{k}\,. \tag{27}$$

Noting that

$$\boldsymbol{\nabla}_{\boldsymbol{k}} n^0(\boldsymbol{r},\boldsymbol{k},t) = \frac{\partial n^0(\boldsymbol{r},E,t)}{\partial E}\boldsymbol{\nabla}_{\boldsymbol{k}} E = \frac{\partial n^0(\boldsymbol{r},E,t)}{\partial E}\hbar\langle \boldsymbol{v}\rangle_{\boldsymbol{k}}\,, \tag{28}$$

we eventually obtain

$$j_i^{\mathrm{q}} = e^2 \int \tau(\boldsymbol{k}) v_i v_j E_j \left(-\frac{\partial n^0}{\partial E}\right)\frac{\mathrm{d}^3\boldsymbol{k}}{4\pi^3}\,. \tag{29}$$

It is important to note the derivative of the Fermi–Dirac distribution in this integral. This function is sharply peaked around the chemical potential and has a rather small width of the order of a few kT. This reflects the fact that only electrons with energy close to the Fermi level actually contribute to conduction. It is clear that, if the Fermi level falls within a band gap, the material will be a poor conductor, i.e., a good insulator.

1.6 Electrical Conduction in the Ballistic Regime

We shall now consider a specific geometry. A cylindrical wire of length less than the mean free path connects two electron reservoirs. The electrons can therefore pass through the wire without undergoing collisions. In this regime, called the ballistic regime, the wavelike behaviour of the electrons plays a fundamental role. To write down an expression for the flux, we go back to the form established above in the wave approach. It is important to remember that the sum over the states is no longer a sum over plane waves, but a sum over the electronic states in a cylindrical wire, i.e., in a wave guide. As the radius of the wire is reduced, the number of modes also decreases. This means that the electric current will decrease too. The conductance becomes quantised in proportion to the number of modes. Each reservoir is described by a different chemical potential which takes into account the electric potential. We cannot use the relaxation time approximation. However, we may describe the electrons coming from the positive or negative x directions by

Fermi–Dirac distributions with chemical potentials μ^+ and μ^-, respectively. We then have

$$\boldsymbol{j}^{\mathrm{q}} = \int -e\langle \boldsymbol{v}\rangle_{\boldsymbol{k}} \frac{1}{4\pi^3}\left[n^+(\boldsymbol{k}) - n^-(-\boldsymbol{k})\right] \mathrm{d}^3\boldsymbol{k}\,, \tag{30}$$

where the integral is carried out over half a Brillouin zone with $k_x > 0$. This form is analogous to the expression for the flux of molecules in the rarefied gas regime derived in the Chapter on Transport in Dilute Media by *Carminati* in this volume.

Carrying out the calculation for a wire of cross-sectional area πa^2 connecting two metals with Fermi energy E_{F}, across which we apply a potential difference U, we find that the current can be written [4]

$$I = GU = \frac{2e^2}{h}\frac{k_{\mathrm{F}}^2 a^2}{4}\,, \tag{31}$$

where $2e^2/h$ is called the conductance quantum. The number multiplying it is the number of modes that can propagate in the wire. The product of the two numbers is called the Sharvin conductance. In the above calculation, it is assumed that the metal can be described by a free electron gas. It is also assumed that the radius a is big enough to avoid confinement effects. Summing over all states amounts to summing over all solid angles. This calculation is therefore very close to the calculation for ballistic transfer in gases.

2 Phonons

In this section, we shall discuss how to calculate the thermal energy flux due to the motion of atoms in a crystal.

2.1 Vibrational Modes in a Lattice

We begin by recalling the main ideas required to discuss heat conduction phenomena in crystals.

Vibrational Modes. 1D Case

The system considered here is a linear chain of N atoms. The interatomic distance is a. The position of the nth atom is $x_n(t) = na + u_n(t)$. We only consider motion parallel to the Ox axis (longitudinal motion). It is assumed that each atom only interacts with its nearest neighbours so that the energy can be written in the form

$$E = T + V = \sum_{n=1}^{N} \frac{1}{2} m\dot{u}_n^2 + \frac{1}{2}\sum_{n,m=1}^{N} \Phi(x_n - x_m)\,. \tag{32}$$

Moreover, when the amplitude of the motions is small enough, a truncated expansion of the interaction potential can be used, so that the energy becomes

$$E = \sum_{n=1}^{N} \frac{1}{2} m\dot{u}_n^2 + \frac{1}{2} \sum_{n=1}^{N} K(u_n - u_{n+1})^2 \,, \tag{33}$$

where K is the second derivative of the potential. This yields a set of coupled dynamical equations for each atom, viz.,

$$m\ddot{u}_n = -K(2u_n - u_{n-1} - u_{n+1}) \,. \tag{34}$$

Description of Modes. Dispersion Relation

The system can be decoupled by seeking solutions of type

$$u_n(t) = u \exp(\mathrm{i}kna - \mathrm{i}\omega t) \,.$$

Substituting this into the equation of motion, we obtain the dispersion relation

$$\omega(k) = 2\omega_0 \sin \frac{ka}{2} \,, \tag{35}$$

where $\omega_0^2 = K/m$. This type of solution must satisfy a boundary condition when the chain has finite length. Applying the periodic boundary condition, i.e., $u_n = u_{n+N}$, as before, we obtain $\exp(\mathrm{i}kNa) = 1$, implying that k must assume one of the discrete values $k = 2\pi p/Na$. We can now represent the displacement of each atom by adding the possible contributions of each of these solutions in the form

$$u_n(t) = \frac{1}{\sqrt{N}} \sum_{p=1}^{N} X_p \exp\Big[\mathrm{i}k_p na - \mathrm{i}\omega(k_p)t\Big] \,. \tag{36}$$

We recognise here the same structure as a discrete Fourier transform. Inverting, this yields

$$X_p \exp\Big[-\mathrm{i}\omega(k_p)t\Big] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} u_n(t) \exp(-\mathrm{i}k_p na) \,. \tag{37}$$

The last two equations show that the displacements of the atoms can be represented by normal variables X_p. Each normal variable is a normal mode of vibration of the chain, characterised by a wave vector $k_p = 2\pi p/Na$ and a normal frequency $\omega_p = \omega(k_p)$. In the same way, we can introduce a normal variable for the momentum, viz.,

$$P_p = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} m\dot{u}_n(t) \exp(-\mathrm{i}k_p na) \,. \tag{38}$$

2.2 Phonon Energy

Quantising the Energy

Replacing $u_n(t)$ and $\dot{u}_n(t)$ by their expressions in terms of normal variables in the expression for the energy of the system, it can be shown[2] that the energy is the sum of the energies associated with each normal mode of frequency ω_p:

$$E = \frac{1}{2}\sum_{p=1}^{N}\left(\frac{1}{m}P_pP_p^* + m\omega_p^2 X_pX_p^*\right) . \tag{39}$$

We see that the energy of the system is the sum of the energy of each mode. Moreover, the structure of the energy of each mode is precisely that of a harmonic oscillator of natural frequency ω_p. Classically, the sum of the kinetic energy and the potential energy of a harmonic oscillator with complex amplitude X_p and angular frequency ω_p is

$$E_p = \tfrac{1}{2}m\omega_p^2|X_p|^2 . \tag{40}$$

We can now use the canonical quantisation procedure to show that the energy associated with each mode is quantised. We know that the position and momentum variables must be replaced by operators. The result is a discrete energy spectrum of the form $E_p = (m + 1/2)\hbar\omega_p$. The vibration amplitude and the quantum number m can be related by equating the two expressions for the energy.

A phonon is a normal mode of vibration of the atomic chain, characterised by frequency ω_p and wave vector $k_p = 2\pi p/Na$ and with energy of the form $(m + 1/2)\hbar\omega_p$. It is clear that the particle aspect only refers to the energy. The phonon is a delocalised vibrational mode. In the following, we shall introduce the idea of the density of states and group velocity associated with the phonon.

Average Energy of a Mode

It is easy to calculate the average energy of a mode by applying the standard results of statistical physics. We consider the situation in which the crystal is in equilibrium with a heat bath at temperature T. We then apply the standard procedure of statistical physics to find the heat energy of the crystal. The starting point is the observation that we must use the canonical ensemble. The second important point is to give a careful definition of the state of the system. It is defined as soon as the state of each mode is known, i.e., as soon as the quantum numbers m associated with each mode are known. The

[2] Simply insert (36) into (33) and use the identity $\sum_{i=1}^{N}\exp(\mathrm{i}k_pna) = N\delta_{p,0}$.

calculation shows that the energy of the crystal can be expressed in the form

$$E = \sum_{p=1}^{N} \left(\bar{m}_p + \frac{1}{2} \right) \hbar \omega_p \,, \tag{41}$$

where

$$\bar{m}_p = \frac{1}{\exp \dfrac{\hbar \omega_p}{kT} - 1} \,. \tag{42}$$

2.3 Density of States. Optical and Acoustic Modes

The density of states in the reciprocal space is easily calculated. The wave vectors of the modes have the form $2\pi p/L$ so that the number of modes in an interval Δk is $L\Delta k/2\pi$. This is easily generalised to three dimensions.

In the case of a 3D crystal, one must also take into account the fact that the atoms can move, not only parallel to the wave vector (longitudinal mode), but also perpendicular to the wave vector (transverse mode). There is therefore one longitudinal mode and two transverse modes for each wave vector.

Finally, we must also consider the case of crystals that are not monatomic. When a crystal contains several atoms per unit cell, we can consider the centre of mass of the unit cell in the way we have just described. Then we also find three modes for each wave vector. These modes are called acoustic modes. To these vibrational modes of the centre of mass, one must add the vibrational modes associated with the relative motions of the atoms within the unit cells, referred to as optical modes. Thus, when a unit cell contains two atoms, there are three optical modes. Since the acoustic modes correspond to the motion of the centre of mass of the unit cell, the phase of all the atoms within a unit cell is the same. In contrast, for an optical mode, the atoms are out of phase. For example, consider a crystal of NaCl. In the transverse acoustic mode, the Na^+ and Cl^- ions move perpendicularly to the wave vector and in the same direction. In contrast, in the transverse optical mode, the ions move perpendicularly to the wave vector and in opposite directions. In the first case, the relative distance between the ions is constant, whilst in the second, the optical mode, it oscillates. When the optical mode is excited, a dipole moment appears in the unit cell, which thus becomes an oscillating electric dipole. This explains why it is called an optical mode.

It should be stressed here that the existence of discrete normal modes of vibration with discrete frequencies is a purely classical feature. What is quantum mechanical here is that the energy can only assume a discrete set of values. From a classical point of view, one could say that the amplitude of vibration cannot vary continuously.

2.4 Calculating the Heat Flux

The heat flux can be calculated in the following way. The energy flux along the monatomic chain is given by the work done by the atom at x_{n-1} on the atom at x_n. The power due to these forces is simply

$$P = \boldsymbol{F} \cdot \boldsymbol{v} = -K(u_n - u_{n-1})\dot{u}_n \,. \tag{43}$$

If the motion is due to a single mode, then

$$u_n(t) = \frac{1}{\sqrt{N}} X_p \exp(\mathrm{i}k_p na - \mathrm{i}\omega_p t) \,. \tag{44}$$

We can then calculate the time average of the power due to forces exerted by atom $n-1$ on atom n, viz.,

$$\begin{aligned} P = \frac{1}{2}\Re(F u_n^*) &= \frac{1}{2}\frac{K|X_p|^2\omega_p}{N}\Re\Big[-\mathrm{i}(1-\mathrm{e}^{-\mathrm{i}ka})\Big] \\ &= \frac{K|X_p|^2\omega_p \sin(ka)}{2N} \,. \end{aligned} \tag{45}$$

Observing that

$$\begin{aligned} \sin(ka) &= 2\sin\frac{ka}{2}\cos\frac{ka}{2} \,, \\ K &= m\omega_0^2 \,, \\ v_p &= \frac{\partial\omega}{\partial k} = a\omega_0 \cos\frac{ka}{2} \,, \\ \omega_p &= \omega_0 \sin\frac{k_p a}{2} \,, \end{aligned}$$

we may rewrite the result in the form

$$P = \frac{E_p v_p}{L} \,, \tag{46}$$

where v_p is the group velocity associated with mode p and E_p is the energy of mode p. The energy flux can be set in the usual form of a product of an energy per unit length and the velocity of the energy flow. It follows from the above calculation that the energy flow velocity is the group velocity $\boldsymbol{\nabla}_{\boldsymbol{k}}\omega$.

In the general case, we must sum the contributions from all modes (all polarisations and all wave vectors). To do this, we must know the average occupation numbers of the various modes. At equilibrium, the modes propagating from left to right and those propagating from right to left have the same amplitude, so the resulting flux is zero. In any case, the general form of the flux is

$$\boldsymbol{J} = \sum_s \int_{\mathrm{BZ}} \frac{\mathrm{d}^3\boldsymbol{k}}{8\pi^3} \left(n_{\boldsymbol{k},s} + \frac{1}{2}\right) \hbar\omega_{\boldsymbol{k},s} \boldsymbol{\nabla}_{\boldsymbol{k}}\omega_s \,, \tag{47}$$

where the index s specifies the mode (optical or acoustic, longitudinal or transverse).

2.5 Calculating the Thermal Conductivity

Introduction

In order to use the formula established above for the heat flux, we need to be able to determine the average occupation number of the states that are not symmetric and which take into account the perturbation due to a temperature gradient. The dynamical equation for the average occupation number of the vibrational states is the Boltzmann equation. Using this equation amounts to assuming that we can define a local temperature with which we can associate a local population of phonons. At first glance, this contradicts the idea that the phonons are modes extending over the whole crystal. In reality, the modes do not actually extend over the whole crystal, and this for two reasons. Firstly, there are defects in the crystal, e.g., interstitial atoms, dislocations, substitution impurities, etc., and these all tend to scatter the waves. Secondly, the potential is not strictly harmonic. Anharmonic deviations can be treated as scattering mechanisms for the phonons. The result of these various phenomena is that a phonon will decay as it propagates, over a characteristic distance called the mean free path which is typically of the order of 10 nm. The theory discussed above which neglects these effects must be considered as a theory that is only valid on length scales shorter than the mean free path.

We may now ask how energy is transferred on length scales that are greater than the mean free path. This brings us back to the idea of local thermodynamic equilibrium. We consider that a region with dimensions extending over a multiple of the mean free path can be characterised by a temperature. The modes are in equilibrium within this region because of collisions. We can then define the average occupation number of the modes using the Bose–Einstein distribution with a locally defined temperature.

Calculating the Average Occupation Number out of Equilibrium

Due to the temperature gradient, the average number of particles in a given state is perturbed. We thus seek a solution close to equilibrium. We use the Boltzmann equation in the relaxation time approximation to obtain this correction. In the stationary regime this leads to

$$\boldsymbol{v}\cdot\boldsymbol{\nabla}_{\boldsymbol{r}} n = -\frac{n-n^0}{\tau}\,, \tag{48}$$

where the velocity is the group velocity. Seeking a solution of the form $n = n^0 + n^0\phi$, we find to first order,

$$n^0\phi = -\tau\frac{\partial n^0}{\partial T}\boldsymbol{v}\cdot\boldsymbol{\nabla}_{\boldsymbol{r}} T\,. \tag{49}$$

The flux is obtained by substituting this expression into the integral form of the flux, which yields

$$\boldsymbol{J} = -\lambda \boldsymbol{\nabla}_{\boldsymbol{r}} T \,, \tag{50}$$

where the conductivity λ is given by

$$\lambda = \int \frac{\mathrm{d}^3 \boldsymbol{k}}{8\pi^3} \tau v^2 \frac{\partial n^0}{\partial T} \hbar\omega \,. \tag{51}$$

Observing that the contribution of a mode $\boldsymbol{k}$ to the specific heat capacity is

$$c_v(\boldsymbol{k}) = \frac{\partial n^0}{\partial T} \hbar\omega(\boldsymbol{k}) \,,$$

and using $\Lambda(\boldsymbol{k})$ to denote the mean free path of the mode, we may rewrite the conductivity in the form

$$\lambda = \int \frac{\mathrm{d}^3 \boldsymbol{k}}{8\pi^3} \Lambda(\boldsymbol{k}) v(\boldsymbol{k}) c_v(\boldsymbol{k}) \,. \tag{52}$$

References

[1] J. M. Ziman: *Electrons and Phonons* (Clarendon Press, Oxford 1960) 40, 41
[2] J. M. Ziman: *Principles of the Theory of Solids* (Cambridge University Press, Cambridge 1965) 40, 41
[3] N. Ashcroft, D. Mermin: *Solid State Physics* (Harcourt College Publishers, Fort Worth 1976) 40, 41, 43
[4] N. Agraït, A. Levy Yeyati, J. M. van Ruitenbeek: Phys. Rep. **337**, 81 (2003) 47

Index

Introduction to Radiative Transfer

Rémi Carminati

Ecole Centrale Paris, Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion (EM2C), Centre National de la Recherche Scientifique, 92295 Châtenay-Malabry Cedex
remi.carminati@em2c.ecp.fr

Abstract. The aim of this Chapter is to introduce concepts and methods for modelling radiative transfer on short length scales. Electromagnetic radiation propagating in an arbitrary medium is characterised by various length scales: wavelength, coherence length, mean free path (of scattering, transport, or absorption), and skin depth. In order to cover several areas of interest for micro and nanoheat transfer, the discussion is divided into two parts.

In the first part (Sects. 1–3), we consider the propagation of radiation in scattering and absorbing media. The basic tool is the equation of radiative transfer. We shall consider in particular the case of systems with characteristic sizes of the order of the mean free path l (or in which the evolution time is of the order of l/c, where c is the energy propagation speed), but which remain large compared with the wavelength and the coherence length. We describe the various transport regimes (ballistic, multiple scattering and diffusive) and stress the analogy between this situation and the problem of heat conduction.

In the second part (Sect. 4), we treat the case of systems with characteristic sizes less than the wavelength and the coherence length. An electromagnetic formalism is then essential for modelling radiative transfer. We introduce a general calculational method based on the fluctuation–dissipation theorem. This method will be used in the Chapter by *Joulain* in this volume, which is devoted to the study of radiative transfer in nanostructures.

1 Radiative Transfer Equation

The radiative transfer equation (RTE) was introduced in astrophysics to describe the propagation of radiation in interstellar media [1], and in nuclear physics to describe the propagation of neutrons in reactors [2]. The RTE is an equation for the transport of specific intensity. In this part, we introduce the specific intensity in a phenomenological way, along with the phenomena of absorption, scattering, and thermal emission. We then establish the RTE by considering the energy balance.

1.1 Specific Intensity, Flux, Energy Density

The specific intensity $L_\nu(\boldsymbol{r}, \boldsymbol{r}, t)$ is defined so that the monochromatic radiative energy flux P_ν crossing a surface element $\mathrm{d}S$ centered at the point

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 55–76 (2007)

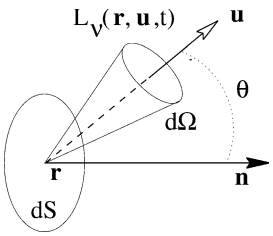

Fig. 1. Notation for defining specific intensity

$\boldsymbol{r} = (x, y, z)$, in an element of solid angle $\mathrm{d}\Omega$ centered on the direction $\boldsymbol{u}$, in the frequency range $[\nu, \nu + \mathrm{d}\nu]$, and at time t is given by

$$P_\nu(\boldsymbol{r}, \boldsymbol{u}, t) = L_\nu(\boldsymbol{r}, \boldsymbol{u}, t)\boldsymbol{u} \cdot \boldsymbol{n}\, \mathrm{d}S\, \mathrm{d}\Omega\, \mathrm{d}\nu\,. \tag{1}$$

The notation is illustrated in Fig. 1. The specific intensity is expressed in units of $\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mathrm{sr}^{-1} \cdot \mathrm{Hz}^{-1}$.

We introduce the radiative flux vector $\boldsymbol{q}_\nu$, a non-directional quantity expressed in terms of the specific intensity L_ν by

$$\boldsymbol{q}_\nu(\boldsymbol{r}, t) = \int_{4\pi} L_\nu(\boldsymbol{r}, \boldsymbol{u}, t)\boldsymbol{u}\, \mathrm{d}\Omega\,. \tag{2}$$

The vector $\boldsymbol{q}_\nu$, defined here in the purely radiometric context, can be identified with the Poynting vector defined in the context of electromagnetic theory. Its flux through an area S gives the global flux per unit frequency crossing this area in units of W/Hz. However, it is more difficult to define the specific intensity in the framework of electromagnetism, requiring the theory of coherence (see for example [3]). This difficulty arises because it is not obvious how to translate the phenomenological idea of a flux crossing a surface *in a given direction* into the language of waves (and indeed the Poynting vector does not provide this information).

The energy density per unit frequency at point $\boldsymbol{r}$ and time t is given by

$$u_\nu(\boldsymbol{r}, t) = \int_{4\pi} \frac{L_\nu(\boldsymbol{r}, \boldsymbol{u}, t)}{c}\, \mathrm{d}\Omega\,. \tag{3}$$

The speed c in this equation is the speed of propagation of the energy. In a particle view, the energy density $u_\nu(\boldsymbol{r}, t)$ can be related to the photon density $n_\nu(\boldsymbol{r}, t)$ per unit volume and per unit frequency ν by $u_\nu(\boldsymbol{r}, t) = n_\nu(\boldsymbol{r}, t)h\nu$ where h is Planck's constant. Note that, if the specific intensity is isotropic, we have $u_\nu(\boldsymbol{r}, t) = 4\pi L_\nu(\boldsymbol{r}, t)/c$.

1.2 Absorption, Scattering and Thermal Emission

Absorption

Consider a monochromatic radiative energy flux P_ν propagating in an absorbing medium in a direction $\boldsymbol{u}$ that is perpendicular to the surface element $\mathrm{d}S$.

Let s be the curvilinear coordinate in the $\boldsymbol{u}$ direction. Over an element of length $\mathrm{d}s$, a fraction $\mathrm{d}P_\nu$ of the energy is absorbed. Quite generally, we may write

$$\mathrm{d}P_\nu(s+\mathrm{d}s,\boldsymbol{u},t) = -\kappa_\nu P_\nu(s,\boldsymbol{u},t)\,\mathrm{d}s\,. \tag{4}$$

The coefficient κ_ν introduced here is the monochromatic absorption coefficient. Its reciprocal $l_\mathrm{a} = 1/\kappa_\nu$ is the absorption mean free path or absorption length.

Extinction by Scattering

In the scattering process, a fraction of the energy initially propagating in the direction $\boldsymbol{u}$ is scattered in a different direction $\boldsymbol{u}'$, which leads to a reduction in the energy flux in the direction $\boldsymbol{u}$. We thus proceed as for absorption, introducing a scattering coefficient σ_ν. Its reciprocal $l_\mathrm{s} = 1/\sigma_\nu$ is the scattering mean free path or scattering length.

If we consider the energy propagating in a given direction, e.g., in a collimated beam, then both the absorption and the scattering process contribute to extinction of the incident energy. This can be written symbolically in the form

$$\text{extinction} = \text{absorption} + \text{scattering}\,.$$

We speak of an absorbing medium when the absorption process dominates over the scattering process, and conversely. For example, the particles suspended in Indian ink or soot particles in smoke will tend to absorb visible radiation rather than scatter it and hence appear black, whereas a cloud is a highly scattering medium (and poorly absorbent) in the visible, so that it appears white.

We thus introduce the extinction coefficient

$$\beta_\nu = \sigma_\nu + \kappa_\nu\,. \tag{5}$$

Its reciprocal $l_\mathrm{ext} = 1/\beta_\nu$ is called the extinction length. In particular, the energy flux associated with a beam collimated in a given direction in a medium decreases according to a power law

$$P_\nu(s) = P_\nu(0)\exp(-\beta_\nu s)\,,$$

known as the Beer–Lambert law.

To characterise the scattering power of a medium, we define the albedo by

$$a_\nu = \frac{\sigma_\nu}{\beta_\nu} = \frac{\sigma_\nu}{\sigma_\nu + \kappa_\nu}\,. \tag{6}$$

If the albedo is zero, the medium is purely absorbing like Indian ink in the visible. If it is equal to unity, the medium is purely scattering, like a cloud in the visible.

Scattering. Phase Function

During scattering, energy is redistributed in all directions. In particular, if we are concerned with the change in the flux propagating in direction $\boldsymbol{u}$ between s and $s+\mathrm{d}s$, energy initially propagating in a different direction $\boldsymbol{u}'$ can, after scattering, find itself propagating in the direction $\boldsymbol{u}$. It then contributes to an increase in the flux in the direction $\boldsymbol{u}$. To describe this phenomenon, we introduce the phase function $p_\nu(\boldsymbol{r},\boldsymbol{u},\boldsymbol{u}')$. It represents the fraction of the energy flux which, arriving at point $\boldsymbol{r}$ in the direction $\boldsymbol{u}'$, is scattered in the direction $\boldsymbol{u}$. Some examples of phase functions are given in Appendix. The increase in the flux due to scattering between s and $s+\mathrm{d}s$ can then be written

$$\mathrm{d}P_\nu(s+\mathrm{d}s,\boldsymbol{u},t) = \frac{\sigma_\nu}{4\pi}\int_{4\pi} p_\nu(s,\boldsymbol{u},\boldsymbol{u}')P_\nu(s,\boldsymbol{u}',t)\,\mathrm{d}\Omega'\,\mathrm{d}s\,. \tag{7}$$

It is useful to make the following remarks:

- When the medium is homogeneous, the phase function is independent of the position $\boldsymbol{r}$, an assumption we shall make in the following.
- There are several normalisations for the phase function in the literature. We shall use

$$\frac{1}{4\pi}\int_{4\pi} p_\nu(\boldsymbol{u},\boldsymbol{u}')\,\mathrm{d}\Omega' = 1\,. \tag{8}$$

  If the phase function is constant, we speak of isotropic scattering. This means that scattering leads to an equiprobable energy distribution in all directions. Otherwise we speak of anisotropic scattering.
- The phase function depends on the direction of incidence $\boldsymbol{u}'$ and the direction of scattering $\boldsymbol{u}$. For particles with some kind of symmetry, e.g., spheres, the phase function only depends on the relative angle Θ between the two directions, and more precisely, on the cosine of this angle. We then write $p_\nu(\boldsymbol{u},\boldsymbol{u}') = p_\nu(\cos\Theta)$.

Thermal Emission

Over an element of length $\mathrm{d}s$ in the $\boldsymbol{u}$ direction, a fraction $\mathrm{d}P_\nu$ of the energy can be emitted thermally. We then write

$$\mathrm{d}P_\nu(s+\mathrm{d}s,\boldsymbol{u},t) = \eta_\nu\,\mathrm{d}s\,, \tag{9}$$

where η_ν is the monochromatic emission coefficient. If the medium is in local thermodynamic equilibrium (LTE), the specific intensity at any point is the equilibrium specific intensity $n^2L^0_\nu(T)$ at the local temperature T, where n is the real part of the refractive index of the medium and $L^0_\nu(T)$ is the Planck function. Moreover, in LTE, the absorbed flux is equal to the emitted flux at every point. We thus have [1]

$$\eta_\nu = \kappa_\nu n^2 L^0_\nu(T)\,. \tag{10}$$

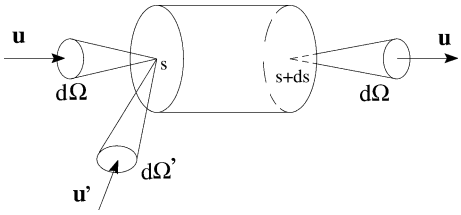

Fig. 2. Notation for describing the radiative energy balance. $\mathrm{d}s = c\,\mathrm{d}t$ in the direction $\boldsymbol{u}$

1.3 Establishing the RTE. Radiative Energy Balance

We now examine the radiative energy balance for a volume element $\mathrm{d}V$ as illustrated in Fig. 2. The negative contribution to the balance comes from extinction by absorption and scattering, whilst the positive contribution comes from scattering and thermal emission.

Consider the directional energy density $u_\nu(s, \boldsymbol{u}, t)$ which is related to the specific intensity by

$$u_\nu(s, \boldsymbol{u}, t) = \frac{L_\nu(s, \boldsymbol{u}, t)}{c}\,, \tag{11}$$

where c is the speed of energy propagation. The energy balance for the element $\mathrm{d}V$ can then be written

$$\begin{aligned} u_\nu(s + c\,\mathrm{d}t, \boldsymbol{u}, t + \mathrm{d}t)\,\mathrm{d}V = {} & u_\nu(s, \boldsymbol{u}, t)\,\mathrm{d}V \\ & - (\kappa_\nu + \sigma_\nu) u_\nu(s, \boldsymbol{u}, t)\,\mathrm{d}V c\,\mathrm{d}t \\ & + \kappa_\nu n^2 \frac{L^0_\nu(T)}{c}\,\mathrm{d}V c\,\mathrm{d}t \\ & + \frac{\sigma_\nu}{4\pi} \int_{4\pi} p_\nu(\boldsymbol{u}, \boldsymbol{u}') u_\nu(s, \boldsymbol{u}', t)\,\mathrm{d}\Omega'\,\mathrm{d}V c\,\mathrm{d}t\,. \end{aligned}$$

Rearranging the terms in this equation and using (11), we obtain

$$\begin{aligned} \frac{L_\nu(s + c\,\mathrm{d}t, \boldsymbol{u}, t + \mathrm{d}t) - L_\nu(s, \boldsymbol{u}, t)}{c\,\mathrm{d}t} = {} & -(\kappa_\nu + \sigma_\nu) L_\nu(s, \boldsymbol{u}, t) + \kappa_\nu n^2 L^0_\nu(T) \\ & + \frac{\sigma_\nu}{4\pi} \int_{4\pi} p_\nu(\boldsymbol{u}, \boldsymbol{u}') L_\nu(s, \boldsymbol{u}', t)\,\mathrm{d}\Omega'\,. \end{aligned} \tag{12}$$

In the limit as $\mathrm{d}t$ tends to zero, the left-hand side of this equation tends to the total derivative of the specific intensity, viz.,

$$\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t} L_\nu(s, \boldsymbol{u}, t) = \left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L_\nu = \left(\frac{1}{c}\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\right) L_\nu\,, \tag{13}$$

where $\boldsymbol{\nabla}$ is the gradient operator with respect to the position variable $\boldsymbol{r}$. Finally, we obtain the radiative transfer equation in the form

$$\frac{1}{c}\frac{\partial}{\partial t}L_\nu(\boldsymbol{r},\boldsymbol{u},t)+\boldsymbol{u}\cdot\boldsymbol{\nabla}L_\nu(\boldsymbol{r},\boldsymbol{u},t)=-(\kappa_\nu+\sigma_\nu)L_\nu(\boldsymbol{r},\boldsymbol{u},t)+\kappa_\nu n^2 L^0_\nu(T)+\frac{\sigma_\nu}{4\pi}\int_{4\pi}p_\nu(\boldsymbol{u},\boldsymbol{u}')L_\nu(\boldsymbol{r},\boldsymbol{u}',t)\,\mathrm{d}\Omega'\,. \tag{14}$$

1.4 Discussion

The RTE is an integro-differential equation describing the transport of specific intensity, a quantity depending on position, direction, frequency and time. There is an analogy between this equation and the Boltzmann equation introduced in the kinetic transport theory for particles. This analogy will be important in the Chapter by *Lemonnier* in this volume, where phonon transport is used to model heat conduction in solids. In fact, many techniques have been developed to solve the RTE in the context of radiative transfer [1, 4, 5]. These techniques can be extended to the case of heat conduction by phonons.

When deriving the RTE, we mentioned several length scales, namely the absorption and scattering lengths. Depending on the ratio of the characteristic length scales in the systems under investigation and these quantities, various transfer regimes are observed: ballistic regime, multiple scattering regime, and diffusive regime. We shall now describe these regimes. By virtue of the analogy between the RTE and the Boltzmann equation for phonon transport, it will be possible to transpose the physical phenomena that we shall discuss directly to the case of heat conduction. We shall begin by considering the diffusive regime, then the regime in which the characteristic length scales of the systems are of the order of (or shorter than) the mean free paths, these being the relevant regimes in micro and nanoheat transfer.

2 From the RTE to the Diffusion Approximation

In this section, we shall show that, on large length and time scales, the RTE simplifies to a transport equation for the energy density. This quantity no longer depends on the direction of propagation, which greatly simplifies the problem. The transport equation for the energy density, obtained in the large scale limit, is a diffusion equation with the same form as the heat equation. In the diffusion regime, the radiative flux is proportional to the gradient of the energy density. The multiplicative coefficient is called the diffusion coefficient.

Once we have established the diffusion equation, we shall examine the analogy between the diffusive regime for radiative transfer and heat conduction on large scales in a solid or gas (Fourier law). Note that the diffusion

equation is encountered in many transport phenomena and is widely used in practice in a range of fields from heat transfer, through photothermal imaging, to imaging in biological media [6].

We shall then discuss transitions between the various regimes, on the basis of the full RTE, as a function of the characteristic size of the system under investigation. We shall be concerned in particular with the transition to the non-diffusive (non-Fourier) regime, stressing the key behaviour and orders of magnitude. The results obtained can be transposed directly (at least in a qualitative manner) to the case of heat conduction.

To simplify the discussion, we consider a cold medium, in which thermal emission in the medium can be neglected at the relevant frequencies. This corresponds, for example, to visible radiation transfer in a medium at 300 K.

2.1 From the P_1 Approximation to the Diffusion Equation

Consider a scattering medium in which the absorption length l_a is very long compared with the scattering length l_s (albedo close to unity). The diffusion approximation is applicable a priori for such a medium if its characteristic length is greater than l_s. We are thus dealing with a medium in which there is multiple scattering. During scattering, the specific intensity tends to become isotropic. One can then seek an approximation for the specific intensity in the form of a sum of an isotropic term and a correction depending on the direction. This is the basic idea of the P_1 approximation. This approximation can be approached systematically by expanding the specific intensity in terms of the Legendre polynomials. We speak of the P_1 approximation when we retain only the first two terms in such an expansion, and the P_n approximation when we go to order n. The P_1 approximation is the usual starting point for establishing the diffusion approximation [4]. Note that, if the absorption is not low, this approach is no longer valid. We shall return to this point later. One can also use another approach, based on asymptotic analysis of the normal modes of the RTE, originally developed to handle neutron transport [2].

In this section, we shall outline the approach based on the P_1 approximation. The starting point is the time-dependent radiative transfer equation. The aim is to relate the radiative flux to the energy density. Using the Einstein notation (summation over repeated indices), the RTE (14) becomes (in a cold medium)

$$\frac{1}{c}\partial_t L(\boldsymbol{r},\boldsymbol{u},t) = -u_i\partial_i L(\boldsymbol{r},\boldsymbol{u},t) - \beta L(\boldsymbol{r},\boldsymbol{u},t) + \frac{\sigma}{4\pi}\int_{4\pi} p(\boldsymbol{u},\boldsymbol{u}')L(\boldsymbol{r},\boldsymbol{u}',t)\,\mathrm{d}\Omega' \,. \quad (15)$$

To simplify the notation, we have dropped the subscript ν on all radiative quantities, bearing in mind that we are working with monochromatic radiation throughout.

Energy Conservation Equation

One can establish an energy conservation equation relating the energy density defined by (3) and the radiative flux vector defined by (2). The RTE is an energy conservation equation for that part of the radiation that propagates in a given direction $\boldsymbol{u}$. To obtain an equation involving the energy density, we simply integrate (15) over all directions $\boldsymbol{u}$, i.e., over $\mathrm{d}\Omega$. This yields

$$\partial_t u(\boldsymbol{r},t) = -\partial_i q_i(\boldsymbol{r},t) - c\kappa u(\boldsymbol{r},t)\,. \tag{16}$$

Expression for the Radiative Flux

We now seek a general expression for the radiative flux vector. To do so, we multiply (15) by u_j and integrate once again over the directions $\boldsymbol{u}$. This yields

$$\frac{1}{c}\partial_t q_j(\boldsymbol{r},t) = -\int_{4\pi} u_j u_i \partial_i L(\boldsymbol{r},\boldsymbol{u},t)\,\mathrm{d}\Omega - \beta q_j(\boldsymbol{r},t) + \frac{\sigma}{4\pi}\int_{4\pi} p(\boldsymbol{u},\boldsymbol{u}')u_j L(\boldsymbol{r},\boldsymbol{u}',t)\,\mathrm{d}\Omega'\,\mathrm{d}\Omega\,. \tag{17}$$

From this equation, we obtain an expression for the radiative flux vector:

$$q_j(\boldsymbol{r},t) = -\frac{1}{\beta c}\partial_t q_j(\boldsymbol{r},t) - \frac{1}{\beta}\int_{4\pi} u_j u_i \partial_i L(\boldsymbol{r},\boldsymbol{u},t)\,\mathrm{d}\Omega + \frac{a}{4\pi}\int_{4\pi} p(\boldsymbol{u},\boldsymbol{u}')u_j L(\boldsymbol{r},\boldsymbol{u}',t)\,\mathrm{d}\Omega'\,\mathrm{d}\Omega\,. \tag{18}$$

The last term of this equation can be written more simply by using the relation

$$\int_{4\pi} p(\boldsymbol{u},\boldsymbol{u}')u_j\,\mathrm{d}\Omega = 4\pi g u_j'\,, \tag{19}$$

where g is the average value of $\cos\Theta = \boldsymbol{u}\cdot\boldsymbol{u}'$, viz.,

$$g = \langle\cos\Theta\rangle = \frac{1}{4\pi}\int_{4\pi} p(\cos\Theta)\cos\Theta\,\mathrm{d}\Omega\,. \tag{20}$$

This result can be shown provided that the phase function depends only on $\cos\Theta$, as happens for example in the case where the scattering particles are spherical. We then obtain for the radiative flux

$$q_j(\boldsymbol{r},t) = -\frac{1}{\beta c}\partial_t q_j(\boldsymbol{r},t) - \frac{1}{\beta}\int_{4\pi} u_j u_i \partial_i L(\boldsymbol{r},\boldsymbol{u},t)\,\mathrm{d}\Omega + agq_j(\boldsymbol{r},t)\,, \tag{21}$$

which implies that

$$q_j(\boldsymbol{r},t) = -\frac{1}{1-ag}\left[\frac{1}{\beta c}\partial_t q_j(\boldsymbol{r},t) + \frac{1}{\beta}\int_{4\pi} u_j u_i \partial_i L(\boldsymbol{r},\boldsymbol{u},t)\,\mathrm{d}\Omega\right]\,. \tag{22}$$

This expression is quite general. The only assumption made so far concerns the phase function, which must only depend on the cosine of the relative scattering angle Θ.

Diffusion Approximation

In order to obtain an expression for the radiative flux in the form of a diffusion equation, we introduce the P_1 approximation, which amounts to assuming that the specific intensity is quasi-isotropic. This assumption is valid if there are sufficient scattering events to ensure isotropy, which is only possible if the absorption is low enough. The specific intensity can then be represented in terms of a basis of Legendre polynomials, in an expansion to order one [4]. This yields

$$L(\boldsymbol{r},\boldsymbol{u},t) = \frac{1}{4\pi}L^0(\boldsymbol{r},t) + \frac{3}{4\pi}\boldsymbol{q}(\boldsymbol{r},t)\cdot\boldsymbol{u}\,, \tag{23}$$

where $\boldsymbol{q}$ is the radiative flux vector. Using this form for the specific intensity, the second term in (22) can be written down explicitly. Finally,

$$q_j(\boldsymbol{r},t) = -\frac{c}{3[\kappa+\sigma(1-g)]}\partial_j u(\boldsymbol{r},t) - \frac{1}{c[\kappa+\sigma(1-g)]}\partial_t q_j(\boldsymbol{r},t)\,. \tag{24}$$

In this expression there is a term $\sigma(1-g)$ whose reciprocal defines a new length scale. This is the transport mean free path, or transport length:

$$l_{\mathrm{tr}} = \frac{1}{\sigma(1-g)} = \frac{l_{\mathrm{s}}}{1-g}\,. \tag{25}$$

This length scale is characteristic of variations in the radiative flux (and the energy density) in the diffusive regime.

In many cases, the term $\partial_t q_j$ in (24) is negligible compared with the gradient of the energy density. To see this, we simply compare the two terms $[\kappa+\sigma(1-g)]q_j$ and $\partial_t q_j/c$. If τ is the characteristic time of flux variations and q the order of magnitude of the flux, the order of magnitude of the first term is $[\kappa+\sigma(1-g)]q \sim q/l_{\mathrm{tr}}$ (for a low absorption medium), whilst that of the second term is $q/c\tau$. The second term is thus negligible if the condition $l_{\mathrm{tr}} \ll c\tau$ holds. Even when l_{tr} has a value of a few millimeters, the condition is satisfied provided that $\tau > 0.1$ ns.

Neglecting the term $\partial_t q_j$ in (24), we obtain a radiative flux proportional to the gradient of the energy density:

$$q_j(\boldsymbol{r},t) = -\frac{c}{3[\kappa+\sigma(1-g)]}\partial_j u(\boldsymbol{r},t) = -D\partial_j u(\boldsymbol{r},t)\,, \tag{26}$$

where we have introduced the diffusion coefficient D. This expression for the flux has the form of the Fourier law, or Fick's law, or Ohm's law, which are all diffusion equations.

Having obtained this relation, it is easy to show that the energy density u satisfies a diffusion equation. To do so, we substitute (26) into the energy conservation equation (16). Neglecting the term in $\partial_t q_j$ in that equation too, we find that

$$D\Delta u(\boldsymbol{r},t) = \partial_t u(\boldsymbol{r},t) + \kappa c u(\boldsymbol{r},t)\,. \tag{27}$$

2.2 Discussion

The result we have just established shows that, in the limit of long length and time scales, the radiative energy density in a scattering medium obeys a diffusion equation of the same type as the heat equation. The energy flux is proportional to the gradient of the energy density. The diffusive regime thus obtained is analogous to the regime in which the Fourier law is applicable in the context of heat conduction. In this case, the distribution function characterising the energy carriers (molecules in a gas and phonons in a crystal) obeys Boltzmann's equation (just as the specific intensity obeys the RTE). The approach used to model heat conduction is therefore perfectly analogous to the one used to model the propagation of radiation in a scattering medium [7, 8]. The Fourier regime is obtained in the limit of small Knudsen numbers (see the Chapter on Transport in Dilute Media by *Carminati* in this volume), just as the diffusive regime is obtained in the limit $l_{\rm tr} \ll L$ and $l_{\rm tr}/c \ll \tau$, where L characterises the size of the system and τ the dynamical evolution of the system.

Two comments are in order concerning the derivation of the diffusion equation (27):

- The approach using the P_1 approximation leads to the diffusion coefficient

  $$D = \frac{c}{3[\kappa + \sigma(1-g)]} = \frac{c}{3(l_{\rm a} + l_{\rm tr})}\,.$$

  However, other approaches are possible, leading to diffusion equations, but with different diffusion coefficients. In particular, the dependence of D on the absorption function varies from one approach to another. This observation has been the subject of recent investigations due to its importance in the field of imaging in a scattering medium, where absorption cannot generally be ignored. See for example [9–11].
- If we keep the time-dependent term in (24), we obtain the so-called telegrapher's equation. This equation has the advantage of giving rise to a bounded propagation speed equal to $c/\sqrt{3}$, and hence removes the main disadvantage with the diffusion equation, which instantaneously gives a nonzero response at an arbitrarily great distance from the source. This disadvantage is well known in the theory of heat conduction on very short time scales, since the heat equation exhibits the same problem of instantaneous response.

2.3 Rosseland Approximation

To end this section on the diffusion approximation, we shall now show how to obtain the Rosseland approximation (allowing one to consider radiative conductivity in an absorbing, non-scattering medium) from (26). We consider an absorbing and emissive medium (dropping the hypothesis of a cold medium here) which is optically thick locally (l_{a} less than the length scale on which the specific intensity varies in the medium). At any point a local thermodynamic equilibrium is then set up. The specific intensity is thus the equilibrium specific intensity at the local temperature of the point under consideration. At each frequency ν, we thus have

$$L_\nu(\boldsymbol{r},\boldsymbol{u},t) \approx L_\nu^0\Big(T(\boldsymbol{r})\Big)\,. \tag{28}$$

According to (26), the monochromatic radiative flux is then

$$q_{\nu,j}(\boldsymbol{r}) = -\frac{4\pi}{3\kappa}\partial_j L_\nu^0\Big(T(\boldsymbol{r})\Big)\,, \tag{29}$$

using the fact that $u_\nu = 4\pi L_\nu^0/c$. Integrating over the frequency, we obtain the radiative flux as a function of the local temperature of the medium:

$$q_j(\boldsymbol{r}) = -\frac{4}{3\kappa}\partial_j\left[\sigma_{\mathrm{S}}T^4(\boldsymbol{r})\right]\,, \tag{30}$$

where σ_{S} is the Stefan constant. Hence,

$$q_j(\boldsymbol{r}) = -\frac{16}{3\kappa}\sigma_{\mathrm{S}}T^3(\boldsymbol{r})\partial_j T(\boldsymbol{r}) = -k_{\mathrm{rad}}\partial T_j(\boldsymbol{r})\,. \tag{31}$$

This expression for the flux is formally analogous to the Fourier conduction law. It reveals an equivalent radiative conductivity k_{rad} called the Rosseland conductivity, which is strongly dependent on the temperature. The existence of such a conductivity depends on the same assumptions as the diffusion approximation investigated earlier.

3 Transport Regimes

In this section, we shall discuss the various transport regimes as a function of the length and time scales characterising the system: ballistic regime, intermediate regime of multiple scattering, and diffusive regime. To do this, we shall appeal to experimental results [12] and numerical simulation of the RTE in plane geometries (transmission through a slab of scattering medium of thickness L). The numerical method used is detailed in [13].

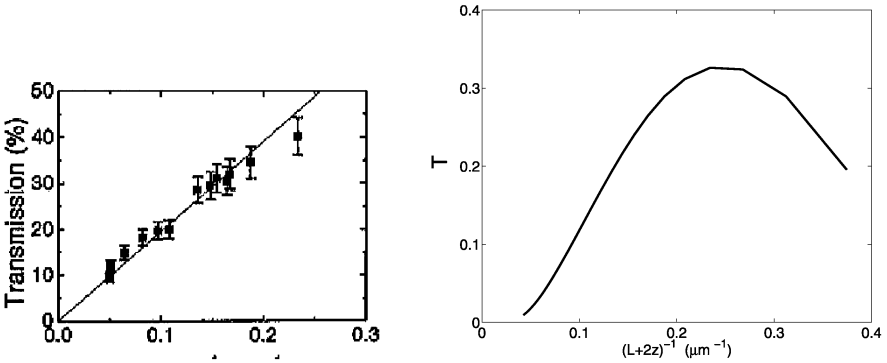

Fig. 3. Flux transmitted in static regime through a slab of TiO_2 particles with refractive index $n = 2.8$ and average diameter 220 nm. The wavelength of the incident light flux is 780 nm. The transport length is $l_{\rm tr} = 0.95\,\mu$m. *Left*: Measurements [12]. The *straight line* is the linear dependence expected in the diffusive regime according to (32). *Right*: RTE calculation [14]

3.1 Static Transmission. Ohmic Conductance and Short-Scale Deviations

The experimental results represented in Fig. 3 (left) show diffuse transmissivity measurements through a slab of different thicknesses L [12]. The sample is made of TiO_2 particles with refractive index $n = 2.8$ and average diameter 220 nm. The slab thickness varies from 1.43 μm to 18 μm. The wavelength of the incident light flux is $\lambda = 780$ nm. The transport length is estimated to be $l_{\rm tr} = 0.95\,\mu$m. Figure 3 (right) shows the results obtained from a static RTE calculation.

In the diffusive regime, one can calculate the diffuse transmissivity T of a slab in the static regime. This yields

$$T = \frac{l_{\rm tr} + z}{L + 2z}\,, \tag{32}$$

where z is a length of the order of $l_{\rm tr}$ which comes from the expression for the boundary conditions on the faces of the slab in the diffusion approximation [4, 15]. This explains why the experimental results in Fig. 3 are represented as a function of $(L + 2z)^{-1}$. Note that, since the result (32) is only valid a priori when $L \gg l_{\rm tr}$, we have $T \propto 1/L$. This $1/L$ dependence of the diffuse transmission is characteristic of the diffusive regime. Such behaviour occurs in particular in the context of Ohm's law (the electrical conductance of a wire of length L is proportional to $1/L$) and also Fourier's law (the thermal conductance of a flat wall of thickness L is proportional to $1/L$).

Diffusive behaviour is clearly visible in Fig. 3 for large values of L (linear regime). As L decreases, i.e., when $1/(L + 2z)$ increases, the experimental points deviate from diffusive behaviour. This deviation is still more clearly

visible in the numerical calculation, which was carried out for even smaller values of L. This result is a first example of non-diffusive behaviour at short length scales. In this example in the static regime, the deviation occurs for scales $L \sim L_{\mathrm{tr}}$, which is of the order of $1\,\mu\mathrm{m}$ in this example.

3.2 Transitions Between Regimes in the Dynamic Case

The question of transitions between regimes has been studied in the context of radiative propagation in scattering and aborbing media by various authors, using the RTE as transport equation [16–20]. In the dynamical regime, the analysis presented here is based upon the measurement (or calculation) of the decay time of light pulses transmitted through a slab. The total hemispherical transmissivity (integrated over angles) $T(t)$ decreases exponentially over long times. From this $\exp(-t/\tau)$ decay, we can extract a decay time τ. One can then introduce an effective diffusion coefficient D by identifying the decay time with the one obtained in the diffusive regime. For a scattering and absorbing medium, we then have

$$D = \frac{L_{\mathrm{eff}}^2}{\pi^2}\left(\frac{1}{\tau} - \kappa c\right) . \tag{33}$$

The effective length associated with the slab in this case is $L_{\mathrm{eff}} = L+2z$. As in the static case, this comes from writing the boundary conditions on the faces of the slab in the diffusion approximation. We then have $z = z_0(1+R)/(1-R)$, where $z_0 = 0.7l_{\mathrm{tr}}$ is the so-called extrapolation distance and R the average reflection coefficient at the interface between the scattering medium and the medium outside the slab. It will not be relevant to the present discussion to go into detailed discussions about how the boundary condition is obtained. It is in any case approximate, since the diffusion approximation is not a priori valid near the walls. For more detail, see for example [21].

Figure 4 (left) shows measurements [12] and Fig. 4 (right) shows a numerical simulation based on solution of the time-dependent RTE [14,20]. The quantity represented is the effective diffusion coefficient as a function of the thickness L of the system. The coefficient D is normalised by its asymptotic value, taken at $L = 25\,\mu\mathrm{m} \approx 27l_{\mathrm{tr}}$. We observe a large deviation from the diffusive regime when $L < 8l_{\mathrm{tr}}$. The agreement between the model using the RTE and experiment is very good. Both show a deviation of more than 50 % at small thicknesses. As L decreases, we move progressively from a diffusive transport regime of Fourier type, through an intermediate multiple scattering (non-diffusive) regime, to a regime where ballistic transport dominates in the slab. Note that the RTE allows one to deal unambiguously with the transitions between these regimes. The separation between the ballistic and multiple scattering regimes in the RTE is discussed further in Sect. 3.3.

The numerical simulation also allows one to show that the change in the effective diffusion coefficient at short length scales is extremely sensitive to

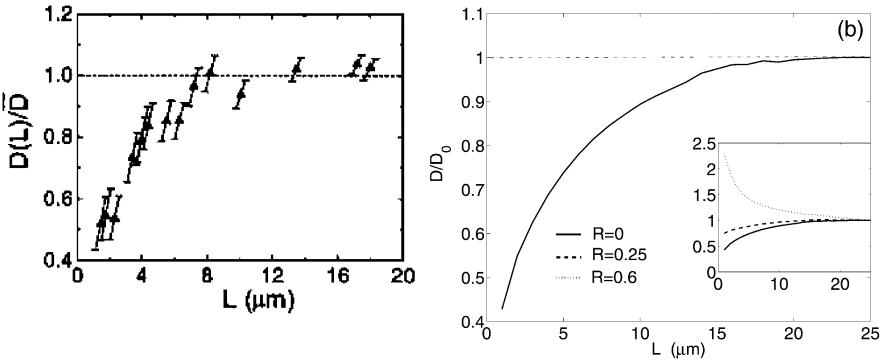

Fig. 4. Effective diffusion coefficient of a slab as a function of its thickness L. The slab contains TiO_2 particles and is illuminated at wavelength $\lambda = 780\,\text{nm}$. Other parameters are $g = 0.27$, $l_s = 0.65\,\mu\text{m}$, and $l_a = 200\,\mu\text{m}$ (albedo $a = 0.997$). The effective index of the slab is $n = 1.39$. (**a**) Measurements [12]. (**b**) Diffusion coefficient D defined by (33), normalised by its asymptotic value D_0. *Continuous curve*: Reflection coefficient averaged over angles $\bar{R} = 0$. The *inset* shows the results obtained for different levels of internal reflection

the level of internal reflection within the slab. The inset in Fig. 4 (right) shows that, by modifying the average reflection coefficient R, the shape of the curve $D(L)$ is radically changed. One can even move from a decrease to an increase in D at short length scales. This example thus shows that reflection at the interfaces of a system becomes a dominating phenomenon when the system size is of the order of, or smaller than, the transport mean free path l_{tr}. This phenomenon dominates bulk scattering and therefore drives the transmission decay (and hence the value of the effective diffusion coefficient) over long times.

It is worth stressing once again the analogy between the transmission of radiation through a scattering slab (which can be treated using the RTE) and the conduction of heat through a solid film (which can be modelled using the Boltzmann equation). Figure 5 shows the equivalent thermal conductivity of a Si film as a function of its thickness L (measurements and models [22]). Note the reduction in equivalent conductivity at short length scales. The resemblance with the results of Fig. 3 is striking. In particular, this result brings out the role of reflection at the interfaces. Indeed, the experimental results can only be explained if this phenomenon is taken into account in the model.

3.3 Ballistic and Multiple Scattering Components in the RTE

To end this section on the RTE and transitions between regimes, we shall show that the ballistic and multiple scattering contributions can be clearly separated in the RTE. To do so, we assume that the system is illuminated

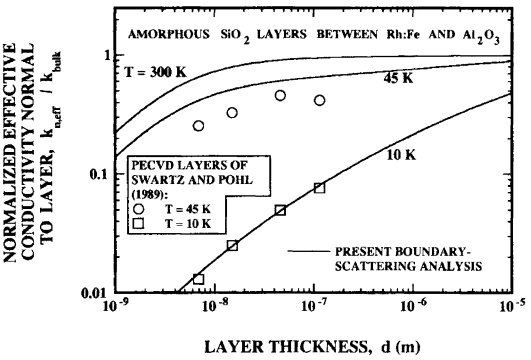

Fig. 5. Thermal conductivity of an amorphous silicon dioxide layer as a function of its thickness L. Note the reduction in the effective conductivity at small thicknesses [22]

by a collimated beam propagating in the direction $\boldsymbol{u}_0$. We then express the specific intensity at any point in the form of a sum of a collimated (ballistic) component and a multiply scattered component:

$$L(\boldsymbol{r},\boldsymbol{u},t) = L_{\mathrm{bal}}(\boldsymbol{r},t)\delta(\boldsymbol{u}-\boldsymbol{u}_0) + L_{\mathrm{diff}}(\boldsymbol{r},\boldsymbol{u},t)\,, \tag{34}$$

where the Dirac distribution $\delta(\boldsymbol{u}-\boldsymbol{u}_0)$ is taken in the sense of an integral over the solid angle $\mathrm{d}\Omega$, so that

$$\delta(\boldsymbol{u}-\boldsymbol{u}_0) = \frac{\delta(\theta-\theta_0)\delta(\phi-\phi_0)}{|\sin\theta_0|}\,.$$

Inserting the decomposition (34) into the RTE (14), we obtain two equations. For the ballistic component, the RTE becomes (assuming a cold medium)

$$\frac{1}{c}\frac{\partial}{\partial t}L_{\mathrm{bal}}(\boldsymbol{r},t) + \boldsymbol{u}\cdot\boldsymbol{\nabla}L_{\mathrm{bal}}(\boldsymbol{r},t) = -(\kappa+\sigma)L_{\mathrm{bal}}(\boldsymbol{r},t)\,. \tag{35}$$

Integrating this, we obtain the Beer–Lambert equation for the evolution of the ballistic specific intensity. For a non-absorbent medium, we retrieve the fact that the ballistic specific intensity decreases over a characteristic length $l_{\mathrm{s}} = 1/\sigma$. For the multiply scattered component, the RTE becomes

$$\begin{aligned}\frac{1}{c}\frac{\partial}{\partial t}L_{\mathrm{diff}}(\boldsymbol{r},\boldsymbol{u},t) + \boldsymbol{u}\cdot\boldsymbol{\nabla}L_{\mathrm{diff}}(\boldsymbol{r},\boldsymbol{u},t) &= -(\kappa+\sigma)L_{\mathrm{diff}}(\boldsymbol{r},\boldsymbol{u},t)\\ &\quad + \frac{\sigma}{4\pi}\int_{4\pi} p(\boldsymbol{u},\boldsymbol{u}')L_{\mathrm{diff}}(\boldsymbol{r},\boldsymbol{u}',t)\,\mathrm{d}\Omega'\\ &\quad + \frac{\sigma}{4\pi}p(\boldsymbol{u},\boldsymbol{u}_0)L_{\mathrm{bal}}(\boldsymbol{r},t)\,.\end{aligned} \tag{36}$$

Here we have a source term proportional to the ballistic component at the relevant point and time. This term describes the fact that the energy leaving the ballistic component by scattering is redistributed in the diffuse component. The relative weight of ballistic transport and diffuse transport drives the overall regime. It is only when all the energy is transported by the diffuse component *and* the length and time scales are large compared with l_{tr} and l_{tr}/c, respectively, that transport becomes diffusive.

4 Electromagnetic Approach to Thermal Emission

In this last section, we leave the problem of transfer in scattering and absorbing media to discuss the case of systems with dimensions smaller than the wavelength, and possibly also the coherence length, e.g., two solid bodies at typical temperatures, a submicron distance apart, and exchanging thermal radiation. To handle this kind of situation, one is compelled to use the formalism of electromagnetism. The Chapter by *Joulain* in this volume is devoted to a detailed study of this problem. Here we shall simply introduce the general calculational method, based on the fluctuation–dissipation theorem.

4.1 Intuitive View of the Thermal Emission Mechanism

The mechanism of thermal emission can be understood and the emitted fluxes calculated by appealing to elementary notions of electromagnetic radiation. When a body is held at temperature T, the charges in it (essentially free electrons for a metal or ions for a polar crystal) have disordered motion due to thermal agitation. In each volume element, there is a fluctuating current per unit volume (or fluctuating dipole moment) that can be given a statistical description. As this current is time dependent, it radiates an electromagnetic field which is the field emitted thermally by a heated body.

Given the current, the thermally emitted field can be calculated by solving the problem of radiation by an antenna. However, the current is only known statistically. Its average value is zero, and so therefore is the average of the emitted field. (As Maxwell's equations are linear, the relationship between the current and the radiated field is linear.) However, the mean squared value of the fluctuating currents is not zero. The mean power of the thermally emitted field is therefore nonzero, as all the interesting quantities in heat transfer, e.g., the power absorbed in a solid, the flux of the Poynting vector, the energy density, etc., are quadratic quantities in the electromagnetic field.

4.2 Principle Underlying the Calculation of Thermal Emission. Fluctuation–Dissipation Theorem

We shall describe here the basic idea behind the calculation of a field emitted thermally at some fixed frequency ω by a body held at temperature T and

described by its dielectric constant $\varepsilon(\omega) = n^2(\omega)$, where n is the complex refractive index of the medium. The fluctuating current (and the field) are stationary random variables. To work in the frequency domain, we introduce the power spectral density.

Power Spectral Density. Spectral Correlation

Consider the electric field $\boldsymbol{E}(\boldsymbol{r}, t)$. We introduce the correlation function of the k and l components of the field:

$$C_{kl}(\boldsymbol{r}, \boldsymbol{r}', t - t') = \langle E_k(\boldsymbol{r}, t) E_l(\boldsymbol{r}', t') \rangle \,,$$

which depends only on the difference $\tau = t - t'$ because the field is assumed to be stationary in the statistical sense. The power spectral density is then the temporal Fourier transform of the correlation function:

$$W_{kl}(\boldsymbol{r}, \boldsymbol{r}', \omega) = \int_{-\infty}^{+\infty} C_{kl}(\boldsymbol{r}, \boldsymbol{r}', \tau) \exp(\mathrm{i}\omega\tau) \, \mathrm{d}\tau \,. \tag{37}$$

This quantity is a measure of the spatial correlation of the field at the given frequency. Indeed, we may write the spectral correlation function in the form

$$\langle E_k(\boldsymbol{r}, \omega) E_l(\boldsymbol{r}', \omega') \rangle = 2\pi W_{kl}(\boldsymbol{r}, \boldsymbol{r}', \omega) \delta(\omega - \omega') \,. \tag{38}$$

The presence of the Dirac distribution $\delta(\omega - \omega')$ shows that there is no correlation between the components at two different frequencies, which is a direct consequence of the stationarity assumption. The same two expressions also hold for the current density $\boldsymbol{j}(\boldsymbol{r}, t)$.

Fluctuation–Dissipation Theorem

We have seen that the interesting quantities in heat transfer are quadratic in the field, and hence also in the current density (the relationship between the latter being linear). The fundamental quantity is therefore the spectral correlation function of the currents in the emitting medium. When this medium is in thermodynamic equilibrium at temperature T, this correlation function is given by a general theorem of statistical physics called the fluctuation–dissipation theorem [23]. The result can be written in the form

$$\langle j_k(\boldsymbol{r}, \omega) j_l(\boldsymbol{r}', \omega') \rangle = 4\pi\omega\varepsilon_0 \varepsilon''(\omega) \Theta(\omega, T) \delta_{kl} \delta(\boldsymbol{r} - \boldsymbol{r}') \delta(\omega - \omega') \,, \tag{39}$$

where δ_{kl} is the Kronecker delta and $\Theta(\omega, T)$ is the mean energy of a harmonic oscillator in equilibrium at temperature T. The quantity $\varepsilon''(\omega)$ is the imaginary part of the dielectric constant which describes absorption in the medium. We observe here that, if the medium is non-absorbent so that $\varepsilon'' = 0$, it does not emit thermal radiation.

Relationship Between Field and Currents. Green Function

Once the current correlation function is known, the only remaining problem is to relate the radiated field to the currents, whereupon all the important quantities, e.g., energy density, exchanged fluxes, etc., can be calculated. Since Maxwell's equations are linear, the most general relation is

$$E_k(\boldsymbol{r},\omega) = \mathrm{i}\mu_0\omega \int_V G_{kl}(\boldsymbol{r},\boldsymbol{r}',\omega) j_l(\boldsymbol{r}',\omega)\,\mathrm{d}^3\boldsymbol{r}'\,, \tag{40}$$

where the integral is taken over the whole volume of the body and summation is implied over repeated indices (Einstein convention). The quantity $G_{kl}(\boldsymbol{r},\boldsymbol{r}',\omega)$ is a tensor because the field is not generally collinear with the current. This is the Green tensor, or simply the Green function. To calculate it, we must calculate the radiation from a current element (or point dipole) in the given configuration. The problem of thermal radiation is thus reduced to the purely classical problem of radiation from an antenna. This extremely powerful method was developed by *Rytov* in the 1950s [23]. In the context of short-scale radiative transfer, it has the great advantage that it involves no limiting assumptions about the radiation. Indeed, problems of interference, diffraction, near field, and polarisation are all rigorously accounted for in this approach. The only restriction is the use of a dielectric constant to describe the medium, whence the medium is treated macroscopically.

This method can be used to calculate the near-field thermal emission from a flat source [24] or the radiative transfer between bodies separated by nanometric distances [25]. The method is described and exemplified in [26], for example. In the context of the present book, a detailed discussion of the application to radiative transfer in nanostructures will be given in the Chapter by *Joulain* in this volume.

Appendix. Examples of Phase Functions

Angular Dependence

The phase function depends on the angles of incidence and scattering. For particles with some symmetry, e.g., spheres, the phase function depends only on the relative angle between the two directions. In spherical coordinates, we have $\boldsymbol{u}(\theta,\phi)$ and $\boldsymbol{u}'(\theta',\phi')$. The cosine $\cos\Theta = \boldsymbol{u}\cdot\boldsymbol{u}'$ of the relative angle is given by

$$\cos\Theta = \mu\mu' + \sqrt{1-\mu^2}\sqrt{1-\mu'^2}\cos(\phi-\phi')\,, \tag{41}$$

where $\mu = \cos\theta$ and $\mu' = \cos\theta'$.

Constant Phase Function

The simplest phase function is the constant phase function $p(\cos\Theta) = 1$. Scattering is isotropic and the anisotropy factor g is exactly zero.

Rayleigh Phase Function

If the particle is non-absorbent and very small compared with the wavelength of the incident wave (a so-called Rayleigh particle), it behaves as an electric dipole. If the incident wave is non-polarised, the phase function is then

$$p(\cos\Theta) = \tfrac{3}{4}(1 + \cos^2\Theta)\,. \tag{42}$$

Henyey–Greenstein Phase Function

A widely used model is that associated with the Henyey–Greenstein phase function. This function depends only on the two parameters $\cos\Theta$ and g:

$$p(\cos\Theta) = \frac{1 - g^2}{\sqrt{(1 + g^2 - 2g\cos\Theta)^3}}\,. \tag{43}$$

In particular, this phase function satisfies the normalisation specified in (8).

Mie Phase Function

The Mie theory proceeds by exact solution of the scattering of a plane electromagnetic wave by a homogeneous and optically isotropic spherical particle. Given the refractive index n_p of the particle, the index n_h of the host medium, the particle radius r, and the wavelength $\lambda = \lambda_0/n_\mathrm{h}$ of the incident wave in the host medium, the Mie theory delivers the extinction, scattering and absorption cross-sections of the particle, together with the albedo a, the phase function $p(\cos\Theta)$, and the anisotropy parameter g, all as a function of the particle size parameter $X = 2\pi r/\lambda$. These are the properties of a single particle. If the scattering by different particles is assumed to be independent and we know the concentration of particles in the medium, we can obtain all the parameters of the RTE, e.g., extinction length, scattering length, absorption length, and phase function of the medium, in the form of simple relations.

Expansion in Terms of Legendre Polynomials

In the general case, the phase function $p_\nu(\cos\Theta)$ can be expanded in terms of Legendre polynomials:

$$p(\cos\Theta) = \sum_{n=0}^{M} a_n P_n(\cos\Theta)\,. \tag{44}$$

The a_n are expansion coefficients and the $P_n(\cos\Theta)$ are the Legendre polynomials of degree n. The series converges after a finite number of terms depending on the required accuracy. The number of terms is determined by the anisotropy of the phase function. The more anisotropic it is, the more terms are required in the expansion. For example, if we expand the Henyey–Greenstein phase function in terms of Legendre polynomials, we obtain the relation $a_n = g^n$.

References

[1] S. Chandrasekhar: *Radiative Transfer* (Dover, New York 1960) 55, 58, 60

[2] K. M. Case, P. F. Zweifel: *Linear Transport Theory* (Addison-Wesley, Reading, Massachusetts 1967) 55, 61

[3] L. A. Apresyan, Y. A. Kravtsov: *Radiation Transfer: Statistical and Wave Aspects* (Gordon and Breach, Amsterdam 1996) 56

[4] A. Ishimaru: *Wave Propagation and Scattering in Random Media* (IEEE Press, Piscataway 1997) 60, 61, 63, 66

[5] G. E. Thomas, K. Stamnes: *Radiative Transfer in the Atmosphere and Ocean* (Cambridge University Press, Cambridge 1999) Chap. 8 60

[6] A. Mandelis: Diffusion of waves and their uses, Phys. Tod. **53**, 29–34 (2000) 61

[7] A. Majumdar: Microscale heat conduction in dielectric thin films, J. Heat Transfer **115**, 7–16 (1993) 64

[8] G. Chen: Size and interface effects on thermal conductivity of superlattices and periodic thin-film structures, J. Heat Transfer **119**, 220–229 (1997) 64

[9] R. Aronson, N. Corngold: Photon diffusion coefficient in an absorbing medium, J. Opt. Soc. Am. A **16**, 1066–1071 (1999) 64

[10] R. Graaf, J. J. Ten Bosch: Diffusion coefficient in photon diffusion theory, Opt. Lett. **25**, 43–45 (2000) 64

[11] R. Pierrat, J.-J. Greffet, R. Carminati: Photon diffusion coefficient in scattering and absorbing media, J. Opt. Soc. Am. A **23**, 1106–1110 (2006) 64

[12] R. H. J. Kop, P. de Vries, R. Sprik, A. Lagendijk: Observation of anomalous transport of strongly multiple scattered light in thin disordered slabs, Phys. Rev. Lett. **79**, 4369–4372 (1997) 65, 66, 67, 68

[13] R. Elaloufi, R. Carminati, J.-J. Greffet: Time-dependent transport through scattering media: From radiative transfer to diffusion, J. Opt. A: Pure Appl. Opt. **4**, S103–S108 (2002) 65

[14] R. Elaloufi: *Propagation du rayonnement en milieu diffusant. Etude de la transition en régime balistique et régime diffusif et des fluctuations temporelles d'intensité*, Ph.D. thesis, Ecole Centrale Paris (2003) 66, 67

[15] R. Aronson: Boundary conditions for diffusion of light, J. Opt. Soc. Am. A **12**, 2532–2539 (1995) 66

[16] A. D. Kim, A. Ishimaru: Optical diffusion of continuous wave, pulsed and density waves in scattering media and comparisons with radiative transfer, Appl. Opt. **37**, 5313–5319 (1998) 67

[17] Z. Q. Zhang, I. P. Jones, H. P. Schriemer, J. H. Page, D. A. Waitz, P. Sheng: Wave transport in random media: The ballistic to diffusive transition, Phys. Rev. E **60**, 4843–4850 (1999) 67
[18] K. Mitra, S. Kumar: Development and comparison of models for light-pulse transport through scattering–absorbing media, Appl. Opt. **38**, 188–196 (1999) 67
[19] X. Zhang, Z. Q. Zhang: Wave transport through thin slabs of random media with internal reflection: Ballistic to diffusive transition, Phys. Rev. E **66**, 016612 (2002) 67
[20] R. Elaloufi, R. Carminati, J.-J. Greffet: Diffusive-to-ballistic transition in dynamic light transmission through thin scattering slabs: A radiative transfer approach, J. Opt. Soc. Am. A **21**, 1430–1437 (2004) 67
[21] J. X. Zhu, D. J. Pine, D. A. Weitz: Internal reflection of diffusive light in random media, Phys. Rev. A **44**, 3948–3959 (1991) 67
[22] K. E. Goodson, M. I. Flik, L. T. Su, D. A. Antoniadis: Prediction and measurement of the thermal conductivity of amorphous dielectric layers, J. Heat Transfer **116**, 317–324 (1994) 68, 69
[23] S. M. Rytov, Y. A. Kravtsov, V. I. Tatarskii: *Principles of Statistical Radiophysics*, vol. 3 (Springer, Berlin 1989) Chap. 3 71, 72
[24] R. Carminati, J.-J. Greffet: Near-field effects in spatial coherence of thermal sources, Phys. Rev. Lett. **82**, 1660 (1999) 72
[25] J. P. Mulet, K. Joulain, R. Carminati, J.-J. Greffet: Enhanced radiative heat transfer at nanometric distances, Microscale Thermophysical Eng. **6**, 209–222 (2002) 72
[26] K. Joulain, J. P. Mulet, F. Marquier, R. Carminati, J. J. Greffet: Surface electromagnetic waves thermally excited: Radiative heat transfer, coherence properties and Casimir forces revisited in the near field, Surf. Sci. Rep. **57**, 59–112 (2005) 72

Index

Solution of the Boltzmann Equation for Phonon Transport

Denis Lemonnier

Laboratoire d'Etudes Thermiques (UMR CNRS 6608), ENSMA, Téléport 2, BP 40109, 86961 Futuroscope Chasseneuil Cedex
denis.lemonnier@ensma.fr

Abstract. We discuss two popular methods for solving the radiative transfer equation in the field of thermal radiation, which can be used to calculate conduction on nanoscales under certain hypotheses. After a brief summary of the theory leading to a radiative transfer equation for phonons, we present the P_1 method and the discrete ordinate method. The first is based on a global approach to transfer in which the only unknown is the local internal energy. In particular, it uses an approximate treatment of the boundary conditions and for this reason becomes somewhat inaccurate when transfer is dominated by ballistic phonons from the bounding surfaces. The second method takes into account the directional aspect of the transfer and yields better results than the P_1 method, except near the diffusive regime. It solves a transport equation in a discrete set of directions. Integrated quantities such as the internal energy and flux are evaluated using quadrature formulas. Whereas the partial differential equation derived in the P_1 approach can be solved by standard methods, the numerical system associated with the discrete ordinate method is more specific, particularly in cylindrical geometries.

1 Introduction

Heat conduction is no longer correctly described by the Fourier law in submicron scale objects (10–100 nm). One must go back to the basic physics of the phenomenon which, in non-metallic materials, is the physics of vibrations in a crystal lattice. In this case, the main energy carriers are phonons. These are sufficiently analogous[1] to photons to justify defining a phonon intensity as

[1] The behaviour of phonons and photons is similar in several ways, whether the similarity be real or contrived in some sense by hypothesis. In particular, they are both bosons, i.e., their equilibrium distribution is the Bose–Einstein distribution. These quasi-particles are treated as classical particles beyond a certain length scale, i.e., the coherence length for phonons and the wavelength for photons, and their propagation is described in each case by a Boltzmann equation. Moreover, the relaxation time approximation used to describe collisions between phonons leads to the definition of a kind of absorption coefficient, as for photons, despite the different physics. (Photons interact together, whereas photons interact with matter.) Finally, when we introduce the Debye model, the equilibrium intensity of the phonons is analogous to a Planck function, with the speed of sound rather than the speed of light (and up to a factor of 3/2).

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 77–106 (2007)

well as a transfer equation that is identical under certain hypotheses to the one used for thermal radiation in absorbing–emitting media. For this reason, several computational methods taken over from radiative transfer (or neutronics, where these methods were originally developed) can be used to make predictions concerning conduction on nanoscales.

In this Chapter, we present two of the most popular of these methods: the P_1 method and the discrete ordinate method. The first is more succinct. It does not directly use the intensity as variable. As we shall see, the intensity is a quantity depending on the direction of propagation. Instead, it uses an integrated quantity proportional to the internal energy. The method leads to a simple second order partial differential equation, which is easy to solve by standard numerical methods. However, this succinctness is obtained at the expense of the ballistic component of transfer (arising from the bounding surfaces), and for this reason the P_1 method is only applicable in media with dimensions significantly greater than the mean free path of the energy carriers. On the other hand, the discrete ordinate method completely accounts for the directional aspect of the intensity. It involves solving the transfer equation in a finite number of directions and calculating the integrated quantities, such as the internal energy or the flux, by quadrature formulas. Although this method has a reputation for accuracy, and economy in terms of computer time, these general considerations must be reassessed in each case. However, it has become very popular and is today the most widely used method in the context of thermal radiation.

Before going into the details of the two methods, we shall outline the main features of the theoretical model used to describe heat conduction on nanoscales as energy transport by phonons. The reader wishing to go further with this theory should refer to the basic textbooks on solid state physics, e.g., [1–4].

2 Theoretical Model

2.1 Intensity. Internal Energy. Flux

In insulating and semiconducting materials at room temperature, heat propagates mainly by vibrations of the atoms making up the crystal lattice. Quantum mechanics shows that the energy of each mode is a multiple of $\hbar\omega$, where ω is the angular frequency and $\hbar$ is Planck's constant divided by 2π. This introduces the idea of the phonon as both a vibrational quantum of the lattice and a quasi-particle of zero mass carrying energy $\hbar\omega$. Beyond a certain length known as the coherence length, phonons behave like classical

particles[2] and, by analogy with thermally radiated photons, one can define a phonon radiative intensity [5]:

$$I_\omega(\boldsymbol{x}, t, \boldsymbol{s}) = \frac{1}{4\pi} D(\omega) n_\omega(\boldsymbol{x}, t, \boldsymbol{s}) \hbar\omega v_\omega \,, \tag{1}$$

where $D(\omega)$ is the number of modes per unit volume with angular frequency ω, while v_ω is the group velocity and n_ω the average number of phonons[3] with angular frequency ω moving in the direction of the unit vector $\boldsymbol{s}$. At any point $\boldsymbol{x}$ in this phonon gas, the internal energy density and the flux can be calculated as spectral quantities and as total quantities by:

$$e_\omega(\boldsymbol{x}, t) = \frac{1}{v_\omega} \int_{4\pi} I(\boldsymbol{x}, t, \omega, \boldsymbol{s}) \, \mathrm{d}\Omega \,, \tag{2}$$

$$e(\boldsymbol{x}, t) = \int_0^{\omega_{\max}} e_\omega(\boldsymbol{x}, t) \, \mathrm{d}\omega \,, \tag{3}$$

$$\boldsymbol{q}''_\omega(\boldsymbol{x}, t) = \int_{4\pi} I(\boldsymbol{x}, t, \omega, \boldsymbol{s}) \boldsymbol{s} \, \mathrm{d}\Omega \,, \tag{4}$$

$$\boldsymbol{q}''(\boldsymbol{x}, t) = \int_0^{\omega_{\max}} \boldsymbol{q}''_\omega(\boldsymbol{x}, t) \, \mathrm{d}\omega \,. \tag{5}$$

The integrals are carried out over all directions in space, i.e., 4π steradians, so that $\mathrm{d}\Omega$ denotes an elementary solid angle around the direction $\boldsymbol{s}$, and in (3) and (5), over the whole spectrum up to the cutoff frequency $\omega_{\max}$ imposed by the periodicity of the crystal lattice.

When the medium is locally in equilibrium at temperature T, the number of phonons has a Bose–Einstein distribution and the internal energy is given by

$$e \equiv e^0(T) = \int_0^{\omega_{\max}} \left(\mathrm{e}^{\hbar\omega / k_\mathrm{B} T} - 1 \right)^{-1} D(\omega) \hbar\omega \, \mathrm{d}\omega \,, \tag{6}$$

[2] One then considers that the phonons have lost all information concerning their wavelike character, and in particular concerning the phase differences from one phonon to another. This loss of coherence is observed physically over distances of the order of the nanometer due to random superposition of a large number of interference patterns, whereupon it is no longer possible to observe any interference fringe.

[3] To simplify the notation, we assume that the quantities appearing here, i.e., D, n_ω, and v_ω, are averaged over all branches of the dispersion curves of the material.

which yields in the Debye model[4] [2]

$$e^0(T) = \frac{4\sigma_{\mathrm{ph}} T^4}{c_{\mathrm{s}}} F(\Theta_{\mathrm{D}}/T)\,, \quad F(z) = \frac{\displaystyle\int_0^z \frac{x^3\,\mathrm{d}x}{\mathrm{e}^x - 1}}{\displaystyle\int_0^\infty \frac{x^3\,\mathrm{d}x}{\mathrm{e}^x - 1}}\,, \tag{7}$$

where Θ_{D} is the Debye temperature of the material, F is the fraction of the equilibrium radiation of phonons between zero and the cutoff frequency, c_{s} is the speed of sound in the material, and σ_{ph} is the Stefan–Boltzmann constant of the phonons given by

$$\sigma_{\mathrm{ph}} = \frac{\pi^2 k_{\mathrm{B}}^4}{40 c_{\mathrm{s}}^2 \hbar^3}\,. \tag{8}$$

2.2 Transfer Equation

Phonon propagation is governed by the transport equation

$$\frac{\partial n_\omega}{\partial t} + v_\omega \boldsymbol{s}\cdot\boldsymbol{\nabla} n_\omega = \left.\frac{\mathrm{d}n_\omega}{\mathrm{d}t}\right|_{\mathrm{col}} \approx \frac{n_\omega^0 - n_\omega}{\tau_\omega}\,, \tag{9}$$

in which the right-hand side accounts for the rates of creation and destruction of phonons after collisions. Since it is difficult to obtain an exact expression for this term, one appeals almost exclusively to the relaxation time approximation [2, 4]. This consists in assuming that the source term is proportional to the difference $n_\omega^0 - n_\omega$, where n_ω^0 is the equilibrium distribution of phonons with angular frequency ω at the given local temperature. The proportionality factor is the reciprocal of a time τ_ω which, multiplied by the speed of propagation, gives the phonon mean free path between consecutive collisions, viz., $\Lambda_\omega = v_\omega \tau_\omega$. Multiplying (9) by $D\hbar\omega/4\pi$, we obtain

$$\frac{1}{v_\omega}\frac{\partial I_\omega}{\partial t} + \boldsymbol{s}\cdot\boldsymbol{\nabla} I_\omega = \kappa_\omega (I_\omega^0 - I_\omega)\,, \tag{10}$$

where we have set $\kappa_\omega = \Lambda_\omega^{-1}$. This new equation is formally identical to the radiative transfer equation in an absorbing–emitting material and is often called the phonon radiative transfer equation. Here, κ_ω plays the role of

[4] This model amounts to assuming that the group velocity v_ω is independent of ω and equal to the speed of sound c_{s} in the material, and also that [2]

$$D(\omega) = \frac{3\omega^2}{2\pi^2 c_{\mathrm{s}}^3}\,, \quad \omega \le \omega_{\mathrm{D}}\,,$$

where ω_{D} is the cutoff frequency related to the Debye temperature Θ_{D} of the material by $\omega_{\mathrm{D}} = k_{\mathrm{B}}\Theta_{\mathrm{D}}/\hbar$ and k_{B} is the Boltzmann constant.

an absorption coefficient[5] and $I^0_\omega(T) = D(\omega)n^0_\omega(T)v_\omega/4\pi$ is the equilibrium radiation intensity at the local temperature T. Integrating (10) over the whole spectrum and all directions in space, it follows that

$$\frac{\partial e}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{q}'' = \int_0^{\omega_{\max}} \left(4\pi I^0_\omega - \int_{4\pi} I_\omega \,\mathrm{d}\Omega\right) \mathrm{d}\omega\,. \tag{11}$$

When there is no internal heat source, the left-hand side of this equation is zero, implying the same of the right-hand side. This fixes the radiative equilibrium conditions for phonons:[6]

$$\int_0^{\omega_{\max}} \int_{4\pi} I_\omega \,\mathrm{d}\Omega\,\mathrm{d}\omega = 4\pi \int_0^{\omega_{\max}} I^0_\omega \,\mathrm{d}\omega\,, \tag{12}$$

with, in the Debye model,

$$\int_0^{\omega_{\max}} I^0_\omega \,\mathrm{d}\omega = \frac{\sigma_{\mathrm{ph}} T^4}{\pi} F(\Theta_{\mathrm{D}}/T)\,. \tag{13}$$

Boundary Conditions

In order to solve the transfer equation (10), one needs to know the intensity entering the medium at its boundaries. This amounts to fixing I_ω for all directions $\boldsymbol{s}$ such that $\boldsymbol{s}\cdot\boldsymbol{n} > 0$, where $\boldsymbol{n}$ is the local inward normal. There are essentially two types of boundary condition. The first consists in attributing to I_ω a known value I^*_ω that depends only on the properties of the starting surface, and in particular its temperature. In this case, we set

$$I_\omega(\boldsymbol{x}_{\mathrm{w}}, \boldsymbol{s}, t) = I^*_\omega(\boldsymbol{x}_{\mathrm{w}}, \boldsymbol{s}, t)\,, \quad \text{for} \quad \boldsymbol{s}\cdot\boldsymbol{n} > 0\,, \tag{14}$$

which also fixes the normal component of the flux leaving the boundary:

$$q''_{\mathrm{out},\omega}(\boldsymbol{x}_{\mathrm{w}}, t) = \int_{\boldsymbol{s}\cdot\boldsymbol{n}>0} I^*_\omega(\boldsymbol{x}_{\mathrm{w}}, \boldsymbol{s}, t)\boldsymbol{s}\cdot\boldsymbol{n}\,\mathrm{d}\Omega\,. \tag{15}$$

The outgoing intensities are usually assumed to be isotropic, i.e., I^*_ω independent of $\boldsymbol{s}$, in particular if one chooses I^*_ω equal to the equilibrium intensity

[5] Unlike photons, phonons do not interact with matter. Strictly speaking they are not therefore either emitted or absorbed. However, they are created or destroyed during collisions, through mechanisms that essentially involve three bodies: two phonons give a phonon (destruction) or a phonon gives two phonons (creation). The term 'absorption coefficient' for κ_ω does not refer to a physical reality but arises from a formal analogy between the phonon and photon transfer equations.

[6] Note that this equality is only valid after integration over the whole spectrum. It cannot be applied separately at each frequency, since generally $\int_{4\pi} I_\omega \,\mathrm{d}\Omega\,\mathrm{d}\omega \neq 4\pi I^0_\omega$.

at the temperature of one of the walls,[7] i.e., $I_\omega^* = I_\omega^0(T_\mathrm{w})$. In this case, since $\int_{\boldsymbol{s}\cdot\boldsymbol{n}>0} \boldsymbol{s}\cdot\boldsymbol{n}\,\mathrm{d}\Omega = \pi$, we have the simple result

$$I_\omega(\boldsymbol{x}_\mathrm{w}, \boldsymbol{s}, t) = \frac{q''_{\mathrm{out},\omega}(\boldsymbol{x}_\mathrm{w}, t)}{\pi} \quad \text{independent of } \boldsymbol{s}\,. \tag{16}$$

A pure reflection condition can also be applied at the wall, which is then treated as an adiabatic surface. Phonon reflection is general considered to be either diffuse, i.e., the reflected flux is uniformly spread over all directions, or specular, i.e., mirror-type reflection, or indeed a combination of the two. In the general case, this leads to

$$I_\omega(\boldsymbol{x}_\mathrm{w}, \boldsymbol{s}, t) = \frac{p}{\pi}\int_{\boldsymbol{s}'\cdot\boldsymbol{n}<0} I_\omega(\boldsymbol{x}_\mathrm{w}, \boldsymbol{s}')|\boldsymbol{s}'\cdot\boldsymbol{n}|\,\mathrm{d}\Omega' + (1-p)I_\omega(\boldsymbol{x}_\mathrm{w}, \hat{\boldsymbol{s}})\,, \tag{17}$$

where p is the level of diffuse reflection and $\hat{\boldsymbol{s}}$ the direction symmetric to $\boldsymbol{s}$ with respect to the normal $\boldsymbol{n}$.

2.3 Diffusive Regime

When the characteristic size L of the medium becomes large compared with the mean free path of the phonons ($\kappa_\omega L \gg 1$), the intensity can be expanded in a Taylor series:

$$I_\omega = I_0 + \frac{1}{\kappa_\omega L} I_1 + \frac{1}{(\kappa_\omega L)^2} I_2 + \cdots\,. \tag{18}$$

Substituting this into (10) gives to first order,

$$I_\omega = I_\omega^0 - \frac{1}{\kappa_\omega}\boldsymbol{s}\cdot\boldsymbol{\nabla} I_\omega^0\,. \tag{19}$$

The standard thermal radiation calculation which leads to the definition of the Rosseland conductivity can be used here to give

$$\boldsymbol{q}'' \approx -\frac{4\pi}{3}\int_0^{\omega_\mathrm{max}} \frac{1}{\kappa_\omega}\nabla I_\omega^0\,\mathrm{d}\omega\,, \quad \kappa L \gg 1\,, \tag{20}$$

or, since I_ω^0 depends only on the local temperature,

$$\boldsymbol{q}'' \approx -\lambda\nabla T\,, \quad \lambda = \frac{4\pi}{3}\int_0^{\omega_\mathrm{max}} \frac{1}{\kappa_\omega}\frac{\partial I_\omega^0}{\partial T}\,\mathrm{d}\omega\,. \tag{21}$$

We thus retrieve the Fourier law in the limit of the diffuse regime which prevails on length scales much greater than the phonon mean free path.

[7] This corresponds to the black body assumption in thermal radiation.

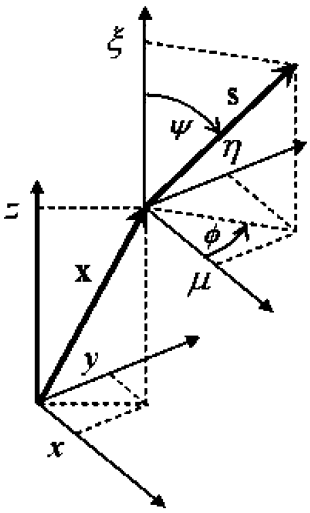

Fig. 1. Axes used to specify the position $\boldsymbol{x}$ of the field point and the direction of propagation $\boldsymbol{s}$ (Cartesian coordinates). ψ and ϕ are the polar angles specifying $\boldsymbol{s}$ and μ, η, and ξ are the direction cosines

3 The P_1 Method

One of the main difficulties when modelling radiative transfer of phonons, photons or any other particle stems from the directional nature of this mode of energy propagation. For this reason, one often seeks through approximate methods to treat the directional aspect in the most global manner possible in order to limit computation time. This is exemplified by the Rosseland model in the diffusive regime. The P_1 method presented in this section achieves this aim of avoiding directional calculations whilst remaining relatively reliable. It was first developed to model stellar radiation in astrophysics [6], then in neutronics [7], where the discrete ordinate and Monte Carlo methods were also devised. It takes its name from a class of methods known as the P_N methods, based on an expansion of the specific intensity in terms of a basis of orthogonal functions (the spherical harmonics associated with the Legendre polynomials P_N), truncated at order N. Experience shows that the order 1 expansion already gives good results, whereas higher order expansions often considerably complicate the calculations without significantly improving accuracy. A complete description of the method in the context of thermal radiation can be found in the book by *Modest* ([8], Chap. 15). An example application to phonon transfer is described in [9].

3.1 General Idea

In the stationary regime, the transfer equation (10) reduces to

$$\boldsymbol{s} \cdot \boldsymbol{\nabla} I_\omega + \kappa_\omega I_\omega = \kappa_\omega I_\omega^0 \,. \tag{22}$$

Solving this equation for a given angular frequency ω amounts to handling a five-variable integro-differential equation. There are three position variables,

i.e., the components of the vectors $\boldsymbol{x}$, and two variables to characterise the directions $\boldsymbol{s}$, i.e., two polar angles, or two of its three direction cosines, the third being related to the first two since the sum of their squares is equal to unity. The idea of the spherical harmonic method is to decompose the specific intensity relative to a basis of orthogonal functions (generalised Fourier series expansion):

$$I_\omega(\boldsymbol{x},\boldsymbol{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} I_l^m(\boldsymbol{x}) Y_l^m(\boldsymbol{s})\,. \tag{23}$$

Here, Y_l^m are the spherical harmonics defined by

$$Y_l^m(\Omega) = (-1)^{(m+|m|)/2} \left[\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} \mathrm{e}^{\mathrm{i}m\psi} P_l^{|m|}(\cos\psi)\,, \tag{24}$$

where ψ and ϕ are the polar angles (zenith and azimuth) characterising $\boldsymbol{s}$ in the local frame attached to the field point (see Fig. 1), and $P_l^{|m|}$ are the associated Legendre polynomials.[8] The point about this decomposition is that it replaces the unknown $I_\omega(\boldsymbol{x},\boldsymbol{s})$ by the coefficients $I_l^m(\boldsymbol{x})$ which do not depend on the direction.[9]

In theory, we now simply replace I_ω by its expansion in the transfer equation (22), then multiply the resulting equation successively by each of the functions Y_l^m and integrate the result over all directions. The orthogonality of the spherical harmonics means that we obtain as many partial differential equations as unknowns I_l^m, i.e., theoretically, an infinite number! In practice, of course, the expansion must be truncated. To first order, this yields[10]

$$I_\omega(\boldsymbol{x},\boldsymbol{s}) = I_0^0(\boldsymbol{x})Y_0^0(\boldsymbol{s}) + I_1^{-1}(\boldsymbol{x})Y_1^{-1}(\boldsymbol{s}) + I_1^0(\boldsymbol{x})Y_1^0(\boldsymbol{s}) + I_1^1(\boldsymbol{x})Y_1^1(\boldsymbol{s})\,, \tag{25}$$

[8] The associated Legendre polynomial (of the first kind) P_n^m is defined to be one of the solutions of the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left[(1-\xi^2)\frac{\mathrm{d}M}{\mathrm{d}\xi}\right] + \left[n(n+1) - \frac{m^2}{1-\xi^2}\right]\,, \quad -1 \le \xi \le 1\,.$$

When $m = n$, we obtain the standard Legendre polynomial P_n. The first polynomials P_n^m are mutually orthogonal in the sense that

$$\int_{-1}^{1} P_n^m(\xi) P_{n'}^m(\xi)\,\mathrm{d}\xi = 0\,, \quad \text{if} \quad n \neq n'\,.$$

[9] However, they do depend on the frequency. The subscript ω has been omitted here to simplify the notation.

[10] Experience shows that even order truncations are not very accurate. So to improve on the P_1 method, one must go to P_3. This brings in 16 unknown functions instead of 4 for P_1. Beyond this, calculation time becomes prohibitive.

or, since $P_0^0(\cos\psi) = 1$, $P_1^0(\cos\psi) = \cos\psi$, and $P_1^1(\cos\psi) = \sin\psi$,

$$I_\omega(\boldsymbol{x},\boldsymbol{s}) = I_0^0(\boldsymbol{x}) + \frac{I_1^{-1}(\boldsymbol{x})}{\sqrt{2}}\mathrm{e}^{-\mathrm{i}\phi}\sin\psi + I_1^0(\boldsymbol{x})\cos\psi - \frac{I_1^1(\boldsymbol{x})}{\sqrt{2}}\mathrm{e}^{\mathrm{i}\phi}\sin\psi\,. \quad (26)$$

This is the P_1 approximation. There are four unknown functions: I_0^0, I_1^{-1}, I_1^0, and I_1^1, and the expansion (23) amounts to representing the intensity by a relation of the form[11]

$$I_\omega(\boldsymbol{x},\boldsymbol{s}) \approx a_\omega(\boldsymbol{x}) + \boldsymbol{b}_\omega(\boldsymbol{x})\cdot\boldsymbol{s}\,, \quad (27)$$

where the four unknowns are now the scalar a_ω and the three components of the vector $\boldsymbol{b}_\omega$. Integrating I_ω then $\boldsymbol{s}I_\omega$ over all directions of space, we obtain the relations[12]

$$a_\omega(\boldsymbol{x}) = \frac{1}{4\pi}\int_{4\pi} I_\omega(\boldsymbol{x},\boldsymbol{s})\,\mathrm{d}\Omega = \frac{G_\omega(\boldsymbol{x})}{4\pi}\,, \quad (28)$$

$$\boldsymbol{b}_\omega(\boldsymbol{x}) = \frac{3}{4\pi}\int_{4\pi} I_\omega(\boldsymbol{x},\boldsymbol{s})\boldsymbol{s}\,\mathrm{d}\Omega = \frac{3}{4\pi}\boldsymbol{q}''_\omega(\boldsymbol{x})\,, \quad (29)$$

where G_ω is the incident spectral radiation, related to the internal energy by

$$e(\boldsymbol{x}) = \int_0^{\omega_{\max}} \frac{G_\omega(\boldsymbol{x})}{v_\omega}\,\mathrm{d}\omega\,, \quad (30)$$

and $\boldsymbol{q}''_\omega$ is the spectral flux. At this stage, we still have four unknowns, i.e., G_ω and the three components of $\boldsymbol{q}''_\omega$. To proceed from here, we return to the local energy balance (11) which gives, in the stationary regime,

$$\boldsymbol{\nabla}\cdot\boldsymbol{q}''_\omega = \kappa_\omega(4\pi I_\omega^0 - G_\omega)\,. \quad (31)$$

[11] Since $\mathrm{e}^{\mathrm{i}\phi} = \cos\phi + \mathrm{i}\sin\phi$, (26) can be rewritten in the form

$$I_\omega(\boldsymbol{x},\boldsymbol{s}) \approx \underbrace{I_0^0(\boldsymbol{x})}_{a_\omega(\boldsymbol{x})} + \underbrace{\frac{I_1^{-1}(\boldsymbol{x}) - I_1^1(\boldsymbol{x})}{\sqrt{2}}}_{b_{\omega1}(\boldsymbol{x})}\underbrace{\sin\psi\cos\phi}_{\mu} - \underbrace{\mathrm{i}\frac{I_1^{-1}(\boldsymbol{x}) + I_1^1(\boldsymbol{x})}{\sqrt{2}}}_{b_{\omega2}(\boldsymbol{x})}\underbrace{\sin\psi\sin\phi}_{\mu} + \underbrace{I_1^0(\boldsymbol{x})}_{b_{\omega3}(\boldsymbol{x})}\underbrace{\cos\psi}_{\xi}\,,$$

which reads $I_\omega(\boldsymbol{x},\boldsymbol{s}) \approx a_\omega(\boldsymbol{x}) + \boldsymbol{b}_\omega(\boldsymbol{x})\cdot\boldsymbol{s}$, where (μ,η,ξ) are the components of $\boldsymbol{s}$ and $(b_{\omega1}, b_{\omega2}, b_{\omega3})$ are the components of the vector $\boldsymbol{b}_\omega$.

[12] Note that $\int_{4\pi}\mathrm{d}\Omega = 4\pi$, $\int_{4\pi}\boldsymbol{s}\,\mathrm{d}\Omega = 0$, and $\int_{4\pi}\boldsymbol{s}\boldsymbol{s}\,\mathrm{d}\Omega = 4\pi\boldsymbol{\delta}/3$, where $\boldsymbol{\delta}$ is the unit tensor with components $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. Here $\boldsymbol{s}\boldsymbol{s}$ is the symmetric second rank tensor whose components are the products of the direction cosines of $\boldsymbol{s}$ taken in pairs, viz., μ^2, $\mu\eta$, $\mu\xi$, η^2, $\eta\xi$, and ξ^2.

This also corresponds to the transfer equation (22) integrated over all directions. Repeating this integration but after multiplying (22) by the vector $\boldsymbol{s}$, we obtain[13]

$$\int_{4\pi} \boldsymbol{s}\boldsymbol{\nabla}\cdot(\boldsymbol{s}I_\omega)\,\mathrm{d}\Omega + \kappa\boldsymbol{q}''_\omega = 0\,. \tag{32}$$

Now since I_ω is approximated by $(G_\omega + 3\boldsymbol{s}\cdot\boldsymbol{q}''_\omega)/4\pi$,

$$\boldsymbol{\nabla}\cdot(\boldsymbol{s}I_\omega) = \frac{1}{4\pi}\boldsymbol{\nabla}\cdot(\boldsymbol{s}G_\omega) + \frac{3}{4\pi}\boldsymbol{\nabla}\cdot\left[\boldsymbol{s}(\boldsymbol{s}\cdot\boldsymbol{q}''_\omega)\right]\,, \tag{33}$$

and a rather tedious expansion shows that[14]

$$\int_{4\pi} \boldsymbol{s}\boldsymbol{\nabla}\cdot(\boldsymbol{s}I_\omega)\,\mathrm{d}\omega = \frac{1}{3}\boldsymbol{\nabla}G_\omega\,. \tag{34}$$

Hence (32) implies that

$$\boldsymbol{q}''_\omega(\boldsymbol{x}) = -\frac{1}{3\kappa_\omega}\boldsymbol{\nabla}G_\omega(\boldsymbol{x})\,, \tag{35}$$

which means that the flux derives from a potential $-G_\omega/3\kappa_\omega$. Substituting this in (31), we have

$$\boldsymbol{\nabla}\cdot\left[-\frac{1}{3\kappa_\omega}\boldsymbol{\nabla}G_\omega\right] = \kappa\left(4\pi I^0_\omega - G_\omega\right)\,, \tag{36}$$

so that, when κ_ω is independent of $\boldsymbol{x}$,

$$\nabla^2 G_\omega(\boldsymbol{x}) - 3\kappa_\omega^2 G_\omega(\boldsymbol{x}) = -12\pi\kappa_\omega^2 I^0_\omega\,. \tag{37}$$

This is a Helmholtz equation. Its solution provides the incident radiation field over the whole region. Once G_ω is known, the radiative flux is obtained from (35) and the internal energy from (30).

In a grey medium, i.e., one in which κ_ω does not depend on the frequency so that $\kappa_\omega = \kappa$ is constant, (37) applies to total quantities and we may drop the subscript ω. The equilibrium condition (12) then gives

$$\int_{4\pi} I\,\mathrm{d}\Omega = G = 4\pi I^0\,. \tag{38}$$

The P_1 model then reduces to the simple Laplace equation

$$\nabla^2 G = 0\,, \quad \text{or} \quad \nabla^2 I^0 = 0\,, \tag{39}$$

given that $I^0(T) = e^0(T)/4\pi$.

[13] Note that we can write $\boldsymbol{\nabla}\cdot(\boldsymbol{s}I_\omega)$ in the place of $\boldsymbol{s}\cdot\boldsymbol{\nabla}I_\omega$, since $\boldsymbol{s}$ is a constant vector.

[14] We use once again the results $\int_{4\pi}\mathrm{d}\Omega = 4\pi$, $\int_{4\pi}\boldsymbol{s}\,\mathrm{d}\Omega = 0$, and $\int_{4\pi}\boldsymbol{ss}\,\mathrm{d}\Omega = 4\pi\boldsymbol{\delta}/3$, together with $\int_{4\pi}\boldsymbol{sss}\,\mathrm{d}\Omega = \boldsymbol{0}$, where $\boldsymbol{sss}$ is the rank three tensor whose components are the products of the direction cosines of $\boldsymbol{s}$ taken in threes, i.e., μ^3, $\mu^2\eta$, $\mu^2\xi$, $\mu\eta\xi$, etc.

3.2 Boundary Conditions

In order to solve the differential equation (37), one needs to be able to specify boundary conditions for G_ω. Now those we have, viz., (16) and (17), are given in terms of intensities and there is no exact way of expressing these conditions in terms of incident radiation. The most widely used approximate model is the Marschak model [10] based on global conservation of the normal component of the flux through the boundary. Hence using the approximate expression

$$I_\omega(\boldsymbol{x}, \boldsymbol{s}) \approx \frac{1}{4\pi}\Big[G_\omega(\boldsymbol{x}) + 3\boldsymbol{q}''_\omega(\boldsymbol{x}) \cdot \boldsymbol{s}\Big]$$

for the intensity, the flux entering into the medium along the normal $\boldsymbol{n}$ at a point $\boldsymbol{x}_\mathrm{w}$ on the bounding surface is given by[15]

$$J_\mathrm{w} = \int_{\boldsymbol{s}\cdot\boldsymbol{n}>0} I_\omega(\boldsymbol{x}_\mathrm{w}, \boldsymbol{s})\boldsymbol{s}\cdot\boldsymbol{n}\,\mathrm{d}\Omega = \frac{1}{4}\Big[G_\omega(\boldsymbol{x}_\mathrm{w}) + 2\boldsymbol{q}''_\omega(\boldsymbol{x}_\mathrm{w})\cdot\boldsymbol{n}\Big]\,.$$

Then, eliminating $\boldsymbol{q}''_\omega$ with the help of (35), we have

$$4J_\mathrm{w} = G_\omega(\boldsymbol{x}_\mathrm{w}) - \frac{2}{3\kappa_\omega}\frac{\partial G_\omega}{\partial n}(\boldsymbol{x}_\mathrm{w})\,, \tag{40}$$

where $\partial G_\omega/\partial n = \boldsymbol{n}\cdot\boldsymbol{\nabla}G_\omega$ denotes the derivative of G_ω along the normal to the wall. If (16) is applicable, then $J_\mathrm{w} = q''_{\mathrm{emit},\omega}(\boldsymbol{x}_\mathrm{w})$, which implies that

$$\frac{2}{3\kappa_\omega}\frac{\partial G_\omega}{\partial n}(\boldsymbol{x}_\mathrm{w}) - G_\omega(\boldsymbol{x}_\mathrm{w}) = -4q''_{\mathrm{emit},\omega}(\boldsymbol{x}_\mathrm{w})\,. \tag{41}$$

For a perfectly reflecting wall (without being able to distinguish the diffuse and specular cases, which are treated in the same way in the P_1 method), we have simply[16]

$$\frac{\partial G_\omega}{\partial n}(\boldsymbol{x}_\mathrm{w}) = 0\,. \tag{42}$$

3.3 Numerical Solution

No particular problem arises in solving the Helmholtz equation (37) with boundary conditions of type (41) or (42). It is a linear, second order, elliptic partial differential equation with linear boundary conditions. Good algorithms exist for solving this kind of problem, e.g., Gauss–Seidel method, ADI methods, or multigrid algorithms, and so on [11]. Note that we are treating the radiative problem here by solving a single partial differential equation, in contrast to methods where we work directly with the intensity (discrete ordinate method) and where there are as many equations to solve as discrete directions taken into account. This is the great advantage of the P_1 method, justifying its popularity even today.

[15] We use $\int_{\boldsymbol{s}\cdot\boldsymbol{n}>0} \boldsymbol{s}(\boldsymbol{s}\cdot\boldsymbol{n})\,\mathrm{d}\Omega = \boldsymbol{n}\int_{\boldsymbol{s}\cdot\boldsymbol{n}>0}\boldsymbol{s}\boldsymbol{s}\,\mathrm{d}\Omega$ and $\int_{\boldsymbol{s}\cdot\boldsymbol{n}>0}\boldsymbol{s}\boldsymbol{s}\,\mathrm{d}\Omega = 2\pi\boldsymbol{\delta}/3$.

[16] The reflection condition is equivalent to the insulation condition $\boldsymbol{q}''_\omega\cdot\boldsymbol{n} = 0$, or since $\boldsymbol{q}''_\omega \propto \boldsymbol{\nabla}G_\omega$, $\boldsymbol{\nabla}G_\omega\cdot\boldsymbol{n} = \partial G/\partial n = 0$.

3.4 Advantages and Disadvantages

The advantages of the P_1 method were discussed above: simple formulation, linear problem, only one equation to solve, and economical in computation time. However, its main drawback comes from the approximate treatment of the intensity and the fluxes from the bounding walls. The early truncation of the expansion of I_ω in the Marschak model (for the boundary conditions) implicitly assumes that the intensity has a certain degree of isotropy.[17] This is justified if we are close to the diffusive regime because in this case I_ω tends to equal I_ω^0, which is isotropic. However, it is not justified when ballistic transfer dominates. To remedy this situation, a modified version of the P_1 method has been developed, in which radiation is treated by separating the contribution from the walls and the contribution from the medium itself. If we have an exact method for handling wall–wall exchanges (taking into account attenuation in the medium), P_1 is only used to deal with phonons produced in the medium. The two solutions (radiation from the walls and radiation from the medium) are superposed to yield a complete solution [8, 9]. This time, in the limit of a purely ballistic medium, the method will be exact and P_1 maintains its good behaviour for highly absorbent media. This makes the method more precise globally.

4 Discrete Ordinate Method

The discrete ordinate method uses numerical quadrature to calculate integrals along solid angles arising in the calculation of internal energy and radiative fluxes. It begins by discretising the angle space in a finite number of directions in which the radiative transfer equation is then solved, so as to produce several values of the intensity at every point. The weighted sum of these intensities is then used to calculate the values of e_ω and $\boldsymbol{q}''_\omega$ locally.

This method is due to *Chandrasekhar* [12], but it was then largely developed for application to neutronics by *Carlson* and *Lathrop* [13]. Their work in the 1960s is still the reference today. The application of discrete ordinates to thermal radiation is mainly due to *Fiveland* [14–16]. A good description of the method can also be found in the book by *Modest* ([8], Chap. 16). It has also been applied recently to phonon transfer [17–23].

[17] Multiplying the transfer equation (22) by $\boldsymbol{s}$ and integrating over all directions, we obtain the relation $\boldsymbol{\nabla P}_\omega + \kappa_\omega \boldsymbol{q}''_\omega = \boldsymbol{0}$, where $\boldsymbol{P}_\omega$ is the radiative pressure tensor given by $\boldsymbol{P}_\omega = \int_{4\pi} \boldsymbol{ss} I_\omega \, \mathrm{d}\Omega$. If I_ω is isotropic, then $\boldsymbol{P}_\omega = 4\pi I_\omega \boldsymbol{\delta}/3$ (see previous footnotes) and $G_\omega = 4\pi I_\omega$. Hence $\boldsymbol{P}_\omega = G_\omega \boldsymbol{\delta}/3$ and we have

$$\tfrac{1}{3} \boldsymbol{\nabla} G_\omega + \kappa \boldsymbol{q}''_\omega = \boldsymbol{0} \,,$$

which corresponds to (35), used in P_1.

4.1 General Idea

Calculating the Internal Energy and Flux Vector

The discrete ordinate method proceeds as follows:

– The transfer equation is solved in a finite number of directions $\boldsymbol{s}_m$, $m = 1, \ldots, M$. For each direction, the intensity field[18] $I_m(\boldsymbol{x}, t) = I_\omega(\boldsymbol{x}, t, \boldsymbol{s}_m)$ is calculated over the whole region by solving

$$\frac{1}{v_\omega}\frac{\partial I_m}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{s}_m I_m + \kappa_\omega I_m = \kappa I^0(T)\,. \tag{43}$$

– The integrals appearing in (2) and (4) are calculated using quadrature formulas such as

$$e_\omega(\boldsymbol{x}, t) \approx \frac{1}{v_\omega}\sum_{m=1}^{M} w_m I_m(\boldsymbol{x}, t)\,, \tag{44}$$

and

$$\boldsymbol{q}''_\omega(\boldsymbol{x}, t) \approx \sum_{m=1}^{M} w_m I_m(\boldsymbol{x}, t)\boldsymbol{s}_m\,, \tag{45}$$

where w_m are the weights.

Calculating the Incident Boundary Fluxes and Expressing Boundary Conditions

The use of quadrature formulas extends to the calculation of boundary fluxes. Hence the radiative flux incident at a point $\boldsymbol{x}_\mathrm{w}$ on the wall can be calculated using

$$\begin{aligned}\boldsymbol{q}''_{\mathrm{inc},\omega}(\boldsymbol{x}_\mathrm{w}, t) &= \int_{\boldsymbol{s}\cdot\boldsymbol{n}<0} I_\omega(\boldsymbol{x}_\mathrm{w}, t, \boldsymbol{s})|\boldsymbol{s}\cdot\boldsymbol{n}|\,\mathrm{d}\Omega \\ &\approx \sum_{m' \text{ s.t. } \boldsymbol{s}_{m'}\cdot\boldsymbol{n}<0} w_{m'} I_{m'}(\boldsymbol{x}_\mathrm{w}, t)|\boldsymbol{s}_{m'}\cdot\boldsymbol{n}|\,,\end{aligned} \tag{46}$$

where $\boldsymbol{n}$ is the inward normal vector. Note that the quadrature formula only involves some of the M discrete directions here since the domain of integration is no longer 4π steradians, but 2π steradians.

The boundary conditions applying to I_m can be deduced directly from (16) and (17). They concern all directions m leaving the boundary, i.e., such that $\boldsymbol{s}_m\cdot\boldsymbol{n} > 0$. We thus impose

$$I_m(\boldsymbol{x}_\mathrm{w}, t) = I_{\mathrm{emit},\omega}(\boldsymbol{x}_\mathrm{w}, \boldsymbol{s}_m, t)\,, \tag{47}$$

[18] To simplify notation, the index ω is omitted for I_m although it is indeed a spectral quantity.

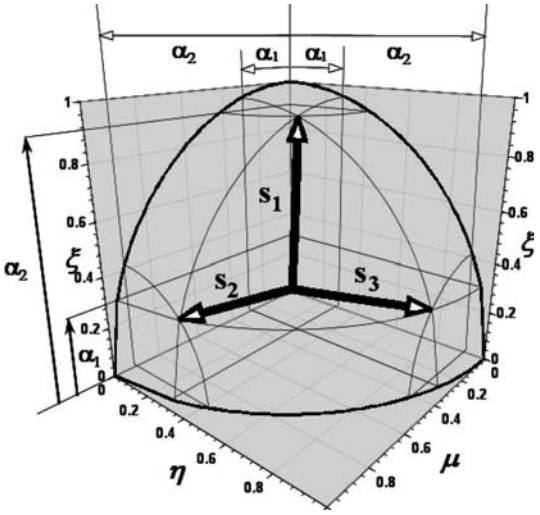

Fig. 2. Directions defining the S_4 quadrature (first octant)

if the emission from the wall is given, and

$$I_m(\boldsymbol{x}_\mathrm{w}, t) = \frac{p}{\pi} \sum_{m' \text{ s.t.} \boldsymbol{s}'_m \cdot \boldsymbol{n} < 0} w_{m'} I_{m'}(\boldsymbol{x}_\mathrm{w}, t) |\boldsymbol{s}'_m \cdot \boldsymbol{n}| + (1-p) I_{\hat{m}}(\boldsymbol{x}_\mathrm{w}, t)\,, \quad (48)$$

for pure reflection. The last expression assumes that the direction $\boldsymbol{s}_{\hat{m}}$ symmetrical to $\boldsymbol{s}_m$ with respect to $\boldsymbol{n}$ is also one of the M discrete directions. This is generally the case by virtue of the rules of symmetry used to construct the quadratures, provided the boundary is parallel to a coordinate plane.

4.2 Choice of Quadratures

S_N Quadrature

In multidimensional problems, the most commonly chosen quadratures are the S_N quadratures [13, 15]. They are constructed by imposing strict rules of symmetry to ensure that no direction of propagation is favoured. Hence, if $\boldsymbol{s}_m = (\mu_m, \eta_m, \xi_m)$ lies in the quadrature, then:

- All directions $(\pm\mu_m, \pm\eta_m, \pm\xi_m)$ also lie in the quadrature and have the same weight.
- All directions found by permuting (μ_m, η_m, ξ_m) also lie in the quadrature and are attributed the same weight.

One consequence of the first constraint is that it suffices to define the directions of the quadrature in one octant of the sphere, e.g., for $\mu, \eta, \xi > 0$, and then all others are determined by symmetry over the full 4π steradians of solid angle.

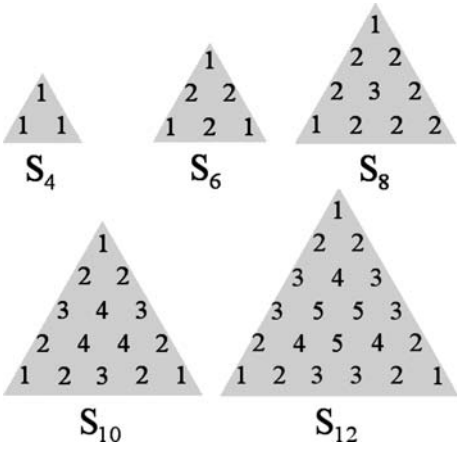

Fig. 3. Schematic representation of quadratures S_4 to S_{12} in the first octant of the sphere. All directions carrying the same number are attributed the same weight

To order N, the S_N quadrature comprises $M = N(N+2)$ directions. Hence S_2 comprises 8 directions, S_4 has 24 and S_6 has 48. However, in 2D configurations, the symmetries of the problem mean that one need only use half of the directions, multiplying each weight by 2.

In addition to the symmetry conditions already mentioned,[19] the weights w_m must satisfy

$$\sum_{m=1}^{M} w_m = \int_{4\pi} \mathrm{d}\Omega = 4\pi\,, \tag{49}$$

in order to provide a good representation of the radiative emission at all points in the medium. Note that another desirable condition,[20] viz.,

$$\sum_{m=1}^{M} \mu_m w_m = \int_{4\pi} \mu\,\mathrm{d}\Omega = 0\,, \tag{50}$$

is automatically satisfied due to the quadrature symmetries. Moreover, we have seen that the incident fluxes on the boundary walls are calculated by integrating, not over the full 4π sr, but over just half of the space. This is why we also generally require

$$\sum_{m\,\mathrm{s.t.}\,\mu_m<0}^{M} |\mu_m| w_m = \int_{2\pi\mathrm{s.t.}\,\mu<0} \mu\,\mathrm{d}\Omega = \pi\,, \tag{51}$$

[19] Any direction deduced from another by symmetry considerations must have the same weight.

[20] This condition refers equally to any of the direction cosines μ, η or ξ.

Table 1. Directions and weights for quadratures S_4, S_6 and S_8 (first octant) [24]

Quadrature	Direction cosines			Weight
	μ	η	ξ	w
S_4	0.295 875 9	0.295 875 9	0.908 248 3	0.523 598 7
	0.295 875 9	0.908 248 3	0.295 875 9	0.523 598 7
	0.908 248 3	0.295 875 9	0.295 875 9	0.523 598 7
S_6	0.183 867 0	0.183 867 0	0.965 601 3	0.160 951 7
	0.183 867 0	0.695 051 4	0.695 051 4	0.362 646 9
	0.183 867 0	0.965 601 3	0.183 867 0	0.160 951 7
	0.695 051 4	0.183 867 0	0.695 051 4	0.362 646 9
	0.695 051 4	0.695 051 4	0.183 867 0	0.362 646 9
	0.965 601 3	0.183 867 0	0.183 867 0	0.160 951 7
S_8	0.142 255 5	0.142 255 5	0.979 554 3	0.171 235 9
	0.142 255 5	0.577 350 3	0.804 008 7	0.099 228 4
	0.142 255 5	0.804 008 7	0.577 350 3	0.099 228 4
	0.142 255 5	0.979 554 3	0.142 255 5	0.171 235 9
	0.577 350 3	0.142 255 5	0.804 008 7	0.099 228 4
	0.577 350 3	0.577 350 3	0.577 350 3	0.461 717 9
	0.577 350 3	0.804 008 7	0.142 255 5	0.099 228 4
	0.804 008 7	0.142 255 5	0.577 350 3	0.099 228 4
	0.804 008 7	0.577 350 3	0.142 255 5	0.099 228 4
	0.979 554 3	0.142 255 5	0.142 255 5	0.171 235 9

in order to ensure accurate flux calculations over surfaces perpendicular to the x axis and, by symmetry, over any wall parallel to a coordinate plane. For S_6 and higher orders, further constraints are required to determine all the weights. These are generally supplied by imposing conservation of moments, e.g.,

$$\sum_{m=1}^{M} \mu_m^2 w_m = \int_{2\pi} \mu^2 \, \mathrm{d}\Omega = \frac{4\pi}{3} \, .$$

The reader will find a complete review of the generation of the S_N quadratures in [24].

Data for quadratures S_4, S_6 and S_8 are given in Table 1. A detailed list of quadratures S_2 to S_{12} can also be found in the appendix to [20].

Other Quadratures

Other quadratures can also be used with the discrete ordinate method. Hence, one can select a set of MN directions by choosing M and N discrete values of the polar angles characterising $\boldsymbol{s}$. They are generally attributed uniform weights $w = 4\pi/MN$. These quadratures may involve an arbitrary number

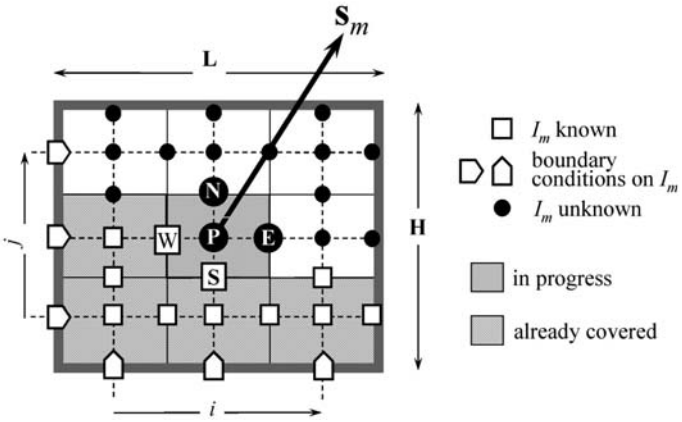

Fig. 4. Control volume for integrating the transfer equation in Cartesian coordinates

of directions[21] (up to restrictions on computation time) and are reputed to reduce certain inaccuracies due to the discretisation of the propagation directions.[22] A review of the various possible quadratures can be found in [21].

Setting up Cartesian Coordinates

For clarity, we restrict the discussion here to the stationary regime[23] and 2D geometries. The extension to three dimensions is immediate, at least in Cartesian geometries. In this case, the transfer equation becomes

$$\mu_m \frac{\mathrm{d}I_m}{\mathrm{d}x} + \eta_m \frac{\mathrm{d}I_m}{\mathrm{d}y} + \kappa_\omega I_m = \kappa_\omega I_\omega^0 \,. \tag{52}$$

Consider a rectangular region $\boldsymbol{x} = (x, y) \in [0, L] \times [0, H]$ covered by a uniform mesh with $I \times J$ cells of size $\Delta x \Delta y$. For each cell, the edges are specified with the usual conventions by indicating the four cardinal points, as shown in Fig. 4.

[21] The S_N quadratures are often limited to S_{20} because beyond this point, the construction rules may involve negative weights w_m.

[22] This concerns in particular the problem known as the ray effect (see Sect. 4.5). It is less important for quadratures with constant weights, but these generally require a greater number of directions to obtain an overall accuracy equivalent to results with the S_N quadratures.

[23] The use of the discrete ordinate method to handle non-stationary problems has recently come to the fore in connection with studies of radiative phenomena on ultrashort time scales (shorter than the time taken for light to cross the medium). Numerical diffusion and stability problems are then generally encountered in the time scheme. For more detail, see [20], for example.

4.3 Integrating the RTE over a Control Volume

Integrating (52) over a control volume gives

$$\mu_m A_x (I_{m,\mathrm{E}} - I_{m,\mathrm{W}}) + \eta_m A_y (I_{m,\mathrm{N}} - I_{m,\mathrm{S}}) + \kappa_\omega \Delta V I_{m,\mathrm{P}} = S_\mathrm{P} \Delta V \,, \quad (53)$$

where

$$A_x = \Delta y \,, \quad A_y = \Delta x \,, \quad \Delta V = \Delta x \Delta y \,, \quad S_\mathrm{P} = \kappa_\omega I^0_\omega(T) \,. \quad (54)$$

When the direction cosines μ_m and η_m are both positive, the direction of propagation is such that, for each control volume, the specific intensities are known on the edges W and S and unknown at the center P of the cell, and on the edges E and N (see Fig. 4). Two further relations are thus required to eliminate $I_{m,\mathrm{E}}$ and $I_{m,\mathrm{N}}$ and thereby calculate $I_{m,\mathrm{P}}$ explicitly. These relations are obtained by interpolation by assuming that

$$I_{m,\mathrm{P}} = I_{m,\mathrm{W}} + a(I_{m,\mathrm{E}} - I_{m,\mathrm{W}}) = I_{m,\mathrm{S}} + b(I_{m,\mathrm{N}} - I_{m,\mathrm{S}}) \,, \quad (55)$$

whereupon

$$I_{m,\mathrm{E}} = I_{m,\mathrm{P}} + \frac{I_{m,\mathrm{P}} - I_{m,\mathrm{W}}}{a} \,, \quad I_{m,\mathrm{N}} = I_{m,\mathrm{P}} + \frac{I_{m,\mathrm{P}} - I_{m,\mathrm{S}}}{b} \,. \quad (56)$$

The coefficients a and b vary between $1/2$ and 1 and may be different from one point to another. When $a = 1/2 = b$, interpolations are globally second order in Δx and Δy. This corresponds to the so-called diamond scheme. For all other values of a and b, accuracy is only first order but the stability of the scheme is better. This is the case in particular when $a = 1 = b$ (step scheme), which simply imposes $I_{m,\mathrm{E}} = I_{m,\mathrm{N}} = I_{m,\mathrm{P}}$.

Having eliminated $I_{m,\mathrm{E}}$ and $I_{m,\mathrm{N}}$ via the interpolations (56), (53) reduces to an explicit relation between the unknown intensity $I_{m,\mathrm{P}}$, the known values of $I_{m,\mathrm{E}}$ and $I_{m,\mathrm{N}}$, and the source term S_P :

$$I_{m,\mathrm{P}} = \frac{\lambda_x I_{m,\mathrm{W}} + \lambda_y I_{m,\mathrm{S}} + \Delta V S_\mathrm{P}}{\lambda_x + \lambda_y + \lambda_0} \,, \quad (57)$$

where

$$\lambda_x = \frac{\mu_m A_x}{a} \,, \quad \lambda_y = \frac{\eta_m A_y}{b} \,, \quad \lambda_0 = \kappa_\omega \Delta V \,. \quad (58)$$

Once $I_{m,\mathrm{P}}$ has been calculated, the values of the other unknowns $I_{m,\mathrm{E}}$ and $I_{m,\mathrm{N}}$ can be deduced from the interpolation formulas (56). Hence, assuming that we are working on cell (i, j), we can move to the calculation in cell $(i + 1, j)$ using:

– $I_{m,\mathrm{W}(i+1,j)} = I_{m,\mathrm{E}(i,j)}$, a value just calculated or else imposed by boundary conditions if $i = 0$.

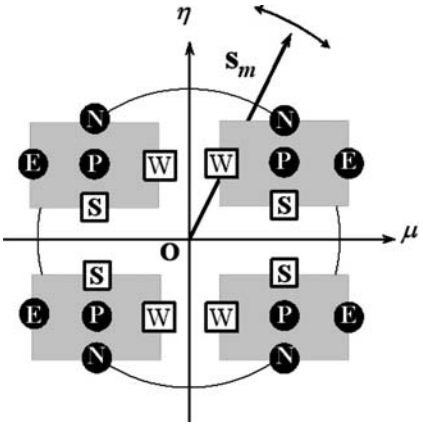

Fig. 5. Orientation of control volumes in the direction of $\boldsymbol{s}_m$

– $I_{m,\mathrm{S}(i+1,j)} = I_{m,\mathrm{N}(i+1,j-1)}$, a value calculated previously at level $j-1$ or else imposed by boundary conditions if $j = 0$.

This assumes that the region of calculation is swept out, when $\mu_m > 0$ and $\eta_m > 0$, by varying j from 1 to J and that, for each value of j, i then varies from 1 to I.

If the direction cosines are not positive, the above formulas remain valid provided that the orientation WESN of the cells is taken in the direction of propagation as shown in Fig. 5, and that $|\mu_m|$ and $|\eta_m|$ are used in the definitions of λ_x and λ_y instead of μ_m and η_m, respectively. The region of calculation is also swept out differently, but one can account for all cases in a global way using the following sequence:

– vary j from J_1 to J_2 with an increment k_j,
– for each value of j, vary i from I_1 to I_2 with an increment k_i,

with the correspondence

	I_1	I_2	k_i
$\mu > 0$	1	I	1
$\mu < 0$	I	1	-1

	J_1	J_2	k_j
$\eta > 0$	1	J	1
$\eta < 0$	J	1	-1

. (59)

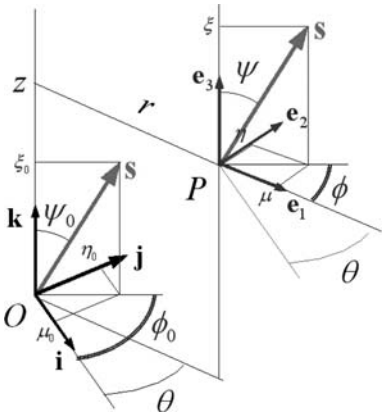

Fig. 6. Axes for curvilinear coordinates

Discrete Ordinates in Cylindrical Coordinates

Angular Redistribution

For a region described in cylindrical coordinates r, θ, z, the transport term on the left-hand side of the transfer equation is expressed in the form[24]

$$\boldsymbol{\nabla} \cdot (\boldsymbol{s} I_\omega) = \frac{\mu}{r} \frac{\partial (r I_\omega)}{\partial r} + \frac{1}{r} \frac{\partial (\eta I_\omega)}{\partial \theta} + \xi \frac{\partial I_\omega}{\partial z} \,, \tag{60}$$

where μ, η, ξ are the direction cosines of $\boldsymbol{s}$ in the local frame $\mathcal{R}(\boldsymbol{x}) = \{P(\boldsymbol{x}), \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ (see Fig. 6).

In contrast to the Cartesian case, the orientation of the coordinate axes changes with the position of the point $P(\boldsymbol{x})$. Hence, a direction $\boldsymbol{s}$ that is the same at every point with respect to a fixed frame $\mathcal{R}_0 = \{O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ will be viewed at different angles ψ, ϕ in each frame $\mathcal{R}(\boldsymbol{x})$. It follows that the direction cosines μ, η, ξ, which are themselves functions of ψ and ϕ, depend on the coordinates of the position $\boldsymbol{x} = (r, \theta, z)$. In fact, in cylindrical coordinates, only ϕ varies with θ, in such a way that

$$\phi\Big|_{\mathcal{R}(\boldsymbol{x})} = \phi\Big|_{\mathcal{R}_0} - \theta \quad \Longrightarrow \quad \frac{\partial \phi}{\partial \theta} = -1 \,. \tag{61}$$

Hence the derivative with respect to θ appearing in (60) must be decomposed into

$$\frac{\partial}{\partial \theta}\bigg|_{\boldsymbol{s}\,\text{const.}/\mathcal{R}_0} = \frac{\partial}{\partial \theta}\bigg|_{\mu,\eta,\xi} + \frac{\partial \phi}{\partial \theta}\bigg|_{r,z} \frac{\partial}{\partial \phi} = \frac{\partial}{\partial \theta}\bigg|_{\mu,\eta,\xi} - \frac{\partial}{\partial \phi} \,. \tag{62}$$

[24] Recall that, since $\boldsymbol{s}$ is a constant vector (it does not change with the position $\boldsymbol{x}$ when we move along a given trajectory or optical path), the two expressions $\boldsymbol{s} \cdot \boldsymbol{\nabla} I_\omega$ and $\boldsymbol{\nabla} \cdot (\boldsymbol{s} I_\omega)$ are equal. These two terms have the same expansion in Cartesian coordinates, but not in cylindrical coordinates. The term $\boldsymbol{\nabla} \cdot (\boldsymbol{s} I_\omega)$ arises when the transport term in the transfer equation is expressed in conservative form, which lends itself well to numerical treatment by a finite element method.

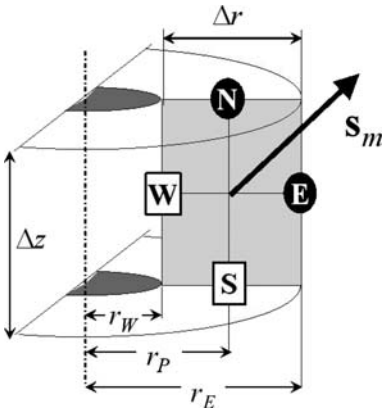

Fig. 7. Control volume for integrating the transfer equation in cylindrical coordinates

However, the derivatives with respect to r or z do not introduce extra terms since the orientation of the frame does not change when we move along r or z keeping θ constant. Hence the full expression for the transport term is, in cylindrical coordinates,

$$\boldsymbol{\nabla}\cdot(\boldsymbol{s}I)\Big|_{\mu,\eta,\xi} - \frac{1}{r}\frac{\partial(\eta I)}{\partial\phi} = \frac{\mu}{r}\frac{\partial(rI)}{\partial r} + \frac{\eta}{r}\frac{\partial I}{\partial\theta} + \xi\frac{\partial I}{\partial z} - \underbrace{\frac{1}{r}\frac{\partial(\eta I)}{\partial\phi}}_{\text{A.R.}}, \tag{63}$$

where the term marked A.R. corresponds to the so-called angular redistribution. Finally, the transfer equation can be written

$$\frac{\mu}{r}\frac{\partial(rI_\omega)}{\partial r} + \frac{\eta}{r}\frac{\partial I_\omega}{\partial\theta} + \xi\frac{\partial I_\omega}{\partial z} - \frac{1}{r}\frac{\partial(\eta I_\omega)}{\partial\phi} + \kappa_\omega I_\omega = \kappa_\omega I^0_\omega(T)\,. \tag{64}$$

4.4 Integrating over a Control Volume

To simplify the discussion, we consider here a problem with azimuthal symmetry, i.e., $\partial/\partial\theta \equiv 0$. Consider a uniform mesh in the plane $\boldsymbol{x} = (r, z) \in [R_1, R_2] \times [0, L]$ with $I \times J$ cells of size $\Delta r\Delta z$. For an annular space, R_1 and R_2 are the small and large radii, respectively, and for a hollow cylinder, $R_1 = 0$ and R_2 is the internal radius of the cylinder.

The faces of each cell are specified by analogy with the Cartesian case (see Fig. 7). The transfer equation expressed in a direction m of the quadrature and integrated over a volume element $\Delta V = r\Delta r\Delta z$ centered on P then gives

$$\mu_m(A_{\mathrm{E}}I_{m,\mathrm{E}} - A_{\mathrm{W}}I_{m,\mathrm{W}} + \xi_m A_z(I_{m,\mathrm{N}} - I_{m,\mathrm{S}}) - \frac{\partial(\eta_m I_{m,\mathrm{P}})}{\partial\phi}\Delta r\Delta z + \kappa_\omega\Delta V I_{m,\mathrm{P}} = S_{\mathrm{P}}\Delta V\,, \tag{65}$$

where

$$A_{\mathrm{E}} = r_{\mathrm{E}}\Delta z\,, \quad A_{\mathrm{W}} = r_{\mathrm{W}}\Delta z\,, \quad A_z = r_{\mathrm{P}}\Delta r\,, \quad S_{\mathrm{P}} = \kappa_\omega I^0(T_{\mathrm{P}})\,. \tag{66}$$

Let us now integrate (64) over an element of solid angle $\delta\Omega_m$ centered on the direction $\boldsymbol{s}_m$. The integral

$$\int_{\delta\Omega_m} \frac{\partial(\eta_m I_{m,\mathrm{P}})}{\partial\phi}\,\mathrm{d}\Omega\,,$$

which arises here cannot be evaluated analytically because the domain over which ψ and ϕ vary to describe $\delta\Omega_m$ is generally unknown.[25] However, when the directions $m-1$, m and $m+1$ lie on the same latitude, we may consider that the discretised expression for this integral takes the form

$$\int_{\delta\Omega_m} \frac{\partial(\eta_m I_{m,\mathrm{P}})}{\partial\phi}\,\mathrm{d}\Omega \approx \alpha_{m+1/2}L_{m+1/2,\mathrm{P}} - \alpha_{m-1/2}L_{m-1/2,\mathrm{P}}\,, \tag{67}$$

where it remains to define the coefficients $\alpha_{m\pm1/2}$. For all other terms in (64), the integration over $\delta\Omega_m$ is carried out using the approximation

$$\int_{\delta\Omega_m} f(\boldsymbol{\Omega})\,\mathrm{d}\Omega \approx w_m f(\boldsymbol{\Omega}_m)\,. \tag{68}$$

Dividing the whole equation by w_m, this leads to

$$\begin{aligned}&\mu_m(A_{\mathrm{E}}I_{m,\mathrm{E}} - A_{\mathrm{W}}I_{m,\mathrm{W}} + \xi_m A_z(I_{m,\mathrm{N}} - I_{m,\mathrm{S}})\\ &-\frac{\Delta r\Delta z}{w_m}\left(\alpha_{m+1/2}I_{m+1/2,\mathrm{P}} - \alpha_{m-1/2}I_{m-1/2,\mathrm{P}}\right) + \kappa_\omega\Delta V I_{m,\mathrm{P}} = S_{\mathrm{P}}\Delta V\,.\end{aligned} \tag{69}$$

Conservation of Intensity

The coefficients $\alpha_{m\pm1/2}$ are determined in such a way as to guarantee that the discrete form (69) of the transfer equation does indeed vanish when we set $\kappa_\omega = 0 = S_{\mathrm{P}}$ and $I_{m,\mathrm{E,W,S,N,P}} = I_0$ constant. Noting that $A_{\mathrm{E}} - A_{\mathrm{W}} = \Delta r\Delta z$, we see immediately that this condition requires

$$I_0\Delta r\Delta z\left(\mu_m - \frac{\alpha_{m+1/2} - \alpha_{m-1/2}}{w_m}\right) = 0\,, \tag{70}$$

or

$$\alpha_{m+1/2} - \alpha_{m-1/2} = \mu_m w_m\,. \tag{71}$$

[25] In fact, even when this integration is possible analytically, e.g., for a quadrature generated by direct discretisation of the polar angles, the result has to be modified anyway to guarantee conservation of the intensity (see below).

This recurrence relation is used to calculate the coefficients α step by step for directions situated on the same latitude, once the starting value is known. Suppose that the directions $M_1 \leq m \leq M_2$ are all on the same latitude $\xi = \xi_l$ $(l = 1, N)$.[26] Since the angular redistribution term vanishes when integrated over $\phi \in [0, 2\pi]$, the same must be true of the discretised expression. This implies that

$$\sum_{m=M_1}^{M_2} \left(\alpha_{m+1/2} I_{m+1/2,\mathrm{P}} - \alpha_{m-1/2} I_{m-1/2,\mathrm{P}}\right) = \alpha_{M_2+1/2} I_{M_2+1/2,\mathrm{P}} - \alpha_{M_1-1/2} I_{M_1-1/2,\mathrm{P}} = 0\,. \quad (72)$$

By periodicity, we have $I_{M_2+1/2,\mathrm{P}} = I_{M_1-1/2,\mathrm{P}}$. Hence, $\alpha_{M_2+1/2}$ and $\alpha_{M_1-1/2}$ must be equal, and furthermore, physically, they can only be zero. If we now observe that the index $M_1 - 1/2$ of latitude l coincides with the index $M_2 + 1/2$ of the previous latitude $l-1$, we simply require $\alpha_{1/2} = 0$ for the first of them $(M_1 = 1)$ in order to ensure the same for all the values $\alpha_{M_1-1/2}$. Finally, the recurrence relation (71) can be used to find $\alpha_{m\pm1/2}$ for $m = 1, \ldots, M$, starting from $\alpha_{1/2} = 0$.

Discretised Expression

From the above discussion, the discretisation of the transfer equation over a control volume gives

$$\mu_m (A_{\mathrm{E}} I_{m,\mathrm{E}} - A_{\mathrm{W}} I_{m,\mathrm{W}} + \xi_m A_z (I_{m,\mathrm{N}} - I_{m,\mathrm{S}}) - A_{\mathrm{R}} I_{m+1/2,\mathrm{P}} + A_{\mathrm{L}} I_{m-1/2,\mathrm{P}} + \kappa_\omega \Delta V I_{m,\mathrm{P}} = S_{\mathrm{P}} \Delta V\,, \quad (73)$$

where

$$A_{\mathrm{E}} = r_{\mathrm{E}} \Delta z\,, \quad A_{\mathrm{W}} = r_{\mathrm{W}} \Delta z\,, \quad A_z = r_{\mathrm{P}} \Delta r\,, \quad (74)$$

$$A_{\mathrm{R}} = \Delta r \Delta z \frac{\alpha_{m+1/2}}{w_m}\,, \quad A_{\mathrm{L}} = \Delta r \Delta z \frac{\alpha_{m-1/2}}{w_m}\,. \quad (75)$$

Let us assume that the intensities are known not only on faces W and S, but also at P for the direction $m-1$. The interpolation relations (56) apply with

$$I_{m+1/2,\mathrm{P}} = I_{m-1/2,\mathrm{P}} + \frac{I_{m,\mathrm{P}} - I_{m-1/2,\mathrm{P}}}{c}\,, \quad (76)$$

and, as for a and b, $1/2 \leq c \leq 1$. We can then obtain an explicit relation between $I_{m,\mathrm{P}}$ and the known values of the specific intensities in the form

$$I_{m,\mathrm{P}} = \frac{\lambda_r I_{m,\mathrm{W}} + \lambda_z I_{m,\mathrm{S}} - \lambda_d I_{m-1/2,\mathrm{P}} + S_{\mathrm{P}} \Delta V}{\mu_m \dfrac{A_{\mathrm{E}}}{a} + \lambda_z - \dfrac{A_{\mathrm{R}}}{c} + \lambda_0}\,, \quad (77)$$

[26] For example, in the S_4 quadrature, the directions $5 \leq m \leq 12$ are all at latitude $\xi = 0.9082483$.

where

$$\lambda_r = \mu_m A_{\mathrm{E}} \left(\frac{1}{a} - 1 \right) + A_W \,, \quad \lambda_z = \xi_m \frac{A_z}{b} \,, \tag{78}$$

$$\lambda_0 = \kappa_\omega \Delta V \,, \quad \lambda_d = \frac{A_{\mathrm{R}}}{c} - \mu_m \Delta r \Delta z \,. \tag{79}$$

As in the Cartesian case, these expressions remain valid when μ_m or ξ_m are negative. One then orients the cell WESN according to the direction of propagation of the radiation, replacing μ_m and ξ_m by $|\mu_m|$ and $|\xi_m|$ in (77) and (78).

First Direction on Each Latitude

The values of $I_{m-1/2,\mathrm{P}}$ must be initialised before beginning the calculation in the first direction of each latitude. To do this, we consider each time the particular direction $\eta = 0$ for which there is no angular redistribution, for one is then in the plane of symmetry (r, z) in which the orientation of the local frame does not change. The specific intensities in the direction $\eta = 0$ are calculated by simply setting $\lambda_d = 0$ in (77) and (79). The values thereby obtained are only used to initialise the field $I_{m-1/2}$. They are not taken into account for the calculation of e or the fluxes since this particular direction does not belong to the quadrature set.

4.5 Advantages and Disadvantages

The discrete ordinate method is certainly the most widely used method today for thermal radiation calculations, but also in neutronics, the field where it was first devised. The only real challenge in terms of accuracy comes from the Monte Carlo method,[27] but the latter requires considerably longer computation times. Although the transfer equation must be solved in each of the discrete directions of the quadrature (so that the same calculation is repeated up to 24 times with S_8, for example), this solution is completely explicit (no matrices need to be inverted) and hence very fast. Moreover, the discrete ordinate method can take into account all directional aspects of the transfer and in particular it can handle non-isotropic boundary conditions, such as specular reflection, or the Snell–Descartes law at the interface of two media with different group velocities. Globally, the ratio of accuracy to computation time is excellent. However, there are certain limitations, e.g., for applications to arbitrary geometries and/or non-rectangular meshes, as well as in essentially diffusive media. In addition, there are two specific numerical problems that may impair the accuracy of the method.

[27] In the hierarchy of methods used for radiative calculations, the zone method ([8], Chap. 17) has a reputation for great accuracy, but requires an enormous calculational effort for the evaluation of exchange factors and hence a considerable computation time. There is little doubt that the Monte Carlo approach is preferable.

The first relates to the interpolation (56) of the intensity within the unit cell. The centered (diamond) scheme can produce negative intensities locally, resulting in undesirable fluctuations in the solution. The step scheme is more stable from this point of view, but it is also less accurate because it introduces a great deal of numerical diffusion. A simple solution to this problem is to calculate the interpolation coefficients at each point using [22]

$$a = \max\left(\frac{1}{2}, 1 - \frac{|\mu_m| A_x}{\kappa_\omega \Delta V + 2|\eta_m| A_y}\right), \tag{80}$$

$$b = \max\left(\frac{1}{2}, 1 - \frac{|\eta_m| A_y}{\kappa_\omega \Delta V + 2|\mu_m| A_x}\right), \tag{81}$$

in Cartesian geometry, and [23]

$$a = \max\left(\frac{1}{2}, 1 - \frac{|\mu_m| A_\mathrm{W}}{\kappa_\omega \Delta V + 2|\xi_m| A_z + 2|A_\mathrm{R}|}\right), \tag{82}$$

$$b = \max\left(\frac{1}{2}, 1 - \frac{|\xi_m| A_z}{\kappa_\omega \Delta V + 2|\mu_m| A_\mathrm{E} + 2|A_\mathrm{R}|}\right), \tag{83}$$

$$c = \max\left(\frac{1}{2}, 1 - \frac{|A_\mathrm{L}|}{\kappa_\omega \Delta V + 2|\mu_m| A_\mathrm{E} + 2|\xi_m| A_z}\right), \tag{84}$$

in cylindrical geometries. Negative intensities are then avoided, whilst maintaining the accuracy of the scheme as close as possible to second order.

The other problem is related to the discretisation of the directions in space, which leads to the so-called ray effect. This occurs for instance when the medium has a small region (surface or volume) at high temperature in relatively cold surroundings. If the medium is homogeneous, this source would normally affect all points at similar distances from its center in roughly the same way. Now among these points, there are some from which none of the discrete directions actually intercepts the source. The calculation thus attributes no flux to them. Furthermore, a point which sees the source in a solid angle $\delta\Omega$ in one of the directions $\boldsymbol{s}_m$ will be attributed a flux calculated with the solid angle w_m (the weight of $\boldsymbol{s}_m$) rather than $\delta\Omega$. Hence, the energy from the source, even if it is globally conserved by the numerical scheme, is incorrectly distributed. This leads to results fluctuating around the exact solution. There is no good remedy for this problem, inherent in the discretisation of the propagation directions, and particularly relevant in a dominantly ballistic regime. It diminishes (without theoretically disappearing altogether) when the number of discrete directions is increased.[28]

[28] However, for a given number of directions, it increases when the spatial mesh is refined.

Appendix. P_1 Method and Hyperbolic Heat Equation

In Sect. 3.1, we discussed the P_1 method in the stationary regime. In this appendix, we shall outline the more general case of a variable regime and show how to recover the hyperbolic heat equation from P_1, at least under certain conditions.

P_1 Method in Non-Stationary Regime

Integrating (10) and also (10) multiplied by $\boldsymbol{s}$ over the whole 4π sr solid angle, we obtain

$$\frac{1}{v_\omega}\frac{\partial G_\omega}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{q}''_\omega + \kappa_\omega G_\omega = 4\pi\kappa_\omega I^0_\omega\,, \tag{85}$$

$$\frac{1}{v_\omega}\frac{\partial \boldsymbol{q}''_\omega}{\partial t} + \boldsymbol{\nabla} P_\omega + \kappa_\omega \boldsymbol{q}''_\omega = 0\,, \tag{86}$$

respectively, where P_ω is the tensor $\int_{4\pi} \boldsymbol{ss} I_\omega \,\mathrm{d}\Omega$. As before (see Sect. 3.1), the P_1 approximation amounts to assuming that

$$I_\omega(\boldsymbol{x}, t, \boldsymbol{s}) \approx \frac{G_\omega(\boldsymbol{x}, t)}{4\pi} + \frac{3}{4\pi}\boldsymbol{q}''_\omega(\boldsymbol{x}, t)\cdot\boldsymbol{s}\,, \tag{87}$$

which yields[29]

$$\boldsymbol{\nabla} P_\omega \approx \tfrac{1}{3}\boldsymbol{\nabla} G_\omega\,. \tag{88}$$

Equation (86) then becomes

$$\frac{1}{v_\omega}\frac{\partial \boldsymbol{q}''_\omega}{\partial t} + \frac{1}{3}\boldsymbol{\nabla} G_\omega + \kappa_\omega \boldsymbol{q}''_\omega = 0\,. \tag{89}$$

Hyperbolic Transport Equation

Differentiating (85) with respect to t and dividing by v_ω, we obtain

$$\frac{1}{v_\omega^2}\frac{\partial^2 G_\omega}{\partial t^2} + \boldsymbol{\nabla}\cdot\left(\frac{1}{v_\omega}\frac{\partial \boldsymbol{q}''_\omega}{\partial t}\right) + \frac{\kappa_\omega}{v_\omega}\frac{\partial G_\omega}{\partial t} = 4\pi\frac{\kappa_\omega}{v_\omega}\frac{\partial I^0_\omega}{\partial t}\,. \tag{90}$$

Hence, according to (89),

$$\frac{1}{v_\omega^2}\frac{\partial^2 G_\omega}{\partial t^2} - \frac{1}{3}\nabla^2 G_\omega - \kappa_\omega \boldsymbol{\nabla}\cdot\boldsymbol{q}''_\omega + \frac{\kappa_\omega}{v_\omega}\frac{\partial G_\omega}{\partial t} = 4\pi\frac{\kappa_\omega}{v_\omega}\frac{\partial I^0_\omega}{\partial t}\,. \tag{91}$$

[29] We have $\int_{4\pi} \boldsymbol{ss}\,\mathrm{d}\Omega = 4\pi\boldsymbol{\delta}/3$ and $\int_{4\pi} \boldsymbol{ss}[\boldsymbol{q}''_\omega(\boldsymbol{x})\cdot\boldsymbol{s}]\,\mathrm{d}\Omega = \boldsymbol{q}''_\omega(\boldsymbol{x})\int_{4\pi}\boldsymbol{sss}\,\mathrm{d}\Omega$, where $\int_{4\pi}\boldsymbol{sss}\,\mathrm{d}\Omega = 0$ (zero tensor). This gives $P_\omega = G_\omega(\boldsymbol{x})\boldsymbol{\delta}/3$, and hence $\boldsymbol{\nabla}P_\omega = \boldsymbol{\nabla}G_\omega/3$.

Using (85) to reexpress $\boldsymbol{\nabla}\cdot\boldsymbol{q}''_\omega$, we find

$$\frac{1}{v_\omega^2}\frac{\partial^2 G_\omega}{\partial t^2}-\frac{1}{3}\nabla^2 G_\omega-\kappa_\omega\left(4\pi\kappa_\omega I^0_\omega-\frac{1}{v_\omega}\frac{\partial G_\omega}{\partial t}-\kappa_\omega G_\omega\right)+\frac{\kappa_\omega}{v_\omega}\frac{\partial G_\omega}{\partial t}=4\pi\frac{\kappa_\omega}{v_\omega}\frac{\partial I^0_\omega}{\partial t}\,, \quad (92)$$

which simplifies to

$$\frac{1}{v_\omega^2}\frac{\partial^2 G_\omega}{\partial t^2}+2\frac{\kappa_\omega}{v_\omega}\frac{\partial G_\omega}{\partial t}-\frac{1}{3}\nabla^2 G_\omega+\kappa_\omega^2 G_\omega=4\pi\kappa_\omega\left(\kappa_\omega I^0_\omega+\frac{1}{v_\omega}\frac{\partial I^0_\omega}{\partial t}\right)\,, \quad (93)$$

or, dividing by κ_ω^2,

$$\tau_\omega^2\frac{\partial^2 G_\omega}{\partial t^2}+2\tau_\omega\frac{\partial G_\omega}{\partial t}-\frac{\Lambda_\omega^2}{3}\nabla^2 G_\omega+G_\omega=4\pi\left(I^0_\omega+\tau_\omega\frac{\partial I^0_\omega}{\partial t}\right)\,, \quad (94)$$

where

$$\Lambda_\omega=\frac{1}{\kappa_\omega}\,,\quad \tau_\omega=\frac{1}{v_\omega\kappa_\omega}\,, \quad (95)$$

so that Λ_ω is in fact the mean free path and τ_ω the relaxation time introduced in Sect. 2.2.

Diffusive Regime

The approximation of purely diffusive transport amounts to assuming that[30]

$$G_\omega\approx 4\pi I^0_\omega\,. \quad (96)$$

In this case, using the notation introduced for (95), (89) divided by κ_ω together with (94) become

$$\tau_\omega\frac{\partial \boldsymbol{q}''_\omega}{\partial t}+\boldsymbol{q}''_\omega=-\frac{4\pi\Lambda_\omega}{3}\boldsymbol{\nabla}I^0_\omega\,, \quad (97)$$

$$\tau_\omega^2\frac{\partial^2 I^0_\omega}{\partial t^2}+\tau_\omega\frac{\partial I^0_\omega}{\partial t}-\frac{\Lambda_\omega^2}{3}\nabla^2 I^0_\omega=0\,. \quad (98)$$

[30] From (19), $I_\omega=I^0_\omega-(1/\kappa_\omega)\boldsymbol{s}\cdot\boldsymbol{\nabla}I^0_\omega$ and hence

$$\int_{4\pi} I_\omega\,\mathrm{d}\Omega=4\pi I^0_\omega-\frac{1}{\kappa_\omega}\boldsymbol{\nabla}I^0_\omega\cdot\underbrace{\int_{4\pi}\boldsymbol{s}\,\mathrm{d}\Omega}_{=0}\,.$$

Since the internal energy density (per unit frequency) is[31] $e_\omega = 4\pi I^0_\omega / v_\omega$, the last equation can also be written as

$$\tau_\omega^2 \frac{\partial^2 e_\omega}{\partial t^2} + \tau_\omega \frac{\partial e_\omega}{\partial t} - \frac{\Lambda_\omega^2}{3} \nabla^2 e_\omega = 0 \,. \tag{99}$$

If we now assume that τ_ω, Λ_ω, and v_ω are independent of the frequency (analogous to the assumption of a grey medium in thermal radation), i.e., $\tau_\omega = \tau$, $\Lambda_\omega = \Lambda$, and $v_\omega = v$, we can express the preceding equations in terms of total quantities, i.e., integrated over the whole spectrum:

$$\tau \frac{\partial \boldsymbol{q}''}{\partial t} + \boldsymbol{q}'' = -\frac{4\pi\Lambda}{3} \boldsymbol{\nabla} I^0 = -\frac{\Lambda v}{3} \boldsymbol{\nabla} e \,, \tag{100}$$

$$\tau^2 \frac{\partial^2 e}{\partial t^2} + \tau \frac{\partial e}{\partial t} - \frac{\Lambda^2}{3} \nabla^2 e = 0 \,, \tag{101}$$

with

$$e = \frac{4\pi I^0}{v} = \frac{4\pi}{v} \int_0^{\omega_{\max}} I^0_\omega \, \mathrm{d}\omega \,. \tag{102}$$

Comparison with the Hyperbolic Heat Equation

The hyperbolic heat equation (HHE), viz.,

$$\tau \frac{\partial^2 e}{\partial t^2} + \frac{\partial e}{\partial t} - \boldsymbol{\nabla} \cdot (\lambda \boldsymbol{\nabla} T) = 0 \,, \tag{103}$$

where $e = \rho c T$ is the internal energy density, follows from the modified Fourier law:

$$\tau \frac{\partial \boldsymbol{q}''}{\partial t} + \boldsymbol{q}'' = -\lambda \boldsymbol{\nabla} T \,. \tag{104}$$

These two equations should be compared with (101) and (100), respectively, which can also be written in the form

$$\tau \frac{\partial^2 e}{\partial t^2} + \frac{\partial e}{\partial t} - \underbrace{\frac{\Lambda}{\tau v}}_{=1} \boldsymbol{\nabla} \cdot \left(\frac{\Lambda v}{3} \frac{\partial e}{\partial T} \boldsymbol{\nabla} T \right) = 0 \,, \tag{105}$$

$$\tau \frac{\partial \boldsymbol{q}''}{\partial t} + \boldsymbol{q}'' = -\frac{\Lambda v}{3} \frac{\partial e}{\partial T} \boldsymbol{\nabla} T \,. \tag{106}$$

The analogy between HHE and the non-stationary P_1 model is then established by setting

$$\lambda = \frac{\Lambda v}{3} \frac{\partial e}{\partial T} = \frac{4\pi\Lambda}{3} \frac{\partial I^0}{\partial T} \,, \tag{107}$$

in the diffusive regime and for a medium whose characteristics (relaxation time, mean free path, and propagation speed of phonons) are independent of the frequency.

[31] From (2). See last footnote.

References

[1] N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Harcourt College Publishers 1976) 78

[2] C. Kittel: *Introduction to Solid State Physics* (John Wiley, New York 1996) 78, 80

[3] J. Ziman: *Electrons and Phonons. The Theory of Transport Phenomena in Solids* (Clarendon Press, Oxford 1960) 78

[4] G. P. Srivastava: *The Physics of Phonons* (Adam Hilger 1990) 78, 80

[5] A. Majumdar: Microscale heat conduction in dielectric thin films, J. Heat Transf. **115**, 7–16 (1993) 79

[6] J. H. Jeans: The equation of radiative transfer of energy, Monthly Notices, Roy. Astronom. Soc. **78**, 28–36 (1917) 83

[7] R. L. Murray: *Nuclear Reactor Physics* (Prentice Hall 1957) 83

[8] M. F. Modest: *Radiative Heat Transfer*, 2nd ed. (Academic Press 2003) 83, 88, 100

[9] G. Chen: Ballistic-diffusive equations for transient heat conduction from nano to microscales, J. Heat Transf. **124**, 320–328 (2002) 83, 88

[10] R. E. Marchak: Note on the spherical harmonic method as applied to the Milne problem for a sphere, Phys. Rev. **71**, 443–446 (1947) 87

[11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, Cambridge 1992) 87

[12] S. Chandrasekar: *Radiative Transfer* (Dover Publications, New York 1960) 88

[13] B. G. Carlson, K. D. Lathrop: Transport theory. The method of discrete ordinates, in H. Greenspan, C. V. N. Kelber, D. Okrent (Eds.): *Computing Methods in Reactor Physics* (Gordon and Breach 1968) 88, 90

[14] W. A. Fiveland: A discrete ordinate method for predicting radiative heat transfer in axisymmetric enclosures, ASME Paper 82-HTD-20 (1982) 88

[15] W. A. Fiveland: Discrete ordinate solutions of the radiative transport equation for rectangular enclosures, J. Heat Transf. **106**, 699–706 (1984) 88, 90

[16] W. A. Fiveland: Three-dimensional radiative heat transfer solutions by the discrete ordinate method, J. Thermophys. Heat Transfer **2**, 309–316 (1988) 88

[17] A. A. Joshi, A. Majumdar: Transient ballistic and diffusive phonon heat transport, Appl. Phys. **74**, 31–39 (1993) 88

[18] J. D. Chung, M. Kaviany: Effect of phonon pore scattering and pore randomness on effective conductivity of porous silicon, Int. J. Heat Mass Transf. **43**, 521–538 (2000) 88

[19] S. Volz, D. Lemonnier, J. B. Saulnier: Clamped nanowire thermal conductivity based on phonon transport equation, Microscale Thermophys. Eng. **5**, 191–207 (2001) 88

[20] D. Balsara: Fast and accurate discrete ordinate methods for multidimensional radiative transfer. Part I: Basic methods, J. Quant. Spectrosc. Rad. Transf. **69**, 671–707 (2001) 88, 92, 93

[21] R. Koch, R. Becker: Evaluation of quadrature schemes for the discrete ordinate method, J. Quant. Spectrosc. Rad. Transf. **84**, 423–435 (2004) 88, 93

[22] K. D. Lathrop: Spatial differencing of the transport equation. Positivity versus accuracy, J. Comp. Phys. **4**, 475–498 (1969) 88, 101

[23] S. Jendoubi, H. S. Lee, T. K. Kim: Discrete ordinate solutions for radiatively participating media in cylindrical enclosures, J. Thermophys. Heat Transfer **7**, 213–219 (1993) 88, 101
[24] W. A. Fiveland: The selection of discrete ordinate quadrature sets for anisotropic scattering, ASME HTD **160**, 89–96 (1991) 92

Index

Radiative Transfer on Short Length Scales

Karl Joulain

Laboratoire d'Etudes Thermiques, ENSMA, BP 109, 86961 Futuroscope Cedex
karl.joulain@ensma.fr

Abstract. Heat transfer of radiative origin is treated classically by the transport equation of a phenomenological quantity called the specific intensity (see the Chapter on Radiative Transfer by *Carminati* in this volume). This quantity and its dynamical equation are based on a radiometric approach (incoherent addition of fluxes, geometrical optics). When the relevant distances become as small as the wavelength or less, such hypotheses are no longer valid. One must then go back to the more general ideas of electromagnetic theory.

The aim of this Chapter is to show how to calculate heat transfer of radiative origin from the equations governing the electromagnetic field. We begin by reviewing these equations. We then show how to calculate the quantities relevant to heat transfer, such as the radiative flux. We also review the theory of dipole radiation, which will prove helpful in the ensuing radiative transfer calculations.

In Sect. 2, we examine how a small dipolar sphere at temperature T will radiate in vacuum, and how transfer occurs between two such dipolar spheres at different temperatures. We then calculate the electromagnetic energy density near a surface at temperature T (Sect. 3). We shall see how this energy density is modified in the near field, particularly when the materials in place can carry surface waves. Finally, in Sect. 4, we investigate the near-field radiative transfer between two plane surfaces, bringing out the key role played once again by surface waves.

1 Review of Electromagnetism

1.1 Maxwell's Equations and Constitutive Relations

In matter, the electromagnetic field is completely determined by the four quantities $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{B}$, and $\boldsymbol{H}$, where $\boldsymbol{E}$ is the electric field, $\boldsymbol{D}$ is the electric displacement, $\boldsymbol{B}$ is the magnetic induction, and $\boldsymbol{H}$ is the magnetic field. In the presence of free charge and current densities, ρ and $\boldsymbol{j}$, the fields obey Maxwell's equations:

$$\boldsymbol{\nabla}\cdot\boldsymbol{B}=0\,,\qquad \boldsymbol{\nabla}\cdot\boldsymbol{D}=\rho\,, \tag{1}$$

$$\boldsymbol{\nabla}\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t}\,,\qquad \boldsymbol{\nabla}\times\boldsymbol{H}=\boldsymbol{j}+\frac{\partial\boldsymbol{D}}{\partial t}\,. \tag{2}$$

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 107–131 (2007)

The local conservation of charge is ensured by the second equations of (1) and (2), which together imply that

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0\,. \tag{3}$$

If we now append the Lorentz force law describing the force on a particle of charge q, moving with velocity $\boldsymbol{v}$, viz., $\boldsymbol{f} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$, we then have a complete dynamical description of the charges and electromagnetic fields (classical and non-relativistic). We shall often be dealing with monochromatic quantities in this Chapter. Any time signal can be decomposed into monochromatic signals by the Fourier transform:

$$\psi(\boldsymbol{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\boldsymbol{r}, \omega) \mathrm{e}^{-\mathrm{i}\omega t}\, \mathrm{d}\omega\,.$$

The fields are related by constitutive relations. For linear, isotropic media, these relations are

$$\boldsymbol{D}(\boldsymbol{r}, \omega) = \varepsilon_0 \varepsilon(\boldsymbol{r}, \omega) \boldsymbol{E}(\boldsymbol{r}, \omega)\,, \quad \boldsymbol{B}(\boldsymbol{r}, \omega) = \mu_0 \mu(\boldsymbol{r}, \omega) \boldsymbol{H}(\boldsymbol{r}, \omega)\,,$$

where ε is the dielectric constant and μ the relative permeability. These are local relations in space and in frequency. They are therefore a priori nonlocal in the reciprocal space, i.e., the space of wave vectors $\boldsymbol{k}$, and in time. In the following, we shall restrict to the case where the dielectric constant and permeability are local quantities in the direct space. However, in the temporal space, we will have relations like

$$\boldsymbol{D}(\boldsymbol{r}, t) = \int \varepsilon_0 \varepsilon(\boldsymbol{r}, t - t') \boldsymbol{E}(\boldsymbol{r}, t')\, \mathrm{d}t'\,.$$

The medium is then dispersive. When there is no free current, as we shall assume throughout this Chapter, the Fourier frequency components of the electromagnetic fields obey a wave equation of the form

$$\nabla^2 \psi(\boldsymbol{r}, \omega) + \mu\varepsilon \frac{\omega^2}{c^2} \psi(\boldsymbol{r}, \omega) = 0\,, \tag{4}$$

known as the Helmholtz equation. The product $\varepsilon_0 \mu_0 = 1/c^2$ is the reciprocal of the square of the speed of an electromagnetic wave in vacuum, whereas $\varepsilon\mu = n^2$ is the square of the refractive index n of the medium. Equation (4) has solutions of the form

$$\psi(\boldsymbol{r}, \omega) = A\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} + B\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\,, \tag{5}$$

where $\boldsymbol{k} = k\boldsymbol{u}$ and $k = \sqrt{\mu\varepsilon}\omega/c$.

1.2 Plane Wave Expansion

Any electromagnetic wave can be expanded in plane waves. Consider a monochromatic component of an electromagnetic wave propagating in the positive z direction. If we know this component at a point $\boldsymbol{r} = (x, y, z)$ and if this wave is square summable in the plane (x, y), then this component can be written in the form of a Fourier transform

$$\psi(\boldsymbol{r}, \omega) = \int \frac{\mathrm{d}\alpha\, \mathrm{d}\beta}{4\pi^2} \psi(\alpha, \beta, z, \omega) \mathrm{e}^{\mathrm{i}(\alpha x + \beta y)} \,. \tag{6}$$

Substituting this into the Helmholtz equation (4), we find that $\psi(\alpha, \beta, z, \omega)$ satisfies

$$\frac{\partial^2 \psi(\alpha, \beta, z, \omega)}{\partial z^2} + \left(\mu\varepsilon \frac{\omega^2}{c^2} - \alpha^2 - \beta^2 \right) \psi(\alpha, \beta, z, \omega) = 0 \,, \tag{7}$$

whence,

$$\psi(\boldsymbol{r}, \omega) = \int \frac{\mathrm{d}\alpha\, \mathrm{d}\beta}{4\pi^2} \psi(\alpha, \beta, z = 0, \omega) \mathrm{e}^{\mathrm{i}(\alpha x + \beta y + \gamma z)} \,, \tag{8}$$

where $\gamma^2 = \mu\varepsilon\omega^2/c^2 - \alpha^2 - \beta^2$, $\Im(\gamma) > 0$, and $\Re(\gamma) > 0$ if $\Im(\gamma) = 0$. Any electromagnetic wave is then expressed as a sum of plane waves with wave vector $\boldsymbol{k} = (\alpha, \beta, \gamma)$ and amplitude $\psi(\alpha, \beta, z = 0, \omega)$, and this sum is none other than the Fourier transform of $\psi(\boldsymbol{r}, \omega)$ in the plane $z = 0$.

We now consider the case of a two-dimensional light beam, i.e., in the (x, z) plane, with width Δx in the plane $z = 0$, propagating in the positive z direction. Let us investigate the amplitudes in its plane wave expansion. Consider the three cases:

- $\Delta x \gg \lambda$ ($\lambda = 2\pi c/\omega$). Writing the Fourier transform of the field in the plane $z = 0$, we note that its spatial Fourier components will assume large values for spatial frequencies α lying in the range from $-2\pi/\Delta x$ to $2\pi/\Delta x$. In this case, the plane wave expansion will be limited to components for which γ^2 remains positive and will be very close to ω^2/c^2. The wave vector $\boldsymbol{k}$ will thus lie very close to the z direction.
- $\Delta x \sim \lambda$. The spatial frequencies α lying in the range from $-2\pi/\Delta x$ to $2\pi/\Delta x$ ensure that γ^2 will be close to zero. Now when γ is close to zero, we have a component making an angle close to $\pi/2$ with respect to the normal. The beam will have a very large opening angle (diffraction).
- $\Delta x \ll \lambda$. The spatial frequencies are essentially those for which $\gamma^2 < 0$, i.e., γ is pure imaginary and the spatial frequencies correspond to exponentially decaying evanescent waves. Very little light is transmitted through a hole with dimensions less than the wavelength.

One of the useful aspects of this expansion is that Maxwell's equations are particularly simple for plane waves. Hence the expansion (8) generalises to a

vectorial form in three dimensions, and for all the fields. For each plane wave, Maxwell's equations become

$$\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{B}(\boldsymbol{k},\omega)=0\,,\qquad \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{D}(\boldsymbol{k},\omega)=0\,, \tag{9}$$

$$\mathrm{i}\boldsymbol{k}\times\boldsymbol{E}(\boldsymbol{k},\omega)=\mathrm{i}\omega\boldsymbol{B}(\boldsymbol{k},\omega)\,,\qquad \mathrm{i}\boldsymbol{k}\times\boldsymbol{H}(\boldsymbol{k},\omega)=-\mathrm{i}\omega\boldsymbol{D}(\boldsymbol{k},\omega)\,, \tag{10}$$

when there is no free current. The plane waves are therefore transverse, i.e., the fields are orthogonal to $\boldsymbol{k}$. The third Maxwell equation also shows that the vector triple $\big(\boldsymbol{k},\boldsymbol{E}(\boldsymbol{k},\omega),\boldsymbol{B}(\boldsymbol{k},\omega)\big)$ is right-handed.

1.3 Energy Conservation, Poynting Vector, and Energy Density

We shall now calculate the power dissipated in a volume V containing charges and currents. To do so, we must calculate the work done per unit time by the Lorentz force. Consider the non-dispersive case. For a continuous current distribution, the dissipated power is given by [1]

$$P=\int_V \boldsymbol{j}(\boldsymbol{r},t)\cdot\boldsymbol{E}(\boldsymbol{r},t)\,\mathrm{d}^3\boldsymbol{r}\,. \tag{11}$$

Eliminating $\boldsymbol{j}$ with the help of the fourth Maxwell equation, we obtain

$$P=\int_V\left[\boldsymbol{E}\cdot(\boldsymbol{\nabla}\times\boldsymbol{H})-\boldsymbol{E}\cdot\frac{\partial\boldsymbol{D}}{\partial t}\right]\mathrm{d}^3\boldsymbol{r}\,. \tag{12}$$

Using $\boldsymbol{\nabla}\cdot(\boldsymbol{E}\times\boldsymbol{H})=\boldsymbol{H}\cdot(\boldsymbol{\nabla}\times\boldsymbol{E})-\boldsymbol{E}\cdot(\boldsymbol{\nabla}\times\boldsymbol{H})$ and the third Maxwell equation,

$$\begin{aligned}P&=\int_V \boldsymbol{j}(\boldsymbol{r},t)\cdot\boldsymbol{E}(\boldsymbol{r},t)\,\mathrm{d}^3\boldsymbol{r}\\&=-\int_V\left[\boldsymbol{\nabla}\cdot(\boldsymbol{E}\times\boldsymbol{H})+\boldsymbol{E}\cdot\frac{\partial\boldsymbol{D}}{\partial t}+\boldsymbol{H}\cdot\frac{\partial\boldsymbol{B}}{\partial t}\right]\mathrm{d}^3\boldsymbol{r}\,.\end{aligned} \tag{13}$$

We now set

$$u=\tfrac{1}{2}(\boldsymbol{E}\cdot\boldsymbol{D}+\boldsymbol{H}\cdot\boldsymbol{B})\quad\text{and}\quad\boldsymbol{S}=\boldsymbol{E}\times\boldsymbol{H}\,,$$

where u is the electromagnetic energy density associated with the free charges [2] and $\boldsymbol{S}$ is the Poynting vector. We deduce the local form of the conservation of energy, viz.,

$$\frac{\partial u}{\partial t}=-\boldsymbol{\nabla}\cdot\boldsymbol{S}-\boldsymbol{j}\cdot\boldsymbol{E}\,. \tag{14}$$

Hence the variation of the electromagnetic energy in a volume is equal to minus the sum of the work done by the Lorentz force and the flux of the Poynting vector. In permanent regime, the work done by the Lorentz force is

minus the flux of the Poynting vector. We shall use these relations in what follows.

When there is dispersion, it is essential to work with monochromatic quantities. In this case, the relevant quantities for transfer calculations, such as the power dissipated, the Poynting vector, or the energy density, will be oscillating quantities. They must therefore be averaged over a period. Consider, for example, the power dissipated at frequency ω. We have

$$\boldsymbol{E}(\boldsymbol{r},t) = \Re\left[\boldsymbol{E}(\boldsymbol{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t}\right] = \tfrac{1}{2}\left[\boldsymbol{E}(\boldsymbol{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t} + \boldsymbol{E}^*(\boldsymbol{r},\omega)\mathrm{e}^{\mathrm{i}\omega t}\right],$$

with an analogous expression for the current $\boldsymbol{j}$. We now calculate

$$\begin{aligned}\boldsymbol{j}(\boldsymbol{r},t)\cdot\boldsymbol{E}(\boldsymbol{r},t) &= \tfrac{1}{4}\left[\boldsymbol{j}(\boldsymbol{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t} + \boldsymbol{j}^*(\boldsymbol{r},\omega)\mathrm{e}^{\mathrm{i}\omega t}\right]\\ &\quad\cdot\left[\boldsymbol{E}(\boldsymbol{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t} + \boldsymbol{E}^*(\boldsymbol{r},\omega)\mathrm{e}^{\mathrm{i}\omega t}\right]\\ &= \tfrac{1}{2}\Re\left[\boldsymbol{j}(\boldsymbol{r},\omega)\cdot\boldsymbol{E}^*(\boldsymbol{r},\omega) + \boldsymbol{j}(\boldsymbol{r},\omega)\cdot\boldsymbol{E}(\boldsymbol{r},\omega)\mathrm{e}^{-2\mathrm{i}\omega t}\right]. \quad (15)\end{aligned}$$

The second term averages to zero over a period. Therefore the average dissipated power is

$$P(\boldsymbol{r},\omega) = \tfrac{1}{2}\Re\Big[\boldsymbol{j}(\boldsymbol{r},\omega)\cdot\boldsymbol{E}^*(\boldsymbol{r},\omega)\Big]. \quad (16)$$

In the same way, the Poynting vector is

$$\boldsymbol{S}(\boldsymbol{r},\omega) = \tfrac{1}{2}\Re\Big[\boldsymbol{E}(\boldsymbol{r},\omega)\times\boldsymbol{H}^*(\boldsymbol{r},\omega)\Big], \quad (17)$$

and the energy density is

$$u = \tfrac{1}{4}\Re\Big[\boldsymbol{E}(\boldsymbol{r},\omega)\cdot\boldsymbol{D}^*(\boldsymbol{r},\omega) + \boldsymbol{B}(\boldsymbol{r},\omega)\cdot\boldsymbol{H}^*(\boldsymbol{r},\omega)\Big]. \quad (18)$$

1.4 Potentials

Since $\boldsymbol{B}$ satisfies $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$, it can be derived from a vector potential $\boldsymbol{A}$, i.e.,

$$\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}. \quad (19)$$

Substituting into the third Maxwell equation, we obtain

$$\boldsymbol{\nabla}\times\left(\boldsymbol{E} + \frac{\partial\boldsymbol{A}}{\partial t}\right) = 0.$$

The quantity in brackets thus derives from a scalar potential V and we obtain

$$\boldsymbol{E} = -\boldsymbol{\nabla}V - \frac{\partial\boldsymbol{A}}{\partial t}. \quad (20)$$

The choice of potentials $\boldsymbol{A}$ and V is not unique. Hence, if we take any function $\Phi(\boldsymbol{r},t)$ and add $\boldsymbol{\nabla}\Phi$ to $\boldsymbol{A}$ and $-\partial\Phi/\partial t$ to V, the fields derived from these new potentials will be the same. We are free to choose $\boldsymbol{A}$ and V such that

$$\boldsymbol{\nabla}\cdot\boldsymbol{A}+\frac{1}{c^2}\frac{\partial V}{\partial t}=0\,. \tag{21}$$

The choice of such a relation between $\boldsymbol{A}$ and V is called a gauge condition and in particular, (21) is known as the Lorentz gauge condition. In this gauge, the potentials $\boldsymbol{A}$ and V satisfy very similar equations and have simple expressions in terms of the charge and current distributions [1]:

$$\boldsymbol{A}(\boldsymbol{r},t)=\frac{\mu_0}{4\pi}\int\frac{\boldsymbol{j}(\boldsymbol{r}',t-|\boldsymbol{r}-\boldsymbol{r}'|/c)}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3\boldsymbol{r}'\,, \tag{22}$$

$$V(\boldsymbol{r},t)=\frac{1}{4\pi\varepsilon_0}\int\frac{\rho(\boldsymbol{r}',t-|\boldsymbol{r}-\boldsymbol{r}'|/c)}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3\boldsymbol{r}'\,, \tag{23}$$

where the integrals are taken over the region occupied by sources. These relations are the expressions for the retarded potentials, which are similar in form to the formulas in magnetostatics and electrostatics. The main difference comes from the fact that, in these expressions, the influence of a charge or current element at $\boldsymbol{r}'$ at time t is only felt at $\boldsymbol{r}$ at a later time $t+|\boldsymbol{r}-\boldsymbol{r}'|/c$. Hence the term 'retarded' potentials.

1.5 Dipole Radiation

We shall now calculate the electromagnetic field in vacuum created at $\boldsymbol{r}$ by a monochromatic electric dipole at the origin. The polarisation density vector, or dipole density vector, is given in complex notation by

$$\boldsymbol{P}(\boldsymbol{r},t)=\boldsymbol{p}_0\mathrm{e}^{-\mathrm{i}\omega t}\delta(\boldsymbol{r})\,.$$

The current density is related to the dipole density by $\boldsymbol{j}=\partial\boldsymbol{P}/\partial t$, whence $\boldsymbol{j}(\boldsymbol{r},t)=-\mathrm{i}\omega\boldsymbol{p}_0\mathrm{e}^{-\mathrm{i}\omega t}\delta(\boldsymbol{r})$. We set $\boldsymbol{r}=r\boldsymbol{u}_r$. From the expression for the retarded potentials, we write the vector potential in the form $\boldsymbol{A}(\boldsymbol{r},t)=\boldsymbol{A}(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}\omega t}$, where

$$\boldsymbol{A}(\boldsymbol{r})=-\frac{\mathrm{i}\omega\mu_0}{4\pi}\frac{\boldsymbol{p}_0\mathrm{e}^{\mathrm{i}kr}}{r}\,, \tag{24}$$

and $k=\omega/c$. From the expression for the vector potential $\boldsymbol{A}$ and the general relation $\boldsymbol{\nabla}\times(f\boldsymbol{A})=f\boldsymbol{\nabla}\times\boldsymbol{A}+(\boldsymbol{\nabla}f)\times\boldsymbol{A}$, we obtain the following formula for the magnetic induction $\boldsymbol{B}$:

$$\boldsymbol{B}(\boldsymbol{r})=-\frac{\mathrm{i}\omega\mu_0}{4\pi}\mathrm{e}^{\mathrm{i}kr}\left(\frac{\mathrm{i}k}{r}-\frac{1}{r^2}\right)\boldsymbol{u}_r\times\boldsymbol{p}_0\,. \tag{25}$$

In vacuum, the fourth Maxwell equation can be used to relate the electric field to the magnetic induction by

$$\boldsymbol{E} = -\frac{c^2}{\mathrm{i}\omega}\boldsymbol{\nabla} \times \boldsymbol{B}\,.$$

Using the general relations

$$\boldsymbol{\nabla}\times(\boldsymbol{u}_r\times\boldsymbol{p}_0) = -(\boldsymbol{p}_0\cdot\boldsymbol{u}_r)\frac{\boldsymbol{u}_r}{r} - \frac{\boldsymbol{p}_0}{r}\,, \quad \boldsymbol{u}_r\times(\boldsymbol{u}_r\times\boldsymbol{p}_0) = (\boldsymbol{p}_0\cdot\boldsymbol{u}_r)\boldsymbol{u}_r - \boldsymbol{p}_0\,,$$

we obtain the following expression for the electric field $\boldsymbol{E}$:

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{\mathrm{e}^{\mathrm{i}kr}}{4\pi\varepsilon_0} \times \left[\left(-\frac{k^2}{r} - \frac{\mathrm{i}k}{r^2} + \frac{1}{r^3}\right)\boldsymbol{u}_r\times(\boldsymbol{u}_r\times\boldsymbol{p}_0) + 2\boldsymbol{u}_r(\boldsymbol{p}_0\cdot\boldsymbol{u}_r)\left(\frac{1}{r^3} - \frac{\mathrm{i}k}{r^2}\right)\right], \quad (26)$$

or alternatively,

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{k^3\mathrm{e}^{\mathrm{i}kr}}{4\pi\varepsilon_0}\left\{\left[\frac{1}{kr} + \frac{\mathrm{i}}{(kr)^2} - \frac{\mathrm{i}}{(kr)^3}\right]\boldsymbol{p}_0 + \boldsymbol{u}_r(\boldsymbol{p}_0\cdot\boldsymbol{u}_r)\left[\frac{3}{(kr)^3} - \frac{3\mathrm{i}}{(kr)^2} - \frac{1}{kr}\right]\right\}. \quad (27)$$

This electric field can also be written in the form of a contraction of a second rank tensor (matrix) $\overleftrightarrow{\boldsymbol{G}}_0$ with the vector $\boldsymbol{p}_0$, viz.,

$$\boldsymbol{E}(\boldsymbol{r}) = \overleftrightarrow{\boldsymbol{G}}_0(\boldsymbol{r})\cdot\boldsymbol{p}_0\,, \quad (28)$$

where

$$\overleftrightarrow{\boldsymbol{G}}_0(\boldsymbol{r}) = \frac{k^3\mathrm{e}^{\mathrm{i}kr}}{4\pi\varepsilon_0}\left\{\left[\frac{1}{kr} + \frac{\mathrm{i}}{(kr)^2} - \frac{\mathrm{i}}{(kr)^3}\right]\overleftrightarrow{\boldsymbol{I}} + (\boldsymbol{u}_r\otimes\boldsymbol{u}_r)\left[\frac{3}{(kr)^3} - \frac{3\mathrm{i}}{(kr)^2} - \frac{1}{kr}\right]\right\}, \quad (29)$$

with $\overleftrightarrow{\boldsymbol{I}}$ the identity matrix and $\boldsymbol{u}_r\otimes\boldsymbol{u}_r$ the tensor product of $\boldsymbol{u}_r$ with itself.

To clarify these expressions for $\boldsymbol{E}$ and $\boldsymbol{B}$, take $\boldsymbol{p}_0 = p_0\boldsymbol{z}$ and consider spherical coordinates $\boldsymbol{u}_r$, $\boldsymbol{u}_\theta$ and $\boldsymbol{u}_\phi$. In this case, $\boldsymbol{u}_r\times\boldsymbol{p}_0 = -p_0\sin\theta\boldsymbol{u}_\phi$ and $\boldsymbol{u}_r\times(\boldsymbol{u}_r\times\boldsymbol{p}_0) = p_0\sin\theta\boldsymbol{u}_\theta$. Then $\boldsymbol{E}$ and $\boldsymbol{B}$ can be written

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{\mathrm{i}\mu_0\omega p_0\sin\theta\mathrm{e}^{\mathrm{i}kr}}{4\pi}\left(\frac{\mathrm{i}k}{r} - \frac{1}{r^2}\right)\boldsymbol{u}_\phi\,, \quad (30)$$

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{p_0\sin\theta\mathrm{e}^{\mathrm{i}kr}}{4\pi\varepsilon_0}\left(-\frac{k^2}{r} - \frac{\mathrm{i}k}{r^2} + \frac{1}{r^3}\right)\boldsymbol{u}_\theta + \frac{2p_0\cos\theta\mathrm{e}^{\mathrm{i}kr}}{4\pi\varepsilon_0}\left(\frac{1}{r^3} - \frac{\mathrm{i}k}{r^2}\right)\boldsymbol{u}_r\,. \quad (31)$$

There are two important regimes here. The first is the short-range regime for which $r \ll \lambda$, i.e., $kr \ll 2\pi$, and the fields become

$$\boldsymbol{B}(\boldsymbol{r}) = -\frac{\mathrm{i}\mu_0\omega p_0 \sin\theta}{4\pi r^2}\boldsymbol{u}_\phi\,, \tag{32}$$

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{2p_0\cos\theta}{4\pi\varepsilon_0 r^3}\boldsymbol{u}_r + \frac{p_0\sin\theta}{4\pi\varepsilon_0 r^3}\boldsymbol{u}_\theta\,. \tag{33}$$

We recognise here the expression for the electrostatic electric field.

The other regime, of particular importance in the present context, is the far-field regime, for which $r \gg \lambda$. The fields then become

$$\boldsymbol{B}(\boldsymbol{r}) = -\frac{\mu_0\omega^2 \mathrm{e}^{\mathrm{i}kr} p_0 \sin\theta}{4\pi rc}\boldsymbol{u}_\phi\,, \tag{34}$$

$$\boldsymbol{E}(\boldsymbol{r}) = -\frac{\omega^2 \mathrm{e}^{\mathrm{i}kr} p_0 \sin\theta}{4\pi\varepsilon_0 rc^2}\boldsymbol{u}_\theta\,. \tag{35}$$

We observe that the fields $\boldsymbol{B}$ and $\boldsymbol{E}$ go as $1/r$ in the far field. They are also orthogonal and in phase. Moreover, we have $|E|/|B| = c$. Hence, far from the dipole, at a point $\boldsymbol{r} = r\boldsymbol{u}_r$, the electromagnetic field behaves as a plane wave propagating in the direction $\boldsymbol{u}_r$. The direction of the plane wave and the structure of the field depend on the observation point. We can say that the electromagnetic field has a local plane wave structure.

We obtain the power radiated by the dipole by calculating the flux of the Poynting vector through a sphere of infinite radius. At an infinite distance from the dipole, we keep only the terms of the far-field approximation in the Poynting vector $\boldsymbol{S} = \Re(\boldsymbol{E}\times\boldsymbol{B}^*)/2\mu_0$. The latter points in the direction $\boldsymbol{u}_r$. Hence,

$$\boldsymbol{S} = \frac{p_0^2\omega^4 \sin^2\theta}{32\pi^2\varepsilon_0 c^3 r^2}\boldsymbol{u}_r\,. \tag{36}$$

We recover the classical dipole radiation pattern. In particular, the emission is maximal in the direction perpendicular to the dipole axis and zero along the axis. The total power is obtained by integrating the flux to obtain

$$P = \int \boldsymbol{S}\cdot\boldsymbol{r}r\,\mathrm{d}\Omega = \frac{p_0^2\omega^4}{32\pi^2\varepsilon_0 c^3}\int_0^\pi \sin^3\theta\,\mathrm{d}\theta\int_0^{2\pi}\mathrm{d}\phi = \frac{p_0^2\omega^4}{12\pi\varepsilon_0 c^3}\,. \tag{37}$$

2 Calculating Radiative Transfer on Short Length Scales

In this section, we shall calculate radiative transfer in contexts involving nanoparticles, treating the latter as electric dipoles. To begin with, we calculate the thermal radiation from a dipole particle throughout space. We then calculate the short-range radiative transfer between two dipole particles at different temperatures.

2.1 Thermal Emission from a Nanoparticle

Consider a particle with characteristic size of the order of ten to a hundred nanometers. We shall be concerned with heat transfer at standard temperatures, so that the electromagnetic field associated with heat exchange will have a typical wavelength of the order of ten microns. We may therefore expect the electromagnetic field to be constant within the particle. We will be able to apply the dipole approximation. This nanoparticle will be characterised by a polarisability $\alpha(\omega)$ relating the electric dipole moment $\boldsymbol{p}$ to the electric field $\boldsymbol{E}$ by $\boldsymbol{p}(\omega) = \varepsilon_0 \alpha(\omega) \boldsymbol{E}(\omega)$. The fluctuation–dissipation theorem (see Sect. 4.2 of the Chapter on Radiative Transfer by *Carminati* in this volume and [3]) tells us that the fluctuations in the dipole moment are related to the imaginary part of the polarisability and to the temperature T according to

$$\langle p_i(\omega) p_j^*(\omega') \rangle = \frac{4\pi}{\omega} \varepsilon_0 \Im\big[\alpha(\omega)\big] \Theta(\omega, T) \delta_{ij} \delta(\omega - \omega') \,, \tag{38}$$

where the energy $\Theta(\omega, T)$ of an oscillator[1] at equilibrium is

$$\Theta(\omega, T) = \frac{\hbar\omega}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1} \,. \tag{39}$$

From the power radiated by an elementary dipole oscillating at frequency ω as given by (37), we shall calculate the power $P(\omega)$ radiated by the particle in a frequency interval $\mathrm{d}\omega$, whereupon the total radiated power will be $P = \int_0^\infty P(\omega)\,\mathrm{d}\omega$. In the present case the dipole is given by

$$\boldsymbol{p}(t) = \int_{-\infty}^{+\infty} \boldsymbol{p}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \frac{\mathrm{d}\omega}{2\pi} \,.$$

In order to calculate the radiation in a frequency band $\mathrm{d}\omega$, we take the expression for a monochromatic dipole and replace p_0 by $p(\omega)\,\mathrm{d}\omega/2\pi$ and $\delta(\omega - \omega')$ by $1/\,\mathrm{d}\omega$. Note that, since thermal dipoles are random, one must sum the contributions from the three polarisation directions ($|p_i|^2$, $i = x, y, z$).

[1] The term $\hbar\omega/2$ does not appear in this expression because we are concerned with energy transfer. For a more detailed discussion, see [4].

Moreover, in the heat transfer context, we are only concerned[2] with positive frequency quantities [4]. Hence,

$$P(\omega)\,\mathrm{d}\omega = 4 \times \frac{\omega^4}{12\pi\varepsilon_0 c^3} \times 3\langle |p_x(\omega)|^2 \rangle \left(\frac{\mathrm{d}\omega}{2\pi}\right)^2 = \frac{\omega^3}{\pi^2 c^3} \Im[\alpha(\omega)] \Theta(\omega, T)\,\mathrm{d}\omega\,. \tag{40}$$

The power radiated into space is proportional to the imaginary part of the polarisability, which is related to the scattering cross-section of the particle.

2.2 Radiative Power Exchanged Between Two Spherical Nanoparticles

Consider two spherical nanoparticles labelled 1 and 2, with dielectric constants $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ and temperatures T_1 and T_2, respectively. We shall now calculate the power dissipated in each of them. The heat exchange will be the difference between these two powers. We begin by calculating the power of the electromagnetic field of particle 1 dissipated by particle 2, viz.,

$$P_{1\to 2}(\omega) = \frac{1}{2}\Re \int_2 \boldsymbol{j}(\boldsymbol{r}', \omega) \cdot \boldsymbol{E}^*(\boldsymbol{r}', \omega)\,\mathrm{d}^3\boldsymbol{r}'\,. \tag{41}$$

Now in particle 2, $\boldsymbol{j}(\boldsymbol{r}', \omega) = -\mathrm{i}\omega\varepsilon_0(\varepsilon_2 - 1)\boldsymbol{E}_{\mathrm{loc}}(\boldsymbol{r}', \omega)$, where $\boldsymbol{E}_{\mathrm{loc}}$ is the local field[3] in particle 2. Hence,

$$P_{1\to 2}(\omega) = \frac{\omega\varepsilon_0}{2} \int_2 \Im(\varepsilon_2) |\boldsymbol{E}_{\mathrm{loc}}(\boldsymbol{r}', \omega)|^2\,\mathrm{d}^3\boldsymbol{r}'\,. \tag{42}$$

We are still in the context of the dipole approximation. The spherical nanoparticle is considered to be immersed in a uniform incident field $\boldsymbol{E}_{\mathrm{inc}}$. We then know [1] that, to a first approximation, the incident field is related to the local field by

$$\boldsymbol{E}_{\mathrm{loc}}(\boldsymbol{r}') = \frac{3}{\varepsilon_2 + 2}\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}')\,. \tag{43}$$

[2] The quantities $x(t)$ we are concerned with here are real. Their Fourier transforms therefore satisfy the relation $x(-\omega) = x^*(\omega)$. All spectral information in real signals is thus contained in the positive frequencies. We introduce analytic signals x^{a} which, for positive frequencies, have the same value as the Fourier transform of the signal and, for negative frequencies, are zero. We have $x(t) = 2\Re[x^{\mathrm{a}}(t)]$. Here, $P(\omega)$ is the analytic signal associated with the real power. Since it is a quadratic quantity, it follows that $P(\omega)$ is equal to four times the value of the power calculated with the complex quantities.

[3] The electric displacement is $\boldsymbol{D} = \varepsilon_0\varepsilon\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P}$. We deduce that $\boldsymbol{P} = \varepsilon_0(\varepsilon - 1)\boldsymbol{E}$. Since $\boldsymbol{j} = \partial\boldsymbol{P}/\partial t = -\mathrm{i}\omega\boldsymbol{P}$, we deduce that $\boldsymbol{j}(\omega) = -\mathrm{i}\omega\varepsilon_0(\varepsilon - 1)\boldsymbol{E}(\omega)$ for the spectral quantities.

Since the polarisability of a spherical particle of radius a is given by

$$\alpha_2 = 4\pi a^3 \frac{\varepsilon_2 - 1}{\varepsilon_2 + 2}\,, \tag{44}$$

we find that

$$P_{1\to 2}(\omega) = \varepsilon_0 \frac{\omega}{2} \Im(\alpha_2) |\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)|^2\,, \tag{45}$$

where $\boldsymbol{r}_2$ is the position of particle 2. We now calculate the incident field on particle 2 produced by the fluctuating currents of particle 1 at temperature T_1. Let $\overleftrightarrow{\boldsymbol{G}}(\boldsymbol{r}_2, \boldsymbol{r}_1)$ be the vacuum Green tensor giving the electric field at $\boldsymbol{r}_2$ produced by an elementary dipole placed at $\boldsymbol{r}_1$. With the usual conventions for Green functions in electromagnetism, we have

$$\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega) = \mu_0 \omega^2 \overleftrightarrow{\boldsymbol{G}}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega) \cdot \boldsymbol{p}\,.$$

Equation (45) involves the squared modulus of $\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)$. To use the fluctuation–dissipation theorem (38), we evaluate $\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)\boldsymbol{E}^*_{\mathrm{inc}}(\boldsymbol{r}_2, \omega')$ (scalar product of two vectors). Hence,

$$\begin{aligned}
&\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)\boldsymbol{E}^*_{\mathrm{inc}}(\boldsymbol{r}_2, \omega') \\
&= \mu_0 \omega^2 \omega'^2 \sum_{n,m,l=x,y,z} G_{nm}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega) G^*_{nl}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega') p_m(\boldsymbol{r}_1, \omega) p_l(\boldsymbol{r}_1, \omega')\,.
\end{aligned} \tag{46}$$

Taking the ensemble average of $\boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)\boldsymbol{E}^*_{\mathrm{inc}}(\boldsymbol{r}_2, \omega')$ and using (38), we obtain

$$\begin{aligned}
&\left\langle \boldsymbol{E}_{\mathrm{inc}}(\boldsymbol{r}_2, \omega)\boldsymbol{E}^*_{\mathrm{inc}}(\boldsymbol{r}_2, \omega') \right\rangle \\
&\qquad = 4\pi \mu_0^2 \varepsilon_0 \omega^3 \Im[\alpha_1(\omega)] \Theta(\omega, T_1) \sum_{n,m} |G_{nm}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega)|^2 \delta(\omega - \omega')\,.
\end{aligned} \tag{47}$$

It remains only to find the various components of the Green tensor which were almost calculated in the last section. Indeed, (27) gives the expression for the electric field produced by a monochromatic elementary dipole $\boldsymbol{p}_0$. To obtain the Green function in a frequency band $\mathrm{d}\omega$, we follow the same procedure as in the last section. We replace $\delta(\omega - \omega')$ by $1/\,\mathrm{d}\omega$ and $\overleftrightarrow{\boldsymbol{G}}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega)$ by $\overleftrightarrow{\boldsymbol{G}}_0(\boldsymbol{r}_2 - \boldsymbol{r}_1) \times (\mathrm{d}\omega/2\pi)^2 \times 1/\mu_0\omega^2$. Taking into account the fact that we are using analytic signals to restrict to positive frequencies [4] and keeping only terms in $1/(kr)^3$, the power transferred from particle 1 to particle 2 over a frequency band $\mathrm{d}\omega$ is

$$\begin{aligned}
P_{1\to 2}(\omega)\,\mathrm{d}\omega = 4 \times \varepsilon_0 \frac{\omega}{2} \Im(\alpha_2) \times 4\pi \mu_0^2 \varepsilon_0 \omega^3 \Im[\alpha_1(\omega)] \Theta(\omega, T_1) \\
\times \frac{\mathrm{d}\omega}{4\pi^2} \times \sum_{n,m} |G_{nm}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega)|^2\,.
\end{aligned} \tag{48}$$

Equation (29) implies that

$$\sum_{n,m} |G_{nm}(\boldsymbol{r}_2, \boldsymbol{r}_1, \omega)|^2 = \frac{3}{8\pi^2 k^4 |\boldsymbol{r}_2 - \boldsymbol{r}_1|^6} \,. \tag{49}$$

In the same way, we find an analogous equation for the power transferred from nanoparticle 2 to nanoparticle 1. Finally, after simplifying all the factors, the power exchanged between the two nanoparticles becomes

$$P_{1\leftrightarrow 2} = \frac{3}{4\pi^3} \frac{\Im[\alpha_1(\omega)]\Im[\alpha_2(\omega)]}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|^6} \Big[\Theta(T_1, \omega) - \Theta(T_2, \omega)\Big] \,. \tag{50}$$

Note the spatial dependence of this transfer, going as $1/d^6$, typical of the induced dipole–induced dipole interaction. These are interactions of van der Waals type. Microscopically, they can be interpreted in the following way. Thermal or quantum fluctuations distort the charge distributions in a nanoparticle, thereby generating a fluctuating dipole. This fluctuating dipole induces a field on the other nanoparticle which distorts its charge distribution, thereby generating a second dipole. The interaction between these two correlated dipoles induces an energy transfer and a force, the so-called van der Waals force.

In addition to this, (50) shows that the transfer depends on the imaginary part of the polarisation of each nanoparticle. For spherical particles of radius a, it can be shown that

$$\Im[\alpha(\omega)] = 4\pi a^3 \frac{3\Im[\varepsilon(\omega)]}{|\varepsilon(\omega) + 2|^2} \,. \tag{51}$$

We thus observe that, if the real part of the dielectric constant is close to -2 at some frequency and if its imaginary part is small, $\Im(\alpha)$ will be large. There is then a resonance at the surface of the sphere. Such resonances occur in metals at visible wavelengths and in polar materials (like glass, SiC, and III–V and II–VI semiconductors) at infrared wavelengths. Figure 1 shows the power exchanged between two spherical nanoparticles of amorphous SiO_2 (glass) of radius 5 nm, set 100 nm apart. One of the particles is at 500 K and the other at 300 K. Note that power is only exchanged at certain wavelengths corresponding to resonances of the material. The total power exchanged is 1.69×10^{-15} W. This leads to a time scale of 10^{-2} s for changes induced in the temperature of a nanoparticle.

3 Thermal Near-Field Emission from a Plane Surface

In this section, we shall study the electromagnetic field above a flat interface separating a material at temperature T (medium 2, permittivity ε) from a vacuum (medium 1) (see Fig. 2). We investigate this electromagnetic field

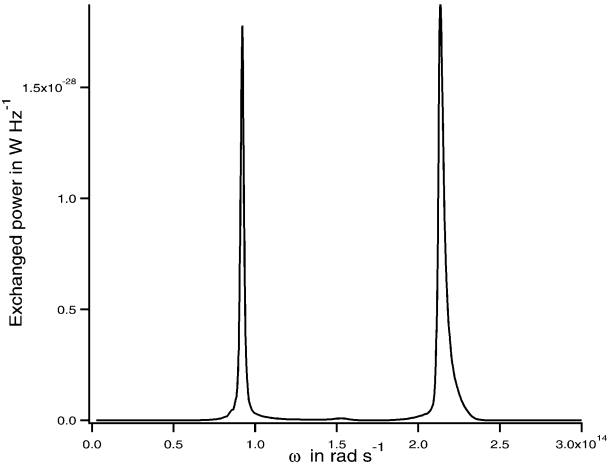

Fig. 1. Power exchanged between two spherical nanoparticles of amorphous SiO_2 (glass) of radius 5 nm, set 100 nm apart

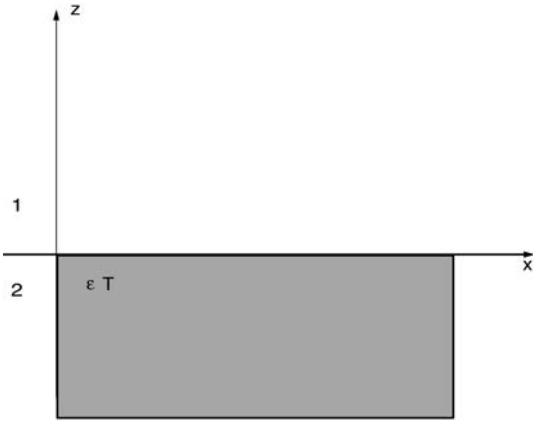

Fig. 2. Plane interface separating a semi-infinite material (permittivity ε) at temperature T from the vacuum

via two quantities, the radiative flux and the electromagnetic energy density, which will be calculated using the Green functions for a system with an interface. Consider the energy density, for example. To obtain this quantity, we must calculate $\langle \boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle$ and $\langle \boldsymbol{H}(\boldsymbol{r},\omega)\boldsymbol{H}^*(\boldsymbol{r}',\omega')\rangle$. If we know the Green functions $\overleftrightarrow{\boldsymbol{G}}^E$ and $\overleftrightarrow{\boldsymbol{G}}^H$ relating the current in the material to the

field above the interface, then we can use the fluctuation–dissipation theorem to calculate the required quantities. If there is an electric current $\boldsymbol{j}$ at $\boldsymbol{r}_2$, then

$$\boldsymbol{E}(\boldsymbol{r}_1,\omega) = \mathrm{i}\mu_0\omega \int \overleftrightarrow{\boldsymbol{G}}^{E}(\boldsymbol{r}_1,\boldsymbol{r}_2,\omega)\cdot\boldsymbol{j}(\boldsymbol{r}_2,\omega)\,\mathrm{d}^3\boldsymbol{r}_2\,, \tag{52}$$

$$\boldsymbol{H}(\boldsymbol{r}_1,\omega) = \int \overleftrightarrow{\boldsymbol{G}}^{H}(\boldsymbol{r}_1,\boldsymbol{r}_2,\omega)\cdot\boldsymbol{j}(\boldsymbol{r}_2,\omega)\,\mathrm{d}^3\boldsymbol{r}_2\,. \tag{53}$$

For example, let us calculate $\langle \boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle$, the averaged value of the scalar product of the fields:

$$\langle \boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle = \left\langle \mu_0^2\omega\omega' \int \overleftrightarrow{\boldsymbol{G}}^{E}(\boldsymbol{r},\boldsymbol{r}',\omega) \right.$$
$$\left. \cdot\,\boldsymbol{j}(\boldsymbol{r}',\omega)\,\mathrm{d}^3\boldsymbol{r}'\overleftrightarrow{\boldsymbol{G}}^{E^*}(\boldsymbol{r}_1,\boldsymbol{r}'',\omega')\cdot\boldsymbol{j}(\boldsymbol{r}'',\omega')\,\mathrm{d}^3\boldsymbol{r}''\right\rangle. \tag{54}$$

The fluctuation–dissipation theorem applied to the electric currents tells us that

$$\langle j_m(\boldsymbol{r}',\omega)j_l^*(\boldsymbol{r}'',\omega')\rangle = 4\pi\varepsilon_0\Im(\varepsilon)\omega\Theta(\omega,T)\delta_{ml}\delta(\boldsymbol{r}'-\boldsymbol{r}'')\delta(\omega-\omega')\,. \tag{55}$$

Substituting this into (54) yields

$$\langle|\boldsymbol{E}(\boldsymbol{r},\omega)|^2\rangle$$
$$= 4\pi\mu_0^2\omega^3\varepsilon_0\Theta(\omega,T)\sum_{m,l=x,y,z}\int \Im[\varepsilon''(\boldsymbol{r}',\omega)]|G^E_{ml}(\boldsymbol{r},\boldsymbol{r}')|^2\,\mathrm{d}^3\boldsymbol{r}'\,. \tag{56}$$

The last integration is only carried out over the lower half-space because the imaginary part of the dielectric constant is zero in the upper half-space. The Green functions for a system with an interface can be found in the literature [5]. When the currents are in medium 2 and the field is required in medium 1, the Green functions are given by

$$\overleftrightarrow{\boldsymbol{G}}^{E}(\boldsymbol{r},\boldsymbol{r}') = \frac{\mathrm{i}}{16\pi^3}\int\frac{1}{\gamma_2}(\hat{s}t_{21}^s\hat{s}+\hat{p}_1^+t_{21}^p\hat{p}_2^+)\mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}')}\mathrm{e}^{\mathrm{i}(\gamma_1 z-\gamma_2 z')}\,\mathrm{d}^2K\,, \tag{57}$$

$$\overleftrightarrow{\boldsymbol{G}}^{H}(\boldsymbol{r},\boldsymbol{r}') = -\frac{k_0}{16\pi^3}\int\frac{1}{\gamma_2}(-\hat{p}_1^+t_{21}^s\hat{s}+\hat{s}t_{21}^p\hat{p}_2^+)\mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}')}\mathrm{e}^{\mathrm{i}(\gamma_1 z-\gamma_2 z')}\,\mathrm{d}^2K, \tag{58}$$

where $\boldsymbol{r}=(\boldsymbol{R},z)$, the wave vector is $\boldsymbol{k}=(\boldsymbol{K},\gamma)$, and

$$\gamma_i = \varepsilon_i k_0^2 - K^2\,,\quad \boldsymbol{K}=K\hat{K}\,,\quad \hat{s}=\hat{K}\times\hat{z}\,,\quad \hat{p}_i^{\pm}=\frac{|K|\hat{z}\mp\gamma\hat{K}}{n_i k_0}\,,$$

with

$$t_{21}^{s} = \frac{2\gamma_2}{\gamma_1 + \gamma_2}\,, \quad t_{21}^{p} = \frac{2n_1 n_2 \gamma_2}{\varepsilon_1\gamma_2 + \varepsilon_2\gamma_1}\,,$$

the Fresnel transmission factors in polarisations s or p. These Green functions are expressed as integrals of elementary plane waves. From the expressions for the Green functions, we see that the behaviour of the Fresnel factors will play a dominating role in the behaviour of the energy density.

The calculations leading to the energy density, and later the Poynting vector, are somewhat tedious. Since they all follow the same pattern, we shall only go into all the details for $\langle|\boldsymbol{E}(\boldsymbol{r},\omega)|^2\rangle$. To begin with, we calculate

$$\int_{z<0} \Im(\varepsilon)|G_{ml}^{E}(\boldsymbol{r},\boldsymbol{r}')|^2\,\mathrm{d}^3\boldsymbol{r}'\,.$$

Substituting in the expression for the Green tensor of the electric field, we obtain

$$\begin{aligned}
&\int_{z<0} \Im(\varepsilon)|G_{ml}^{E}(\boldsymbol{r},\boldsymbol{r}')|^2\,\mathrm{d}^3\boldsymbol{r}' \\
&\quad = \frac{1}{256\pi^6}\int_{-\infty}^{0}\mathrm{d}z'\int \mathrm{d}^2\boldsymbol{R}'\,\mathrm{d}^2\boldsymbol{K}\,\mathrm{d}^2\boldsymbol{K}'\frac{\mathrm{e}^{\mathrm{i}(\boldsymbol{K}-\boldsymbol{K}')\cdot(\boldsymbol{R}-\boldsymbol{R}')}}{\gamma_2(\boldsymbol{K})\gamma_2^*(\boldsymbol{K}')} \\
&\qquad \times \mathrm{e}^{\mathrm{i}z[\gamma_1(\boldsymbol{K})-\gamma_1^*(\boldsymbol{K}')]}\mathrm{e}^{\mathrm{i}z'[\gamma_2^*(\boldsymbol{K}')-\gamma_2(\boldsymbol{K})]} \\
&\qquad \times \Big[(\hat{e}_m\cdot\hat{s})t_{21}^{s}(\hat{s}\cdot\hat{e}_l) + (\hat{e}_m\cdot\hat{p}_1^{+})t_{21}^{p}(\hat{p}_2^{+}\cdot\hat{e}_l)\Big] \\
&\qquad \times \Big[(\hat{e}_m\cdot\hat{s})t_{21}^{s*}(\hat{s}\cdot\hat{e}_l) + (\hat{e}_m\cdot\hat{p}_1^{+*})t_{21}^{p*}(\hat{p}_2^{+*}\cdot\hat{e}_l)\Big]\,. \qquad (59)
\end{aligned}$$

The integration over $\boldsymbol{R}'$ produces a Dirac distribution in the integral equal to $4\pi^2\delta(\boldsymbol{K}-\boldsymbol{K}')$. The integration over $\boldsymbol{K}'$ is then immediate. The integration over z' is also simple when the dielectric constant is uniform throughout the material. One then integrates $\exp[-2\Im(\gamma_2 z')]$ from $-\infty$ to 0. One can also use the relation $\omega^2/c^2\varepsilon'' = 2\Re(\gamma_2)\Im(\gamma_2)$ in medium 2. Finally, the squared modulus of the electric field can be written

$$\begin{aligned}
\langle\boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle = \frac{\mu_0\omega\Theta(\omega,T)}{16\pi^3}\int \mathrm{d}^2\boldsymbol{K}\frac{\Re(\gamma_2)}{|\gamma_2|^2}\mathrm{e}^{-2\Im(\gamma_1)z} \\
\times \Big[(\hat{e}_m\cdot\hat{s})t_{21}^{s}(\hat{s}\cdot\hat{e}_l) + (\hat{e}_m\cdot\hat{p}_1^{+})t_{21}^{p}(\hat{p}_2^{+}\cdot\hat{e}_l)\Big] \\
\times \Big[(\hat{e}_m\cdot\hat{s})t_{21}^{s*}(\hat{s}\cdot\hat{e}_l) + (\hat{e}_m\cdot\hat{p}_1^{+*})t_{21}^{p*}(\hat{p}_2^{+*}\cdot\hat{e}_l)\Big]\,. \qquad (60)
\end{aligned}$$

It remains to calculate all the terms in the second and third lines of the integrand. This is a straightforward but lengthy exercise. After grouping all the terms, we observe that the cross terms involving products of Fresnel

coefficients for different polarisations contribute nothing to the final result. Hence,

$$\langle \boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle = \frac{\mu_0\omega\Theta(\omega,T)}{8\pi^2}\int_0^\infty K\,\mathrm{d}K\,\frac{\Re(\gamma_2)}{|\gamma_2|^2}\mathrm{e}^{-2\Im(\gamma_1)z} \times\left[|t_{21}^s|^2+\frac{|t_{21}^p|^2(K^2+|\gamma_1|^2)(K^2+|\gamma_2|^2)}{|n_1|^2|n_2|^2k_0^4}\right]. \quad (61)$$

We also have the following relations:

$$\Re(\gamma_1)(1-|r_{12}^s|^2)+2\Im(\gamma_1)\Im(r_{12}^s)=\Re(\gamma_2)\frac{|\gamma_1|^2}{|\gamma_2|^2}|t_{12}^s|^2\,, \quad (62)$$

$$\Re(\varepsilon_1^*\gamma_1)(1-|r_{12}^s|^2)+2\Im(\varepsilon_1^*\gamma_1)\Im(r_{12}^s)=\Re(\varepsilon_2^*\gamma_2)\frac{|n_1|^2}{|n_2|^2}\frac{|\gamma_1|^2}{|\gamma_2|^2}|t_{12}^s|^2\,, \quad (63)$$

$$\Re(\varepsilon_1^*\gamma_1)=\Re(\gamma_1)\frac{|\gamma_1|^2+K^2}{k_0^2}\,, \quad (64)$$

$$\Im(\varepsilon_1^*\gamma_1)=\Im(\gamma_1)\frac{K^2-|\gamma_1|^2}{k_0^2}\,. \quad (65)$$

The expression for the squared modulus of the electric field now becomes

$$\begin{aligned}&\langle \boldsymbol{E}(\boldsymbol{r},\omega)\boldsymbol{E}^*(\boldsymbol{r}',\omega')\rangle\\ &\quad=\frac{\mu_0\omega\Theta(\omega,T)}{8\pi^2}\left\{\int_0^{\omega/c}\frac{K\,\mathrm{d}K}{\gamma_1}\left(\frac{1-|r_{12}^s|^2}{2}+\frac{1-|r_{12}^p|^2}{2}\right)\right.\\ &\qquad\left.+\int_{\omega/c}^\infty\frac{2K\,\mathrm{d}K}{\gamma_1}\left[\Im(r_{12}^s)+\Im(r_{12}^p)\left(\frac{2K^2}{k_0^2}-1\right)\right]\mathrm{e}^{-2|\gamma_1|z}\right\}. \end{aligned}\quad (66)$$

From (19), multiplied by 4 because we are using analytic signals, and feeding in the above calculation together with the result for the squared modulus of the magnetic field, we have

$$u(\omega,T)=\frac{\Theta(\omega,T)\omega}{2\pi^2c^2}\left\{\int_0^{\omega/c}\frac{K\,\mathrm{d}K}{\gamma_1}\left(\frac{1-|r_{12}^s|^2}{2}+\frac{1-|r_{12}^p|^2}{2}\right)\right.\\ \left.+\int_{\omega/c}^\infty\frac{K^3\,\mathrm{d}K}{k_0^2|\gamma_1|}\left[\Im(r_{12}^s)+\Im(r_{12}^p)\right]\mathrm{e}^{-2|\gamma_1|z}\right\}. \quad (67)$$

The expression for the energy density is thus the sum of two terms. In the first, the parallel wave vector lies between 0 and ω/c, i.e., it corresponds to a wave propagating in vacuum. In the second, the parallel wave vector is greater than ω/c, i.e., it corresponds to a wave decaying exponentially, known as an evanescent wave. Since these waves decay, they are not expected to contribute a long way from the interface. In the presence of a 'black'

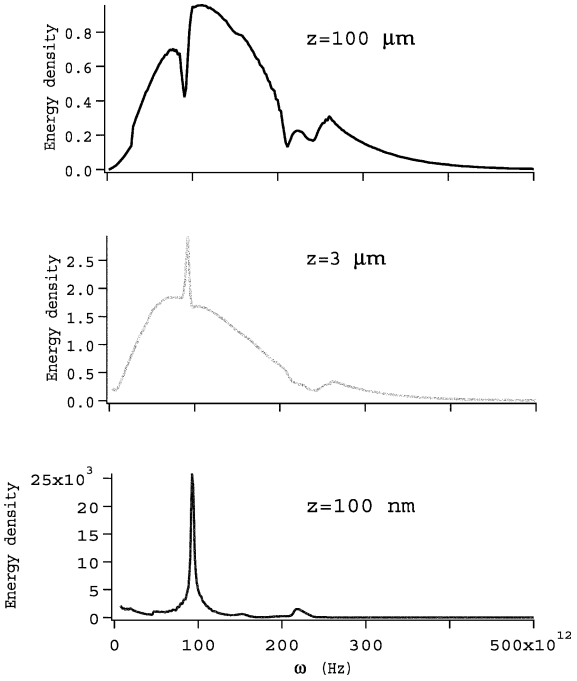

Fig. 3. Energy density for three different heights above an interface separating a semi-infinite medium of amorphous SiO_2 at temperature $T = 300\,\mathrm{K}$ from the vacuum

material, i.e., one in which the Fresnel reflection factors r_{12}^s and r_{12}^p are zero, we find an energy density equal to half of what would be found inside a cavity in thermodynamic equilibrium. This is comforting, because we have indeed calculated the electromagnetic energy density due to the presence of thermal currents in a *half*-space.

Figure 3 shows the electromagnetic energy density at various distances from a half-space filled with amorphous silica at $T = 300\,\mathrm{K}$. This material is polar and has surface resonances, something we shall return to shortly. At a distance of $100\,\mu\mathrm{m}$ from the interface, i.e., a greater distance than the characteristic wavelength of thermal radiation, we observe a curve for the electromagnetic energy density that is very similar to that of a black body, except that there are 'gaps' at certain wavelengths. These correspond to frequencies for which glass is highly reflective. When we move closer to the surface, at distances shorter than the typical wavelength of thermal radiation, we find that the energy density increases and that peaks appear at those same wavelengths for which there were gaps further out. Finally, at very short distances, we find that the energy density has increased at certain wavelengths by more than 4 orders of magnitude, and that it has become almost monochromatic. This behaviour contrasts starkly with that of a black

body. It is due to the presence of surface waves close to the interface, which decrease as one moves perpendicularly away from the interface.

Surface waves are electromagnetic waves which, physically, are associated with a collective oscillation in the material. In a metal, surface waves are associated with a collective oscillation of electrons. One then speaks of surface plasmon polaritons. These resonances occur in the visible wavelength range. In a polar dielectric material, surface waves are associated with a collective oscillation of optical phonons. One then has phonon polaritons. From the standpoint of electromagnetic theory, these waves only exist in polarisation p. The dispersion relation between their wave vector parallel to the interface and the angular frequency ω is

$$K = \frac{\omega}{c}\sqrt{\frac{\varepsilon}{\varepsilon+1}}\,. \tag{68}$$

A necessary condition for the existence of such waves is that the dielectric constant be less than -1. Figure 4 shows the dispersion relation of a surface wave. Note that the modulus of K is always greater than ω/c. This implies that the surface waves are evanescent waves and that they will play a negligible role far from the interface. We also observe the presence of an asymptote at a certain angular frequency corresponding to $\Re[\varepsilon(\omega)] = -1$. For slightly lower frequencies than this critical value, there are a great many surface waves for an interval $\mathrm{d}\omega$. This implies that the density of states of the surface waves will be large close to the resonance frequency. Since the energy density is the product of the density of states and the mean energy of a state, we may predict that the peaks in the energy density will correspond to peaks in the density of states, i.e., to asymptotes of the dispersion relations of the surface waves. These peaks also correspond to the resonances of $\Im(r_{12}^p)$. Indeed,

$$\Im(r_{12}^p) = \frac{2\Im(\varepsilon)}{|\varepsilon+1|^2}$$

has a peak when the dielectric constant is close to -1 and its imaginary part is not too large. Note that, from a practical point of view, the local detection of the energy density above a surface can be used to measure the local electromagnetic density of states [6].

We can also calculate the radiative flux from a semi-infinite space at temperature T. To do so, we must calculate the ensemble average of the Poynting vector due to the presence of fluctuating currents. Given the geometry of the system, we only need to evaluate

$$\langle S_z(\boldsymbol{r},\omega)\rangle = \tfrac{1}{2}\Re[\langle E_x H_y^* - E_y H_x^*\rangle]\,.$$

By analogous calculations to those used to find the energy density, it can be shown that this quantity has the form

$$\langle S_z(\boldsymbol{r},\omega)\rangle = \frac{\hbar\omega}{4\pi^2}\frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1}\int_0^{\omega/c} K\,\mathrm{d}K(1-|r_{12}^s|^2+1-|r_{12}^p|^2)\,. \tag{69}$$

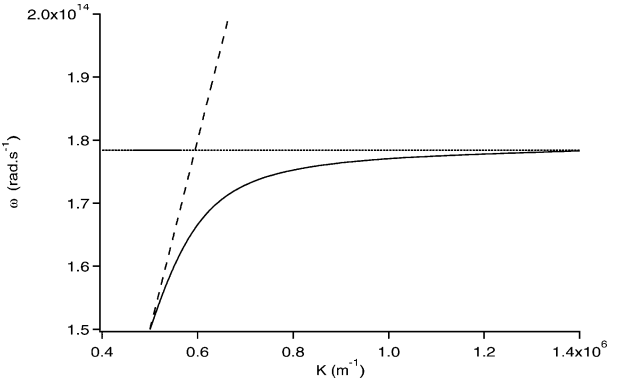

Fig. 4. Dispersion relation for surface waves propagating along an interface separating the vacuum from a polar material. The dispersion relation is located under the light cone (*dotted line*), thereby revealing the evanescent nature of the surface waves

We see that only the propagative waves contribute to the expression for the radiative flux. There is no surprise here, since there are no waves coming in from positive z values.[4] Moreover, it can be shown that, if there is axial symmetry about the direction normal to the interface, then $2\pi K\,\mathrm{d}K = \mathrm{d}\Omega\cos\theta$, and hence,

$$\langle S_z(\boldsymbol{r},\omega)\rangle = \frac{\hbar\omega^3}{4\pi^2c^2}\frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1}\int_{\Omega=2\pi}\frac{\cos\theta\,\mathrm{d}\Omega}{2\pi}(1-|r_{12}^s|^2+1-|r_{12}^p|^2)\,. \tag{70}$$

For a black body, we find that the emitted flux is indeed equal to $\pi L_\omega^0(T)$, where

$$L_\omega^0(T) = \frac{\hbar\omega^3}{4\pi^3c^2}\frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1}$$

is the black body intensity. Otherwise, we observe that the expression for the flux has the same form as in classical theory:

$$q(\omega) = \langle S_z(\boldsymbol{r},\omega)\rangle = \int \mathrm{d}\omega\varepsilon'_\omega(\theta)L_\omega^0(T)\cos\theta\,\mathrm{d}\Omega\,, \tag{71}$$

provided that we identify the emissivity $\varepsilon'_\omega(\theta) = (1-|r_{12}^s|^2+1-|r_{12}^p|^2)/2$.

[4] Consider a scalar field $\psi(z) = a\mathrm{e}^{\mathrm{i}kz}+b\mathrm{e}^{-\mathrm{i}kz}$. For such a field, the current is given by $j = \Im(\psi\,\mathrm{d}\psi^*/\,\mathrm{d}z)$. If k is real, then $j = k(|a|^2-|b|^2)$, whereas if k is pure imaginary, the current is proportional to $\Im(ab^*)$. We thus see that evanescent waves can only contribute to the current in the presence of an 'advancing' wave and a 'retreating' wave, i.e., both a and b different from zero. This argument also applies to vector fields for which the current is then the Poynting vector.

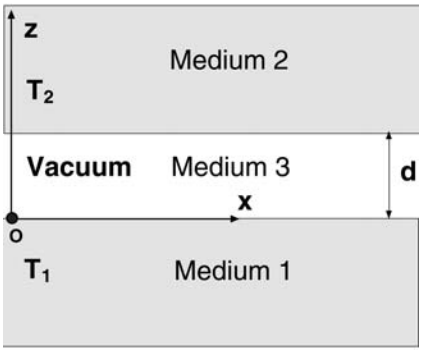

Fig. 5. Two semi-infinite materials 1 and 2, separated by a medium 3 (the vacuum) of thickness d

4 Near-Field Radiative Transfer Between Two Planes

In this section, we shall be concerned with the radiative transfer occurring between two semi-infinite spaces at different temperatures T_1 and T_2, separated by a vacuum, as shown in Fig. 5. The media 1 and 2 are characterised by dielectric constants ε_1 and ε_2, and medium 3 is the vacuum. The radiative transfer between the two media is characterised by the radiative flux. Recall that, according to the classical theory of radiation, this radiative flux can be written

$$q(\omega) = \int_0^{2\pi} \cos\theta \, \mathrm{d}\Omega \int_0^{\infty} \mathrm{d}\omega \frac{\varepsilon'_{1\omega}\varepsilon'_{2\omega}}{1-\rho'_{1\omega}\rho'_{1\omega}} \left[L^0_\omega(T_1) - L^0_\omega(T_2) \right], \tag{72}$$

where the $\varepsilon'_{i\omega}$ are the directional emissivities and the $\rho'_{i\omega}$ are the directional reflectances.

To carry out a full calculation, we must once again calculate the ensemble average of the Poynting vector. The idea is the same as before, except that the Green functions of the system have changed. Indeed, the flux can be written [7, 8]

$$q(\omega) = \langle S_z(d^+, T_1) \rangle - \langle S_z(0^-, T_2) \rangle \,, \tag{73}$$

where $S_z(d^+, T_1)$ is the vertical component of the Poynting vector at $z = d^+$ and due to the fluctuating currents in medium 1. Likewise, $S_z(0, T_2)$ is the vertical component of the Poynting vector at $z = 0$ and due to the fluctuating currents in medium 2. To to this calculation, we thus require the Green

functions relating the currents in medium 1 to the field in medium 2. These Green functions are given by

$$\overleftrightarrow{\boldsymbol{G}}^{E}(\boldsymbol{r},\boldsymbol{r}',\omega)=\frac{\mathrm{i}}{16\pi^3}\int\frac{1}{\gamma_1}(\hat{s}t_{12}^{s}\hat{s}+\hat{p}_1^{+}t_{12}^{s}\hat{p}_2^{+})\mathrm{e}^{\mathrm{i}[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}')]}\mathrm{e}^{\mathrm{i}\gamma_2 z-\mathrm{i}\gamma_1 z'}\,, \tag{74}$$

$$\overleftrightarrow{\boldsymbol{G}}^{H}(\boldsymbol{r},\boldsymbol{r}',\omega)=-\frac{k_0}{16\pi^3}\int\frac{n_2}{\gamma_1}(-\hat{p}_2^{+}t_{12}^{s}\hat{s}+\hat{s}t_{12}^{p}\hat{p}_1^{+})\mathrm{e}^{\mathrm{i}[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}')]}\mathrm{e}^{\mathrm{i}\gamma_2 z-\mathrm{i}\gamma_1 z'}\,, \tag{75}$$

where

$$t_{12}^{s,p}=\frac{t_{13}^{s,p}t_{32}^{s,p}\mathrm{e}^{\mathrm{i}\gamma_3 d}}{1-r_{13}^{s,p}r_{32}^{s,p}\mathrm{e}^{2\mathrm{i}\gamma_3 d}}\,. \tag{76}$$

We observe that the Green function for two interfaces is the same as the Green function for one interface, except that the single-interface transmission coefficient has been replaced by a generalised double-interface transmission coefficient which takes multiple reflections into account. It can be shown that the flux can be expressed as a sum of two terms $q(\omega)=q^{\mathrm{prop}}(\omega)+q^{\mathrm{evan}}(\omega)$. The first term $q^{\mathrm{prop}}(\omega)$ is the contribution from propagating waves:

$$q^{\mathrm{prop}}(\omega)=\sum_{i=s,p}\int\frac{\mathrm{d}\omega\,\mathrm{d}\Omega\cos\theta}{2}\frac{(1-|r_{31}^{i}|^2)(1-|r_{32}^{i}|^2)}{|1-r_{31}^{i}r_{32}^{i}\mathrm{e}^{2\mathrm{i}\gamma_3 d}|^2}\Big[L_\omega^0(T_1)-L_\omega^0(T_2)\Big]\,. \tag{77}$$

Note that $1-|r_{31}^{s,p}|^2$ and $1-|r_{32}^{s,p}|^2$ are the energy transmission factors between media 3 and 1 and media 3 and 2 for polarisations s and p, respectively. As we saw previously, these transmission coefficients can be identified with an emissivity. There is a great similarity between the formula for the flux for propagating waves and the classical expression for the radiative flux between two semi-infinite materials. The denominator seems different. However, note that the flux expression includes an integral over frequency. The function $\mathrm{e}^{\mathrm{i}\gamma_3 d}$ varies much more quickly with ω than the Fresnel factors. We thus find an average of $|1-r_{31}^{i}r_{32}^{i}\mathrm{e}^{2\mathrm{i}\gamma_3 d}|^2$ when we integrate. This average is equal to $1-|r_{31}^{i}|^2|r_{32}^{i}|^2$. Identifying the reflectance with the squared modulus of the Fresnel reflection factor, it follows that the expression for the classical radiative transfer between media 1 and 2 is equal to the contribution of the propagating waves to this transfer.

The second term in the expression for the flux is

$$q^{\mathrm{evan}}(\omega)=\sum_{s,p}\int\mathrm{d}\omega\int_{\omega/c}^{\infty}2K\,\mathrm{d}K\mathrm{e}^{-2\Im(\gamma_3)d}\frac{\Im(r_{31}^{i})\Im(r_{32}^{i})}{|1-r_{31}^{i}r_{32}^{i}\mathrm{e}^{2\mathrm{i}\gamma_3 d}|^2}\frac{L_\omega^0(T_1)-L_\omega^0(T_2)}{k_0^2}\,. \tag{78}$$

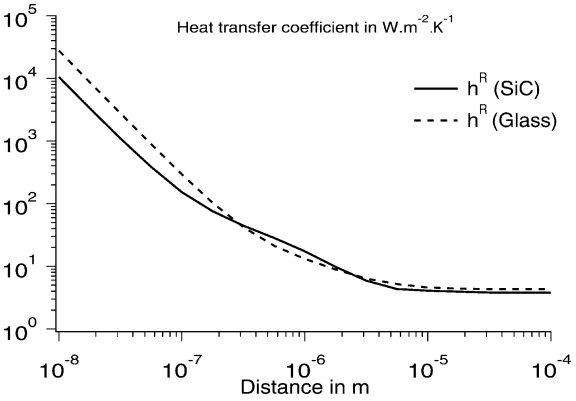

Fig. 6. Radiative transfer coefficient for different separations between the semi-infinite materials at temperature $T = 300\,\mathrm{K}$ [8]

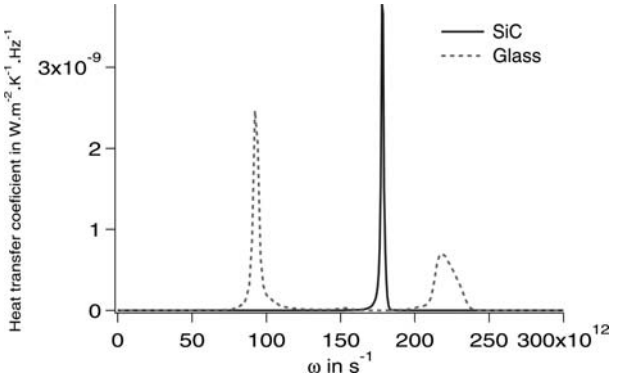

Fig. 7. Monochromatic transfer coefficient for separation $d = 10\,\mathrm{nm}$ at temperature $T = 300\,\mathrm{K}$ [8]

This is the contribution from evanescent waves. This is not zero when there are two interfaces because there are advancing and retreating waves in the cavity. As the separation decreases, this term becomes more and more important, as can be seen from the exponential $\mathrm{e}^{-2\Im(\gamma_3)d}$ in the integrand. When the materials are capable of carrying surface waves, the evanescent waves propagating along each interface can interact. A further transfer can then occur by the tunnel effect.

We define a radiative transfer coefficient as the limit of the ratio of the radiative flux and the temperature difference between the two media when this temperature difference tends to zero, viz.,

$$h^{\mathrm{R}}(\omega) = \lim_{T_1 - T_2 \to 0} \frac{q(\omega)}{T_1 - T_2}\,. \tag{79}$$

Figure 6 shows $h^{\mathrm{R}}(\omega)$ as a function of the distance between two semi-infinite media made of glass or SiC. For a separation greater than the wavelength of the thermal radiation, i.e., for $d > 10\,\mu\mathrm{m}$, the transfer does not depend on the separation. At shorter separations, the transfer increases as $1/d^2$ with decreasing d. For a separation of 10 nm, the transfer coefficient has increased by 4 orders of magnitude relative to its far-field value. If we consider, for a separation of 10 nm, the dependence of $h^{\mathrm{R}}(\omega)$ on the frequency (see Fig. 7), we find that the transfer is large in the near field for frequencies corresponding to surface waves. The heat transfer is thus practically monochromatic in the near field due to the presence of surface waves, as mentioned earlier. We can obtain an asymptotic expression for the transfer:

$$h^{\mathrm{R}}(\omega) \sim \frac{1}{d^2} \frac{\Im(\varepsilon_1)\Im(\varepsilon_2)}{|1+\varepsilon_1|^2|1+\varepsilon_2|^2} \times k_{\mathrm{B}} \left(\frac{\hbar\omega}{k_{\mathrm{B}}T}\right)^2 \frac{\mathrm{e}^{\hbar\omega/k_{\mathrm{B}}T}}{(\mathrm{e}^{\hbar\omega/k_{\mathrm{B}}T}-1)^2}\,. \tag{80}$$

This expression explains the $1/d^2$ dependence of the transfer coefficient as well as its strong frequency dependence. Indeed, when there are surface waves, in particular when the dielectric constant is close to -1, we find that the radiative transfer coefficient has a peak, just like the Fresnel reflection factor.

5 Conclusion

We have seen that the calculation of thermal radiation requires two ingredients. The first is the laws of classical electromagnetism, used in classical radiation calculations for the current in an antenna, for example. The second is the fluctuation–dissipation theorem which relates current fluctuations to thermal or quantum fluctuations in the medium. We have applied these thermal calculations to several simple systems. We observe that, at short distance scales, thermal radiation can behave quite differently from the predictions of classical radiation based on a radiometric approach. We have seen in particular that the presence of surface waves can significantly increase heat transfer on the nanoscale.

References

[1] J. D. Jackson: *Classical Electrodynamics*, 2nd ed. (John Wiley, New York 1975) 110, 112, 116

[2] Y. N. Obukhov, F. W. Hehl: Electromagnetic energy-momentum and forces in matter, Phys. Lett. A **311**, 277–304 (2003) 110

[3] S. M. Rytov, Y. A. Kravtsov, V. I. Tatarskii: *Principles of Statistical Radiophysics*, vol. 3 (Springer, Berlin 1989) 115

[4] K. Joulain, J.-P. Mulet, F. Marquier, R. Carminati, J.-J. Greffet: Surface electromagnetic waves thermally excited: Radiative heat transfer, coherence properties and Casimir forces revisited in the near field, Surf. Sci. Rep. **57**, 59–112 (2005) 115, 116, 117

[5] J. Sipe: New Green-function formalism for surface optics, J. Opt. Soc. Am. B **4**, 481 (1987) 120
[6] K. Joulain, R. Carminati, J.-P. Mulet, J.-J. Greffet: Definition and measurement of the local density of electromagnetic states close to an interface, Phys. Rev. B **68**, 245405 (2003) 124
[7] D. Polder, D. Van Hove: Theory of radiative heat transfer between closely spaced bodies, Phys. Rev. B **4**, 3303 (1971) 126
[8] J. P. Mulet, K. Joulain, R. Carminati, J.-J. Greffet: Enhanced radiative heat transfer at nanometric distances, Microscale Thermophys. Eng. **6**, 209–222 (2002) 126, 128

Index

Monte Carlo Method

Sebastian Volz

Laboratoire d'Energétique Moléculaire et Macroscopique, Combustion (EM2C), Ecole Centrale Paris, 92295 Châtenay-Malabry Cedex, France
sebastian.volz@em2c.ecp.fr

Abstract. The study of heat transfer on short length and time scales involves understanding the behaviour of microscopic heat carriers in order to quantify the average heat and energy flux. In dilute media, an intermediate statistical analysis is based on a local average of these carriers, namely, the velocity distribution function, and the Boltzmann equation. This same equation also governs the dynamical evolution of the local electron density, the number of phonon modes in conduction, and the spectral intensity in radiation. The approach is invaluable, especially when the carrier collision term is easy to integrate and the geometry of the system is simple.

However, if the interaction physics is not given in terms of these densities by a simple relation, there is no choice but to return to the carrier trajectories. This is the goal of the technique known as molecular dynamics. When the geometry is also complex and as a consequence the system size becomes large, the computation time required by molecular dynamics calculations is prohibitive. The Monte Carlo method takes microscopic constituents into account by sampling them in a relevant way, in order to produce the most accurate statistical averages possible on the basis of a limited number of operations. The behaviour of a small number of carriers is thus described independently under realistic physical constraints. Random sampling involves choosing the carriers and the rules governing their behaviour according to probabilistic laws imposed by:

- the laws of physics,
- a reasonable computation time,
- optimal accuracy for the calculated average value.

Today the Monte Carlo method is used in many areas of research. It is a basic tool for statisticians [1] and chemists use it to study equilibrium configurations of structures and molecules [2, 3], or to predict phase change phenomena [4].

Monte Carlo calculations of neutron fluxes [5], radiative fluxes [6–8], and matter fluxes in fluid flows [9] can be found in a correspondingly vast literature. The Monte Carlo technique can even be extremely effective in solving transport problems in complex geometries [10]. It has thus become a standard tool for tackling transport problems.

However, its application to heat fluxes, either on very short length and time scales in gases, or in crystals [11, 12] remains rather novel. In these contexts, it provides a way of taking into account the finer details of physical interaction mechanisms which standard statistical methods could only do at great cost in analytical complexity.

Whereas the Monte Carlo technique only partially describes systems in macroscopic configurations, by selecting certain energy carriers, it is interesting to note

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 133–154 (2007)

that, on short scales and in certain cases, it provides a direct simulation of all elementary carriers and sometimes an even greater number of carriers.

1 Introduction

1.1 Aims

We shall not attempt an exhaustive description of heat flux calculations, since we shall only be concerned with molecular transport in gases and phonon transport in crystals. Short scale electronic and radiative heat flux calculations were discussed in the Chapter by *Lemonnier* and the Chapter by *Joulain* in this volume, respectively. Although the first flux can be deduced from a particle description – it is shown how to treat the electron as a particle in the theoretical part of the book (Part I) – whereupon it may be calculated using the Monte Carlo approach, the second can only be estimated using a wave description and remains inaccessible to this technique.

The solution of the radiative transfer equation by the Monte Carlo method now belongs to the state of the art in this field. Note, however, that it is only valid when the characteristic dimensions are greater than the wavelength of the field, i.e., of the order of a few microns as far as thermal radiation is concerned.

Section 2 describes the basic features of the Monte Carlo technique, and Sect. 3 calculates the heat flux in gases on length scales of the order of the mean free path. An application to the heat exchange between a nanometric tip and a sample surface is discussed. Section 4 discusses phonon transport on length scales of a small multiple of the mean free path.

The approaches adopted in Sects. 3 and 4 are in fact complementary. The two methods apply in the same way to the two types of carrier, but they are based on sampling techniques adapted to the relevant length scales. By organising the Chapter in this way, we incorporate models of two types of carrier: one is the atom or molecule identified with a point mass, and the other is the phonon whose quasi-particle behaviour makes it more like a wave.

1.2 Heat Flux and Energy Carriers

We shall be concerned in this Chapter with the heat flux generated by atomic motions. In gases and dielectric crystals, heat transport is related only to these motions. The atom is treated as a point mass with known trajectory, i.e., position and velocity. We shall show that the heat flux is the sum of a convective flux, in the form of a product of the velocity and the energy, and a work flux of the kind commonly defined in point particle mechanics or electromagnetism.

The expression for this heat flux $\boldsymbol{\phi}$ in terms of the atomic positions $\boldsymbol{r}_i$ and velocities $\boldsymbol{v}_i$, together with the interaction forces $\boldsymbol{F}_{ij}$, is deduced from the energy conservation equation

$$\frac{\partial e}{\partial t} = -\mathrm{div}\boldsymbol{\phi}\,, \tag{1}$$

where e is the energy. This is combined with the microcanonical expression for the instantaneous local energy

$$e = \sum_i e_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)\,, \tag{2}$$

where e_i is the energy of an atom. Multiplying by the position vector $\boldsymbol{r}$ and integrating over the volume V of the system, we arrive at

$$\int_V \boldsymbol{r} \frac{\partial}{\partial t} \sum_i e_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)\,\mathrm{d}V = -\int_V \boldsymbol{r}\mathrm{div}\boldsymbol{\phi}\;\mathrm{d}V\,. \tag{3}$$

Integrating by parts, we obtain

$$\begin{aligned}\frac{\partial}{\partial t} \sum_i \int_V \boldsymbol{r} e_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)\,\mathrm{d}V = &-\int_V \mathrm{div}(r_x\boldsymbol{\phi})\,\mathrm{d}V\boldsymbol{i} - \int_V \mathrm{div}(r_y\boldsymbol{\phi})\,\mathrm{d}V\boldsymbol{j} \\ &- \int_V \mathrm{div}(r_z\boldsymbol{\phi})\,\mathrm{d}V\boldsymbol{k} + \int_V \boldsymbol{\phi}\,\mathrm{grad}r_x\,\mathrm{d}V\boldsymbol{i} \\ &+ \int_V \boldsymbol{\phi}\,\mathrm{grad}r_y\,\mathrm{d}V\boldsymbol{j} + \int_V \boldsymbol{\phi}\,\mathrm{grad}r_z\,\mathrm{d}V\boldsymbol{k}\,,\end{aligned} \tag{4}$$

where subscripts x, y and z refer to the space directions described by unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ making up an orthonormal basis. The first three terms on the right-hand side can be expressed as surface integrals which vanish because the system is assumed to be isolated. The gradient terms are equal to the unit basis vectors, e.g.,

$$\mathrm{grad}r_x = \frac{\partial r_x}{\partial r_x}\boldsymbol{i} = \boldsymbol{i}\,.$$

The right-hand side of (4) thus becomes equal to the average flux over the volume V. On the left, the Dirac distribution localises the integral at $\boldsymbol{r}_i$. Finally, the average flux operator can be written [13]

$$\boldsymbol{\phi}V = \frac{\partial}{\partial t} \sum_i \boldsymbol{r}_i e_i\,. \tag{5}$$

The distribution of the time derivative leads to

$$\boldsymbol{\phi}(t)V = \sum_i \boldsymbol{v}_i e_i + \sum_{i=1}^{n-1} \sum_{j>i} \boldsymbol{r}_{ij}(\boldsymbol{F}_{ij}\boldsymbol{v}_i)\,, \tag{6}$$

when the atomic energy is replaced by its expression in terms of kinetic and potential energies.

In dilute gases, molecules interact on very short time scales compared with the time intervals between collisions. For this reason, the second term on the right-hand side of (6) involving the interaction force becomes negligible on average and the energy e_i reduces to the kinetic energy of the particle. The first term on the right-hand side describes the heat flux stemming from the displacement of an atom carrying only its kinetic energy. This is the convective flux mentioned earlier. We shall calculate this flux by describing the free trajectories between consecutive pairs of instantaneous collisions. These collisions will nevertheless take into account the nature of the interaction potential.

The second term for its part describes the flux stemming from the work done by the motion of an atom in the potential field of its neighbours. This dominates in solid media and hence describes the phonon flux. In the harmonic approximation,[1] and treating the distance between atoms as much greater than their displacements about their equilibrium positions ($\boldsymbol{r}_{ij} = \boldsymbol{r}_{ij}^0 = \text{const.}$), the second term can be rewritten in the form [14, 15]

$$\boldsymbol{\phi}(t)V = \sum_{i=1}^{n-1} \sum_{j>i} \boldsymbol{r}_{ij}^0 \left([\boldsymbol{C}_{ij}] \boldsymbol{r}_{ij} \boldsymbol{v}_i \right) . \tag{7}$$

It can be shown that the product $\boldsymbol{r}_{ij}^0 [\boldsymbol{C}_{ij}]$ is related to the group velocity $\boldsymbol{v}_{\mathrm{g}q}$ and the term $\boldsymbol{r}_{ij}\boldsymbol{v}_i$ is related to the sum of the phonon energies over all modes. These transformations lead to the Peierls expression already obtained in the Chapter on Electrons and Phonons by *Greffet* in this volulme:

$$\boldsymbol{\phi}(t)V = \sum_q \boldsymbol{v}_{\mathrm{g}q} n_q \hbar \omega_q \, . \tag{8}$$

This new expression raises the question of how many phonons there are in the mode q, a number denoted by n_q. These are treated as particles moving at the group velocity $\boldsymbol{v}_{\mathrm{g}q}$. It is only in this framework that one can carry out a statistical calculation of the particle kinetics using Monte Carlo techniques.

After describing the foundations of the Monte Carlo method, we shall show how to use it to calculate the convective flux in Sect. 3 and the work flux in Sect. 4.

[1] The interaction force is assumed linear in the interatomic distance, i.e., $\boldsymbol{F}_{ij} = [\boldsymbol{C}_{ij}] \cdot \boldsymbol{r}_{ij}$, where $\boldsymbol{C}_{ij}$ is the tensor of force constants.

2 Calculating the Heat Flux with the Monte Carlo Method

2.1 Basic Idea

The idea of the Monte Carlo method is to describe the dynamical evolution of objects as a succession of random (or stochastic) processes whose properties are determined by probability distributions. From these independent processes, we extract an average representing a state variable of the system. The probability distributions are designed to optimise the computation time, proportional to the number of trials, and the accuracy. This abstract definition indicates that a very broad class of problems can be handled, from macro- and mesophysical problems, to the social sciences, urban development, stock markets, and so on.

In this book, we shall of course limit the discussion to calculation of the heat flux, considered as a state variable, from the behaviour of the objects that transport the heat. This behaviour is described by a series of processes, the carrier trajectories, governed by probability distributions representing, for example, collisions or interactions with surfaces. Unlike the molecular dynamics technique, these trajectories will not be deterministic processes. Since they obey probabilistic laws, they are called random walks.

2.2 Sampling Random Walks

The energy carrying entities modelled in the Monte Carlo techniques are usually 'packets' of elementary carriers whose properties (direction, velocity, length, etc.) are assumed identical. These are packets of photons in radiation, molecules in fluid mechanics, or phonons for conduction in crystals. However, in the special case of short length scales, the number of elementary carriers can be very small and the true number of these carriers can actually be modelled. The very idea of a packet loses some of its meaning.

In typical models, the trajectories of the packets are assumed to be governed to some extent by the laws determining the behaviour of a single carrier, an assumption that is not always fully justified. In the same way, in the present context, the random walks will be determined by certain physical constraints imposed upon a single molecule or phonon, an assumption that is now easier to justify. For example, the velocity of a packet of molecules emerging from a collision will be determined by the laws of conservation of momentum and energy as defined in point particle mechanics.

These random selection laws could also be identified with some kind of average behaviour obtained by statistical physics. The mean free path of a molecule between two collisions will be chosen according to a probability distribution of Beer type. This guarantees that the mean free path of our carriers is indeed equal to the mean free path measured or predicted by other means.

Apart from their suitability with regard to the constraints of microscopic physics and average characteristics, the probability distributions determining the random walks are also required to shorten computation times. Since the aim is to predict average fluxes or energies, the way the trajectories are sampled will directly determine the rate of convergence of our average to its exact value.

Since these averages are in fact integrals over the parameters of the random walk, let us see what probability distribution we must use to calculate, for example, the flux ϕ_{m} related to the emission of molecules from an infinite plane surface. We define $n_{\mathrm{D}}^{\mathrm{MB}}(v)$ as the number of molecules per unit volume with velocity in the range from v to $v + \mathrm{d}v$. The molecular flux is the sum of the contributions from all molecules:

$$\phi_{\mathrm{m}} = \int_{v_z>0} v_z n^{\mathrm{MB}}(v)\, \mathrm{d}^3 v\,, \tag{9}$$

whose velocity vector $\boldsymbol{v}$ points into the empty half-space. Since the number of molecules $n_{\mathrm{D}}^{\mathrm{MB}}$ is given by the Maxwell–Boltzmann distribution (see the Chapter on Transport in Dilute Media by *Carminati* in this volume), and since also $v_z = v\cos\theta$ and the volume element of the hemisphere is $\mathrm{d}^3 v = v^2 \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}|v|$, it follows that

$$\phi_{\mathrm{m}} = n \int_{v_z>0} v^3 \mathrm{e}^{-mv^2/2k_{\mathrm{B}}T}\, \mathrm{d}|v| \int_0^{\pi/2} \sin\theta\cos\theta\, \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi\,, \tag{10}$$

where n_{D} is the prefactor in the expression for $n_{\mathrm{D}}^{\mathrm{MB}}$ and z is the axis perpendicular to the surface and oriented towards this half-space. The estimation of such a quantity by the Monte Carlo method consists in calculating the three integrals, or averages, which form the expression for the mass flux with the help of a probabilistic approach that we shall now outline.

For example, to calculate the integral over the magnitude of the velocity, we can choose regular intervals between 0 and a cutoff value b. This is the standard rectangle method illustrated in Fig. 1. Another approach involves randomly drawing values for the velocity magnitude and then summing the areas of the rectangles defined by this new sequence of magnitudes. If we use a uniform distribution function, each magnitude is treated as having the same importance. This probability distribution p is given by

$$p(|v|) = \frac{1}{b}\,. \tag{11}$$

If a large enough number of draws is made, the scale on the horizontal axis will be discretised into intervals of identical width and we recover the rectangle method.

The draws are now made in such a way as to associate with a magnitude value v_i a weight equal to the value at v_i of the function to be integrated:

$$p(|v|) = \frac{v^3 \exp(-mv^2/2k_{\mathrm{B}}T)}{\int v^3 \exp(-mv^2/2k_{\mathrm{B}}T)\, \mathrm{d}|v|}\,. \tag{12}$$

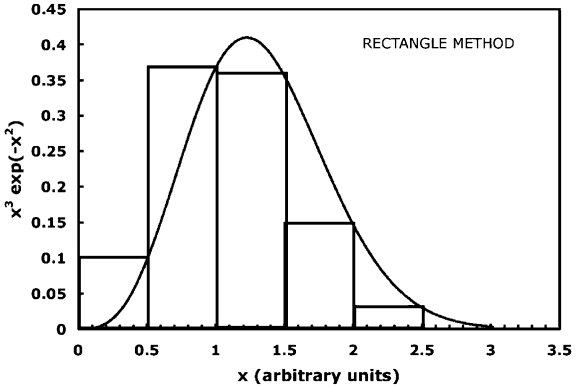

Fig. 1. Calculating an integral by the rectangle method. The accuracy is the same as can be obtained by randomly drawing a sequence of values of x to define the rectangles when the distribution is equiprobable

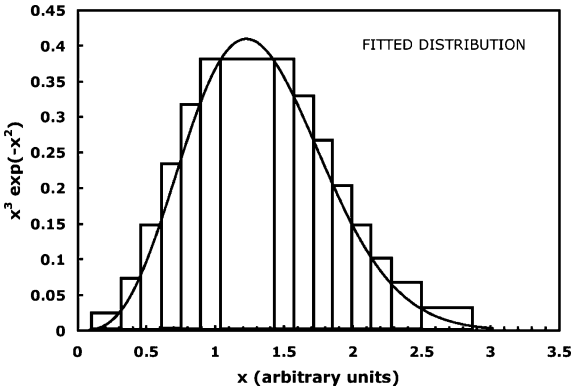

Fig. 2. Calculating an integral by defining rectangles using a random sequence of values of x. The distribution is optimised to calculate the most significant contributions more accurately

Under these conditions, as can be seen from Fig. 2, the greater the contribution of the function to the integral, the more magnitude values will be drawn. Hence, a smaller number of draws will lead to an equal or greater accuracy than can be obtained by the rectangle method. The approach described above is conducted systematically in the treatment of heat transfer by the Monte Carlo method any time a quantity characterising the trajectories has to be summed. Under these conditions, a considerable amount of time can be saved.

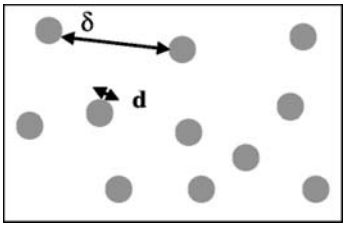

Fig. 3. Dimensions characterising a dilute gas: d is the diameter of a molecule and δ the average distance between two molecules

2.3 Calculating the Statistical Error and Average

Quite generally, the properties of probability functions allow one to establish an expression for the exact integral of a multivalued function f over a volume V in the form

$$I = \int_{\Omega} f(x_1, \ldots, x_N)\, \mathrm{d}\Omega = \frac{V}{N} \sum_{i=1}^{N} f(x_{1i}, \ldots, x_{Ni}) \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}\,, \tag{13}$$

where N is the number of draws. Apart from the gain in computation time, the Monte Carlo method thus provides an exact solution for the value of the integral when the number of draws tends to infinity. This behaviour sets it apart from finite element techniques whose accuracy is limited by the order of the numerical scheme. The last equation shows the possibility of controlling the error in the value of the integral and predicting the gain in accuracy obtained by extending the computation time.

3 Ballistic and Quasi-Ballistic Transport in Gases

Fluid flows are usually modelled using the Navier–Stokes equation, but only if the medium can be considered as a continuous and dilute gas of particles. When the characteristic dimensions are smaller than the mean free path and the continuous medium hypothesis is not valid, the Navier–Stokes equation is no longer applicable. The Monte Carlo method then provides an alternative. After describing the implications of the dilute gas approximation on the nature of collisions, we shall show how to calculate the heat flux without appealing to the continuum approximation.

The Avogadro laws state that the number of molecules per unit volume, or number density n, of a gas depends on the temperature and pressure, but is independent of the composition of the gas. As the volume available to a molecule is $1/n$, the average separation between molecules is the length $\delta = n^{-1/3}$. For low densities, the space available to a molecule is large compared with its diameter d, and only a very small part of space is actually occupied by molecules. Each molecule thus moves around without interacting with

the others. Moreover, when a collision occurs, it is only likely to involve two molecules, i.e., collisions are binary. We have thus defined the condition $\delta \gg d$ for a dilute gas and described its consequences for collisions.

The continuum model is no longer valid when the characteristic length L of the system becomes comparable or less than the mean free path Λ of the molecules [9]

$$\Lambda = \frac{1}{\sqrt{2}\pi(d/2)^2 n}\,.$$

This relation expresses the fact that the volume $\Lambda\pi(d/2)^2$ containing a particle on a trajectory of length equal to the mean free path must contain a collision partner. Λ is the average distance covered by the molecule between two consecutive collisions. When the Knudsen number, defined as the ratio of the mean free path to the characteristic dimension of the system $Kn = \Lambda/L$, is greater than or equal to unity, the molecule moves around without collisions and the transport is said to be quasi-ballistic. For air at atmospheric pressure, the mean free path is 50 nm.

3.1 Molecules and Heat Flux

We now describe the calculation of the heat flux through a gas between two surfaces at different temperatures. In the stationary regime and when the trajectories are purely ballistic (no collisions between molecules), this flux can be expressed as the difference between the fluxes emitted by the two surfaces 1 and 2:

$$\phi_{\mathrm{h}}\Big|_{\alpha} = \int_{v,1\to 2} \frac{1}{2} m v^2 v_\alpha n_{\mathrm{D}}^{\mathrm{MB}}(v)\,\mathrm{d}^3 v - \int_{v,2\to 1} \frac{1}{2} m v^2 v_\alpha n_{\mathrm{D}}^{\mathrm{MB}}(v)\,\mathrm{d}^3 v\,, \qquad (14)$$

where α denotes the direction under consideration. The flux carried by a particle is defined as its kinetic energy multiplied by its speed in direction α. If we now wish to introduce collisions and find out the dynamic evolution and spatial distribution of the flux, the Monte Carlo technique can be used.

In the following, we propose an approach that can be applied when the Knudsen number is greater than 3. Below this, the number of collisions between particles during their motion between the two surfaces becomes too great and the temperatures of the collision partners then depend on their position in the interstice. This situation is discussed in Sect. 4.

We assume that the molecules in the interstice are in equilibrium with their surroundings. The originality of this approach resides in the calculation of a heat flux response that is spatially resolved and temporally impulsive. To estimate the heat flux received by an infinite plane surface in response to a molecular flux ϕ_{m} emitted by a second, finite surface indicated by subscript D, we begin by taking into account the stationary molecular flux emitted:

$$\phi_{\mathrm{m}}^{\mathrm{s}} = S_{\mathrm{D}} \int_{v_z>0} v_z n_{\mathrm{D}}(v)\,\mathrm{d}^3 v = \frac{4 S_{\mathrm{D}} n_{\mathrm{D}}}{(2\pi)^{3/2}} \left(\frac{k_{\mathrm{B}} T_{\mathrm{D}}}{m}\right)^{1/2}. \qquad (15)$$

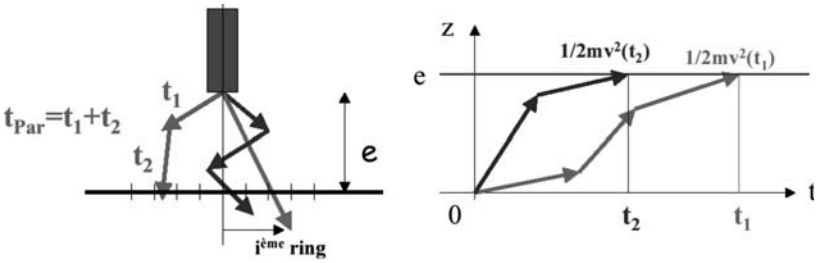

Fig. 4. *Left*: Random walks from a hot to a cold surface. Travel times t_1, t_2, etc., are estimated for each molecule. *Right*: Referring all trajectories to the same time origin, the impulse response is obtained. Times t_1 and t_2 now refer to a time scale in which all molecules leave simultaneously

This has the dimensions of a number of particles per unit time ($\mathrm{d}^3 v = \mathrm{d}|v|\,\mathrm{d}\Omega$, where Ω is the direction variable). S_D denotes the emitting area and $n_\mathrm{D}(v)$ is the number of particles per unit volume with speeds between v and $v + \mathrm{d}v$. The latter is proportional to n_D and the Maxwell distribution. z denotes the vertical axis, orthogonal to the plane surface.

To estimate the response to a pulse of mass flux leaving surface D, treated for example as a disk (represented by the lower part of the cylinder in Fig. 4), trial particles are launched successively from this same surface and the travel time t of each molecule between ejection from the disk and arrival at the receiving surface is calculated. The departure time is then assumed to be the same for all molecules.

When they arrive at the surface, the translational and rotational kinetic energies received are recorded in annular regions of constant width $\mathrm{d}r$ given by the increase in the radius. The result is a time-dependent radial energy function $E(r,t)$ corresponding to the energy deposited between times t and $t + \mathrm{d}t$. The radial heat flux per unit area is then deduced from

$$\langle \phi_\mathrm{h}(r,t) \rangle = \frac{E(r,t)}{\mathrm{d}t\pi(r_i^2 - r_{i-1}^2)N}\,, \tag{16}$$

where N is the total number of particles emitted. This quantity has dimensions of $\mathrm{W}\cdot\mathrm{m}^{-2}$. Physically, this heat flux can be considered as the response to a Dirac pulse of molecular flux coming from the disk. It is therefore the transfer function which relates the excitation (the molecular flux ϕ_m) to the response (the heat flux ϕ_h at the surface):

$$\phi_\mathrm{h}(r,t) = \int_0^t \langle \phi_\mathrm{h}(r, t-\tau) \rangle \phi_\mathrm{m}(\tau)\,\mathrm{d}\tau\,. \tag{17}$$

The right-hand side is a convolution. In the stationary case, the quantity ϕ_m becomes independent of time and the thermal response reduces to the time

integral of the impulse response. This response can thus be deduced from (17) as

$$\phi_{\mathrm{h}}^{\mathrm{s}}(r) = \frac{\int_0^t E(r,\tau)\,\mathrm{d}\tau}{\mathrm{d}t\pi(r_{i+1}^2 - r_i^2)N} S_{\mathrm{D}} \int_{v_z>0}^{t} v_z n(v)\,\mathrm{d}^3 v\,. \tag{18}$$

3.2 Random Walk Distributions

As mentioned in Sect. 2.2, the flux of emitted molecules can be expressed in the form of an integral over the directions of the velocities carrying a term $\cos\theta$, due to the projection of the velocity on the normal, and a term $\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$, which is the element of solid angle:

$$\phi_{\mathrm{m}} = n_{\mathrm{D}} \int_{v_z>0} v^3 \mathrm{e}^{-mv^2/k_{\mathrm{B}}T}\,\mathrm{d}|v| \int_0^{\pi/2} \sin\theta\cos\theta\,\mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi\,. \tag{19}$$

In order to calculate the integrals, the probability distributions of the angles θ and ϕ must be defined by the normalised functions

$$p(\theta)\,\mathrm{d}\theta = \frac{\sin\theta\cos\theta\,\mathrm{d}\theta}{\int_0^{\pi/2}\sin\theta\cos\theta\,\mathrm{d}\theta} = 2\sin\theta\cos\theta\,\mathrm{d}\theta\,, \tag{20}$$

$$p(\phi)\,\mathrm{d}\phi = \frac{\mathrm{d}\phi}{\int_0^{2\pi}\mathrm{d}\phi} = \frac{\mathrm{d}\phi}{2\pi}\,. \tag{21}$$

The angle ϕ is drawn directly as $2\pi R_{\mathrm{f}}$, where R_{f} is a random number drawn equiprobably from the interval $[0,1]$. To obtain the value of $\sin\theta$, we use the relation between the distributions of dependent variables:

$$p(\theta)\,\mathrm{d}\theta = p(R_\theta)\,\mathrm{d}R_\theta\,. \tag{22}$$

By integration, assuming that R_θ is distributed uniformly between 0 and 1, i.e., $p(R_\theta) = 1$, this implies that

$$\int_0^\theta p(\theta')\,\mathrm{d}\theta' = \sin^2\theta = R_\theta\,, \tag{23}$$

and hence that $\sin\theta$ will be obtained by taking the square root of the random number R_θ. The magnitude of the velocity must be drawn with a normalised probability distribution whose primitive can be calculated:

$$p(|v|)\,\mathrm{d}|v| = \frac{v^3\mathrm{e}^{-mv^2/2k_{\mathrm{B}}T}\,\mathrm{d}v}{\int_{v_z>0} v^3\mathrm{e}^{-mv^2/2k_{\mathrm{B}}T}\,\mathrm{d}v} = \left(\frac{k_{\mathrm{B}}T}{m}\right)^2 v^3\mathrm{e}^{-mv^2/2k_{\mathrm{B}}T}\,\mathrm{d}|v|\,, \tag{24}$$

whereupon

$$p(|v|) = 1 - \mathrm{e}^{-mv^2/2k_{\mathrm{B}}T}\left(1 + \frac{mv^2}{k_{\mathrm{B}}T}\right) = R(|v\,. \tag{25}$$

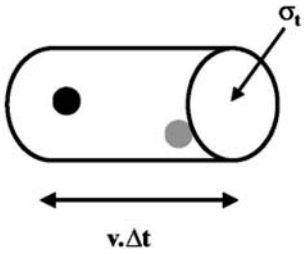

Fig. 5. Interaction volume for two molecules. The *grey molecule* will interact with the *black* one because it is situated within the cylinder determined by the black one

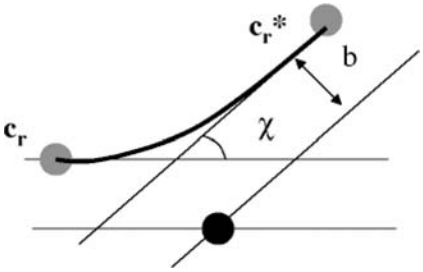

Fig. 6. Centre of mass interaction of the two molecules colliding at the relative velocity with a fixed scattering centre. The angle χ denotes the deflection of the relative velocity vector before and after the collision

To determine the sequence of magnitudes $|v| = P^{-1}(R_{|v|})$, we invert P by seeking the zeros using a technique such as dichotomy.

Finally, the distance travelled l is given by a Beer decay law $p(l) = \mathrm{e}^{-l/\Lambda}$.

3.3 Collision Distributions

In the kinetic approach, a simple model involves treating a molecule as a hard sphere of diameter d. Two molecules collide if they are included in a cylinder of cross-sectional area equal to the total collision cross-section $\sigma_\mathrm{T} = \pi d^2$ and height equal to the product of the relative speed by the free path time Δt, as shown in Fig. 5.

In the Monte Carlo simulation, collision conditions are satisfied for transfer between two surfaces if the molecule is located in the gas interstice at the end of the free path. The velocity at which the test molecule leaves the collision is determined from the energy and momentum conservation laws applied to the relative velocity vector $\boldsymbol{c}_\mathrm{r}$.

The collision distribution for two hard spheres is given by $b = (d/2)\sin\chi$, where b is the impact parameter, i.e., the shortest distance between the two molecules if they do not interact (see Fig. 6). The relative velocity $\boldsymbol{c}_\mathrm{r}$ has conserved magnitude, whilst its direction is determined by the collision dis-

tribution. The velocity $\boldsymbol{c}_1^*$ of the test molecule after the collision is deduced from the conservation laws [9]:

$$\boldsymbol{c}_1^* = \frac{\boldsymbol{c}_1 + \boldsymbol{c}_2 + \boldsymbol{c}_\mathrm{r}^*}{2}\,, \tag{26}$$

where the asterisk indicates the velocity after the collision. The velocity $\boldsymbol{c}_2$ of the collision partner is randomly drawn in the same way as for the initial velocity except that the angle $\theta \in [0, \pi]$ must satisfy $\cos\theta = 1 - 2R_\theta$.

3.4 Transfer Between a Hot Tip and a Surface

We now examine the heat transfer between a hot cylindrical tip with one face parallel to a cold surface. The height e between the tip and surface varies from a small multiple of the mean free path of the gas molecules in the interstice to some fraction of this distance. The standard continuum approach is therefore inapplicable, and so is a purely ballistic calculation. According to the predictions of statistical physics, when the pressure increases, the heat flux must rise in the same proportion [9]. Indeed, for two infinite parallel surfaces held at temperatures T_1 and T_2, this flux is

$$\phi_\mathrm{h} = \frac{P(2k_\mathrm{B})^{3/2}}{k_\mathrm{B}(\pi m)^{1/2}(T_1^{1/2} - T_2^{1/2})}(T_1 - T_2)\,. \tag{27}$$

However, this equation was established for purely ballistic transfer. Now the mean free path, which is inversely proportional to the pressure, will fall when the pressure is raised. Collisions will become more frequent and the flux will drop. The Monte Carlo simulation is able to account for contributions from these two competing effects, in such a way, for example, as to optimise the pressure and the height of the heating element to ensure maximal transfer of heat flux.

We make the following hypothesis. The gas is made up of nitrogen molecules N_2 with collisions governed by the following parameters:

- Diameter of interaction potential at 273 K is $d = 0.415$ nm.
- Molecular mass is 46.5×10^{-27} kg.
- Collision partners are at the temperature T_S of the plane and the surroundings, since it can be shown that they have travelled on average a distance equal to the mean free path.
- Collisions are of hard sphere type.

Moreover, the density per unit volume of particles leaving the disk for the surface is assumed to be given by conservation of mass and the number of molecules in the case of two parallel plates:

$$n_\mathrm{D} = n\frac{T_\mathrm{S}^{1/2}}{T_\mathrm{D}^{1/2} + T_\mathrm{S}^{1/2}}\,, \tag{28}$$

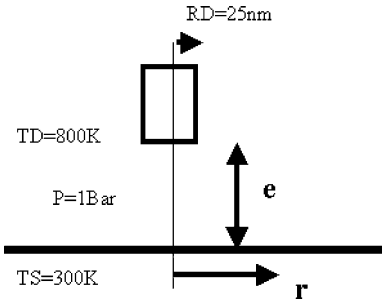

Fig. 7. Schematic view of simulated system. A disk of radius $r_{\mathrm{D}} = 25\,\mathrm{nm}$ and temperature $T_{\mathrm{D}} = 800\,\mathrm{K}$ is parallel to a surface at temperature $T_{\mathrm{S}} = 300\,\mathrm{K}$. The height e of the disk above the surface is variable. The pressure of the exchange gas is imposed in the range 1–30 bar

where n is the average density per unit volume of molecules in the interstice, viz.,

$$n = \frac{P}{k_{\mathrm{B}}(T_{\mathrm{D}}T_{\mathrm{S}})^{1/2}}\,, \tag{29}$$

with k_{B} the Boltzmann constant and P the pressure. The mean free path Λ is simply defined by the relation $\Lambda = 1/\sqrt{2}\pi(d/2)^2 n$. It is assumed again that the molecule is fully adapted to the surface before leaving it, i.e., it is at the same temperature as the surface.

The calculations have all been carried out for $N = 10^6$ trials, a number shown by tests to lead to reasonable accuracy. The computation time varies between about 10 s in the ballistic case to a few minutes when the height e is equal to three times the mean free path. The radius of the disk is equal to 25 nm.

The method is validated by calculating the flux from the disk to the surface in ballistic transfer conditions to begin with, then when the tip–surface separation is 10 nm. We had previously calculated analytic solutions in this configuration. A satisfactory agreement was obtained between the result of the Monte Carlo simulation and the calculation in the ballistic case, with an observed deviation of 0.1 %. Figure 8 shows the values of the exchanged fluxes in molecular nitrogen gas at 1 bar between a disk of radius 25 nm at temperature 800 K and a surface at 300 K.

It can be shown that the heat flux drops off sharply when the tip moves away from the surface, since it has the value $40\,\mathrm{MW}\cdot\mathrm{m}^{-2}$ in the ballistic regime and twenty times less when the tip is located at a height of 100 nm.

We have described the method for calculating the heat flux in a gas when the dimensions of the gas interstice are of the same order as the mean free path and the number of collisions suffered by a test molecule remains less than 3. These dimensions are of the order of a hundred nanometers at normal pressure. We have shown by calculation that the heat flux drops off sharply

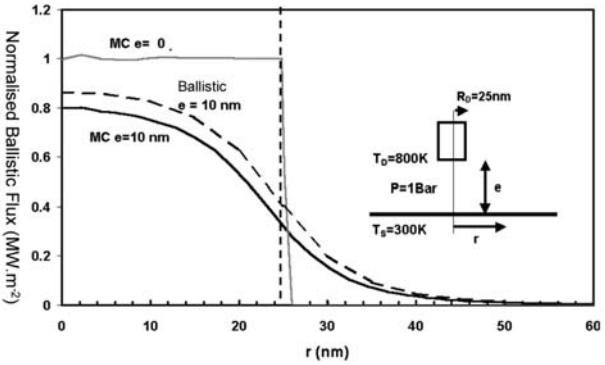

Fig. 8. Radial fluxes exchanged between a hot disk ($T = 800\,\mathrm{K}$) of radius 25 nm placed at variable height e above a cold surface (300 K). The ambient temperature was 1 bar. The *grey continuous curve* corresponds to the pure ballistic case and the configuration with infinite parallel planes. The *black dashed curve* corresponds to the pure ballistic case for the disk and an infinite parallel plane, and the *black continuous curve* to the predictions of the Monte Carlo simulation

when the separation between the heat exchanging surfaces becomes of the order of the mean free path. We shall now examine the problem of heat transfer in solids. The heat carriers are phonons which can be treated as particles whose kinetics is to be determined.

4 Ballistic and Quasi-Ballistic Transport in Insulating Crystals

The problem here is to calculate the phonon flux given in (8) for transport scales varying between a few phonon wavelengths (10 nm) to macroscopic dimensions. Particle kinetics on these scales has already been described for gases [9] and we shall put it to use here. However, there are fundamental differences between phonons in a solid and molecules in a dilute gas. A phonon is a particle characterised by a frequency, a polarisation, a wave vector, a group velocity and a position. Moreover, interactions with other phonons are 3-body interactions, i.e., each involves three such quasi-particles, and these interactions obey rules depending on the above characteristics. The large number of relevant parameters makes the Monte Carlo method a prime candidate. In the Chapter by *Chantrenne* in this volume, it is shown that the domain of application of the molecular dynamics technique is restricted to conditions where the classical limit is observed. In the present case, with realistic phonon populations obeying a Bose–Einstein distribution, heat transfer involves quantum phenomena, in particular at low temperatures. We emphasise that the simulation to be described now has only been achieved recently [12, 13]. Although it could be improved in certain ways, it lays the foundations for

more sophisticated methods in the future. The theoretical points used here can be found in the Chapter on Electrons and Phonons by *Greffet* in this volume.

4.1 Phonons, Temperature and Heat Flux

Equation (8) provides the definition of the heat flux as a function of the number of phonons, the energy quantum, and the group velocity. This definition was introduced in the Chapter on Electrons and Phonons by *Greffet* in this volume.

4.2 Isothermal Cell Technique

Since the characteristic length scale of the system to be described is a small multiple of the mean free path, this system can be subdivided into small cells in which there is local thermodynamic equilibrium. These cells are assumed to be isothermal.

The number of phonons to be handled in such a system is equal to the product of the number of modes, i.e., three times the number of atoms, and the Bose–Einstein statistic (~ 1). Clearly, the simulation method could not handle all the particles over regions with dimensions a few times the mean free path. As in simulation techniques for dilute gases, reference particles are therefore chosen in a proportion R of approximately 10^{-5}, reducing the number of phonons that are actually analysed to about 10^6 if, for example, the simulated volume is a cube of side approximately one micron.

This proportion will thus determine the number of phonon trials in each cell. Before selecting the reference particles, the number of phonons in the cell is given by the product of the total number of phonons N_{tot} and the ratio of the volume V_{C} of the cell to the total volume V_{tot} of the system. Finally, the number N_{C} of phonons selected in the cell can be written

$$N_{\mathrm{C}} = N_{\mathrm{tot}} R \frac{V_{\mathrm{C}}}{V_{\mathrm{tot}}} \sim 10^4 \,. \tag{30}$$

4.3 Modelling Random Walks

The random walk of a phonon is selected by choosing the characteristics of this phonon, i.e., position, frequency, polarisation, wave vector, and group velocity:

– The initial position is drawn uniformly from the volume or surface element.

– The frequency is drawn by dividing the Brillouin zone into regular intervals of width $\Delta\omega$. The probability of drawing a given interval is given by the normalised number of phonons

$$p(\omega) = \frac{n^0(\omega, \mathrm{LA})D(\omega, \mathrm{LA})\Delta\omega + 2n^0(\omega, \mathrm{TA})D(\omega, \mathrm{TA})\Delta\omega}{\sum\limits_{p=\mathrm{LA,2TA}} \int_0^{\omega_\mathrm{D}} n^0(\omega, p)D(\omega, p)\,\mathrm{d}\omega}\,, \tag{31}$$

where n^0 is the average number of phonons given by the Bose–Einstein distribution, $D(\omega)$ is the density of states, i.e., the number of states per frequency interval, and ω_D is the Debye frequency. The factor of 2 corresponds to the two branches of transverse acoustic modes assumed non-degenerate (identical). These are modes in which atoms move perpendicularly to the direction of propagation. The optical branches are neglected here to simplify the discussion. They are only present in crystals containing more than one atom per unit cell in the lattice. They contribute little in themselves, but they can play a role via their interactions with the acoustic modes. Integration of p is carried out numerically and it is inverted in the same way as the velocity magnitude (see Sect. 3.2).

– The polarisation is selected from the average fraction of phonons with longitudinal acoustic (LA) polarisation:

$$p(N_\mathrm{LA}) = \frac{n^0(\omega, \mathrm{LA})D(\omega, \mathrm{LA})}{n^0(\omega, \mathrm{LA})D(\omega, \mathrm{LA}) + 2n^0(\omega, \mathrm{TA})D(\omega, \mathrm{TA})}\,, \tag{32}$$

at frequency ω. TA denotes the two branches of transverse phonons. To obtain the polarisation, a random number can be drawn uniformly between 0 and 1. If this number is less than $p(N_\mathrm{LA})$, the polarisation is longitudinal.

– The wave vector can be calculated in the isotropic approximation. Its magnitude is given by the frequency, assuming that the polarisation and dispersion curve are known. The direction of the vector is drawn in the same way as the direction of the velocity vector in Sect. 2.

– The norm of the group velocity $\boldsymbol{v}_\mathrm{g}$ is found from the dispersion curves, given the frequency and polarisation. The direction of the group velocity is identical to the direction of the wave vector.

The test phonon is now launched. It will move ballistically between two consecutive interactions with other phonons, or with the system surfaces, interfaces, impurities, charges and defects.

Trajectories are described in two stages:

– The position vector is modified by the increment $\boldsymbol{v}_\mathrm{g}\Delta t$. The length of this increment must not exceed the average cell size. This restricts the time increment Δt. When the phonon goes from its initial cell to a neighbouring cell, the energy of the particle is subtracted from the total energy E_i

of the first cell and added to that of the second cell. The temperatures of the cells are then deduced from their energies as time passes, by inverting the formula

$$\frac{E_i}{VR} = \sum_{p,\omega_j} n^0(\omega_j, p, T_i) D(\omega_j, p) \hbar \omega_j \Delta\omega \,, \tag{33}$$

where V denotes the volume of the system and it should be remembered that the number of phonons n^0 depends on the equilibrium temperature. The temperature obtained is injected into the probability distributions (31) and (32).

– The interaction probability is selected. The relaxation time τ_i corresponding to each interaction must be established over the spectrum for the two polarisations and as a function of temperature. The resulting average interaction time is deduced from the Matthiessen rule:

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} \,. \tag{34}$$

The law giving the probability that there is *no* interaction during the time interval Δt is found from the total number of particles $N_{\rm nc}$ that have not suffered collisions at time $t + \Delta t$:

$$N_{\rm nc}(t + \Delta t) = N_{\rm nc}(t) \left(1 - \frac{\Delta t}{\tau}\right) , \tag{35}$$

since $\Delta t/\tau$ gives the probability of collision during the interval Δt. It follows that

$$\frac{N_{\rm nc}(t + \Delta t) - N_{\rm nc}(t)}{\Delta t} = \frac{{\rm d}N_{\rm nc}(t)}{{\rm d}t} = \frac{N_{\rm nc}(t)}{\tau} \,, \tag{36}$$

and hence a time dependence $\mathrm{e}^{-t/\tau}$ for the number of particles that have not undergone collisions. This is therefore the time dependence of the non-interaction probability distribution for our test particle. It is the same as the one used to select the free path length of gas molecules in Sect. 2. Here a uniformly distributed random number greater than $\mathrm{e}^{-\Delta t/\tau}$ will impose interaction.

The interaction leads to a modification of the frequency, polarisation and wave vector of the test phonon. One method involves selecting the frequency as done initially but with the temperature of the cell in which the particle is now located. Likewise for the polarisation and wave vector.

An advanced model [13] makes a distinction between 'normal' contributions, for which the momentum is exactly conserved during the collision, and 'umklapp' interactions, where the sum of the two wave vectors of the interacting phonons cannot be represented in the Brillouin zone. In the first case,

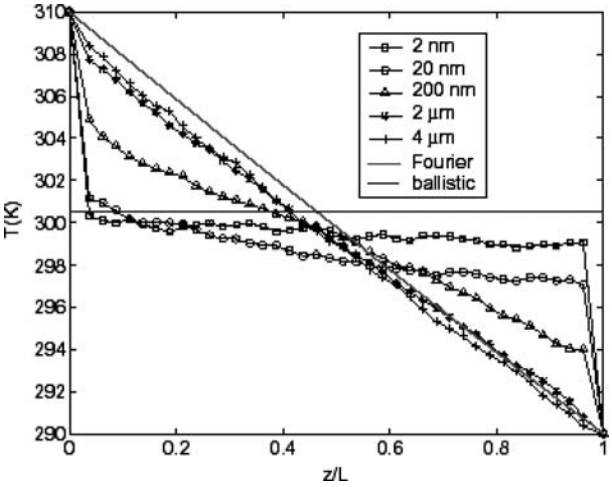

Fig. 9. Temperature profile in a silicon film with totally absorbent walls. *Symbols* correspond to different film thicknesses as indicated

the direction can be assumed unchanged, whereas in the second, it can be assumed random.

Following the interaction processes, the cell energies must be modified whenever energy conservation has not been respected by the frequency correction. Phonons must therefore be added or removed so as to recover the energy calculated for the cell before the interaction. The energy of these created phonons must depend on the cell temperature but also on the probability of destruction of phonons in a mode.

4.4 Conduction in a Thin Film

Consider a silicon film whose faces are held at two different temperatures. Phonons will tend to move from the hot end towards the cold end by reflecting off the lateral boundaries of the system. The temperature profile can be calculated in the stationary regime using the Monte Carlo method. Figure 9 shows the results when the walls are absorbent, for thicknesses between 2 nm and 4 µm (adapted from [13]).

The heat flux crossing the film in its plane is found by calculating the amount of energy crossing a reference plane per unit time. In the stationary regime, the ratio of this flux to the temperature gradient leads to an estimate of the thermal conductivity using the Fourier law. This calculated conductivity, shown in Fig. 10, is in good agreement with measured values [13].

These results show that the Monte Carlo approach described here can be used up to the diffusion limit, i.e., characteristic lengths greater than ten times the mean free path.

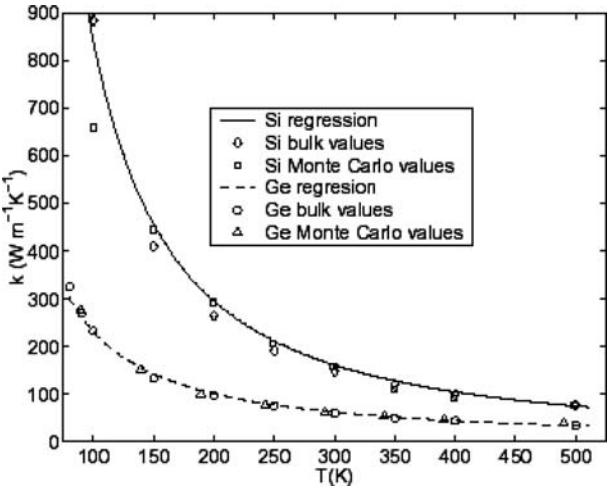

Fig. 10. Conductivity of silicon and germanium at different temperatures. Experimental measurements and Monte Carlo predictions

5 Conclusion

In this Chapter, we have introduced the Monte Carlo method for calculating heat fluxes in gases and dielectrics in terms of the behaviour of elementary heat carriers (molecules or phonons). We have shown how this method can be used to simulate not only the ballistic and quasi-ballistic regimes, but also the diffusive transfer regime. Apart from the possibility of representing most length and time scales, it has been shown that this approach is well suited to distinguishing the various types of carrier and integrating key physical interaction mechanisms. It can handle molecules and phonons just as well as photons and electrons. Finally, it can describe quantum population effects. By virtue of its flexibility and ease of implementation, it can be considered as an essential method for simulating heat transfer on the micro- and nanoscales.

References

[1] C. P. Robert, G. Casella: *Monte Carlo Statistical Methods* (Springer, New York 1999) 133
[2] M. P. Allen, D. J. Tildesley: *Computer Simulation of Liquids* (Clarendon Press, Oxford Science 1987) 133
[3] K. Binder: *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford University Press, Oxford 1995) 133
[4] D. P. Landau, K. Binder: *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge 2000) 133

[5] N. Metropolis, R. Bivins, M. Storm, A. Turkevich, J. M. Miller, G. Friedlander: Monte Carlo calculations on intranuclear cascades. I. Low-energy studies, Phys. Rev. **110**, 185–203 (1958) 133
[6] D. Lemonnier: Méthodes des ordonnées discrètes et de Monte Carlo, in *Cours de Rayonnement* (Ecole de Printemps de Rayonnement, Parent 1996) Chap. 7, pp. 1–66 133
[7] R. Siegel, J. R. Howell: *Thermal Radiation Heat Transfer* (Taylor and Francis-Hemisphere, Washington 2002) 133
[8] M. Modest: *Radiative Heat Transfer*, 2nd ed. (Academic Press, San Diego 2003) 133
[9] G. A. Bird: *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon Press, Oxford 1987) 133, 141, 145, 147
[10] K. K. Sabelfeld: *Monte Carlo Methods in Boundary Value Problems* (Springer, Berlin 1991) 133
[11] F. J. Alexander, A. L. Garcia: *The Direct Simulation Monte Carlo Method* (American Institute of Physics, New York 1997) 133
[12] S. Mazumder, A. Majumdar: Monte Carlo study of phonon transport in solid thin films including dispersion and polarization, J. Heat Transfer **123**, 749 (2001) 133, 147
[13] D. Lacroix, K. Joulain: Monte Carlo transient phonon transport in silicon and germanium at nanoscales, Phys. Rev. B **72**, 064305 (2005) 135, 147, 150, 151
[14] G. Chen: *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules, Phonons, and Photons* (Oxford University Press, Oxford 2005) 136
[15] W. G. Hoover, B. Moran, R. M. More, A. J. C. Ladd: A comparison of molecular dynamics and anharmonic lattice dynamics, Phys. Rev. B **34**, 5058–5064 (1986) 136

Index

Molecular Dynamics

Patrice Chantrenne

Centre de Thermique de Lyon, UMR 5008 CNRS/INSA/UCBL, Institut National des Sciences Appliquées de Lyon, Bâtiment Sadi Carnot, 20 Avenue A. Einstein, 69621 Villeurbanne Cedex
patrice.chantrenne@insa-lyon.fr

Abstract. The aim in this Chapter is to show how molecular dynamics can be used to study conductive heat transfer in matter in terms of an atomic description of that matter. Molecular dynamics can only be used to study heat transfer by phonons, i.e., vibrations of the atomic lattice. It therefore only applies to dielectric materials, i.e., electrical insulators and semiconductors, in which the concentration of free electrons in the lattice is low enough to ensure that heat transfer by electrons is negligible compared with heat transfer by phonons.

There will be two applications here:

- Prediction of the thermal conductivity of macroscopic materials on the basis of a description of their atomic structure: crystals, amorphous materials, with and without defects (voids, substitution defects, interstitial defects, dislocations, and grain boundaries), multilayer composite materials, superlattices, and so on.
- Prediction of the thermal conductivity of nanostructures: nanoparticles, nanowires, nanofilms, nanotubes, and so on.

Since molecular dynamics is not widely used in the heat transfer community, the first part of this Chapter presents the basic principles and implementation of the technique. Section 2 discusses the methods most widely used to calculate the thermal conductivity with the help of molecular dynamics simulations. The thermal conductivity can be calculated on the basis of behavioural models. We shall see in Sect. 3 that molecular dynamics can be used as a tool for determining the vibrational properties of the materials required in these models.

1 Principles of Molecular Dynamics

In this section, we discuss the essential ideas required to understand the principles, implementation and limitations of molecular dynamics. For more details, the reader is referred to the many textbooks devoted to this subject, e.g., [1–4], which served as a basis for the present discussion.

1.1 Definitions and Notation

Molecular dynamics is a numerical computation technique that can simulate the behaviour of materials on the atomic scale. In its simplest version, the atom is treated as an elementary particle. The system studied is thus made

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 155–180 (2007)

up of atoms modelled by point masses, each with three degrees of freedom corresponding to the three space directions (x, y, z). To each atom, labelled by an index i, is associated a mass m_i, a position $\boldsymbol{r}_i$, a velocity $\boldsymbol{v}_i$, and an acceleration $\boldsymbol{a}_i$. Bold face symbols thus denote 3-component vectors corresponding to the three space directions.

The total energy E_i of an atom is equal to the sum of its kinetic energy $E_{\mathrm{k}i}$ and its potential energy $E_{\mathrm{p}i}$. This potential energy is partly due to a field originating outside the system and acting in some specific way on each particle, denoted by $E_{\mathrm{p}i}^{\mathrm{ext}}(\boldsymbol{r}_i)$, e.g., electromagnetic field or gravity, and it is partly due to the interaction potential due to all other atoms or molecules within the system, denoted by $E_{\mathrm{p}i}^{\mathrm{int}}$. The total potential energy of a system is the sum of the external and internal potential energies, viz.,

$$E_{\mathrm{p}} = \sum_i \left(E_{\mathrm{p}i}^{\mathrm{int}} + E_{\mathrm{p}i}^{\mathrm{int}} \right) . \tag{1}$$

The total force $\boldsymbol{F}_i$ on atom i is equal to the gradient of the total potential energy of the system with respect to the position of this atom, i.e.,

$$\boldsymbol{F}_i = -\boldsymbol{\nabla} E_{\mathrm{p}}(\boldsymbol{r}_i) . \tag{2}$$

It can be expressed as the sum $\boldsymbol{F}_i^{\mathrm{ext}} + \boldsymbol{F}_i^{\mathrm{int}}$ of an external force resulting from the external potential field and an internal force due to the potential energy of interaction between atoms:

$$\boldsymbol{F}_i^{\mathrm{ext}} = -\sum_j \boldsymbol{\nabla} E_{\mathrm{p}j}^{\mathrm{ext}}(\boldsymbol{r}_i) , \tag{3}$$

$$\boldsymbol{F}_i^{\mathrm{int}} = -\sum_j \boldsymbol{\nabla} E_{\mathrm{p}j}^{\mathrm{int}}(\boldsymbol{r}_i) , \tag{4}$$

The total force $\boldsymbol{F}_i$ is related to the acceleration $\boldsymbol{a}_i$ by Newton's law

$$\boldsymbol{F}_i = m_i \boldsymbol{a}_i . \tag{5}$$

The velocity and position vectors of all particles are found by integrating this equation with respect to time for all atoms.

When the system is in equilibrium, molecular dynamics provides an ensemble of states of the system described by the positions, velocities, and forces associated with all particles as a function of time. Macroscopic properties of the system can then be calculated using statistical physics. These include elastic constants, binding energies, surface energies, phonon dispersion curves and densities of states, etc. When the system is not in equilibrium, the dynamical evolution of the set of positions and velocities is used to determine transport properties such as viscosity, thermal conductivity, or diffusion coefficient, or to study phenomena such as the creation and propagation of defects, diffusion, phase transformations, surface reconstructions, etc. The choice of system and the way molecular dynamics is implemented will depend on the properties of the phenomenon under investigation. The various systems can be categorised as follows:

- microcanonical if the particle number N, the volume V, and the total energy E are all constant during the simulation,
- canonical if it is the particle number, volume, and temperature T that are constant,
- isobaric–isothermal if it is the particle number, temperature, and pressure P that are constant,
- grand canonical if it is the volume, temperature, and potential energy that are constant, but the particle number is allowed to vary.

In molecular dynamics, all numerical values of the quantities used or calculated are small: displacements are less than one nanometer, forces are of the order of the nanonewton, energies are of the order of the electronvolt ($1.6 \times 10^{-19}\,\mathrm{J}$), and the time steps used in the calculation are of the order of the femtosecond. The units usually used in physics are not well-suited to this, since they would lead to large numerical errors. A change of units is therefore necessary. The most widespread choice, although not the only one, is based on the definition of a new time unit:

$$u_\mathrm{t} = \frac{\sqrt{mc}}{10^{10}}\,, \tag{6}$$

where m is the mass of the atom under consideration and c the reciprocal of the electronvolt, viz., $c = 6.24145 \times 10^{18}\,\mathrm{eV/J}$.

With this unit of time, the following relations can be used to convert values of quantities used in molecular dynamics to the values expressed in SI units:

$$\Delta t_\mathrm{r}(\mathrm{s}) = \Delta t_\mathrm{MD}\frac{\sqrt{mc}}{10^{10}} \quad \text{time step}\,, \tag{7}$$

$$\boldsymbol{r}_\mathrm{r}(\mathrm{m}) = \boldsymbol{r}_\mathrm{MD}10^{-10} \quad \text{distance}\,, \tag{8}$$

$$\boldsymbol{v}_\mathrm{r}(\mathrm{m/s}) = \boldsymbol{v}_\mathrm{MD}\frac{1}{\sqrt{mc}} \quad \text{velocity}\,, \tag{9}$$

$$\boldsymbol{F}_\mathrm{r}(\mathrm{N}) = \boldsymbol{F}_\mathrm{MD}\frac{10^{10}}{c} \quad \text{force}\,, \tag{10}$$

$$\boldsymbol{a}_\mathrm{r}(\mathrm{m/s^2}) = \boldsymbol{a}_\mathrm{MD}\frac{10^{10}}{mc} \quad \text{acceleration}\,, \tag{11}$$

$$E_\mathrm{r}(\mathrm{J}) = E_\mathrm{MD}\frac{1}{c} \quad \text{energy}\,, \tag{12}$$

$$T_\mathrm{r}(\mathrm{K}) = T_\mathrm{MD}\frac{1}{ck_\mathrm{B}} \quad \text{temperature}\,, \tag{13}$$

where $k_\mathrm{B} = 1.38066 \times 10^{-23}\,\mathrm{J/K}$ is the Boltzmann constant. In this new system, the temperature is expressed in eV and thus corresponds to an energy. This makes sense, because the temperature and the average kinetic energy $\langle E_\mathrm{k}\rangle$ of the particles are related by

$$T(\mathrm{K}) = \frac{2}{3k_\mathrm{B}}\langle E_\mathrm{k}\rangle\,, \tag{14}$$

in the SI system.

Using this change of units, for a monatomic system, the mass m of the relevant atom does not come into (5). Note, however, that this is no longer true when there is more than one element to be taken into account. In this case, one of the masses is taken as a reference and a mass correction factor is required in the equations. Moreover, all calculated quantities typically have values in the range from 10^{-3} to 10^{+3}. If very large or very small values are obtained during a calculation, then there is likely to be an error somewhere.

1.2 Integrating Newton's Equation

We shall present two of the most widely used algorithms for integrating Newton's equation. It should be borne in mind that $\boldsymbol{r}$, $\boldsymbol{v}$, $\boldsymbol{a}$, and $\boldsymbol{F}$ are 3-component vectors.

Verlet Algorithm

This is the simplest algorithm. The position at time t is approximated by writing the following truncated expansions:

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t \frac{\partial \boldsymbol{r}}{\partial t} + \frac{1}{2}\delta t^2 \frac{\partial^2 \boldsymbol{r}}{\partial t^2} + \frac{1}{2\times 3}\delta t^3 \frac{\partial^3 \boldsymbol{r}}{\partial t^3} + O(\delta t^4)\,, \tag{15}$$

$$\boldsymbol{r}(t-\delta t) = \boldsymbol{r}(t) - \delta t \frac{\partial \boldsymbol{r}}{\partial t} + \frac{1}{2}\delta t^2 \frac{\partial^2 \boldsymbol{r}}{\partial t^2} - \frac{1}{2\times 3}\delta t^3 \frac{\partial^3 \boldsymbol{r}}{\partial t^3} + O(\delta t^4)\,, \tag{16}$$

whereupon

$$\boldsymbol{r}(t+\delta t) + \boldsymbol{r}(t-\delta t) = 2\boldsymbol{r}(t) + \delta t^2 \frac{\partial^2 \boldsymbol{r}}{\partial t^2} + O(\delta t^4)\,. \tag{17}$$

This can also be written

$$\boldsymbol{r}(t+\delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\delta t) + \delta t^2 \frac{\partial^2 \boldsymbol{r}}{\partial t^2} + O(\delta t^4)\,, \tag{18}$$

where

$$\frac{\partial^2 \boldsymbol{r}}{\partial t^2} = \boldsymbol{a}(t) = \frac{\boldsymbol{F}(t)}{m}\,. \tag{19}$$

The approximation is order δt^4. The estimate of the particle positions at time t does not depend on their velocities. However, the particle velocities still have to be calculated in order to find the kinetic energy and hence the temperature. The velocity is obtained by subtracting (16) from (15):

$$\boldsymbol{r}(t+\delta t) - \boldsymbol{r}(t-\delta t) = 2\delta t \frac{\partial \boldsymbol{r}}{\partial t} + O(\delta t^3)\,, \tag{20}$$

where

$$\frac{\partial \boldsymbol{r}}{\partial t} = \boldsymbol{v}(t)\,. \tag{21}$$

The approximation to the velocity is $O(\delta t^2)$. In order to calculate the velocity at time t, the position must already have been calculated at time $t + \delta t$.

Gear Algorithm

The basic idea here is to use the values of the position and its derivatives at time t to predict their value at time $t + \delta t$. With the new values of the particle positions at time $t+\delta t$, the forces and hence the accelerations can be recalculated at time $t + \delta t$. As there is a difference between the values of the predicted and recalculated accelerations, the particle accelerations are then corrected. However, this correction also affects the position and its derivatives according to a scheme devised by Gear.

The algorithm is based on a truncated expansion of the position:

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t)+\delta t\frac{\partial \boldsymbol{r}}{\partial t}+\frac{1}{2}\delta t^2\frac{\partial^2 \boldsymbol{r}}{\partial t^2}+\frac{1}{2\times 3}\delta t^3\frac{\partial^3 \boldsymbol{r}}{\partial t^3}+\frac{1}{2\times 3\times 4}\delta t^4\frac{\partial^4 \boldsymbol{r}}{\partial t^4}+\cdots\,. \tag{22}$$

The expansion has been cut off after the fourth order term as an illustration of the method, but it can be allowed to include more terms. We introduce the notation

$$\boldsymbol{\chi}_0(t) = \boldsymbol{r}(t)\,,\quad \boldsymbol{\chi}_1(t) = \delta t\frac{\partial \boldsymbol{r}}{\partial t} = \delta t\boldsymbol{v}(t)\,,\quad \boldsymbol{\chi}_2(t) = \frac{1}{2}\delta t^2\frac{\partial^2 \boldsymbol{r}}{\partial t^2} = \frac{1}{2}\delta t^2\boldsymbol{a}(t)\,, \tag{23}$$

$$\boldsymbol{\chi}_3(t) = \frac{1}{2\times 3}\delta t^3\frac{\partial^3 \boldsymbol{r}}{\partial t^3}\,,\quad \boldsymbol{\chi}_4(t) = \frac{1}{2\times 3\times 4}\delta t^4\frac{\partial^4 \boldsymbol{r}}{\partial t^4}\,. \tag{24}$$

Predicted values are attributed a superscript p. For $\boldsymbol{\chi}_0(t + \delta t)$ to $\boldsymbol{\chi}_4(t + \delta t)$, they are given by

$$\boldsymbol{\chi}_0^{\mathrm{p}}(t + \delta t) = \boldsymbol{\chi}_0(t) + \boldsymbol{\chi}_1(t) + \boldsymbol{\chi}_2(t) + \boldsymbol{\chi}_3(t) + \boldsymbol{\chi}_4(t)\,, \tag{25}$$

$$\boldsymbol{\chi}_1^{\mathrm{p}}(t + \delta t) = \boldsymbol{\chi}_1(t) + 2\boldsymbol{\chi}_2(t) + 3\boldsymbol{\chi}_3(t) + 4\boldsymbol{\chi}_4(t)\,, \tag{26}$$

$$\boldsymbol{\chi}_2^{\mathrm{p}}(t + \delta t) = \boldsymbol{\chi}_2(t) + 3\boldsymbol{\chi}_3(t) + 4\boldsymbol{\chi}_4(t)\,, \tag{27}$$

$$\boldsymbol{\chi}_3^{\mathrm{p}}(t + \delta t) = \boldsymbol{\chi}_3(t) + 4\boldsymbol{\chi}_4(t)\,, \tag{28}$$

$$\boldsymbol{\chi}_4^{\mathrm{p}}(t + \delta t) = \boldsymbol{\chi}_4(t)\,. \tag{29}$$

The predicted values of the positions $\boldsymbol{\chi}_0^{\mathrm{p}}(t + \delta t) = \boldsymbol{r}^{\mathrm{p}}(t + \delta t)$ can be used to recalculate the forces applied to each atom, and hence the corrected value of the acceleration from

$$\boldsymbol{\chi}_2^{\mathrm{c}}(t + \delta t) = \frac{1}{2}\delta t^2\boldsymbol{a}^{\mathrm{c}}(t + \delta t)\,.$$

All corrected values carry the superscript c. Setting

$$\Delta\boldsymbol{\chi}_2 = \boldsymbol{\chi}_2^{\mathrm{c}}(t+\delta t) - \boldsymbol{\chi}_2^{\mathrm{p}}(t+\delta t)\,,$$

the corrected values of the position and its derivatives are calculated as follows:

$$\boldsymbol{\chi}_n^{\mathrm{c}}(t+\delta t) = \boldsymbol{\chi}_n^{\mathrm{p}}(t) + C_n \Delta\boldsymbol{\chi}_2\,, \tag{30}$$

where

$$C_0 = \frac{19}{120}\,, \quad C_1 = \frac{3}{4}\,, \quad C_2 = 1\,, \quad C_3 = \frac{1}{12}\,, \quad C_4 = \frac{1}{12}\,.$$

Several iterations of corrections are sometimes required. However, in molecular dynamics, since determination of the forces involves a lot of computation time, only one correction stage is carried out. The coefficients C_n proposed by Gear depend on the number of derivatives taken into account and are determined so as to obtain a stable and accurate calculation scheme.

Choice of Integration Algorithm

Several algorithms are described in the literature. The choice is based on a compromise between the following criteria: stability, accuracy, simplicity, computation time, and required memory.

The discretisation error in the Gear algorithm drops off more rapidly than in the Verlet algorithm when the time is reduced. However, the Gear algorithm is more complicated to implement and requires more computation time and memory. The Verlet algorithm is time symmetric. In other words, if the time step becomes negative (or if all velocities are reversed at some time t), then all particles will follow their trajectories in the opposite directions, up to rounding errors. This is not so for the Gear algorithm. This can produce a tendency for the total energy to deviate as time goes by, the discrepancy depending mainly on the truncation order.

1.3 Interaction Potentials

It is essential to calculate the total potential energy (1) of a system in order to carry out a molecular dynamics simulation, because it is a prerequisite for obtaining the forces exerted on each atom. This energy can be written

$$E_{\mathrm{p}} = \sum_i E_1^{\mathrm{e}}(\boldsymbol{r}_i) + \frac{1}{2}\sum_i \sum_{j\neq i} E_2^{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j) + \frac{1}{3}\sum_i \sum_{j\neq i} \sum_{k\neq j} E_3^{ijk}(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k) + \cdots\,. \tag{31}$$

The first term corresponds to the potential energy due to an external field. The second term is the potential energy due to pairwise interactions. Further terms are in some sense corrections to the pairwise interactions, since the potential energy of a pair of atoms or molecules depends on the simultaneous presence of other atoms or molecules. The potential energy due to the third term can represent more than 10 % of the total potential energy, as happens in silicon or germanium. The determination or choice of interaction potential lies at the heart of molecular dynamics. Its complexity and accuracy determine:

- The dynamics of the atoms or molecules. The potential must lead to correct simulation of observed physical phenomena. For example, it has been shown experimentally for different materials, e.g., gold, silicon, that the organisation of surface atoms is a rearrangement of the bulk crystalline configuration. These surface phenomena are generally the most difficult to model. Quantitatively, calculated properties must correspond to experimentally measured effects, up to the error inherent in the numerical and experimental results.
- The computation time. This is mainly taken up by evaluation of interaction forces, i.e., calculation of the potential energy and its derivatives. The more terms are included in the potential, the more time is required for calculations. For this reason, two-body and three-body potentials are most commonly used. In the literature, effective two-body potentials have been put forward to cater for multiple interactions without unacceptable increase in computation time.

The potential energy of external origin (gravity, electromagnetic field) poses no problem because the description of these potential fields is well known. The difficulty arises when calculating the potential energy due to interaction between atoms, which depends partly on short range forces. These are mainly repulsive nuclear forces leading to an exponential decay of the potential energy for increasing separations between the particles. Interactions between atoms also depend on long range forces of electrostatic origin which depend on the local electron density of the system. This is the most difficult term to evaluate because the electron density has complex space dependence. Electronic configurations can be determined from ab initio calculations, whose results can only be used in molecular dynamics if the associated interaction potential is described analytically in a relatively simple way. To a first approximation, the electrostatic potential energy is a sum of three terms:

- Attractive or repulsive electrostatic forces due to the pointlike spatial distribution of the charges. For example, these may be the electric charges on each atom in an ionic crystal, which have the same positions as the atoms. The potential energy due to these forces is additive, i.e., the potential energy of each pair is independent.
- Attractive induction forces due to distortion of the electrostatic field by insertion of a new atom. The potential energy of each atom due to these

forces is not additive, i.e., for each atom, its value depends on the position of all the other atoms.
– Dispersion forces caused by instantaneous fluctuations in the electrostatic field due to electron motions. These are attractive and dominate in the case of metallic materials, for example. As for the induction forces, the potential energy due to dispersion forces is not additive.

A general form for the potential energy of a particle i in the absence of an external potential can be found in the literature:

$$E_{\mathrm{p}}^{i} = \sum_{j \neq i} E_2(\boldsymbol{r}_i, \boldsymbol{r}_j) = \frac{1}{2} \sum_{j \neq i} \phi(r_{ij}) + F\Big(\sum_{j \neq i} f(r_{ij}) \Big) , \tag{32}$$

where r_{ij} is the distance from particle i to particle j. The function ϕ is a pairwise potential which mainly accounts for short range forces and electrostatic forces. The function F represents the effects of induction and dispersion forces. However, the functions F and ϕ can be considered as mathematical objects with no particular physical significance and whose only role is to provide a correct representation of the interactions. When F is zero, E_2 is effectively a pairwise potential. If F is not identically zero, then E_2 is considered to be a pairwise functional.

Whatever potential is considered, it will involve several parameters in a function defined from physical considerations. The parameters can be determined in two ways:

– The first approach is to identify the parameters by minimising the discrepancy between the actual properties of the material and the values of these properties calculated either from the results of molecular dynamics simulations, or from analytical expansions derived from the potential and leading to an expression for the required properties [5–10].
– The second approach is to identify the parameters by minimising the discrepancy between the values of the atomic positions and the forces on these atoms obtained by ab initio calculation and the values obtained by the molecular dynamics calculation [11]. Another solution is to determine the parameters of the potential by minimising the differences between the properties calculated by molecular dynamics (thermophysical constants, dispersion curves, configuration energies) and those obtained by ab initio calculations [12].

Precise knowledge of these parameters is essential since their values will largely determine simulation results.

1.4 Implementing the Simulation

A simulation begins by defining the initial state characterised by the type of particles and potential, the particle positions, the boundary conditions, the

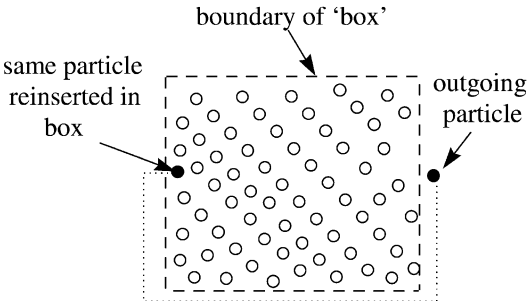

Fig. 1. Periodic boundary condition

velocity field of the particles, and the initial temperature. To simplify the discussion, we consider only an ensemble of atoms of the same element with zero average velocity (fixed centre of gravity) and without rotational motion around the centre of gravity. Only one interatomic potential is required and the choice of potential is directly related to the type of atoms and the physical phenomena under investigation.

In most cases, the initial position of an atom corresponds to the theoretical position in the crystal. For example, gold crystallises in a face-centered cubic configuration, iron in a body-centered cubic configuration, and cobalt in a hexagonal configuration. If the aim is to simulate the behaviour of small crystals, the particles are then positioned in such a way as to constitute the desired monocrystal and the surfaces are left free. If the aim is to study the bulk properties of a material, then a large enough body of particles must be defined to ensure that the presence of free surfaces no longer influences the bulk behaviour of the material. This condition requires the definition of a very large number of particles, and this leads to exorbitant calculational effort. In order to reduce the number of particles, a practical solution is to define periodic boundary conditions. A cubic box is filled with particles. As time goes by, when a particle leaves the box through one face, it automatically reenters it through the opposite face (see Fig. 1). This operation is achieved for each pair of opposite faces of the box. This device is used to simulate materials with infinite dimensions.

In this box, a particle undergoes interactions with all particles included in a sphere of radius equal to the maximal distance of application of the potential field, called the cutoff value (see Fig. 2). For the periodic boundary condition to be consistent, the box must have dimensions such that the sphere defined in this way does not include the same particle twice. This condition is necessary, but not always sufficient to obtain a realistic simulation. The dimension of the box also depends on the physical phenomenon under investigation.

An initial velocity must be attributed to each particle. This velocity is defined by a magnitude and a direction. The initial velocity field serves to

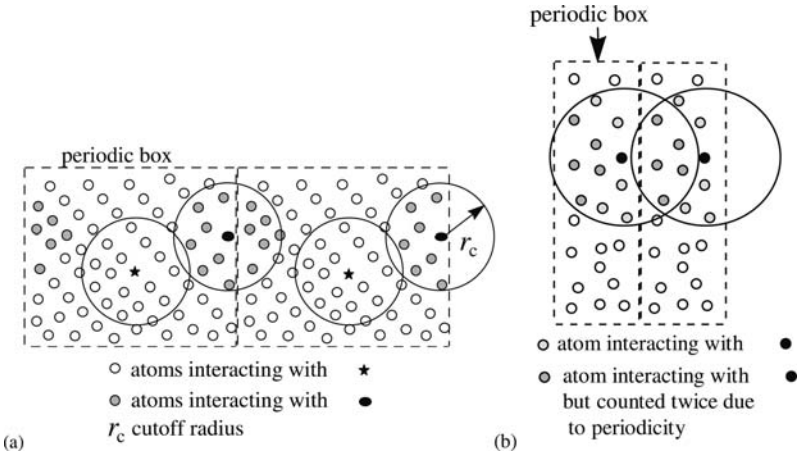

Fig. 2. Determination of interacting atoms under periodic boundary conditions. (**a**) The dimension of the box is greater than the cutoff radius so that interacting atoms are only counted once. (**b**) The dimension of the box is smaller than the cutoff radius in one direction. Some atoms are counted twice

represent the thermal agitation of the particles. The velocity field must therefore satisfy the following criteria:

- The directions of the vectors must be uniformly distributed in space.
- The sum of the velocity vectors must be equal to 0, i.e., no overall motion of the atoms:

$$\sum_{i=1}^{N} v_x^i = 0\,, \quad \sum_{i=1}^{N} v_y^i = 0\,, \quad \sum_{i=1}^{N} v_z^i = 0\,.$$

- The magnitudes of the velocity vectors must obey the Maxwell–Boltzmann distribution corresponding to the temperature T (in K) of the particle ensemble. The distribution function of the magnitude has the form

$$f(v) = cv^2 \exp\left(-\frac{mv^2}{2k_\mathrm{B}T}\right), \tag{33}$$

  where c is a normalising constant such that $\int_0^\infty f(v)\,\mathrm{d}v = 1$.
- The average kinetic energy must satisfy

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}mv_i^2 = \frac{D}{2}k_\mathrm{B}T\,, \tag{34}$$

  where D is the number of dimensions of the problem.

In practice, the initial velocity field satisfies only the first two criteria. The magnitude of all the velocity vectors is the same and equal to

$$v_i = \sqrt{\frac{Dk_{\mathrm{B}}T}{m}}\,, \quad \forall\, i = 1, \ldots, N\,. \tag{35}$$

Indeed, experience shows that, allowing the system to evolve freely from this initial velocity field, the distribution function of the velocity magnitude will evolve naturally after a few tens or hundreds of time steps to the Maxwell–Boltzmann distribution [13]. To calculate the initial velocity field, one must therefore fix the temperature.

1.5 Energy Distribution

Recall that the motion of the atoms as time goes by can be decomposed with respect to a basis of periodic progressive wave functions called vibrational modes, characterised by their angular velocity ω and speed of propagation v. The relation between ω and v is

$$v = \frac{\mathrm{d}\omega}{\mathrm{d}k}\,, \tag{36}$$

where k is a translation vector under which the system is invariant. The various vectors k are defined in the reciprocal space, in the first Brillouin zone of the atomic lattice under consideration. For a wave vector, the angular velocity is given by the dispersion relation. The energy of a vibrational mode is equal to the energy quantum of a phonon at the relevant frequency multiplied by the average number of phonons at this frequency [14, 15]:

$$E_k(T) = \hbar\omega_k \langle n(\omega_k)\rangle\,, \tag{37}$$

where $\langle n(\omega_k)\rangle$ is given by the Planck distribution function

$$\langle n(\omega_k)\rangle = \frac{1}{\exp(\hbar\omega/k_{\mathrm{B}}T) - 1}\,. \tag{38}$$

In molecular dynamics, the behaviour of atoms and molecules is described by classical mechanics. The main consequence of this description concerns the distribution of energy over the various vibrational modes of the system. Indeed, all vibrational modes have the same energy $k_{\mathrm{B}}T$. In molecular dynamics, the average number of phonons is therefore given by

$$\langle n_{\mathrm{MD}}(\omega)\rangle = \frac{k_{\mathrm{B}}T}{\hbar\omega}\,. \tag{39}$$

This is the limiting value of the number of phonons given by the Planck distribution when the temperature tends to infinity.

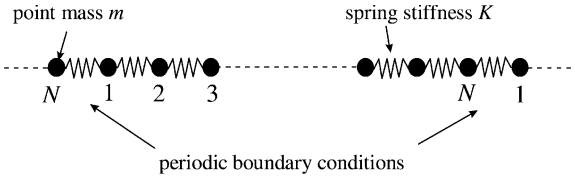

Fig. 3. One-dimensional harmonic system

To sum up, for a system at given temperature T, molecular dynamics overestimates the energy of the system and the phonon population. This important point is illustrated by a chain of harmonic oscillators comprising N point masses m connected by springs of stiffness K, only able to move in the x direction and constrained by periodic boundary conditions (see Fig. 3). At rest, the masses are a distance a apart and the potential energy of the system is zero, i.e., zero force between the masses. The vibrational modes of this oscillator chain are progressive plane waves characterised by the dispersion relation

$$\omega = 2\sqrt{\frac{K}{m}}\left|\sin\frac{ka}{2}\right| , \tag{40}$$

where

$$k = \frac{2\pi n}{aN} , \quad n = 1, \ldots, N . \tag{41}$$

The maximal angular velocity is $\omega_{\mathrm{M}} = 2\sqrt{K/m}$. Figure 4 shows the temperature variation of the dimensionless total energy of the system obtained by dividing the Planck distribution by the energy of the system as calculated for a classical description. The energy of the quantum system tends to that of the classical system for infinite temperatures. Figure 5 compares the densities of states of the classical and quantum systems for two temperature values. At the low temperature ($k_{\mathrm{B}}T/\hbar\omega_{\mathrm{M}} = 0.32$), the classical system has higher density of states than the quantum system. At the high temperature ($k_{\mathrm{B}}T/\hbar\omega_{\mathrm{M}} = 80$), the classical system has identical density of states to the quantum system.

2 Thermal Conductivity Calculation

Three techniques are used to calculate the thermal conductivity from molecular dynamics simulations:

- equilibrium molecular dynamics (EMD),
- non-homogeneous non-equilibrium molecular dynamics (NHNEMD),
- homogeneous non-equilibrium molecular dynamics (HNEMD).

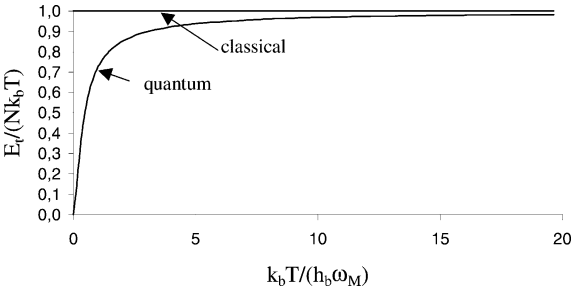

Fig. 4. Temperature variation of the dimensionless total energy of a quantum harmonic system found by dividing by the total energy of the same system considered with a uniform distribution per vibrational mode

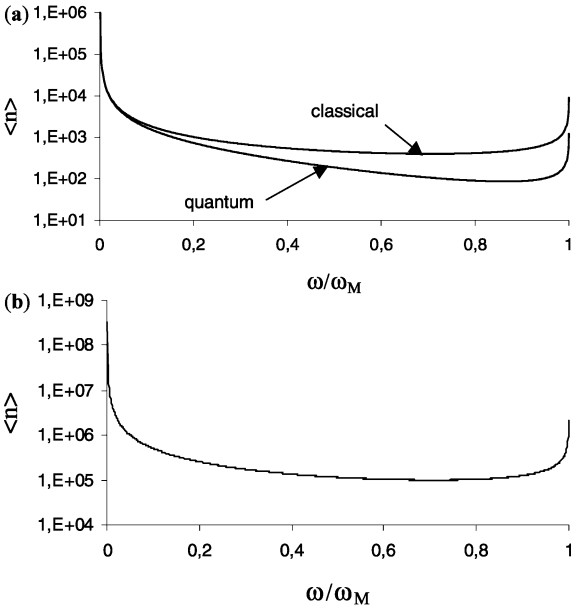

Fig. 5. Phonon number as a function of the dimensionless angular velocity. *Upper*: $k_BT/\hbar\omega_M = 0.32$. *Lower*: $k_BT/\hbar\omega_M = 80$. The average phonon numbers calculated by the classical and quantum approaches are equal in (**b**)

The temperature must be ascertained before the thermal conductivity can be calculated. Indeed, for a system in equilibrium, the thermal conductivity is calculated at a given temperature which must be known, and for a non-equilibrium system, the temperature gradient in the system must be determined.

Whatever technique is used, the temperature calculation is based upon the hypothesis of local thermodynamic equilibrium and equipartition of en-

ergy over all vibrational modes and polarisations. For non-equilibrium systems, the hypothesis of local thermodynamic equilibrium is based upon the assumption of a small difference between the actual distribution function and the Maxwell–Boltzmann velocity distribution [16]. We have checked that, in a non-equilibrium system (solid argon at 25 K) with a temperature gradient of 3×10^8 K/m, the difference between the distribution functions does not exceed 2 %. In non-equilibrium molecular dynamics, due to the small system sizes (a few times the interatomic distance), temperature gradients are much greater than actual temperature gradients. However, the temperature levels always remain reasonable compared with the temperature variations arising in a phase change. The instantaneous temperature of an ensemble of N atoms is given in terms of the average kinetic energy of these atoms by

$$\frac{3}{2}k_{\mathrm{B}}T(t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}m_i v_i^2 \,. \tag{42}$$

The standard deviation of the temperature is proportional to the temperature and inversely proportional to the square root of the number of atoms [17]:

$$\sigma\big(T(t)\big) \propto \frac{T(t)}{\sqrt{N}} \,. \tag{43}$$

Hence, if the number of atoms is large enough (of the order of 10^{23} for a macroscopic system), statistical variations in the temperature are negligible. In molecular dynamics, the number of atoms considered varies from ten to tens of thousands. Then statistical variations in the instantaneous temperature can no longer be ignored. A time average must be calculated to reduce the standard deviation in the temperature:

$$\big\langle T(t)\big\rangle = \frac{1}{N_t}\sum_{i=1}^{N_t} T(t_i) \,, \tag{44}$$

$$\sigma\Big(\big\langle T(t)\big\rangle\Big) \propto \frac{T(t)}{N_t\sqrt{N}} \,. \tag{45}$$

Figure 6 shows the autocorrelation function for instantaneous temperatures calculated at each time step during a molecular dynamics calculation. From the figure it is clear that temperatures remain highly correlated over a time span equal to a few hundred time steps. The time average is then evaluated by using only one instantaneous temperature every 800 time steps, for example. It seems that the definition of the correlation time is still open to debate. Hence, *Lukes* et al. [18] define the autocorrelation time as equal to the average phonon relaxation time.

2.1 Equilibrium Molecular Dynamics

The system considered here is microcanonical, with constant pressure, volume, and energy. Since the system evolves without constraint, after a certain

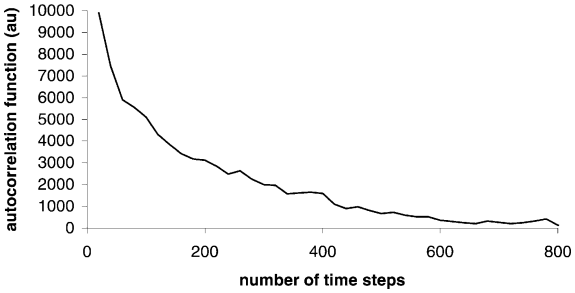

Fig. 6. Autocorrelation function of the local temperature in a system. Results obtained for a carbon nanotube C(5.5) of length 10.4 nm under a temperature gradient of 60 K at an average temperature of 900 K. Brenner interaction potential. Time step 0.0005 ps

number of time steps, it reaches a thermodynamic equilibrium with a constant average temperature. However, due to temperature fluctuations, the instantaneous flux also varies, even though its average value is still zero because the system is at equilibrium. The conductivity calculation uses the fluctuation–dissipation theorem from the linear response theory, which relates the transport properties to instantaneous fluctuations in the system. The thermal conductivity can thus be deduced from variations in the instantaneous flux of an equilibrium system using the Green–Kubo relation [1, 2, 19, 20]

$$\lambda = \frac{V}{3k_{\mathrm{B}}T^2} \int_0^\infty \langle \phi(0)\phi(t) \rangle \, \mathrm{d}t \, . \tag{46}$$

In order to determine the thermal conductivity, one must therefore calculate the instantaneous flux in the system [21]:

$$\boldsymbol{\phi}(t) = \sum_{i=1}^{N} \boldsymbol{v}_i E_i - \frac{1}{2} \sum_{i,j=1}^{N} \boldsymbol{r}_{ij} \boldsymbol{v}_i \boldsymbol{F}_{ij} \, . \tag{47}$$

The first term on the right-hand side represents the energy transported by each atom moving with velocity $\boldsymbol{v}_i$. In solids, this term is generally negligible [22].

One drawback with the method lies in the fact that it is difficult to obtain a sufficiently accurate estimate of the integral in (46). This requires simulations with a relatively large number of time steps, which may become prohibitive for the computation time when sophisticated interaction potentials are used.

2.2 Non-Homogeneous Non-Equilibrium Molecular Dynamics

This is certainly the simplest technique to understand and implement, for it is equivalent to the well known guarded hot plate experiment. The idea is

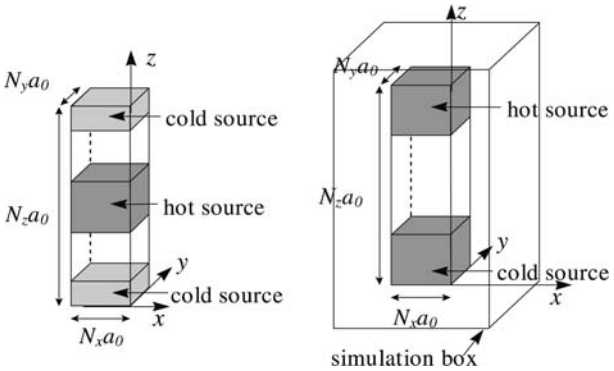

Fig. 7. Configuration for simulating heat transfer. *Left*: Periodic boundary conditions. The simulation enclosure and the system have the same size. *Right*: Free surface conditions. The simulation enclosure is bigger than the system

to simulate one-dimensional heat transfer in a system by inserting a hot and a cold source, then measuring the flux exchanged between the sources and the temperature gradient. The definition of the hot and cold sources depends on the type of boundary conditions used, i.e., periodic or free surface (see Fig. 7). To simulate the heat transfer, one of the techniques proposed in the literature [23] consists in exchanging atoms between the hot and cold sources. Since this technique is rarely used, we shall not go into detail. The most widely used approach consists in modifying the velocity field of the atoms belonging to the heat sources in such a way as to impose either the thermal power exchanged between the hot and cold sources [18, 24, 25], or the temperature of these sources [26–30].

Imposed Power

As the flux ϕ is imposed, the amount of energy to give to the hot source and take from the cold source in each time step is equal to

$$\Phi = \phi S \Delta t \,. \tag{48}$$

This quantity of energy (or its absolute value) is supplied by modifying the kinetic energy of the hot and cold sources. The velocity field is thus multiplied by a weighting coefficient equal to

$$\alpha_{\mathrm{c}}(t) = \sqrt{1 - \Phi / E_{\mathrm{c}}^{\mathrm{k}}(t)} \tag{49}$$

for the cold source and

$$\alpha_{\mathrm{h}}(t) = \sqrt{1 - \Phi / E_{\mathrm{h}}^{\mathrm{k}}(t)} \tag{50}$$

for the hot source, where $E_{\mathrm{c}}^{\mathrm{k}}(t)$ and $E_{\mathrm{h}}^{\mathrm{k}}(t)$ are the instantaneous kinetic energies of the cold and hot sources, respectively.

The exchanged power is perfectly constant. The permanent regime is reached when the average temperature profile becomes stable in time. When the system does not conserve energy, the temperature begins to drift. The average temperature of the system is generally close to its initial temperature.

Imposed Temperatures

The temperatures $T_{\mathrm{h}}^{*}(t)$ and $T_{\mathrm{c}}^{*}(t)$ of the hot and cold sources are calculated after integrating Newton's equations. These temperatures are not the required temperatures T_{h} and T_{c}. The velocity field of the hot and cold sources is thus multiplied by a weighting coefficient to make the correction. The weighting coefficients are

$$\sqrt{\frac{T_{\mathrm{h}}(t)}{T_{\mathrm{h}}^{*}}} \quad \text{and} \quad \sqrt{\frac{T_{\mathrm{c}}(t)}{T_{\mathrm{c}}^{*}}}$$

for the hot and cold sources, respectively. With this method, the energies $\Phi_{\mathrm{h}}(t)$ and $\Phi_{\mathrm{c}}(t)$ released (in absolute value) by the hot and cold sources, respectively, are given by

$$\Phi_{\mathrm{h}}(t) = N_{\mathrm{h}} \left[T_{\mathrm{h}} - T_{\mathrm{h}}^{*}(t)\right] \frac{3k_{\mathrm{B}}}{2} \tag{51}$$

and

$$\Phi_{\mathrm{c}}(t) = N_{\mathrm{c}} \left[T_{\mathrm{c}} - T_{\mathrm{c}}^{*}(t)\right] \frac{3k_{\mathrm{B}}}{2}\,, \tag{52}$$

where N_{h} and N_{c} are the numbers of atoms in the hot and cold sources.

With this method, the temperature gradient between the sources can be perfectly controlled. The instantaneous fluxes exchanged by the heat sources are calculated from the energies $\Phi_{\mathrm{h}}(t)$ and $\Phi_{\mathrm{c}}(t)$ by

$$\phi(t) = \frac{\Phi(t)}{S\Delta t}\,, \tag{53}$$

where S is the cross-sectional area through which the heat flows and Δt is the time step.

In the permanent regime, the average fluxes $\langle \Phi_{\mathrm{h}}(t) \rangle$ and $\langle \Phi_{\mathrm{c}}(t) \rangle$ must be equal in absolute value. This is therefore a simple way to check the principle of energy conservation. To illustrate the method, the temperature profile in a system is shown in Fig. 8. The system is an fcc crystal of solid argon with dimensions $60a_0$ in each of the x, y and z directions. Periodic boundary conditions are imposed in the three directions. The thickness of the hot and cold sources is $12a_0$. Although the average temperatures of the hot and cold

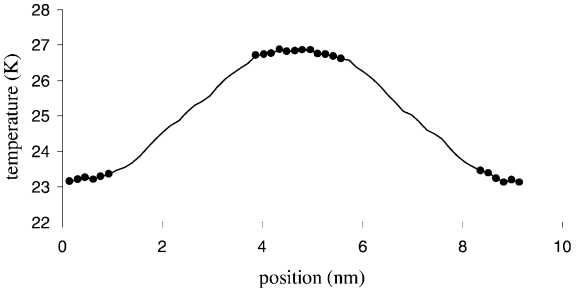

Fig. 8. Temperature profile in a system with thermostatically controlled hot and cold sources. *Black dots* indicate the positions of the hot and cold sources

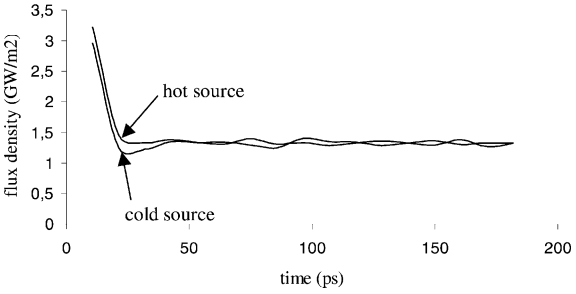

Fig. 9. Evolution of average fluxes exchanged between hot and cold sources

sources are held constant during the simulation, the temperature profile is not uniform in these regions. Rather, it has a parabolic appearance specific to macroscopic media with internal heat sources. Figure 9 shows the evolution of the average fluxes (over 5000 time steps) exchanged by the heat sources. The permanent regime is reached after about 50 ps.

2.3 Homogeneous Non-Equilibrium Molecular Dynamics

This calculational technique was devised by *Evans* [31, 32]. It consists in applying a homogeneous external force field to the system to produce a heat flux. Simultaneously, a Gaussian thermostat [33, 34] is applied to hold the temperature constant. If F_{ext} is the external force, the equation of motion is then

$$m\boldsymbol{a}_i = \frac{1}{2N}\left[\sum_j \boldsymbol{F}_{ij}(E_i - E)\boldsymbol{F}_{\text{ext}} - \sum_j \boldsymbol{F}_{ij}\boldsymbol{r}_{ij}\cdot\boldsymbol{F}_{\text{ext}} + \sum_j \boldsymbol{F}_{jk}\boldsymbol{r}_{jk}\cdot\boldsymbol{F}_{\text{ext}}\right] - \alpha m\boldsymbol{v}_i\,, \quad (54)$$

and α is the coefficient of the Gaussian thermostat given by

$$\alpha = \left(\sum_i m_i^2 \boldsymbol{v}_i \cdot \boldsymbol{v}_i\right) \sum_i \frac{m_i}{2N} \left[\sum_j \boldsymbol{F}_{ij} + (E_i - E)\boldsymbol{F}_{\text{ext}} - \sum_j \boldsymbol{F}_{ij}\boldsymbol{r}_{ij} \cdot \boldsymbol{F}_{\text{ext}} + \sum_j \boldsymbol{F}_{jk}\boldsymbol{r}_{jk} \cdot \boldsymbol{F}_{\text{ext}}\right] \cdot \boldsymbol{v}_i \,. \quad (55)$$

Under these conditions, using perturbation theory, Evans showed that the thermal conductivity can be written

$$\lambda = \lim_{F_{\text{ext}} \to 0} \frac{\langle \phi(t) \rangle}{T F_{\text{ext}}} \,. \quad (56)$$

In order to determine the thermal conductivity, one thus requires the average flux $\langle \phi(t) \rangle$ [the instantaneous flux calculated with (5)] for different values of the force F_{ext}. The value of the latter must be small enough to constitute a small perturbation from equilibrium, i.e., to justify linear perturbation theory.

3 Determining Vibrational Properties

In dielectric materials, heat transfer by conduction is due entirely to vibrations of the atomic lattice. In crystals with a periodic structure, these vibrations can be decomposed with respect to a basis of progressive waves characterised by a wave vector k and a polarisation p. In Sect. 3.1, an expression is established for the thermal conductivity in terms of the vibrational modes, while Sect. 3.2 explains how the properties of these vibrational modes can be determined via molecular dynamics simulations.

3.1 Heat Transfer by Phonons

An analytical formulation of thermal conductivity can be given by decomposing lattice vibrations into vibrational modes [14, 15]. In a system of volume V and temperature T, the number of phonons with wave vector k and polarisation p is given by the Planck distribution function (38), in which the angular speed ω depends on k and p. The relation $\omega(k, p)$ is given by the dispersion curves. The internal energy of the vibrational mode (k, p) is the product of the number of phonons in this vibrational mode and the energy quantum of each phonon:

$$U(k, p) = \hbar\omega \langle n(k, p) \rangle \,. \quad (57)$$

The specific heat of the vibrational mode (k, p) is then given by

$$C(k, p) = \frac{\partial U(k, p)}{\partial T} = k_{\text{B}} x^2 \frac{\mathrm{e}^x}{V(\mathrm{e}^x - 1)^2} \,, \quad (58)$$

with

$$x = \frac{\hbar\omega}{k_{\mathrm{B}}T}\,. \tag{59}$$

Now treating the phonons as particles and using the kinetic theory of gases, the thermal conductivity of phonons associated with a vibrational mode (k,p) in a direction x can be written

$$\lambda_x(k,p) = C(k,p)v^2(k,p)\tau(k,p)\cos^2\big(\theta_x(k)\big)\,, \tag{60}$$

where v is the phonon propagation speed, equal to the group velocity of the vibrational mode (k,p), viz.,

$$v(k,p) = \frac{\mathrm{d}\omega}{\mathrm{d}k}\,, \tag{61}$$

$\tau(k,p)$ is the relaxation time, and $\theta_x(k)$ is the angle between the heat transfer direction and the wave vector k. The total thermal conductivity is thus equal to the sum of the thermal conductivities associated with each vibrational mode (k,p), i.e.,

$$\lambda_x = \sum_k \sum_p \lambda_x(k,p) = \sum_k \sum_p C(k,p)v^2(k,p)\tau(k,p)\cos^2\big(\theta_x(k)\big)\,. \tag{62}$$

The relaxation time depends on all the interactions undergone by the phonons. These fall into three categories:

- Interactions between phonons and the system boundary [35–39].
- Phonon–phonon interactions involving three phonons. To each type of collision, there correspond relaxation times τ_{N} and τ_{U}, for which various relations have been proposed [35, 40–44].
- Phonon–defect interactions. Work by *Klemens* [45, 46] forms the basis for most studies concerning the influence of defects on thermal conductivity [47–50]. The cases investigated correspond to very low defect concentrations (less than 2 %), for which the true vibrational behaviour can be considered as a superposition of the vibrational behaviour of the perfect material and a perturbation.

As mentioned in the last section, the total relaxation time is given by the relation

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}\,. \tag{63}$$

The thermal conductivity can be calculated analytically using (62) provided that the following data are known:

– The different vibrational modes. For a crystal, this is a basic problem of solid state physics. Although trivial for a 1D system, the problem can become quite involved for crystals depending on the geometry of the Brillouin zone.
– The dispersion curves from which the angular speed ω and the group velocity of the vibrational modes can be calculated. The difficulty here lies in modelling them, because the number of dispersion curves is equal to the number of possible directions of the wave vectors.
– The relaxation times due to the various interactions.

3.2 Determining Vibrational Properties

We shall illustrate the determination of vibrational properties via molecular dynamics simulation by considering the example of a Lennard–Jones crystal at constant temperature ($T = 0.215$ LJ units). The system is a face-centered cubic crystal with characteristic dimension Na_0 whose directions ([100], [010], [001]) are aligned with the x, y, and z axes. Periodic boundary conditions are used in all three directions. Phonons are considered as plane waves. For this demonstration, only those vibrational modes with wave vectors in the z direction and with longitudinal polarisation are characterised. There are N atomic planes (indexed by $s = 1, \ldots, N$) perpendicular to the z direction. At each time step, the displacement $u(s,t)$ ($s = 1, \ldots, N$) of these planes in the z direction is calculated relative to their equilibrium position. The Fourier transform of these instantaneous displacements can be used to calculate the instantaneous amplitude $A(k,t)$ for each wave vector. The time average of this vibration amplitude is related to the angular frequency by [51]

$$\left\langle |A_k|^2 \right\rangle = \frac{k_\mathrm{B} T}{m\omega^2} \,. \tag{64}$$

The dispersion curve can then be determined. It can be shown that this curve does not depend on the system size, although the bigger the system, the better the resolution with which this curve can be determined (see Fig. 10). Moreover, we have also checked that the dispersion curve does not depend on the temperature, provided that it remains low. Indeed, the higher the temperature, the greater the enhancement of anharmonic effects. Under the simulation conditions, the dispersion curve has the same shape as would be predicted for a 1D harmonic system, viz., $\omega(k) = \omega_\mathrm{M} \sin(ka_0/2)$, where ω_M depends on the parameters of the potential. This relation can be used to determine the group velocity $v(k) = \mathrm{d}\omega/\,\mathrm{d}k$ (see Fig. 11).

The relaxation time is determined by analysing the autocorrelation of the real part of the amplitude of the vibrational modes, viz., $\left\langle A_k(t)\bar{A}_k(0)\right\rangle$. The amplitude of this function decreases exponentially with a characteristic duration equal to twice the relaxation time of the phonon number (see Fig. 12). The relaxation time can thus be identified for different wave vectors (see

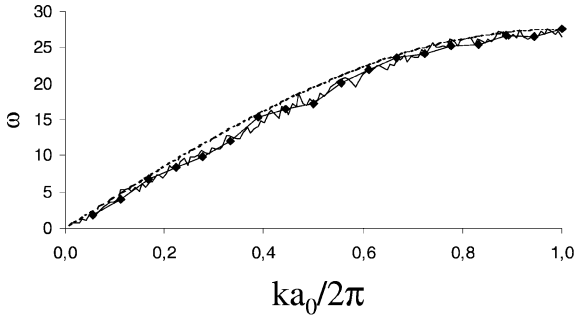

Fig. 10. Dispersion curve for wave vectors in the direction [100] of an fcc Lennard–Jones crystal for two dimensions of the system: $L = 18a_0$ (*continuous curve without symbols*) and $L = 142a_0$ (*curve with symbols*). The *dotted curve* corresponds to the dispersion curve of a 1D harmonic system

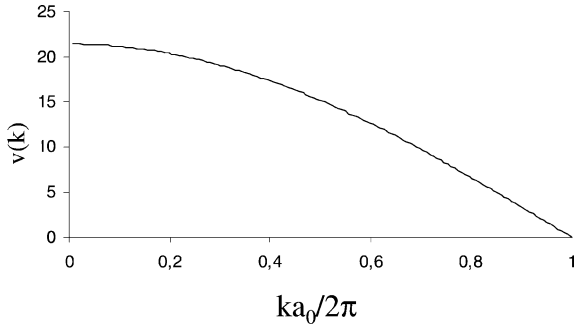

Fig. 11. Group velocity in Lennard–Jones units as a function of the wave vector

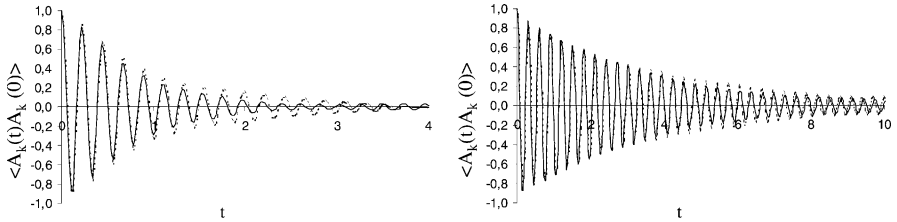

Fig. 12. Autocorrelation $\left\langle A_k(t)\bar{A}_k(0)\right\rangle$ of the vibration amplitude as a function of time, in arbitrary units. *Left*: $k = 2\pi/a$. *Right*: $k = 2\pi/10a$

Fig. 13). The $1/\omega^2$ dependence generally predicted has been recovered to a good level of accuracy.

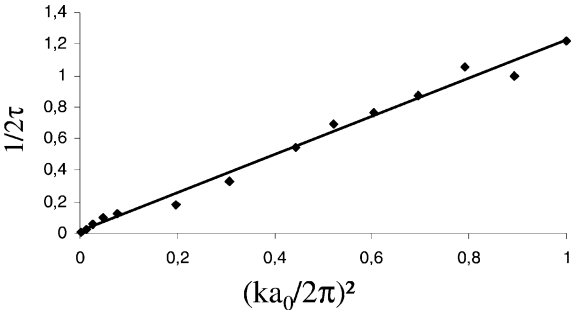

Fig. 13. Reciprocal relaxation time (Lennard–Jones units) as a function of the square of the dimensionless wave vector

Conclusion

We have presented the Molecular Dynamics computational technique and described the different simulation strategies (EMD, NEMD, HNEMD) allowing for the calculation of the thermal conductivity. As MD only involves the classical treatment of the atom dynamics, then only phonon heat transfer in dielectric media at high enough temperatures can be investigated. Due to the computational time, the use of MD is limited to systems which characteristic length is lower than a few nanometers. To estimate the thermal conductivity of large dielectric systems at low temperatures, other method such as the Monte Carlo technique, the P1 and the Discrete Ordinate method which were described in the previous chapters, might be used to solve the Boltzmann equation for phonons. These latter methods require the knowledge of the vibrational properties of the matter, namely, the dispersion curves and the phonon relaxation times. This information can be determined on small systems by using MD and a relevant interatomic potential.

References

[1] M. P. Allen, D. J. Tildesley: *Computer Simulation of Liquids* (Oxford University Press, New York 1997) 155, 169
[2] D. Frenkel, S. Berend: *Understanding Molecular Simulation* (Academic Press, San Diego 1996) 155, 169
[3] D. C. Rapaport: *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge 1995) 155
[4] R. J. Sadus: *Molecular Simulation of Fluids* (Elsevier Science, Amsterdam 1999) 155
[5] D. J. Oh, R. A. Johnson: J. Mater. Res. **3**, 471–478 (1988) 162
[6] F. Ercolessi, M. Parrinello, E. Tosatti: Phil. Mag. A **58**, 213–226 (1988) 162
[7] F. H. Stillinger, T. A. Weber: Phys. Rev. B **31**, 5262–5271 (1985) 162
[8] R. Biswas, D. R. Hamann: Phys. Rev. Lett. **55**, 2001–2004 (1985) 162

[9] D. W. Brenner: Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films, Phys. Rev. B **42**, 9458–9471 (1990) 162
[10] J. Tersoff: Phys. Rev. B **37**, 6991–7000 (1988) 162
[11] F. Ercolessi, J. B. Adams: Europhys. Lett. **26**, 583–588 (1994) 162
[12] M. Gastreich, J. Gale, C. Marian: Phys. Rev. B **68**, 094110 (2003) 162
[13] P. Chantrenne, S. Volz: Techniques de l'Ingénieur, article BE 8290 (2002) 165
[14] C. Kittel: *Introduction to Solid State Physics* (Wiley, New York 1996) 165, 173
[15] N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Harcourt College Publishers, Fort Worth 1976) 165, 173
[16] S. Chapman, T. G. Cowling: *The Mathematical Theory of Non-Uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases* (Cambridge University Press, Cambridge 1970) 168
[17] B. Diu, C. Guthmann, D. Lederer, B. Roulet: *Eléments de physique statistique* (Hermann, Paris 1989) 168
[18] J. R. Lukes, D. Y. Li, X. G. Liang, C. L. Tien: J. Heat Trans. **122**, 536 (2000) 168, 170
[19] R. Zwanzig: Ann. Rev. Phys. Chem. **16**, 67 (1964) 169
[20] A. J. C. Ladd, B. Moran, W. G. Hoover: Phys. Rev. B **34**, 5058 (1996) 169
[21] R. Kubo, M. Toda, N. Hashitsume: *Statistical Physics II*, vol. 31, Springer Series in Solid State Sciences (Springer, Berlin 1991) 169
[22] S. Volz, P. Chantrenne: Techniques de l'Ingénieur, BE 8 291 p. 1 (2002) 169
[23] F. Müller-Plathe: J. Chem. Phys. **106**, 6082 (1997) 170
[24] P. Jund, R. Jullien: Phys. Rev. B. **59**, 13707–13711 (1999) 170
[25] T. Ikeshoji, B. Hafskjold: Mol. Phys. **81**, 251–261 (1994) 170
[26] S. Maruyama: Physica B **323**, 193–195 (2002) 170
[27] B. C. Daly, H. J. Maris: Physica B **316–317**, 247–249 (2002) 170
[28] J. L. Barrat, F. Chiaruttini: Molecular Physics (2003) in press URL `http://xxx.lpthe.jussieu.fr/abs/cond-mat/0209607` 170
[29] C. Olischlger, J. C. Schön: Phys. Rev. B **59**, 4125–4133 (1999) 170
[30] P. Chantrenne, J. L. Barrat: *Eurotherm Seminar No. 75 on Microscale Heat Transfer* (Reims, France 2003) 170
[31] D. J. Evans: Phys. Lett. A **91**, 457 (1982) 172
[32] D. J. Evans: Phys. Rev. A **34**, 1449 (1986) 172
[33] D. J. Evans: J. Chem. Phys. **78**, 3297 (1983) 172
[34] D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran, A. J. C. Ladd: Phys. Rev. **28**, 1016 (1983) 172
[35] I. Rosenblum, J. Adler, S. Brandon: Comput. Mater. Sci. **12**, 9–25 (1998) 174
[36] X. Lu, J. H. Chu: J. Appl. Phys. **83**, 1219–1229 (2003) 174
[37] J. Zou, A. Balandin: J. Appl. Phys. **89**, 2932–2938 (2001) 174
[38] R. Berman, F. E. Simon, J. M. Ziman: Proc. Roy. Soc. (London) A **220**, 171 (1953) 174
[39] M. G. Holland: Phys. Rev. Lett. **132**, 2461–2471 (1963) 174
[40] H. B. G. Casimir: Physica **5**, 595 (1938) 174
[41] P. G. Klemens: *Solid State Physics*, vol. 7 (Academic Press, New York 1963) 174
[42] P. G. Klemens: Proc. Roy. Soc. (London) A **208**, 108 (1951) 174
[43] C. Herring: Phys. Rev. **95**, 954 (1954) 174

[44] J. Callaway: Phys. Rev. **113**, 1046 (1959) 174
[45] P. G. Klemens: *Solid State Physics* (Academic Press, New York 1958) 174
[46] P. G. Klemens: Proc. Phys. Soc. (London) A **48**, 1113 (1955) 174
[47] M. W. Ackerman, P. G. Klemens: Phys. Rev. B Solid State, Third Series **3**, 3 (1971) 174
[48] M. W. Ackerman: Phys. Rev. B **5**, 5 (1972) 174
[49] A. C. Anderson, M. E. Malinowski: Phys. Rev. B **5**, 3199 (1972) 174
[50] D. Kotchetkov, J. Zou, A. A. Balandin, D. I. Florescu, F. H. Pollock: Appl. Phys. Lett. **79**, 4316–4318 (2001) 174
[51] P. M. Chaikin, T. C. Lubensky: *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge 1995) 175
[52] D. J. Quesnel, D. S. Rimai, L. P. DeMejo: Phys. Rev. B **48**, 6795–6807 (1993)
[53] S. Motoyama, Y. Ichikawa, Y. Hiwatari, A. Oe: Phys. Rev. B **60**, 292–298 (1999)

Index

Scanning Thermal Microscopy

Bernard Cretin[1], Séverine Gomès[2], Nathalie Trannoy[3], and Pascal Vairac[1]

[1] FEMTO-ST Institute/Department LPMO, 32 Avenue de l'Observatoire, 25044 Besançon Cedex
{bcretin,pvairac}@femto-st.fr
[2] Centre de Thermique de Lyon, UMR 5008, INSA de Lyon, Bâtiment Sadi Carnot, 9 rue de la Physique, 69621 Villeurbanne Cedex
severine.gomes@insa-lyon.fr
[3] UTAP/ Laboratoire d'Energétique et d'Optique, EA 3802, Université de Reims, Moulain de la Housse, BP 1039, 51687 Reims Cedex
nathalie.trannoy@univ-reims.fr

Abstract. Fundamental research and continued miniaturisation of systems (materials or components) have instigated and still require today the development of specific investigative methods for studying phenomena or properties in many areas of science. This Chapter is concerned with a whole range of methods based on near-field microscopy and developed for the study of micro- and nanoheat transfer. Section 1 describes the working principles of various near-field microscopes as a prerequisite to understanding how they are put to use in the study of heat transfer. Section 2 discusses the type of information than can be obtained and describes the main developments that have given access to such information. Section 3 describes a certain type of local probe microscopy that plays an important part in the science of microheat transfer.

1 Introduction to Near-Field Microscopy

This section[1] discusses recently developed ways of investigating the structure and properties of materials, although it does not claim to be exhaustive. We shall see how the transition was made from the scanning tunneling microscope (STM) to scanning force microscopes (SFM) and microscopes using waves or diffusion effects, in particular, the scanning thermal microscope which is the subject of this Chapter.

1.1 Basic Principles of Near-Field Microscopy

The new high-resolution forms of microscopy known as near-field microscopy all exploit the short-range interaction between a fine tip and the sample under investigation. Depending on the type of investigation, the interaction may be static, e.g., van der Waals force, or it may be dynamical, in which case a wave or dynamical diffusion effect can be used. Resolution can be submicron or even nanometric. In the best cases, atoms can be localised (AFM, STM),

[1] Contributed by B. Cretin and P. Vairac.

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 181–238 (2007)

but as far as the emerging techniques are concerned, resolution is usually in the range 10–100 nm.

To understand the difference between the far field and the near field, we shall consider here only microscopes using propagation or diffusion. In standard microscopes (optical, acoustic, or electronic), the beam is focused on the sample object and detection occurs at distances well above the relevant wavelength. The ultimate resolution of conventional microscopes is imposed by a physical limit, viz., the impossibility of obtaining a spot in the focal plane of the lens with dimensions smaller than half the wavelength. This limit, known as the diffraction limit, follows from the laws of diffraction and is often expressed in the form of the Rayleigh criterion. A simplistic and approximate evaluation of the resolution can be obtained from the Heisenberg uncertainty principle, which relates the spatial localisation to the spectral width of the wave used to investigate the sample:

$$\Delta x \Delta k_x \geq 2\pi \,. \tag{1}$$

This relation shows that, if we use the mathematical spectrum (including negative values),

$$\Delta k_{x\,\max} = \frac{4\pi}{\lambda_{\min}} \,. \tag{2}$$

Then the limiting resolution given by the uncertainty principle in the case where we have equality is

$$\Delta x_{\min} = \frac{\lambda_{\min}}{2} \,. \tag{3}$$

This value does indeed agree with the limit given by the Rayleigh criterion if the propagating medium has a refractive index close to that of the vacuum.

In contrast, near-field microscopes are able to go beyond this physical limit by several orders of magnitude because the interaction they exploit is rather different.

Specific Features of the Near Field in the Case of Propagation or Diffusion

Recall that the one-dimensional propagation equation has the general form

$$\frac{\partial^2 X}{\partial t^2} = v^2 \frac{\partial^2 X}{\partial x^2} \,, \tag{4}$$

a partial differential equation in which X represents the field of the propagating physical quantity, x is the space coordinate, and v is the phase velocity of the wave. As an example, longitudinal elastic waves in a non-absorbent solid medium are solutions to the equation (plane model)

$$\frac{\partial^2 u}{\partial t^2} = \frac{E}{\rho} \frac{\partial^2 u}{\partial x^2} \,, \tag{5}$$

whereupon the speed of the longitudinal wave is

$$v_{\mathrm{L}} = \sqrt{\frac{E}{\rho}}\,. \tag{6}$$

In these formulas, u is the displacement field, E is Young's modulus, and ρ is the density of the medium. In the permanent sinusoidal regime, the solution takes the form

$$u = u_0 \mathrm{e}^{\mathrm{i}(\omega t - kx)}\,, \tag{7}$$

for propagation towards positive x, where u_0 is the amplitude of the elastic wave, ω is the angular frequency, and k is the wave vector parallel to the propagation axis.

Microscopes based on propagation are exemplified by the optical microscope and the acoustic microscope. Considering the wave–particle duality, the scanning electron microscope (SEM) can be classified in the same category.

The diffusion equation has a rather different form, since it only involves a first derivative with respect to time. In 1D, this equation is

$$\frac{\partial X}{\partial t} = a\frac{\partial^2 X}{\partial x^2}\,, \tag{8}$$

where a is the diffusivity. A representative example here is the heat equation

$$\frac{\partial \theta}{\partial t} = a\frac{\partial^2 \theta}{\partial x^2}\,, \quad a = \frac{\kappa}{\rho C}\,, \tag{9}$$

where θ is the temperature, κ is the thermal conductivity, and C is the specific heat capacity. In this case, the solution for a sinusoidal excitation is

$$\theta = \theta_0 \mathrm{e}^{\mathrm{i}\omega t}\mathrm{e}^{-\mathrm{i}x/\mu_{\mathrm{T}}}\mathrm{e}^{-x/\mu_{\mathrm{T}}}\,, \tag{10}$$

where θ_0 is the amplitude of the temperature field, ω the angular frequency, and μ_{T} the thermal penetration depth, sometimes called the skin depth.

As an indication, we give some orders of magnitude for thermally conducting and insulating materials. The values in Table 1 correspond to periodic excitations of period T. The results explain why a probe or other analytical device based on heat diffusion cannot reach deep layers in reasonable observation times. The microscopes using this diffusion are mainly the thermal microscopes. We shall see later that near-field microscopes using evanescent fields are also subject to a diffusion equation.

When the source is very small, the 1D model can no longer be used and the physical behaviour is fundamentally different. Indeed, the main effect leading to attenuation is energy dispersion. Hence, for a point source, a propagating

Table 1. Order of magnitude of the thermal diffusion length

Period T	μ_{T} for brick ($a \sim 10^{-6}\,\mathrm{m}^2/\mathrm{s}$)	μ_{T} for aluminium ($a \sim 8 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$)
1 ms	18 µm	0.16 mm
1 s	0.57 mm	5 mm
1 yr (31.5×10^6 s)	3.2 m	28 m

field can be expressed in the form (approximate form if r is small compared with the wavelength)

$$X_{\mathrm{P}} = \frac{X_0}{r}\mathrm{e}^{\mathrm{i}(\omega t - kr)} \approx \frac{X_0}{r}\mathrm{e}^{\mathrm{i}\omega t}(1 - \mathrm{i}kr)\,, \tag{11}$$

and a diffusive field can be put into the similar form

$$X_{\mathrm{D}} = \frac{X_0}{r}\mathrm{e}^{\mathrm{i}(\omega t - r/\mu_X)}\mathrm{e}^{-r/\mu_X} \approx \frac{X_0}{r}\mathrm{e}^{\mathrm{i}\omega t}\left(1 - \mathrm{i}\frac{r}{\mu_X} - \frac{r}{\mu_X}\right)\,. \tag{12}$$

In practical terms, the $1/r$ term dominates in the diffusive case and it is therefore difficult to distinguish the two types of behaviour when the interaction between the field and the defect occurs at a distance much smaller than $\lambda/2\pi$ (propagation) or μ_X (diffusion).

This phenomenological approach gives an insight into the near field. In practice, the models used are more elaborate, since they must take into account size effects which do not show up in the $1/r$ behaviour of the field.

1.2 Historical Perspective: From Conventional Microscopy to Near-Field Microscopy

The conventional optical microscope was the starting point for the long series of microscopes which followed, or sometimes preceded, the requirements of an industry that sought to miniaturise the products it manufactured. The origin of the microscope comprising two distinct optical systems is attributed to the Dutch spectacle-maker Janssen at the end of the sixteenth century [1, 2]. This instrument was gradually improved and led to many discoveries, e.g., the discovery of micro-organisms by Pasteur. However, by the end of the nineteenth century, it had become accepted, particularly after the work of Lord Rayleigh, that the resolution limit of this type of microscope was something like half the wavelength [3–5]. Optical microscopes have since been perfected and now approach this ultimate resolution, sometimes allowing a 3D reconstruction of certain transparent objects (confocal microscope [6]). Despite this achievement, other physical principles had to be considered in order to gain access to higher resolutions, or to observe within opaque objects.

The second great step in the history of microscopy had its origin in the work of Louis de Broglie, who studied the wave nature of the electron. In

1924, he proposed that the motion of any particle was accompanied by the simultaneous propagation of a wave of wavelength [7–9]

$$\lambda = \frac{h}{mv}\,, \tag{13}$$

where h is Planck's constant, m the relativistic mass of the particle, and v its speed. This discovery made it possible to envisage resolutions that had never been contemplated before. By 1930, the analogy between geometrical optics and 'electronic optics' was already established. The first electron microscope images were obtained by Davisson and Calbick in 1931. Given the very short equivalent wavelength (about 1 Å for an acceleration potential of 100 V), this type of microscope soon became widespread. In practice, the scanning electron microscope achieves resolutions in the range 1–10 nm, but its main advantage is its depth of field. (The depth of field of a microscope increases as the wavelength gets shorter.) On the other hand, there are two drawbacks with this technique: it can only be used under very high vacuum and non-conducting samples must be metallised.

Still by analogy with the optical microscope, the scanning acoustic microscope was developed in 1974. It was based on older work due to Langevin, who invented the sonar in 1917, and Sokolov, who suggested using ultrasonic acoustic waves to detect defects in metals in 1929. The first microscope built by Quate works by transmission [10–12]. The acoustic wave generated by a piezoelectric material is focused by a lens and passes through the object. The attenuation of this wave depends on local inhomogeneities. After scanning, the acoustic wave detected, converted and visualised on a monitor images the acoustic properties of the object. The main advantage of this microscope is its ability to image the interior of opaque objects, e.g., metals, ceramics, or composite materials. One drawback with this microscope arises due to the attenuation of the acoustic wave at high frequencies, which precludes improvement of the resolution. The reason why this microscopy took so long to develop is that the computers required for this level of data acquisition, storage and image restitution were not yet available.

As far as thermal microscopy is concerned, it is difficult to attribute its invention to anyone in particular owing to the wide range of different techniques [13–20]. However, it was progress in optical physics (lasers, modulators) and optoelectronics (infrared detectors) which led to its development in the 1960s and its application to non-destructive testing a decade later. A suitable means of detection combined with a scanning technique was then able to provide a thermal image that was complementary to the images produced by other forms of microscopy. The thermal microscope may be thought of as occupying a position midway between microscopes using waves and near-field microscopes, because the dynamical behaviour of heat transfer corresponds to an evanescent field. The historical development of the scanning thermal microscope is described in detail in Sect. 2.

Many other types of microscopy have been developed, using X or gamma rays, charged particles, or infrared or ultraviolet electromagnetic waves.

A decisive step was taken at the beginning of the 1980s with the introduction of a novel concept: given that the laws of diffraction forbid the observation of details smaller than the wavelength, a tiny probe must be used to detect field perturbations near inhomogeneities and hence improve resolution. The first microscope based on this idea concerning the near field, the electron scanning tunneling microscope, was built by *Binnig* and *Rohrer* in 1982 [21–23]. It was a revolutionary invention that won them the Nobel Prize for Physics. The scanning tunneling microscope detects the tunnel current between a fine metal tip and a conducting material [24]. The lateral resolution of this microscope can reach 0.2 nm and allows observation of the electronic fields around atoms. Individual atoms can be discerned on the images.

The second major family of near-field microscope uses atomic interaction forces. The invention of the atomic force microscope (AFM) is attributed to *Quate*, *Binnig* and *Gerber* in 1986 [25]. Forces are measured by means of a fine tip rigidly connected to a tiny cantilever. The latter has a certain elasticity and bends under stress, so that an optical measurement can then be made. This microscope has the advantage of being able to examine materials that are poor electrical conductors, in contrast to the scanning tunneling microscope. Atomic force microscopes are now widely used to characterise surfaces on the nanoscale (mainly the state of the surface). The high sensitivity of AFM cantilevers makes new force measurements possible (magnetic forces, electrostatic forces, and so on) with a vast range of applications.

At the same time as the AFM was being developed, the optical near field was being used to devise new types of microscope, sensitive to light waves in this case. The first optical near-field microscopes appeared simultaneously in the US and Germany (Massey, Fischer 1983). These microscopes used a hole with diameter smaller than the wavelength. Today, optical near-field microscopes use either a tapered optical fibre, often metallised, or a cantilever including a wave guide and a tip. Several types of near-field optical microscope have already been industrialised. The servo system controlling the tip–sample separation generally uses a vibrational motion of the tip parallel to the sample surface, the so-called dynamic mode AFM. More recently, measurements of tip heating have led to the development of hybrid microscopes combining near-field optics and heat transfer.

The acoustic microscope takes its place among the latest near-field microscopes. The invention of the truncated tip acoustic microscope is due to *Zienuk* and *Latuszeck* in 1986 [26]. However, a resolution better than 1 μm was only achieved in 1989 by *Takata* et al. [27]. Today, commercially produced microscopes are used to carry out topographic studies or combine AFM and acoustic effects. Acoustic techniques using the tip in contact mode are particularly useful because they can identify the local mechanical properties of materials, as does the nanoindenter. Moreover, defects located below the surface can be detected down to a depth that usually exceeds a few microns.

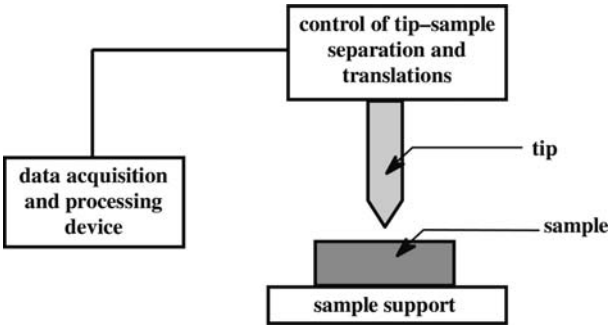

Fig. 1. General layout of a scanning near-field microscope

Over the last few years, new techniques have emerged which now provide a genuine arsenal for applied research. Applications and observation instruments are likely to continue to proliferate over the coming years, going hand in hand with the miniaturisation of many industrial products (microelectronics, data storage for multimedia or computing, local sensors, etc.). The scanning thermal microscope (SThM) is naturally one of the techniques currently under development with a promising future in many areas.

1.3 Scanning Probe Microscopes

We shall now describe the operating principles of the main families of scanning probe microscope, excepting the thermal microscopes. The general scheme of a near-field microscope is shown in Fig. 1.

In this type of microscope, whatever physical interactions are used, data is acquired sequentially, by scanning. At each scan, the relative positions of the sample and tip are controlled by piezoceramics allowing three orthogonal translations. One adjusts the tip–sample distance. It is associated with an electronic circuit which servo controls the height or interaction intensity. Practically speaking, piezoelectric translators generally have a small amplitude motion. Other units must therefore be incorporated to ensure the rough approach of the tip. Moreover, special precautions must be taken to minimise unwanted variations in the tip–sample separation due to external mechanical vibrations or thermal drift. Mechanical antivibration devices use mass–spring systems or damping materials such as rubber. Piezoceramics allow continuous displacements down to the picometer scale but nonlinearities and hysteresis in these materials limit the performance of the microscope unless an accurate (e.g., capacitive) position sensor is integrated into the system.

It should be stressed that scanning microscopes cannot obtain images in real time, in contrast to the kind of full field techniques often available in the far field. In practice, in near-field microscopy, it often takes around a minute to acquire an image. This is a serious limiting factor for observing motion. Scanning rates are usually limited by the relevant detection physics,

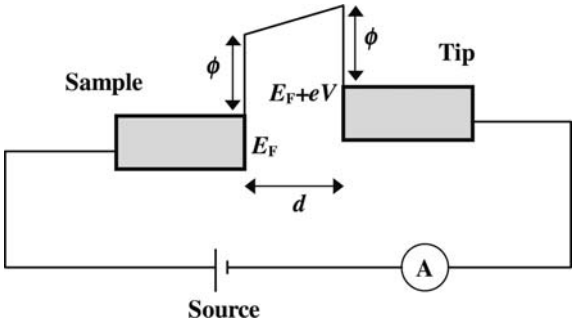

Fig. 2. Schematic view of a tunnel junction

e.g., servo control of the distance, and the acquisition frequency per point is generally less than 10 kHz. In concrete terms, the limiting scan rate can be evaluated once the temporal response of the system is known. Typically, for a measurement time constant τ (sensor and, where necessary, signal processing), and for a pixel size T_p, the maximal displacement speed V_{max} will be of the order of $2T_p/\tau$ (acquisition time corresponding to at least two time constants to reduce measurement error due to integration).

Scanning Tunneling Microscopy (STM)

Although it was first demonstrated theoretically in 1928, the tunnel effect was not put to use until 1982 when it gave birth to the first laboratory microscope, the scanning tunneling microscope (STM) which could form images of atoms. In 1986, its inventors, Binnig and Rohrer of IBM, Zurich, were attributed the Nobel Prize for Physics. The exceptional resolution of the STM inspired the development of a whole range of techniques based on an analogous principle and grouped together under the name of scanning probe microscopy. One of the basic characteristics of the STM is its ability to image conducting surfaces and, in some cases, insulating systems adsorbed onto a conducting surface. (It does not normally work for insulating surfaces.)

The tunnel effect was already well known to physicists at the beginning of the twentieth century. It was investigated more completely with the advent of junction semiconductors and in particular the use of diodes. As a first illustration of the phenomenon, consider a metal–insulator–metal structure. The metal electrodes are assumed to be flat. When one electrode is raised to a potential with respect to the other, a tunnel current can flow if the distance between the electrodes is small enough. Figure 2 illustrates a simple 1D model of a tunnel junction.

A potential barrier that is square at zero bias voltage separates the two metals characterised by their Fermi levels. (If a voltage V is applied across

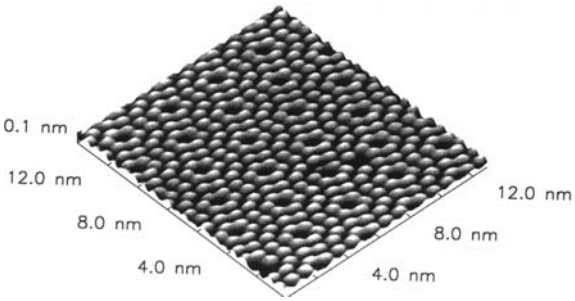

Fig. 3. Example of an STM image with atomic resolution: Si(111)7×7. Courtesy of F. Palmino, MINADO Group, CREST, FEMTO-ST

the electrodes, the Fermi level shifts by an energy eV.) For low bias potentials compared with the barrier height, the tunnel current is given by [24]

$$I \propto \exp\left(-\frac{2}{\hbar}\sqrt{2m\phi}d\right) , \tag{14}$$

where m is the electron mass, ϕ the difference between the kinetic energy and the height of the potential barrier, and d the distance between the tip and sample.

For typical orders of magnitude (energy $\sim$ eV, $d \sim 10\,\text{Å}$), the tunnel current is very small and highly sensitive to the separation between the electrodes, i.e., it can change by an order of magnitude for a variation of a few angstrom units in the separation. This relation also shows that the slightest irregularity on the electrode surface will localise the tunnel current as a result of the smaller separation at this point. It also explains why, for a homogeneous material, holding the voltage and tunnel current constant is equivalent to scanning the surface at a constant height. An example image is shown in Fig. 3. After a specific preparation, the crystal lattice of the silicon surface appears very clearly, demonstrating the ability of the STM to image electronic orbits of atoms.

Atomic Force Microscopy (AFM)

The need for conducting samples seriously limits the field of application of the scanning tunneling microscope. To get round this drawback, Binnig, Quate, and Gerber invented the atomic force microscope, based on the detection of atomic scale forces. The AFM images a surface by means of short range atomic interactions between the tip and sample surface, in particular, the van der Waals attraction. More generally, scanning microscopes that use forces (SFM) can exploit different physical interactions, e.g., electrostatic or magnetic interactions. The advantages of this technique lie mainly in the flexibility of the cantilever which limits damage to the probe tip during contact

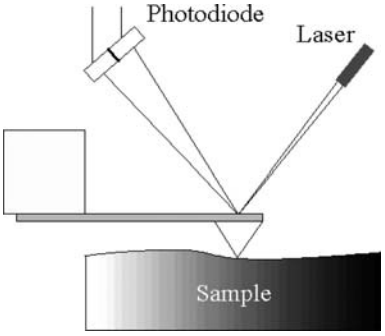

Fig. 4. Schematic view of the atomic force microscope

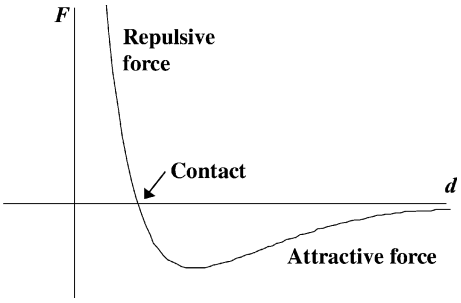

Fig. 5. Atomic force as a function of tip–sample separation d

and the possibility of servo controlling the distance or the force on arbitrary materials, STM being restricted to conductors. On the other hand, the resolution is less good than with an STM because the force–distance dependence is not exponential as in an STM.

The operating principle of the AFM is illustrated in Fig. 4. It consists in bringing a very fine tip close to the surface. The tip is located at the end of a cantilever with low stiffness. The resultant force exerted between the tip and surface causes a deflection of the cantilever which can be measured with an optical deflectometer (laser and photodiodes). The image of the forces (essentially topographic) is obtained by scanning the tip above the surface.

The three main operating modes of the AFM are:

- contact mode,
- non-contact mode,
- tapping® mode.

For the STM, we saw that the dependence was exponential. For the AFM, this dependence is completely different, as can be seen from the sketched graph in Fig. 5. The contact point is usually defined as the distance origin. Short range repulsion between ionic cores give rise to the repulsive part of the forces, while van der Waals forces produce the attractive part.

Van der Waals interactions which bind molecules together can be described by a potential of the form $W = -C/d^6$, where d is the distance be-

tween molecules assumed to be uncharged. From a more macroscopic standpoint, we shall be concerned with the interaction between a sphere and an infinite plane (approximate model for the tip–sample interaction). In this case, the attractive potential can be written in the form [28]

$$W = -\frac{CR}{6d}\,, \tag{15}$$

where R is the radius of the sphere, i.e., the radius of the tip apex. The interaction force is therefore

$$F = -\frac{\partial W}{\partial d} = -\frac{CR}{6d^2}\,. \tag{16}$$

The forces arising in the AFM are typically of the order of nN.

The atomic force microscope usually operates in attractive mode. The repulsive mode corresponds to a pressure of the tip on the surface. In fact, two types of complementary force are involved:

– Contact forces given by the Hertz contact theory which was established for two spherical surfaces in contact and which can be applied to the sphere–plane contact. For an applied load P and a tip radius R, the radius of the contact zone is given by the relation

$$a = \left(\frac{3PR}{4E^*}\right)^{1/3}\,, \quad \frac{1}{E^*} = \frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2}\,, \tag{17}$$

  where E_1 and E_2 are the Young's moduli of the sample and tip, and v_1 and v_2 are their Poisson ratios, respectively.
– Adhesive forces. These tend to prevent the tip and sample from separating after contact. They are given by

$$F_{\mathrm{adh}} = n\pi R\Delta\gamma\,, \tag{18}$$

  where $\Delta\gamma$ is the interface energy (of the order of 1 J/m) and n is a constant between 1.5 and 2.

In practice, the relevant forces are a combination of van der Waals forces and surface tension, in addition to binding forces and repulsion when there is contact. More realistic expressions for contact forces are provided by the Johnson–Kendal–Roberts (JKR) model, the Derjaguin–Muller–Toporov (DMT) model, and the more general model due to Maugis, formulated in 1992 [29].

Note that the atomic force microscope is sometimes coupled with a heat microsensor to carry out a combined topography–temperature survey. The force is servo controlled by the force sensor of the AFM and a feedback loop. In this way, the interaction between the temperature sensor and the sample can be held constant. (In some cases, the heat flux can be servo

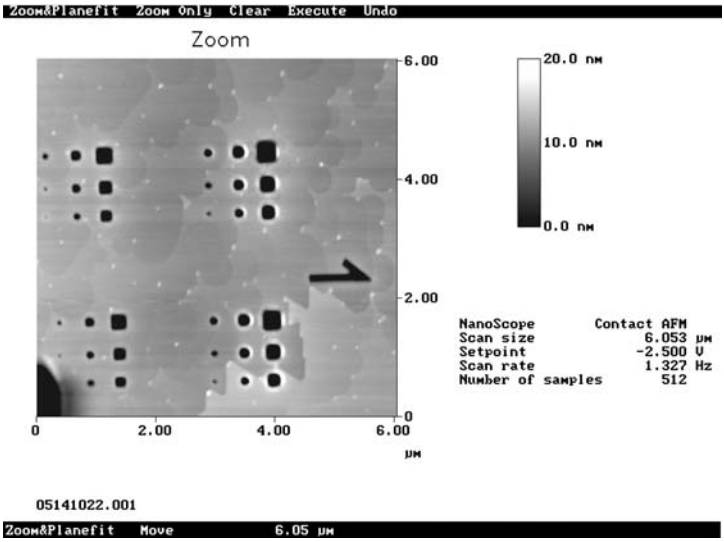

Fig. 6. Example of an AFM image. Silicon surface machined by FIB. Courtesy of F. Palmino, MINADO Group, CREST, FEMTO-ST

controlled because the measured temperature is necessarily inaccurate owing to conduction in the sensor.)

Figure 6 shows an example of an AFM image. The subject is a silicon wafer, machined by focused ion beam (FIB). The image size is $6 \times 6\,\mu\text{m}^2$. The tip goes down to a depth of 20 nm in the holes (machine depth).

Optical Near-Field Microscopy

In 1972, *Ash* and *Nichols* [30] showed that it was possible to work in the near field with microwaves. They obtained a resolution close to $\lambda/60$. The extension to visible wavelengths was achieved by Pohl et al. with a resolution of $\lambda/20$ using an optical stethoscope. These were the first images obtained by scanning near-field optical microscopy (SNOM) [31].

Near-field optical microscopes can be divided into two subfamilies [32,33]:

- Scanning near-field optical microscopy (SNOM or NSOM). The source is an aperture with dimension less than the wavelength. Detection is in the far field, either in transmission or in reflection. This was the first kind of system to be demonstrated.
- Scanning tunneling optical microscopy (STOM) or photon scanning tunneling microscopy (PSTM). In this type of microscope, depicted schematically in Fig. 7, the sample is illuminated by an evanescent light field obtained by total internal reflection. The name 'tunneling optical microscope' comes from the fact that detection uses the tunnel effect.

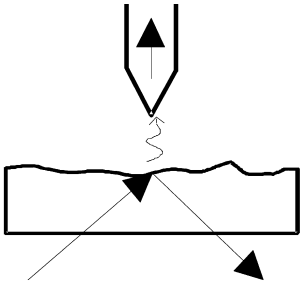

Fig. 7. Schematic illustration of STOM. The sample is illuminated by total internal reflection as indicated by the *lower arrows*

In an optical near-field microscope, the source is a low power laser (a few mW). Detection is made by a photodiode or photomultiplier (one detector suffices in all forms of scanning microscopy). The mechanical setup must have the same quality as an STM setup, i.e., antivibration system, piezoelectric displacement unit, etc.

Applications are similar to those of classical optics but with resolution on a scale of around 10 nm:

- Fine imaging, particularly in biology, and topographical measurements.
- Nanoscale photolithography. This may provide an opportunity for reducing the size of electronic circuits, for example.
- Fluorescence imaging and analysis of local spectra. This can be used to study quantum wells or to detect isolated molecules.
- Spectroscopy. The near-field technique is particularly useful in the infrared (long wavelengths). The main advantage is to be able to characterise a sample locally, even down to the detection of a single molecule.

Scanning Nearfield Acoustic Microscopy

Acoustic microscopy is physically very similar to optical microscopy, since both use waves. However, in practice, the acoustic technique is much more complex because two types of wave are involved: compression waves and shear waves. What is more, they propagate at different speeds, both much slower than the speed of light [34]. In conventional acoustics, to obtain a lateral resolution of the order of one micron, one must work at frequencies greater than GHz. However, one must then use specialised and costly electronics, not to mention the fact that the decay coefficient of acoustic waves increases quadratically with the frequency, severely limiting the depth of penetration.

The scanning acoustic microscope has attracted much interest around the world, although it remains at the laboratory prototype stage. Various acronyms have sprung up to designate vibrating tip acoustic microscopes, including AFAM, SLAM, UFM, SMM and others [35–41]. However, in practice, all devices so far developed use resonant vibrations of cantilevers loaded by the tip, whose frequency depends on the tip–sample contact. The working

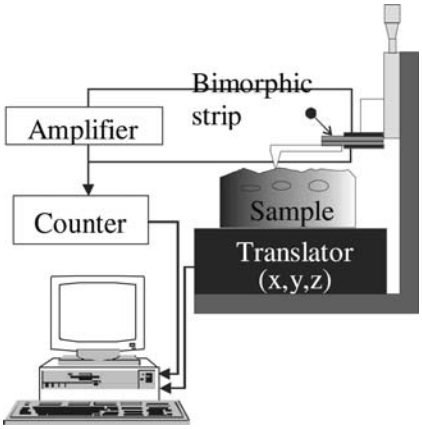

Fig. 8. Schematic illustration of the scanning microdeformation microscope (SMM), developed at LPMO, FEMTO-ST, France

principle illustrated in Fig. 8 is similar in many ways to a dynamic mode AFM whose tip is brought into contact with the sample.

Historically, the force modulation microscope (FMM) was introduced by *Maivald* et al. [42]. A periodic displacement at low frequency (a few kHz) is imposed on the sample by a piezoceramic located under the sample. The tip in contact with the sample follows the harmonic vibrations. Local variations in the elasticity of the sample can be deduced by measuring the amplitude of the cantilever displacements. However, this technique has two disadvantages:

- the stiffness of the cantilever is generally unsuitable,
- one always measures motion resulting from a combination of normal and lateral forces.

An alternative consists in using a mesoscale tip. In this case, the elastic constants of the surface can be measured and defects in the bulk just below the surface can be imaged, to a depth of the order of at most 50 times the contact radius [43].

In practice, the results are not easily analysed and the accuracy of the measurements is low when the resolution is improved. Typically, there is an accuracy of 20–50 % for contact radii of the order of 10 nm, compared with better than 5 % for contact radii of the order of 1 µm. This inaccuracy is related to poor knowledge of the tip geometry and the presence of forces that are difficult to model on the atomic scale because they depend on the surroundings, e.g., temperature, hygrometry.

2 Development of Scanning Thermal Microscopy

The aim here[2] is to describe the development of scanning probe microscopy in the field of heat transfer. This began with the development of a scanning tunneling microscope (STM) that could produce images of insulating materials. Then with the advent of the atomic force microscope (AFM), various techniques were devised, based mainly on the design of new thermal probes, with the aim of improving the resolution obtained by photothermal methods. These probes can measure temperatures or detect heat fluxes locally with high spatial resolution and good sensitivity. In this way a new family was born within the category of scanning probe microscopes (SPM), called scanning thermal microscopes. The sudden expansion of this form of microscopy is due to the fact that it responds to a need to determine thermal properties of materials with submicron resolution, and to detect subsurface defects or hot spots in microsystems.

2.1 Near-Field Microscopy and Heat Transfer

Techniques Based on STM

The scanning tunneling microscope exploits the possibility for electrons to get past a potential barrier between the surface of a sample material and the extremely fine metallic STM tip. The main drawback with the STM is that it can only be used to study electrically conducting materials. Shortly after the invention of the STM, and before the advent of the atomic force microscopes (AFM), Wickramasinghe and coworkers had the idea of using the STM in association with heat detection as a way of making images of insulating materials.

In 1986, this group proposed the scanning thermal profiler in which the standard STM tip is replaced by a thermocouple tip, i.e., a tungsten tip shaped like an STM tip is coated with an insulating layer except at the very apex, followed by a layer of nickel. The end of the probe then constitutes one of the two soldered joints of a thermocouple. The voltage across the tungsten–nickel junction is proportional to the temperature [44]. The tip thus works as a thermocouple, an electromotive force being induced by the temperature difference between two solder joints (hot and cold) constituted by the junction of two different metals. The cold solder joint is the one held at a reference temperature.

The tip is heated by a current of constant strength until its temperature reaches an equilibrium value above room temperature. When this heated probe is brought toward the sample surface, the heat transfer taking place through the air causes the tip to cool down. The resulting variations in the thermoelectric voltage in the thermocouple depend on the separation between the tip and the sample. By controlling these variations, the surface

[2] Contributed by N. Trannoy.

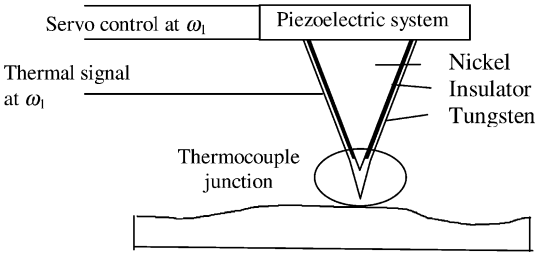

Fig. 9. Schematic diagram of the thermocouple, supported by a piezoelectric element that is used to modulate the tip–sample distance at frequency ω_1 around its average position. The thermoelectric voltage is servo controlled at this frequency to hold the distance constant

relief can be reconstructed as the tip scans across it. To reduce the effects of external perturbation, a major problem in scanning tunneling microscopy, a piezoelectric system is used to generate vertical oscillations of the tip at a frequency of about 1 kHz. This system is associated with a servo loop on the alternating thermoelectric voltage in order to hold the tip–sample distance fixed (see Fig. 9). At this frequency, the amplitude of these oscillations is less than a nanometer and hence smaller than the tip–sample gap (several tens of nanometers). Based as it is on a thermal interaction process between the tip and sample, this microscope can be used to make topographical images of surfaces with a spatial resolution estimated by its authors at 100 nm, without ever coming very near the sample surface.

The method has been modified in order to obtain images of the temperature field at the surface of active samples. The sample is heated by an alternating current at a different frequency (ω_2) to the one used for tip oscillations. During exploration of the sample surface, variations in the thermoelectric voltage are recorded at this new frequency in order to obtain the surface temperature, the tip–sample separation being servo controlled as before. In this device, images cannot be obtained simultaneously. This technique has been tested on an aluminium film deposited on a glass substrate. The surface profile of the aluminium film revealed a particle standing 30 nm high which shows up on the temperature survey. The measured variations in the thermal signal near the particle correspond to temperature variations of the order of millidegree. However, in reality it is difficult to separate thermal from topographic data, since the heat flux required to maintain the tip position depends on the thermal properties of the sample [45].

Attempts have been made to improve this device in order to acquire the topography and temperature map simultaneously. Hence, the rear face of the sample was heated by a laser beam modulated at frequency ω_2 [46]. The authors estimate that the minimal detectable temperature variation is around 1 m°, which corresponds to a flux of a few nanowatts. However, it was observed that heating the sample modifies the map of its surface relief.

This technique has also been adapted to photothermal absorption spectroscopy in order to determine the chemical composition of a material. This adaptation is based on the fact that chemical elements absorb light at characteristic wavelengths. Since an increase in temperature at a certain wavelength indicates that radiation has been absorbed, the sample temperature is measured as a function of the wavelength, using a thermocouple tip, when the sample is illuminated by a tunable laser. The absorption spectrum then reveals the chemical composition of the sample [47]. To enhance the spatial resolution of the microscope, the thermocouple is replaced by the STM tip–sample couple, in which the tip is made from either tungsten or platinum (typical materials for the STM tip) and its apex is assumed to be monoatomic. The sample surface must therefore be metallic. The thermoelectric effect induces a potential difference between the two elements of the thermocouple, and this potential difference depends on the energy absorbed by the sample atoms. By scanning the sample surface, one obtains a map of the optical absorption. This technique, known as the tunneling thermometer, has been used to measure the optical absorption spectrum of a gold film. It has also been able to acquire simultaneously the topography and a thermal map of a quartz sample covered with aluminium dots. Further details appear on the thermal image due to the difference in optical absorption of the materials. The temperature sensitivity of this thermocouple is estimated at 0.01 K. It is limited by Johnson noise in the tunnel resistance and the relatively low thermoelectric coefficient of the tip–sample junction ($3\,\mu V \cdot K^{-1}$).

Replacing the optical excitation by electrical heating, the spatial variation of the thermoelectric voltage across the tip–sample system can be measured. This technique, called scanning chemical potential microscopy (SCPM), has been used to detect the gradient of the chemical potential as a function of the surface temperature of an MoS_2 semiconductor and graphite [48].

Stopka, *Oesterschulze* et al. returned to the idea of the thermocouple tip originated by Wickramasinghe [49–52]. The thermocouple tip used is made of Au/constantan (constantan is a nickel–copper alloy) with sensitivity $43.2\,\mu V \cdot K^{-1}$. The distinguishing feature of their device is that the tip is heated rather than the sample. To this end, the tip is directly irradiated by a laser beam whilst the sample remains at room temperature. Whereas the topographic image is obtained using the standard servo system of the STM to maintain the tip at constant height (whence the tunnel current remains constant too), thermal images are obtained by recording the variations in the thermoelectric voltage by means of a second circuit (see Fig. 10). Depending on whether the laser beam is modulated or not (DC or AC modes), the image will reflect either the thermal conductivity or the thermal diffusivity of the sample. These authors call their microscope the scanning thermal microscope (SThM). The spatial resolution achieved by this method is 30 nm.

This technique is far from simple and the authors themselves recognise that the images are difficult to interpret because the contrast mechanism is

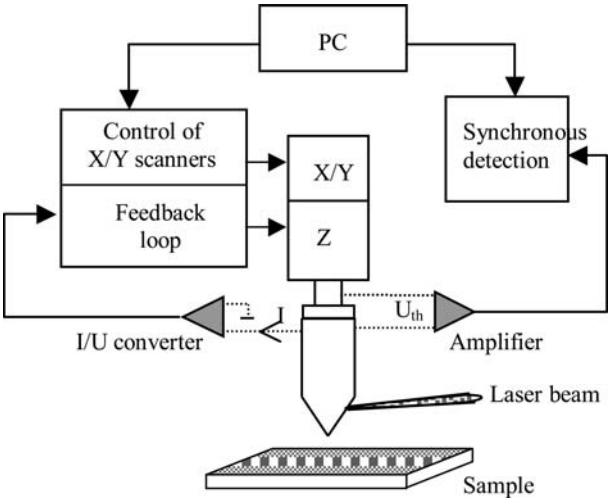

Fig. 10. Schematic view of the experimental arrangement combining the STM (*left*) with the thermocouple tip control system (*right*)

misunderstood. Indeed, this type of device has raised a certain number of questions to which answers have since been provided:

- Modulating the laser source modulates the tunnel current, and this has an effect on the imaging process.
- The thermal response depends on the position of the point on the tip at which the laser beam is focused. Phase analysis of the tunnel signal has allowed a clear separation of effects due to the tip and those due to the sample [53].
- This device operates at low frequency and in this case the thermal diffusion length is rather large, possibly even larger than the thickness of the substrate under investigation.

A study of the thermoelastic behaviour of the tip and sample has been carried out to determine the potential of the scanning thermal microscope. It has been shown that the STM can measure the thermal expansion of the sample quantitatively with a resolution of picometric order in the case where the sample is only subject to laser excitation [53, 54]. The experimental responses in both amplitude and phase correspond to the predictions of a thermoelasticity model, except when the probe is irradiated. In the latter case, the signal amplitude is much larger and, in particular, the responses are different for each tip used. The thermal coupling between tip and sample has also been studied with the aim of determining the influence of tip proximity on the thermal behaviour of the sample. As the distance between the STM tip and the sample is less than 1 nm, hence smaller than the mean free path of the molecules (66 nm), heat exchange by conduction in the ballistic regime can be

considered and described by an exchange coefficient of $40 \times 10^3\,\mathrm{W \cdot m^{-2} \cdot K^{-1}}$. Under these conditions, the influence of the tip on the temperature field of the sample turned out to be rather small for a tunnel effect microscope [55, 56].

A laser excitation is often used to obtain thermal information. Now the photon interaction with the tunnel junction involves several mechanisms that can influence the tunnel current and the tip–sample bias voltage: thermal expansion, thermoelectric voltage due to the temperature difference, optical rectification, and surface photovoltage for semiconductors. The situation is generally recognised to be rather complex. The thermal expansion of the tip has been one of the phenomena often neglected, and yet it was shown to influence the data gathered in a way depending on the geometrical characteristics of the tip [53, 55, 56]. To analyse the thermoelectric voltage phenomenon, a study was carried out in ultrahigh vacuum where only the tip was irradiated [57]. It was shown that, under these conditions, the thermoelectric voltage could be correlated with the crystal orientation of the surface under investigation.

It is thus clear that the results obtained with a scanning tunneling microscope are not easily interpreted. Images are obtained through rather specific types of coupling between the tip and sample [58, 59] and phenomena such as the presence of a water film on the sample surface, for example, have to be taken into account [53]. Moreover, the main disadvantage of the STM is that it requires an electrically conducting surface. To be able to observe integrated circuits, for instance, it would be necessary to coat them with a metallic film which might well hide some crucial data.

Techniques Based on AFM

The advent of AFM provided a way of overcoming these various difficulties. Indeed, quite apart from the fact that any type of sample could be observed, it offered the advantage of two independent systems:

- A device for obtaining the surface topography by controlling the contact force between tip and sample.
- An independent system for controlling thermal measurements, depending on the operating principle of the probe (see Fig. 11).

Even though they involve heat transfer, most of the techniques discussed above are limited to observation of surface topography. Temperature maps obtained in this way do not contain any information intrinsic to the thermal properties of the sample, because this data is mixed in with topographical data. However, since the temperature measurement by a tip–sample junction is related to the thermal conductivity of the sample, it should be possible to obtain thermal conductivity profiles.

It was with this in mind that attempts were made to improve these techniques using the AFM [60]. Indeed, the AFM is sensitive to the electrostatic force field. For an AFM tip coated with a metallic film, this field depends on

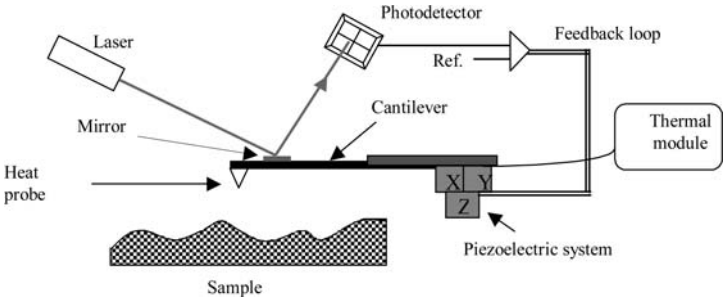

Fig. 11. Schematic view of the experimental arrangement for a scanning thermal microscope based on AFM

the contact potential between tip and sample. The contact potential between two metals depends on various parameters such as the work function or the temperature change of one of the two interacting materials. Hence, in a new setup, the heating of a metallised AFM tip or sample is used to generate a variation in the contact potential, which itself is controlled by measuring the electrostatic force field. This technique has been applied to an Si sample carrying a rectangular SiO_2 structure, the whole thing being coated with a layer of tungsten. The AFM tip was coated with a gold layer. This study succeeded in obtaining a map in which the contrast was related to thermal conductivity variations across the sample.

The methods discussed above are restricted to electrically conducting materials. In 1993, *Majumdar* et al. developed a new technique, also based on the AFM principle, which could study both conducting and non-conducting samples [61]. The AFM tip–sample interaction is used to establish the topography of the sample surface, but the tip is replaced by a thermal probe. The latter is a thermocouple comprising two wires, one made of chromel and the other of alumel, rather like a K-type thermocouple, and these two wires provide a large enough stiffness coefficient to form the cantilever required for an AFM setup. To measure the deflection of this cantilever, a small glass plate coated with a metallic sheet, e.g., aluminium, is used as a tiny mirror that reflects the laser beam (see Fig. 12).

When the probe is in contact with the sample, the temperature difference between the tip and sample generates a thermoelectric voltage in the thermocouple junction. When the sample is explored, the local temperature is measured while the tip–sample force is held constant. By holding the cold junction of the thermocouple at a reference temperature, the measured variations in the thermoelectric voltage provide a thermal image related to the temperature of the sample. This technique has been used to record simultaneously the surface topography and the temperature map of a (metal/semiconductor) field-effect transistor. The images obtained were able to locate regions heated to a few degrees above the average temperature of the transistor (70 °C). Since

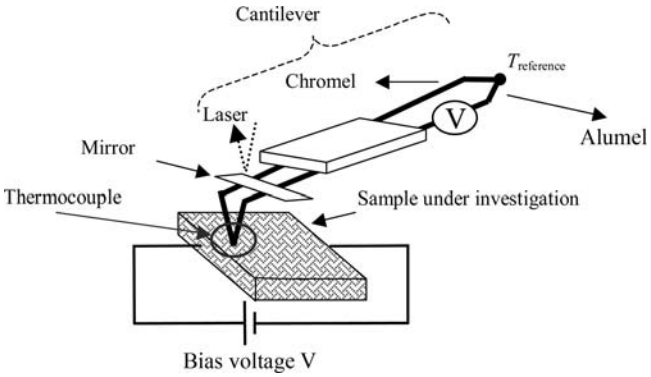

Fig. 12. Schematic view of a thermocouple tip adapted to serve as an AFM probe

the temperature sensitivity depends on the electronic noise, improvements in the detection method made it possible to measure temperature variations of just 0.1 K [62].

Another type of thermal probe that can be adapted to the AFM setup, but not based on a thermocouple, was suggested by *Pylkki* et al., namely the thermoresistor (see Fig. 13) [63, 64]. This probe comprises a Wollaston wire, i.e., a wire made from platinum–rhodium alloy, coated with silver and then electrochemically etched in such as way as to leave the platinum–rhodium wire exposed over a length of 200 µm. The wire is bent into a V shape and forms the thermoresistive element of the probe. When a current flows through it, this probe can serve both as the heat source and the detector, or simply as the detector, depending on the current strength. The idea is to measure the dependence of the electrical resistivity of the tip on the temperature difference between tip and sample. This probe can be used in two different modes:

- A passive mode where a very weak constant current flows in the wire and the probe temperature is recorded during scanning. This mode is rather sensitive to the temperature of the sample surface.
- An active mode where a stronger current serves to initiate the heating of the thermoresistor, thereby generating a loss of flux upon contact with the sample, this heat flux being influenced by the thermal conductivity of the sample.

This microscope, called the scanning thermal microscope (SThM), can operate at constant current or constant temperature, with a feedback loop holding the probe resistance constant. Surface topography and qualitative images of thermal conductivity variations can be obtained simultaneously. Since its invention, this microscope has undergone as many studies concerning the device itself as concerning its applications. Section 3 is devoted to this method.

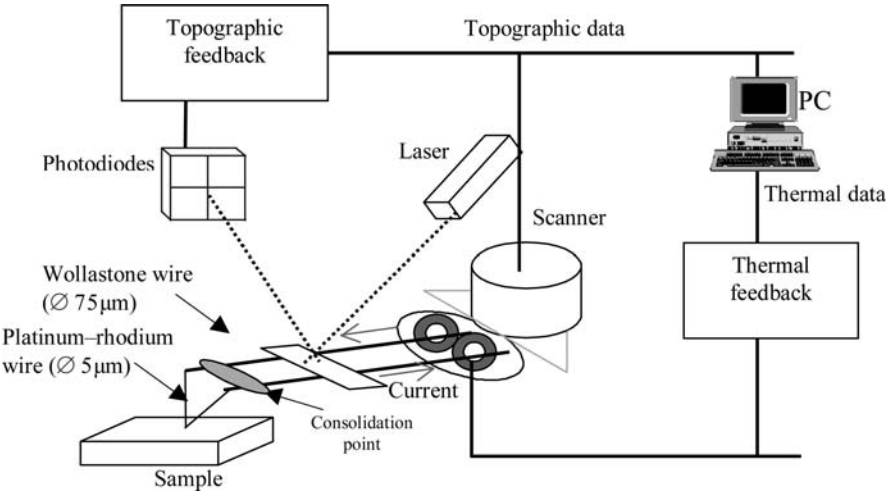

Fig. 13. Experimental arrangement for an AFM equipped with thermoresistive tip

Simply replacing the standard AFM tip by a heat sensitive probe thus gives simultaneous and independent access to topographic and thermal information (temperature and/or conductivity, depending on the type of probe). Since 1996, several thermal probes have been developed although most have remained at the development stage. We shall describe the evolving design of such probes in the next section.

2.2 Thermal Probes

For scanning thermal microscopy in the AFM configuration, the key feature is the thermal probe. New lithographic techniques have facilitated the development of such probes. These tips can be classified into three groups: thermocouple junctions, dynamical cantilevers, and thermoresistive tips.

Thermocouple Tip

In this device, the thermoelectric voltage is directly proportional to the temperature. The first thermocouple tip to be adapted to an AFM was the one due to Majumdar presented earlier. A lateral resolution of 500 nm was announced for a temperature sensitivity of 1 K. The size of the thermocouple junction in this probe is around 25 μm. Heat transfer through the air cannot therefore be neglected and it raises the time constant of the probe–air system, estimated at 56 ms. The water layer present at the sample surface is not taken into account, although it contributes to increasing the exchange area between the probe and sample.

The other problem with this probe is that it is subject to deformation. In 1995, Majumdar and coworkers tried to improve it by bonding a small

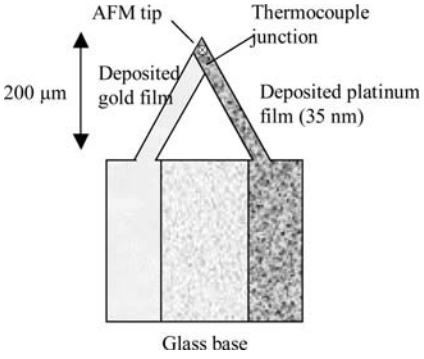

Fig. 14. Schematic diagram of a thermocouple tip formed by depositing gold and platinum films, each 35 nm thick, on different arms of the cantilever of a standard AFM tip (0.5 µm thick)

diamond on the junction at the end of the thermocouple. The facets of this diamond constitute a rather fine AFM tip. This adaptation of the probe provides a non-deformable tip under the action of contact forces and also reduces heat resistance between surface and sample, due to the high thermal conductivity of the diamond. However, this type of probe does not improve lateral resolution, even though it was the first able to detect short-circuit defects in transistors. The main difficulty here is in fabrication, which is difficult and non-reproducible.

To overcome the problem of reproducibility and attempt to improve the spatial resolution of thermal microscopes, various solutions have been proposed using commercial silicon or silicon nitride AFM tips. These AFM tips, often pyramidal in shape, have a radius of curvature of the order of ten nanometers and an apex angle of around 35°. They are held on the end of a cantilever made from a rectangular or V-shaped bar with micrometric length and width for a thickness of a few micrometers. The other end of the cantilever is joined to a parallelipiped measuring a few millimeters, used to manipulate it and place it in the AFM head.

A first probe was fabricated by depositing gold and platinum films, each of thickness 35 nm, on different arms of a V-shaped cantilever (which holds a silicon nitride tip) [62]. The Au−Pt thermocouple junction thereby formed is rather large, i.e., around 40 µm as compared with 25 µm for the previous device, and is thus too sensitive to the heat flux by conduction through the surrounding air (see Fig. 14). The spatial resolution is not better and thermal images are even distorted. However, the temperature sensitivity is improved from 1 K to 0.1 K.

Other proposals are based on specific techniques producing finer thermocouple junctions.

In 1996, Luo et al. succeeded in significantly increasing the spatial resolution of this type of tip by using a metallic evaporation technique to form deposits and by applying strong electric fields to the end of the thermocouple to make it as fine as possible (see Fig. 15). The junction could be made as small as 200 nm for an estimated sensitivity of $6\,\mu V \cdot K^{-1}$. The spatial reso-

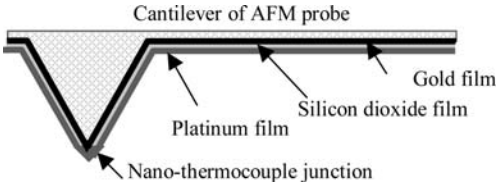

Fig. 15. Schematic diagram of the thermocouple tip proposed by *Luo* et al. [65, 66]

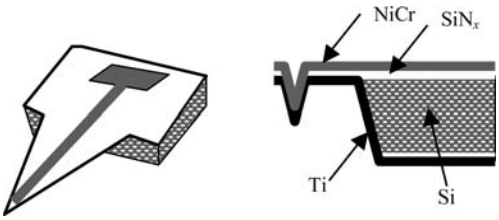

Fig. 16. Schematic diagram of a thermocouple tip directly integrated into the cantilever structure

lution was reduced to 10 nm but topographic images contained unexplained artifacts [65, 66].

In the same year, *Suzuki* et al. suggested directly integrating a thermocouple into the structure of the cantilever itself [67]. As can be seen in Fig. 16, each face of the cantilever arm is made from a metal film. The device is fabricated by successive use of techniques such as low-pressure chemical vapour deposition (LPCVD), photolithography and electrochemical etching. The thermocouple junction created at the tip apex has a sensitivity of $8\,\mu V \cdot K^{-1}$. This junction is extremely fine. The radius of curvature of the tip apex is estimated to be around 10 nm. Despite this performance, images could not be obtained with this probe.

Mills, *Weaver* and coworkers used electron beam lithography to make a tip [68]. They succeeded in obtaining a very fine tip which could acquire topographic images with very high spatial resolution (40 nm) and a thermal image resolution estimated at 80 nm. The thermocouple consisted of a gold film and a palladium film on a pyramidal silicon base coated with silicon nitride and shaped to provide probe and cantilever at the same time (see Fig. 17). The silicon base was then eliminated so that the remaining structure formed a cantilever with an Au/Pd thermocouple at the end. It was shown that, with this tip, all image contrast was due to temperature change in the thermocouple and that there was no artifact due to the topography in the thermal images. The only drawback with this probe is its fragility.

In fact, various deposition methods have been tested for tip design. The development of a new probe is a rather lengthy and difficult business.

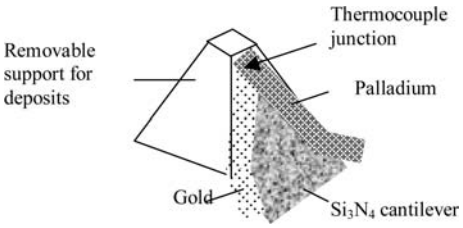

Fig. 17. Schematic diagram of the thermocouple tip proposed by *Mills* et al. [68]

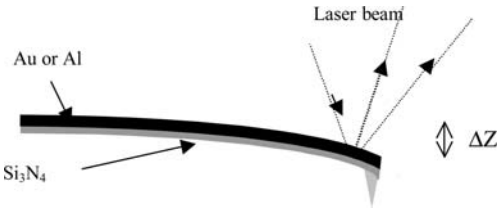

Fig. 18. Bimetallic tip

Dynamic Cantilevers

Thermal probes exploiting thermal expansion have also been proposed. A bimetallic cantilever is made from a silicon nitride AFM tip on which a thin metallic layer of aluminium or gold has been deposited, to a thickness of 50 nm, for example [69]. The sample below the tip is heated by an alternating current and absorption of heat by the tip induces a vertical bending of the cantilever due to the difference in thermal expansion coefficient between the two materials comprising it (see Fig. 18). With this probe, one detects the temperature and thermal expansion of the cantilever, which are related by measuring the deviation of a laser spot. This technique gives an estimated temperature sensitivity of 3 mK and a lateral resolution of 400 nm [69]. Unlike the thermocouple probe, the whole cantilever serves as a temperature sensor and conductive heat exchange through the air therefore dominates.

Another measurement design based on the dynamical behaviour of a standard AFM probe in contact with an active (hence heated) sample was proposed in 1999 to obtain the local thermal expansion and temperature [70]. This new device is called the Scanning Joule Expansion Microscope. It can achieve spatial resolutions of 20 nm. It has been used successfully to study heat generation in a metallic interconnect of diameter 0.5 µm buried under a passivation layer. This idea was put forward to overcome the fragility problem arising in thermal probes. However, it is not simple, because it is based on careful control of cantilever dynamics, still a delicate matter today.

Thermoresistive Probes

The first thermoresistor, described in the last section, was devised by *Pylkki* et al. [63, 64]. The cantilever arm is made from a Wollaston wire ending in a

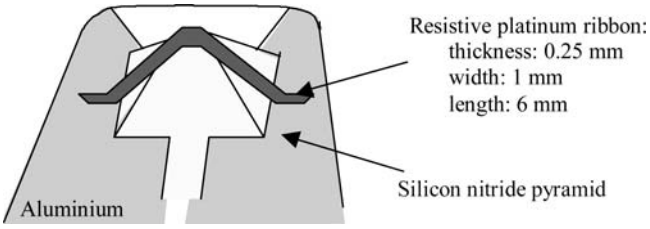

Fig. 19. Schematic diagram of a passive thermoresistive probe based on an AFM tip

platinum–rhodium wire of diameter 5 μm which constitutes the heat sensitive element of the probe (see Fig. 13). The general technique is based on measuring the variation in electrical resistivity of this element. As the contact zone between probe and sample is greater than in the case of a standard pyramidal AFM probe, the topographic images lose resolution. Further studies have been carried out to understand the measurement mechanism of this probe by determining the various modes of heat transfer taking place around the probe. They are described in Sect. 3.

In order to enhance the spatial resolution, a probe was developed by *Lederman* et al. and modified by *Mills* et al. [71, 72]. To make this, the cantilever is first coated with an aluminium film to form two electrodes which are then joined by a resistive platinum ribbon covering a standard pyramidal AFM tip (see Fig. 19). This probe is in fact passive and sensitive to temperature variations on the sample surface. The spatial resolution of this probe is estimated at 100 nm and the temperature sensitivity is of the order of one degree. It has been used to detect voids under a passivation layer of thickness 2 μm coating the aluminium wires of an electronic device.

In contrast to the others, the last two probes just discussed have been commercialised: the Lederman probe from 1999 to 2004, and the Pylkki probe since 1997, originally with the TopoMetrix AFM microscope.

Applications for these thermoresistive probes arose in 1999, but remained rather qualitative until 2000, when a more quantitative approach was discussed in the literature.

More recently (in 2001), building on their experience in CMOS fabrication processes and silicon technology, a German team proposed a thermoresistive nanoprobe integrated into a piezoresistive cantilever (see Fig. 20) [73]. The thermoresistive element is obtained by a novel technique which consists in growing a platinum wire in an electron beam and specific atmosphere. The nanowire thereby obtained is 70 nm in diameter and 5 μm in length. This novel probe operates in the same modes as the probe made by *Pylkki* et al., i.e., in passive mode where the temperature is recorded by measuring the probe resistance, and in active mode where the heated probe is sensitive to the thermal conductivity of the sample. The spatial and thermal resolutions

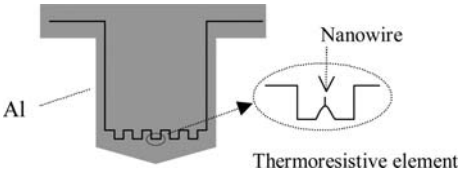

Fig. 20. Schematic diagram of the probe proposed by *Rangelow* et al. [73]

were 20 nm and 10^{-3} K, respectively. To optimise the probe, attempts are now being made to control the cantilever dynamics [74, 75].

Apart from the techniques associated with thermoresistive probes, none of the ingenious ideas described above have witnessed any great increase in their field of applications. Images are hard to interpret physically and require a deeper examination of various modes of heat transfer around these probes in an AFM microscope:

- solid–solid (probe–sample) conduction,
- conduction through the surrounding gas,
- conduction towards the probe support,
- conduction through the water meniscus depending on the probe temperature (or evaporation),
- radiative heat transfer,
- convective heat transfer occurring along the probe.

Depending on the probe configuration, the physical characteristics of the probe, the prevailing temperatures, environmental conditions (temperature, hygrometry, and so on), the measurement will reflect the contribution of each of these transfer mechanisms in unequal proportions. Depending on the state of the sample surface, the probe configuration will also affect the coupling between topographic and thermal data.

The probes proposed up to now are not perfect. A good compromise has to be found between spatial resolution, temperature sensitivity (with a response time suitable for imaging), but also a certain level of feasibility when it comes to reproducibility in fabrication. Research is still under way in this area.

3 SThM with the Micrometric Thermoresistive Wire Probe

This section[3] discusses the current understanding of the first commercialised scanning thermal microscope known as. After describing the device and its main operating modes, we discuss the various parameters controlling the contrast in thermal images that can be obtained with it. We then examine ways of evaluating the impact of these parameters on measurement and in particular on calibration of the microscope for localised thermal metrology. To end, we present a number of examples without attempting to be exhaustive.

[3] Contributed by S. Gomès.

3.1 Goals

The scanning thermal microscope proposed by *Pylkki* et al. [63, 64] was designed to tackle the experimental problems raised in determining the thermophysical properties of microstructures such as grains and grain boundaries, thin films, and powder particles. This determination is essential if we are to understand and model heat transfer and macroscopic properties of heterogeneous materials.

It was also designed to be able to detect local heating at the surface of active samples, in particular in order to study surface temperature fields and the flux of heat transfer inside microelectronic and optronic components.

3.2 Method

Experimental Setup

The basic setup is shown schematically in Fig. 21. There are four main elements:

- A thermal probe. This is the thermoresistive tip developed by *Dinwiddie* et al. in 1994 [76]. This probe comprises a cantilever arm with a thermoresistive element at the end. The cantilever arm is made from a Wollaston wire, i.e., a silver wire of diameter 75 μm containing a platinum–rhodium (10 %) filament of diameter 5 μm. The thermoresistive element is obtained by electrochemically etching this wire over a length of about 200 μm. The thickness of silver surrounding the platinum filament is thereby removed over this length. Once the filament has been exposed, it is bent in a V shape to give the tip its final shape.
- A displacement, positioning and optical control[4] system to guide the thermal tip in its motion relative to the sample surface. The operating principle is the same as in an AFM used mainly in (constant force) contact mode.
- A thermal unit controlling the electrical resistance of the tip (i.e., its average temperature) and the power dissipated within it.
- A data acquisition, visualisation and processing system. As the AFM system and thermal control system are independent, to each value of the tip–sample force there correspond surface functions $S_{\text{topo}}(X, Y)$ and $S_{\text{th}}(X, Y)$ represented by 3D recordings of grey levels or false colours which correspond to the topography and locally measured 'thermal' signals.

The topographic and thermal data acquisition and processing system is provided by a PC which pilots the AFM tip.

[4] A mirror is attached to the cantilever arm of the thermal probe, which has stiffness constant 5–20 $\text{N} \cdot \text{m}^{-1}$.

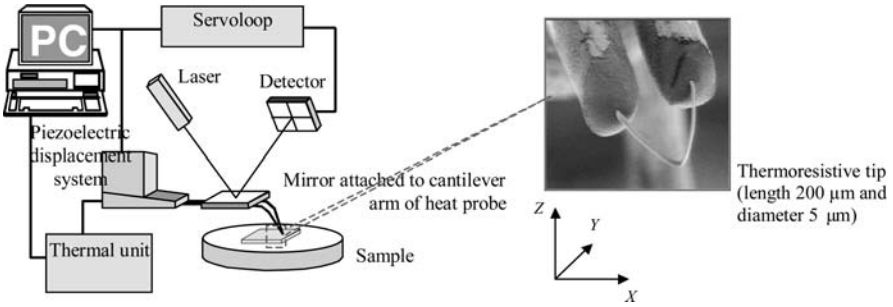

Fig. 21. *Left*: Experimental configuration of the scanning thermal microscope. *Right*: SEM image of the thermoresistive tip

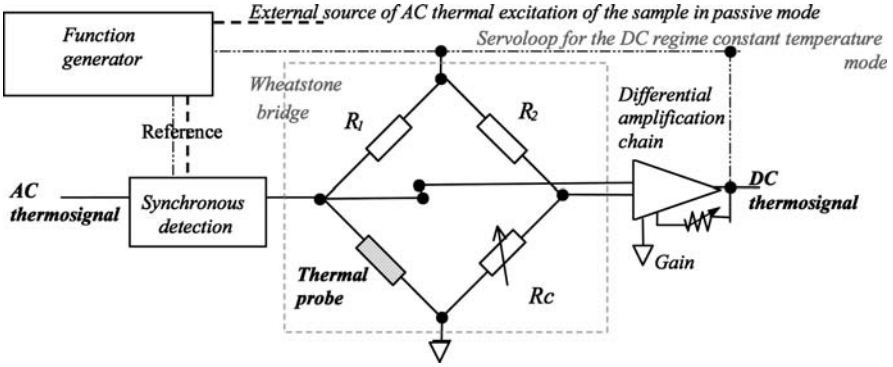

Fig. 22. Electrical circuit for the thermal unit. *Dashed lines*: Components used for the constant temperature mode. *Dotted* and *bold lines*: Electrical connections used in the passive mode

Thermal Measurements

The thermal probe constitutes one leg of a Wheatstone bridge with ratio R_1/R_2 (see the circuit diagram for the thermal unit in Fig. 22). The electrical resistance of the tip is controlled by adjusting the variable resistance R_C of this bridge. At equilibrium,

$$R_C = \frac{R_2}{R_1}(R_P + R_{wire}) , \tag{19}$$

where R_P and R_{wire} are the electrical resistances of the tip and the wires (including the cantilever arm) connecting the probe to the circuit of the thermal unit.

Given the linearity of the relation between the electrical resistance of the thermoresistive element and its temperature, the resistance R_C can be used to control the average temperature T_{op} of the tip:

$$R_C = \frac{R_2}{R_1}\Big\{R_{P0}\big[1 + \alpha_P(T_{op} - T_{P0})\big] + R_{wire}\Big\}, \tag{20}$$

where R_{P0} is the electrical resistance of the tip at a reference temperature T_{P0} and α_P is the temperature coefficient of the electrical resistivity of the platinum–rhodium (10 %) filament, viz., $\alpha_P = 0.00165\,K^{-1}$.

Measurements can be made in DC regime and/or AC regime. For the AC regime, the circuit just described is completed with a synchronous detection device. This provides simultaneous access to the amplitude and phase of the harmonic registered in the AC voltage across the tip (see Fig. 22). Like any thermoresistive sensor, the thermal probe of this SThM can be used in various operating modes, i.e., one passive mode and two active modes:

- **Passive Mode.** This is usually called the temperature contrast mode. The thermal probe is used as a resistive thermometer. A constant current is applied, chosen small enough to ensure that the Joule heating of the tip can be neglected ($I = 1$–$2\,mA$). During a scan, a temperature change at the sample surface can cause a temperature variation ΔT_P in the tip and hence a change in the electrical resistance of the probe given by

  $$\Delta R_P = R_{P0}\alpha_P \Delta T_P\,. \tag{21}$$

  This variation in the electrical resistance implies a change in the differential voltage output from the circuit of the thermal unit. Acquisition of the variations in this voltage thus amounts to detecting local heating at the sample surface. In the AC regime, the useful components of the signal are those extracted at the modulation frequency f of the power of the source exciting the sample.
- **Active Modes.** In this case, a larger current passes through the resistive element of the probe. Heated by the Joule effect, the tip is used as a source to excite the sample. There are two possible active measurement modes: either the current through the probe is held constant (constant current mode), or the electrical resistance of the probe, i.e., the average temperature T_{op} of the thermoresistive element, is held constant (constant temperature mode or thermal conductivity contrast mode).
  - **Constant Current Mode.** As in the passive mode, the differential voltage output from the amplification chain of the thermal unit is recorded. Controlling the variation in the electrical resistance of the heated tip by acquiring this voltage, the evolution of the sensor temperature can be monitored and variations in its flux losses can be detected while the sample surface is being scanned.
  - **Constant Temperature Mode.** In this mode, the Wheatstone bridge is equipped with a feedback loop acting on its equilibrium

voltage in such a way as to hold the average temperature of the tip constant. During an experiment, any flux lost from the heated tip to its surroundings (including the sample) is compensated for electrically by servo controlling the equilibrium voltage of the bridge. If this voltage is recorded when the sample surface is scanned by the probe, one can obtain an image whose contrast contains, among other things, information concerning the local thermal conductivity of the observed materials.

In the AC regime, a sinusoidal current is superposed on the continuous current to heat the tip in a periodic manner. The synchronous detection system is used to extract the amplitude and phase of the useful component of the AC signal. We shall see in Sect. 3.5 how analysis of the third harmonic of this signal has been used to study the local thermophysical properties of samples.

3.3 Thermal Image Contrast

Visualisation is useful for qualitative analysis of any phenomenon. For example, the thermal images shown in Fig. 23b (constant temperature mode) and Fig. 23d (constant current mode) have a contrast which clearly reveals respectively:

- variations in the thermal conductivity at the surface of and just beyond an SiO_2 step on an Si substrate [77],
- the detection of a hot spot inside a MOSFET under breakdown [78].

However, the analysis of SThM images is not always as trivial as in these examples and the thermal image remains a complex ensemble made up of a multitude of entangled data.

Thermal Signal

In imaging mode, the rate at which the surface is scanned is chosen slow enough to justify assuming that the tip can reach its energy equilibrium. In the DC regime, for example, during production of a $100\,\mu\text{m} \times 100\,\mu\text{m}$ image, with a tip speed of $25\,\mu\text{m}\cdot\text{s}^{-1}$ and 512 points per profile, the spatial interval of 195 nm between measurements is covered in 8 ms. This is two orders of magnitude greater than the temporal response of the tip, estimated to be several hundred μs [79].

Whatever SThM mode is used, the thermal signal S_{th} is a voltage whose variations are determined by the variations in the electrical power P dissipated in the tip. P is determined by the various forms of heat transfer to and from the tip. Quite generally, the energy equilibrium of the tip is given in the AC regime by

$$\mathrm{i}\omega m C_{\text{P}} T_{\text{P}\omega} = P_\omega - Q_{\text{W}\omega} - Q_{\text{conv}\omega} - Q_{\text{el}\omega} - Q_{\text{S}\omega}\,, \tag{22}$$

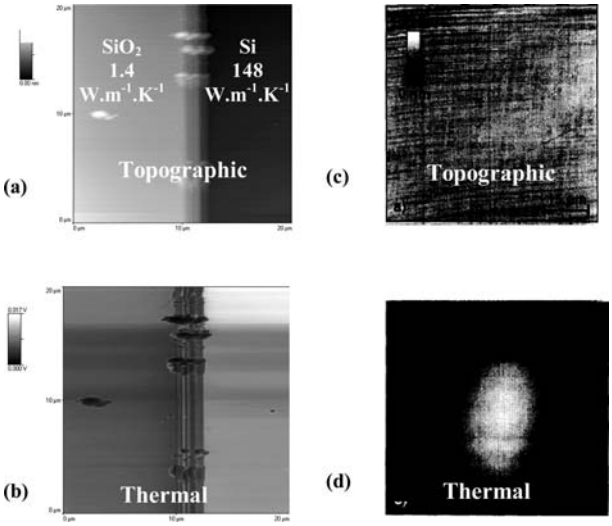

Fig. 23. (**a**) Topographic and (**b**) thermal images of an SiO_2 step of thickness 240 nm on an Si substrate, acquired in constant temperature mode [77]. (**c**) Topographic and (**d**) thermal images of the rear surface of the substrate (thickness 5 µm) of a MOSFET under breakdown, acquired in constant current mode. Reprinted from [78] with permission of Elsevier. Note that the thermal signal is higher, revealed by a brighter shading in the thermal images, when the effective thermal conductivity of the sample is high [see (**b**)] in constant temperature mode, or when the sample temperature is high [see (**d**)] in constant current mode

where m is the mass of the tip, C_P is the specific heat capacity of the tip, $T_{P\omega}$ is the component of the tip temperature with frequency $f = \omega/2\pi$, and the subscript ω denotes the component at frequency f of the (positive or negative) powers Q_W, Q_{conv}, Q_{el}, and Q_S described below (with $\omega = 0$ in DC regime):

- Q_W is the heat rate transferred to the Wollaston support of the filament. The Wollaston, composed mainly of silver, which is a very good conductor, acts as a heat sink for the tip. The flux transferred to it is very large and in particular causes a temperature gradient along the thermoresistive filament. This gradient has been demonstrated in DC regime by local measurements of temperature at different points of a tip heated by the Joule effect (see Fig. 24) [80].
- Q_{conv} is the heat rate exchanged by convection with the surrounding air. Its contribution has been demonstrated experimentally. By analysing the signal obtained in DC regime in the constant temperature mode for a tip without contact immersed in different gases, it has been shown that the power dissipated in the tip (to keep its temperature constant) is higher when the tip temperature and/or the thermal conductivity of the

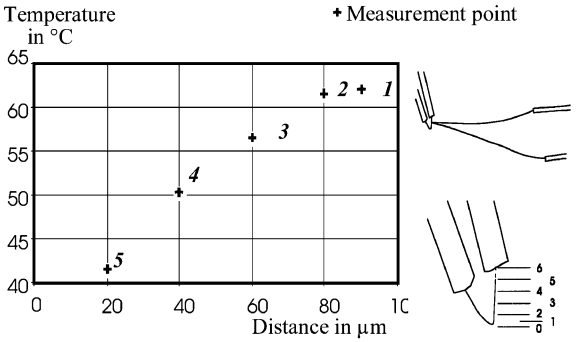

Fig. 24. Temperature profile measured using a thermocouple with junction diameter 1 µm over a half-length of a tip heated by the Joule effect and immersed in air (no contact). Reprinted from [78] with permission of Elsevier

surrounding gas are large (see Fig. 25). This is an indication of convective exchange along the whole length of the filament. Its direct contribution to heat transfer from the tip to its environment is estimated to be of the order of 10 % of P [80].

– Q_{el} is the heat rate resulting from thermoelectric effects such as the Thomson effect within the sensor. This flux has been shown to be negligible [77]. Hence, in the DC regime ($\omega = 0$), we have

$$P = Q_{\mathrm{W}} + Q_{\mathrm{conv}} + Q_{\mathrm{S}} \,. \tag{23}$$

– Q_{S} is the heat rate exchanged with the sample. Exchanged locally between the tip and sample while the sensor is displaced, its variations lead to the observed contrast of the thermal image. Variations in the power dissipated in the tip, and hence in the thermal signal, will be larger when the variations in this flux are greater.

Tip–Sample Heat Transfer

Quite generally, the tip–sample heat transfer is characterised in scanning thermal microscopy by different heat transfer mechanisms which depend heavily on the tip environment and the region under investigation. For measurements carried out in air, these mechanisms include transfer at a distance, i.e., radiation and conduction through the air, and transfer by direct contact between two solids, i.e., direct solid–solid conduction or conduction via a water film adsorbed onto the tip and sample surfaces if the contact temperature is low enough for the water not to evaporate (see Fig. 26) [81].

Short-Range Heat Transfer Between Two Solids

– Radiative transfer. The classical theory of heat exchange by radiation predicts the validity of the Boltzmann law for distances $d > \lambda_{\mathrm{th}}$, where λ_{th}

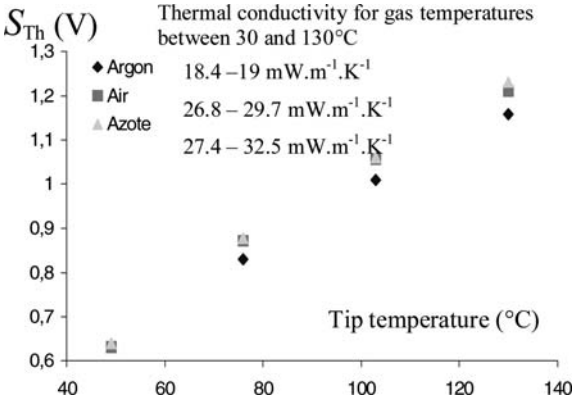

Fig. 25. Variation in the DC signal in constant temperature mode as a function of the tip temperature for measurements carried out in surrounding gases with different thermal conductivities and without contact. Reprinted from [80] with permission of Elsevier

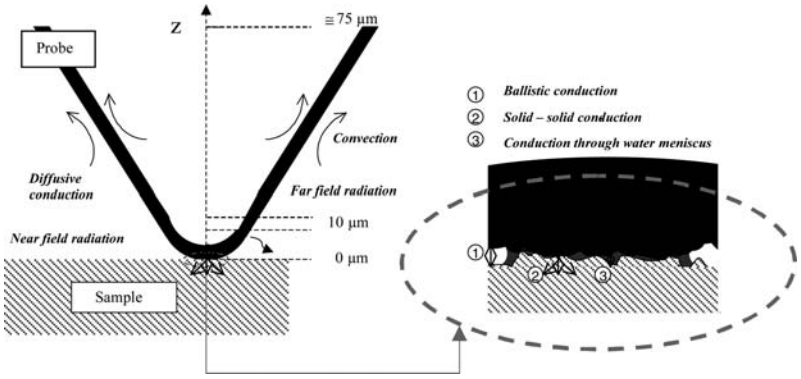

Fig. 26. Different modes of heat transfer between tip and sample

is the wavelength of the thermal radiation given by Wien's law ($\lambda_{\text{th}} = 10\,\mu\text{m}$ for $T = 300\,\text{K}$). This transfer acts over a large part of the surface of the thermoresistive filament. In active mode, its contribution is estimated to be of the order of $10^{-8}\,\text{W}\cdot\text{K}^{-1}$ [82], which is less than 1/100 of the conductances corresponding to other forms of exchange (discussed below). It is usually neglected. In the passive mode, it depends on the temperature and the size of the hot spot on the sample. The end of the tip (regions for which d is less than a few µm) is exposed to near-field radiative heat transfer. According to *Lefèvre* [82], this transfer is theoretically even more negligible than the last.

– Thermal conduction through air. The contribution of thermal conduction through the air is easily observed experimentally in the constant

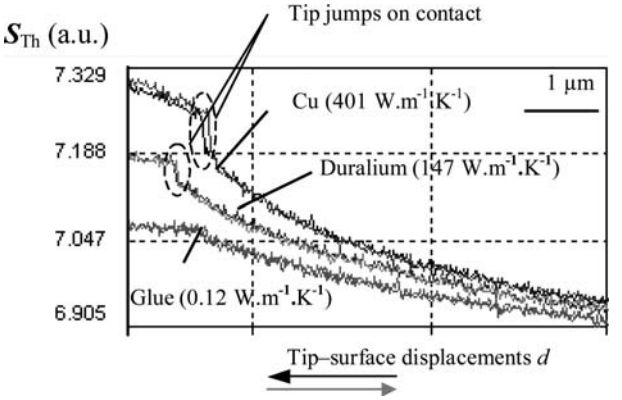

Fig. 27. S_{th} as a function of displacement z obtained in air for copper, glue and duralium in DC regime. The tip temperature T_{P} is 102 °C [83] ©Copyright 1999, IOP Publishing Ltd

temperature mode. When a heated tip approaches the sample, the power dissipated in the filament to hold the temperature constant increases this all the more so as the sample is thermally conductive (see Fig. 27) [83]. During contact measurements, this exchange contributes significantly to tip–sample heat transfer. Its conductance is estimated to be of the order of $2.5 \times 10^{-6}\,\mathrm{W \cdot K^{-1}}$ [82]. There are three transfer regimes via the air depending on the distance z between the relevant region of the tip and the sample surface [84]. In theory, diffusive transfer acts on areas of the tip situated at distances greater than 100 times the mean free path l_{a} of the air molecules ($l_{\mathrm{a}} \sim 50\,\mathrm{nm}$ at atmospheric pressure). Between $100 l_{\mathrm{a}}$ and l_{a}, transfer occurs in the so-called slip regime. The conductive flux goes as $1/z$ as for diffusive transfer, but it is stronger due to the reduced collisional regime (see the Chapter on Transport in Dilute Media by *Carminati* in this volume). More locally, the region of the tip situated at a distance less than l_{a} from the sample undergoes ballistic exchange (Knudsen regime) with the surface. The conductive flux is then independent of z [81, 82]. Simple geometrical considerations lead to an estimate of 1 µm for the radius of the region where ballistic exchange coupling occurs. It is interesting to note that, in the passive mode, the contribution of this transfer to the tip–sample coupling depends on the heated area of the sample.

Contact Heat Transfer Between Two Solids

– Heat transfer via the water film on the sample surface. This transfer has been shown to favour thermal coupling between the tip and sample [81, 83]. The associated thermal conductance and radius of coupling

are estimated at $2 \times 10^{-5}\,\mathrm{W \cdot K^{-1}}$ and 100–300 nm, respectively [80, 84]. For the active modes, the probe temperature can affect the thickness of the water meniscus formed between the tip and sample when contact is made in ambient air with nonzero relative humidity. The heat transfer through this water film falls off as the tip is heated. Although this coupling seems a priori to be able to improve thermal image contrast, the thickness of the adsorbed water layer depends on the hydrophilic properties of the sample. Consequently, the thermal image contrast will be linked to the hydrophilic properties of the sample. To get round this problem, it is recommended when studying a given sample in constant temperature mode to determine the tip temperature at which the water contribution becomes negligible and to work at a higher tip temperature.

– Direct solid–solid thermal conduction. Transfer by direct solid–solid conduction is revealed for good heat conductors by a sudden jump in the heat flux exchanged between the thermal probe and the sample when they come into contact, followed by an increase in this flux as the tip gradually penetrates into the material (for tip temperatures that have been shown to be high enough to evaporate the water meniscus adsorbed onto the tip and sample surfaces). For example, variations in the thermal signal can be observed, the signature of the contribution from direct solid–solid conduction, when contact is made and further pressure applied to a tip heated to 102 °C with samples of copper and duralium (see Fig. 27). A similar experiment has been carried out in nitrogen, showing that the flux then exchanged is independent of the nature of the surrounding gas [83]. The contribution of this exchange to the coupling depends on the thermal conductivity of the sample and the force applied during the measurement. The conductance was reported to vary between 0 and $5 \times 10^{-6}\,\mathrm{W \cdot K^{-1}}$ [82].

Although all these heat transfer mechanisms contribute to the tip–sample coupling, their contributions are closely related to the properties of the sample itself. The cross-section of the cylinder through which the main part of the exchanged heat flux is transferred, dependent as it is on the environmental conditions in which the measurement takes place, will limit the lateral thermal resolution of the tip.

Tip Resolution

This SThM can simultaneously image topographic contrast and thermophysical properties or local heating of sample surfaces. One must therefore distinguish two lateral resolutions: thermal resolution and topographic resolution.

Lateral Topographic Resolution

This corresponds in theory to the smallest distance beyond which two variations of the same height can be distinguished in the topographic image of

a flat sample. Limited by surface expansion phenomena and by the size and geometry of the tip, it will always be less good (of the order of 150 nm for topographic details of height 10 nm) than can be achieved with a typical AFM tip (with a radius of curvature of a few nm).

Lateral Thermal Resolution

This is defined as the smallest distance beyond which two variations with identical temperature or thermophysical property can be distinguished in the thermal image of a plane, mirror-polished sample. In practice, it can be defined in terms of a criterion of type $-n$ dB ($n = 2$ or 3) by analogy with the classical resolution criteria of Rayleigh or Sparrow, for example. It is of course a function of:

- tip temperature and environmental conditions prevailing during measurement (see last section),
- the sensitivity of the thermal signal to variations in the relevant thermal parameter (thermophysical property or temperature),
- the minimum detectable flux,
- characteristics of the sample such as thermophysical properties, size, depth in the sample of surface and/or subsurface details,
- characteristics of the tip, each tip being unique since it is fabricated manually.

This explains the wide spread of values to be found in the literature.

Using a -2 dB criterion, *Hammiche* et al. [85] obtain a lateral resolution of 1.5 µm for micrometric copper particles, coated and emerging from a polymer matrix in the DC regime constant temperature mode. This resolution deteriorates noticeably when the particles are buried at depths of several micrometers ($\leq 5\,\mu$m) in the matrix. Moreover, the experiment showed that the tip can detect surface details with dimensions of a few hundred nanometers [86, 87]. The detection of smaller details (20–30 nm) has also been announced in AC regime [88].

Investigation Depth

The volume of matter probed by a tip in scanning thermal microscopy can be shown to be confined to the tip–surface heat exchange region in the case of purely diffusive materials [89]. For a thermal contact of submicron size characterised by a given power density passing through it, the depth of investigation depends in theory mainly on the size of this contact region and depends only very slightly on the thermal properties of the sample in DC regime [89]. In practice, the contact power density between the tip and sample depends on the thermal properties of the sample. This is due to the very principle of contact measurement (see Sect. 3.5). Like the lateral thermal resolution, the investigation depth of the SThM depends on the five parameters listed above. Given the submicron lateral thermal resolution announced for

the tip we are considering here (assumed to be of the order of magnitude of the tip–sample thermal contact radius), we may consider the SThM to have a 'visibility' to a depth of the order of a few μm in this regime. Studying films of amorphous compounds (C:H/Si:O) deposited on Si(001) substrates in DC regime, *Ruiz* et al. [90] estimated at 5 μm the minimal film thickness for which only the properties of the film and not the effective properties of the film–substrate system are detected. In the AC regime, the investigation depth still depends mainly on the size of the thermal contact region (see the Chapter by *Cretin* and *Vairac* in this volume).

Apart from the tip–sample heat transfer mechanism, the geometrical relationship between the surface of the long thin tip and the surface of the sample also influences thermal image formation. In every case, the reconstruction of the topography of surface features with the SThM tip will be as important as the thermal reconstruction itself. Thermal analysis of the SThM image will only be possible in regions that are resolved clearly enough in the topographic image, the ideal surface being flat and mirror-polished.

Artifacts in Thermal Images

Given the size and shape of the probe, any thermal (or topographic) image will be affected to some extent by the sample topography.

Artifacts Caused by Topography

Variations in the flux between the tip and sample caused by variations in the thermal coupling area when the tip follows the profile of dips and humps (see Fig. 23a and b), or when it is in multiple contact with the sample, will tend to affect the contrast in the thermal image. An extreme case is illustrated in Fig. 28. This shows topographic and thermal images, obtained in DC regime constant temperature mode, of a silicon oxide array ($1.4\,\mathrm{W\cdot m^{-1}\cdot K^{-1}}$) on an Si substrate ($148\,\mathrm{W\cdot m^{-1}\cdot K^{-1}}$). The shape of the holes (diameter 600 nm, depth 240 nm) is completely distorted in the topographic image. The thermal tip cannot enter inside the hole due to its size. The result is a reduction in contact and near-field heat transfer, whence a thermal contrast that does not reflect the contrast in the thermal conductivity of the sample and a loss of information.

It is thus essential to analyse the state of the sample surface with a standard AFM tip in order to obtain a better knowledge of the surface topography to be observed with the SThM, otherwise there is a risk of interpreting topographic artifacts as local variations in thermal properties.

Artifacts Caused by Tip Shape

Owing to the V shape of the tip, the contact area and thermal coupling area on the sample surface have elliptical rather than discoidal shape. As can be seen from the reference frame shown in Fig. 29, this results in a distortion

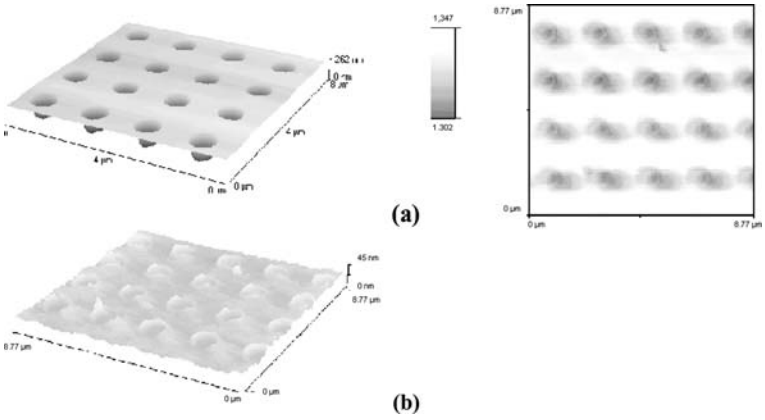

Fig. 28. *Left*: 3D topographic images with an AFM tip (**a**) and with an SThM tip (**b**) of an SiO_2 array on an Si substrate. *Right*: Thermal image of the same. Holes with diameter 600 nm and depth 240 nm are spaced 1.1 μm apart [77]

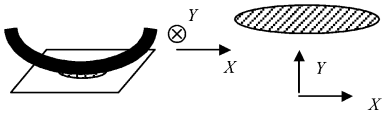

Fig. 29. Effective heat exchange region. *Left*: Side view. *Right*: Top view with frame of reference

of the topographic and/or thermal details which will be more pronounced in the X direction (in both types of image). It is clear that the lateral resolution of the tip will be less good in this same direction.

Other Sources of Error

Other factors can also affect the thermal contrast:

- The tip apex may pick up and carry a dust grain or some other form of contamination during scanning.
- So-called stick-and-slip phenomena, where the tip motion stops and starts in an uncontrollable manner.
- Local variations in mechanical, thermal, and thermomechanical properties of a sample: the contact (and hence the thermal coupling) area, determined for a given tip–sample mechanical stress, can vary significantly depending on the properties of the material making up the sample. It is therefore advisable to establish the mechanical properties of a sample before attempting to analyse the thermal image made from it.

It should also be borne in mind that the geometries of the two surfaces in contact are poorly known and that effects such as local roughness, capillary forces, adhesive forces, and interatomic forces can vary from one material to another (see Sect. 1.3).

Some of these factors can be significantly reduced by optimising the SThM control parameters.

3.4 Controlling and Optimising SThM Functions

As in any experimental device, a certain number of precautions are necessary in order to obtain good measurements. The first stage of the analysis procedure is to acquire the image to be processed. Prior to this operation, samples must always be carefully prepared (cleaning and fine polishing). This facilitates the processing of the thermal image. In particular, the following points should be taken into account:

- The environment of the setup, e.g., ambient temperature, relative humidity of the air, must be perfectly checked and controlled to guarantee stability.
- The contact and displacement of the tip relative to the sample must occur in optimal conditions. Special attention must be given to positioning of the tip in order to minimise the risk of error. This is achieved by suitably adjusting the tip–sample force (which must be small enough to avoid damage to either tip or sample), the gain of the feedback system controlling the force, and the scan rate.
- The response time of the system must be determined and optimised in order to optimise in turn the speed of the sensor, the spatial interval for data acquisition, and the temperature modulation frequency (in AC regime), so as to obtain the most accurate measurement possible [91].

In addition, as for any experimental measurement, one must estimate the experimental error and the influence of the environment and the sensor itself on the measurement.

3.5 Analysing Measurements in Constant Temperature Mode

The constant temperature mode is by far the most widely used. Various approaches have been proposed to clarify the dependence of the thermal signals in this mode on the thermophysical properties of samples.

DC Regime

In the DC tip heating regime, the measured signal (thermal signal) S_{th} is the equilibrium voltage of the Wheatstone bridge, servo controlled to hold the electrical resistance R_{P} of the tip constant (see Fig. 22). This voltage varies linearly with the voltage V_{P} applied to the thermoresistive sensor:

$$V_{\mathrm{P}} = S_{\mathrm{th}} - (R_1 - R_{\mathrm{wire}})I\,, \tag{24}$$

where R_1 is a resistance of the bridge, R_{wire} is the resistance of the wires connecting the tip to the bridge, and $I = S_{\text{th}}/(R_{\text{P}} + R_{\text{wire}} + R_1)$ is the current in the probe.

As discussed in Sect. 3.3, the electrical power P dissipated in the tip, given by $P = V_{\text{P}}^2/R_{\text{P}} = V_{\text{P}}I$ and determined by heat exchange from the heated tip to its environment (its Wollaston support, the surrounding gas, and the sample during a contact), governs the thermal signal:

$$P = \frac{V_{\text{P}}^2}{R_{\text{P}}} = Q_{\text{W}} + Q_{\text{conv}} + Q_{\text{S}} \,. \tag{25}$$

Let P_{J} and P_{C} be the powers dissipated in the tip without and with contact:

$$P_{\text{J}} = \frac{V_{\text{PJ}}^2}{R_{\text{P}}} = Q_{\text{W}} + Q_{\text{conv}} \,, \tag{26}$$

$$P_{\text{C}} = \frac{V_{\text{PC}}^2}{R_{\text{P}}} = Q_{\text{WC}} + Q_{\text{convC}} + Q_{\text{S}} \,. \tag{27}$$

Sensitivity to Thermal Conductivity of the Sample

In 1998, *Ruiz* et al. [90] proposed a first SThM calibration procedure for determining the thermal conductivity of samples. From measurements of the thermal signal made with a given probe for different homogeneous, mirror-polished, rigid samples with known thermal conductivity, they calculated the variation $\Delta P = P_{\text{C}} - P_{\text{J}}$ required to hold the temperature T_{op} of the tip constant for each sample. They then produced a calibration curve giving a perfectly linear dependence of ΔP on the thermal conductivity k_{S} of the samples. Assuming ΔP proportional to the power Q_{S} supplied to the sample by the tip, they suggested that this linearity corresponds to that of Q_{S} as a function of k_{S}, viz., $Q_{\text{S}} = C_{\text{R}}k_{\text{S}}$, where C_{R} is a constant proportional to the product $a_{\text{c}}(T_{\text{op}} - T_{\text{a}})$, where a_{c} is the tip–sample contact radius, T_{op} is the operating temperature of the probe deduced from (19) and (21) (the probe is thus assumed to be isothermal), and T_{a} is the ambient temperature.

Although *Gorbunov* et al. [92] held to this analysis of the measurements, the calibration method was questioned by other research groups [80, 93, 94]. Indeed, *Lefèvre* et al. [94] recently proposed a more detailed study. Figure 30 gives the experimental results they obtained with two microscopes and different tips.

These measurements do not agree with the linearity announced by Ruiz et al., but reveal an asymptotic behaviour in measurements obtained for thermal conductivities greater than a few tens of $\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$. This indicates the much reduced sensitivity of the SThM at high conductivities [80]. The various approaches to the measurements discussed below explain this nonlinearity.

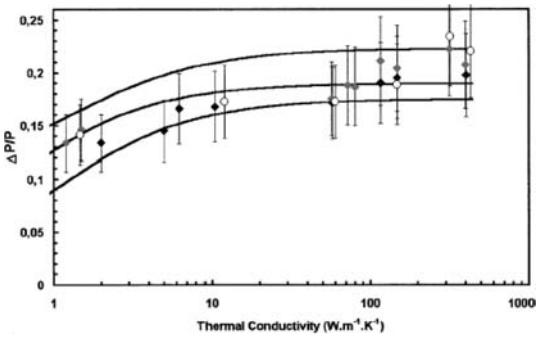

Fig. 30. Dependence of the ratio $\Delta P/P_{\mathrm{J}}$ on the thermal conductivity of standard samples determined from experimental data obtained with two SThMs and different tips. *Smoothing curves* are deduced from (36). Reused with permission from *S. Lefèvre* et al. [94]. ©Copyright 2003, American Institute of Physics

Other Approaches to Measurement

In the first approach, the difference $\Delta P = P_{\mathrm{C}} - P_{\mathrm{J}}$ is assumed to be fully transmitted to the sample and the tip is assumed isothermal (at temperature T_{op}). However, *Callard* et al. [93] and *Gomès* et al. [77, 80] showed in 1999 that these hypotheses cannot reasonably be accepted.

Due to its size, the tip can be treated as a kind of thermal fin of uniform section. The corresponding Biot number is found to be four orders of magnitude less than unity [80]. However, it cannot be treated as isothermal. The heat sink effect of the tip's Wollaston support, which leads to a temperature gradient along the platinum filament, cannot be neglected (see Sect. 3.3).

Moreover, as for any contact method of thermal metrology, a constriction effect is caused by converging thermal current lines in the measurement region. The quantity of heat then exchanged between tip and sample perturbs not only the temperature field in the sample, but also the temperature distribution in the probe. There are two consequences:

– In this measurement region, the temperatures of the sample and the tip apex are no longer the temperatures T_{P} and T_{∞} they had before the tip approached the sample, but the perturbed temperatures T_{PC} and T_{SC}. Assuming heat transfer from the tip to the sample through a projected area on the sample of small radius b (given the submicron resolution of the tip, this hypothesis is justifiable), energy conservation at the tip–sample interface can be expressed in the form [95]

$$Q_{\mathrm{S}} = G(\theta_{\mathrm{PC}} - \theta_0) = 2\pi b k_{\mathrm{S}} \theta_0 \,, \tag{28}$$

where G is the thermal conductance of the tip–sample heat transfer, $\theta_{\mathrm{PC}} = T_{\mathrm{PC}} - T_{\infty}$ and $\theta_0 = T_{\mathrm{SC}} - T_{\infty}$. Note, however, that this equation is only valid for a purely diffusive conductive heat transfer mechanism in the sample.

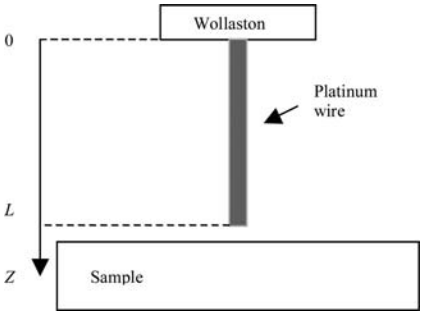

Fig. 31. Description of the tip and its environment

– Losses from the tip to its support vary. Assuming invariant convection losses from the tip,

$$\Delta P = \Delta Q_{\mathrm{W}} + Q_{\mathrm{S}}\,. \tag{29}$$

In order to solve the equation $\Delta P = f(k_{\mathrm{S}})$ relating the measurement to the thermal conductivity of the sample, one must first solve the differential equation expressing the energy balance at the tip given by (26) and (27). The problem now is therefore to solve this equation.

Let A_{P} be the cross-sectional area, p the perimeter, k_{P} the thermal conductivity, and ρ_{P} the electrical resistivity of the tip. These quantities will be assumed invariant under temperature change. By symmetry, the probe will be modelled by two fins of length L ($2L = 200\,\mu\mathrm{m}$) and diameter $a = 5\,\mu\mathrm{m}$ (see Fig. 31).

Under these hypotheses and in the stationary regime, the temperature distribution along the wire heated by the Joule effect satisfies the standard differential equation

$$k_{\mathrm{P}}\frac{\partial^2\theta_{\mathrm{P}}(z)}{\partial z^2} + \frac{ph_{\mathrm{P}}\big(\theta_{\mathrm{P}}(z)\big)}{A_{\mathrm{P}}} + \frac{\rho_{\mathrm{P}}I^2}{A_{\mathrm{P}}^2}\big[1-\alpha_{\mathrm{P}}\theta_{\mathrm{P}}(z)\big] = 0\,, \tag{30}$$

where $\theta_{\mathrm{P}}(z) = T_{\mathrm{P}}(z) - T_{\infty}$ is the axial temperature, h_{P} (estimated to be of the order of $1000\,\mathrm{W}\cdot\mathrm{m}^{-2}\cdot\mathrm{K}^{-1}$ [80]) is the convective transfer coefficient along the tip, z denotes the spatial parameter, and I is the current through the tip, deduced from the measurement to be processed [see (24)].

Direct solution of this equation does not lead to a simple relation $\Delta P = f(Q_{\mathrm{S}})$ to calibrate the SThM. *Lefèvre* et al. [94] therefore simplified it. We now describe their method.

Since the convection loss given by $ph_{\mathrm{P}}\big(\theta_{\mathrm{P}}(z)\big)/A_{\mathrm{P}}$ and the power gain due to the thermoelectric effect given by $-\rho_{\mathrm{P}}I^2\alpha_{\mathrm{P}}\theta(z)/A_{\mathrm{P}}^2$ are estimated at 10–20 % of the power $\rho_{\mathrm{P}}I^2/A_{\mathrm{P}}^2$ supplied by the Joule effect, (28) reduces to the form

$$k_{\mathrm{P}}\frac{\partial^2\theta_{\mathrm{P}}(z)}{\partial z^2} + \frac{\rho_{\mathrm{P}}I_{\mathrm{PC}}^2}{A_{\mathrm{P}}^2} = 0\,. \tag{31}$$

In order to respect the fact that the probe temperature is held constant whatever its environment, the value of the current I_{PC} in this new equation is equated with the value deduced from measurements with or without contact depending on the situation considered. The expression for the local temperature in the probe is then

$$\theta_{\mathrm{P}}(z) = -\frac{\rho_{\mathrm{P}} I_{\mathrm{PC}}^2}{2k_{\mathrm{P}} A_{\mathrm{P}}^2} z^2 + Az + B\,, \tag{32}$$

where A and B are constants determined by the boundary conditions

$$Q_{\mathrm{W}} = 2G_{\mathrm{Ag}}\big(\theta_{\mathrm{P}}(z=0)\big)\,, \tag{33}$$

which expresses the heat rate from the tip to its support at $z = 0$ (see Fig. 32) with $G_{\mathrm{Ag}} = 3.7 \times 10^{-3}\,\mathrm{W \cdot K^{-1}}$ the conductance of the tip support, and

$$Q_{\mathrm{S}} = G(\theta_{\mathrm{PC}} - \theta_0) = 2\pi b k_{\mathrm{S}} \theta_0\,, \tag{34}$$

which expresses the tip–sample coupling, assumed to occur only at the tip–sample interface at $z = L$ (see Fig. 32). The coefficient 2 in (33) accounts for the fact that the tip comprises two branches.

Identifying the tip conductance $G_{\mathrm{Pt}} = k_{\mathrm{P}} A_{\mathrm{P}}/L = 5.9 \times 10^{-6}\,\mathrm{W \cdot K^{-1}}$ in seeking the constants a and b in the parametric form $Q_{\mathrm{S}} = ak_{\mathrm{S}}/(b + k_{\mathrm{S}})$ and considering $G_{\mathrm{Ag}} \gg G_{\mathrm{Pt}}$, Lefèvre et al. obtain

$$\Delta P = \tfrac{3}{2} Q_{\mathrm{S}} \tag{35}$$

and

$$\frac{\Delta P}{P_{\mathrm{C}}} = \frac{V_{\mathrm{PC}}^2 - V_{\mathrm{Pi}}^2}{V_{\mathrm{PC}}^2} = \frac{3}{4} \frac{Gk_{\mathrm{S}}/(G + 2G_{\mathrm{Pt}})}{\dfrac{GG_{\mathrm{Pt}}}{\pi b(G + 2G_{\mathrm{Pt}})} + k_{\mathrm{S}}}\,. \tag{36}$$

Equation (35) shows that ΔP and Q_{S} can be considered proportional. Equation (36) shows the dependence of the measurement on the thermal conductivity k_{S} of the sample, the conductance G of the heat exchange taking place, and the radius b of the tip–sample thermal contact region. Independent of the tip temperature, this relation can be used to calibrate any SThM device provided that the values of G, G_{Pt} and b are accurately known. It can be shown that a precise quantitative characterisation of the thermal conductivity in the DC regime of the constant temperature mode can only be envisaged for poorly conducting materials ($k_{\mathrm{S}} <$ a few tens of $\mathrm{W \cdot m^{-1} \cdot K^{-1}}$) with a lower limit for highly insulating materials. In this last case, the flux released by the sample is very low and this whatever the conductance of the tip–sample heat exchange [94].

AC Regime

The AC regime was developed to improve the sensitivity of the SThM [96] and to be able to control the volume of matter probed by adjusting the modulation frequency of the probe temperature. Various measurement analyses have been carried out in this regime to characterise the local thermal conductivity and diffusivity of materials.

In this regime, measurements are based on the 3ω method developed by *Cahill* et al. in 1987 [97]. This method uses a narrow thermoresistive band (width 100 µm) which simultaneously plays the role of thermometer and modulated excitation source for the sample on which it has been deposited. It has been shown to be a valid way of determining the thermal conductivity of materials.

In the SThM configuration, an alternating current $I_\omega = I_{\omega\mathrm{m}} \cos(\omega t)$ is added to the continuous current in the tip. Modulation frequencies between 100 Hz and several tens of kHz are used. This frequency range is essentially related to the time response of the tip and hence to the finite heat capacity of the sensor and the dimensions of the tip [98, 99]. The signals recorded by synchronous detection are the amplitude and phase of the third harmonic of the alternating voltage then generated across the tip, whose average temperature $\langle T_\mathrm{P} \rangle$ is held constant. Under these conditions, the 2ω component $P_{2\omega}$ of the power dissipated in the probe leads to a modulation of the tip temperature at the same frequency 2ω. This is written

$$\Delta T_{\mathrm{P}2\omega} = T_{\mathrm{P}2\omega\mathrm{m}} \cos(2\omega t)\,, \tag{37}$$

where $T_{\mathrm{P}2\omega\mathrm{m}}$ is the amplitude of the temperature modulation.

From the relation $\Delta R_\mathrm{P} = R_{\mathrm{P}0} \alpha_\mathrm{P} \Delta T_\mathrm{P}$, it can be shown that the amplitude $V_{3\omega\mathrm{m}}$ of the third harmonic of the measured alternating voltage ($V = RI$) has the form

$$V_{3\omega\mathrm{m}} = \frac{I_{\omega\mathrm{m}} R_{\mathrm{P}0} \alpha_\mathrm{P} T_{\mathrm{P}2\omega\mathrm{m}}}{2}\,. \tag{38}$$

Hence $V_{3\omega\mathrm{m}}$ depends only on the variations in the amplitude of the second harmonic of the tip temperature.

Approach to Measurement

In 1999, *Fiege* et al. [99] made a direct application of the approach proposed by *Cahill* et al. [97] to determine the thermal conductivity of samples using the 3ω method to process the SThM measurement. We shall only quote the approximate equation they used. This equation relates the amplitude of the second harmonic of the tip temperature to that of the power per unit length of the tip filament ($p_{2\omega} = P_{2\omega}/2L$):

$$T_{\mathrm{P}2\omega\mathrm{m}} = \frac{P_{2\omega}}{\pi k_\mathrm{S}} (C - \ln \omega)\,, \tag{39}$$

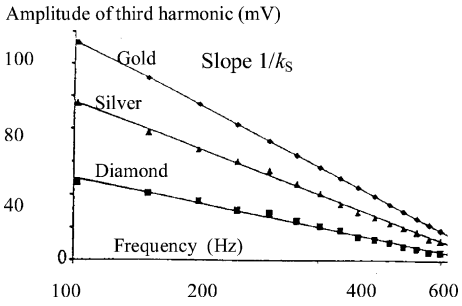

Fig. 32. Amplitude of $V_{3\omega}$ (mV) as a function of f (logarithmic scale). [99] ©Copyright 1999, IOP Publishing Ltd

where C is a term independent of the frequency and a function of the thermal diffusivity of the sample and the effective diameter of the tip–surface thermal contact region.

Figure 32 shows the measurements of the amplitude of the signal obtained by Fiege et al. for samples of known conductivity and various current modulations ω. These curves are linear in the logarithm of the frequency and have a slope inversely proportional to the thermal conductivity of the materials. It can be shown within the framework of the approximations used in this first approach that [99]

$$V_{3\omega}(t) = I_{\omega \mathrm{m}} \alpha_{\mathrm{P}} \frac{P_{2\omega}}{4\pi k_{\mathrm{S}}} (C - \ln \omega) \sin(3\omega t - \phi) \propto \frac{1}{k_{\mathrm{S}}}, \tag{40}$$

where ϕ is the phase difference between this voltage and the current I_{ω}.

Since the constant C can be determined from measurements on reference samples, this approach can be used to calibrate the SThM to carry out thermal conductivity measurements, provided that the samples analysed have similar roughness and elasticity/hardness to the reference samples. Using this method, *Fiege* et al. estimated the thermal conductivity of silver and gold with an error less than 2 % [99].

A quite different approach was developed by *Majumdar* [95] to understand the thermal image contrast in AC regime. He studied the energy balance of the probe in a contact situation. Considering the allowed frequency range of the probe, he noted in particular that, if the Wollaston has much higher AC conductance than the sample, and more precisely, in the case of materials with high thermal diffusivity, the phase will tend to be independent of the properties of the material for high frequencies.

A better understanding of measurements made in the AC regime requires solution of the differential equation expressing the energy balance near the tip. Various groups (LEO in Reims, France [89,100,101] and LET in Poitiers, France [82]) are currently working on this problem.

Some Applications in Constant Temperature Mode

Given the non-ideal size and shape of the tip considered here, quantitative thermophysical characterisation of materials is only currently possible for

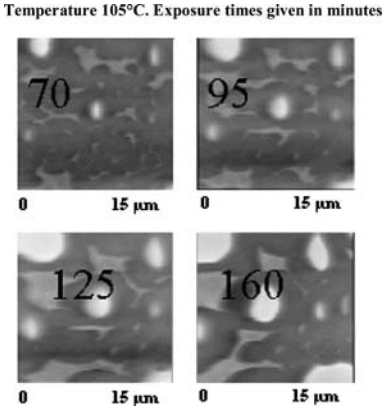

Fig. 33. Thermal images of the same region of a polyvinyl methyl ether/polystyrene blend deposited on glass after various exposure times at a temperature of 105 °C ($T_{\mathrm{op}} = 35°$). ©Copyright 1996 from [107]. Reproduced by permission of Taylor & Francis Group, LLC., http://www.taylorandfrancis.com

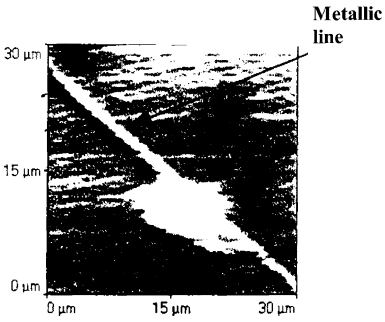

Fig. 34. SThM image of a palladium line 800 nm wide on a glass substrate. Reused with permission from *L. Zhou* et al. [87]. ©Copyright 1997, AVS The Science & Technology Society

plane and mirror-polished surfaces. The SThM is thus used mainly to study nanostructured materials [102, 103] in the form of thin films [90, 93, 104].

In the first work by *Hammiche* et al. [105–107], the SThM is used to specify phase separation processes in polymer blends. For example, spinodal and binodal processes were observed in the polyvinyl methyl ether/polystyrene blend. Figure 33 shows a binodal decomposition. Polystyrene nodules are clearly seen to form and grow in the blend. This work led in particular to a new method for calorimetric analysis with the SThM, known as calorimetric analysis with scanning microscopy (CASM), which should prove useful in the pharmaceutical industry. The probe is then used as a vector for phase separations and changes in volumes of a few μm^3 of matter [100].

Apart from its use for thermal characterisation of materials, the SThM in constant temperature mode has been shown to be a promising tool for writing metallic features of submicron dimensions. *Zhou* et al. [87] have used the heated tip to induce local chemical transformation of a palladium acetate film into metallic palladium. Metallic lines of width 800 nm and thickness 20–35 nm, like the one imaged thermally in Fig. 34, have been written on the surface of a glass substrate by this process.

3.6 Analysing Measurements in Constant Current Mode

The probe is used here as a thermometer. The probe is supplied by a current held constant at a low enough level to avoid heating the tip.

In the DC regime, the signal S_{th} is the differential voltage output by the amplification chain of the thermal unit (see Fig. 22). This voltage is more or less linear in variations of the voltage applied to the tip, and hence also in variations of the average temperature of the tip:

$$S_{\mathrm{th}} \propto A_{\mathrm{amp}} \Big\{ R_{\mathrm{P0}} \big[1 + \alpha_{\mathrm{P}} (T_{\mathrm{P}} - T_{\mathrm{P0}}) \big] + R_{\mathrm{wire}} \Big\} I \,, \tag{41}$$

where I is the current in the probe and A_{amp} is the gain of the amplification chain of the measured differential voltage.

In the AC regime, modulation of the sample temperature at frequency f leads to modulation of the tip temperature at the same frequency. The signals then measured are the amplitude and phase of the alternating component of the voltage across the tip. These signals, in particular the phase, are more sensitive than the DC signal and heavily dependent on the subsurface thermophysical properties of the materials making up the sample [108].

Many groups are currently working to clarify the dependence of the DC and AC signals on the local sample temperature. Since no results have yet been published regarding this work, we shall not discuss the measurement approach for this mode, but simply give some examples of its use.

Some Applications in Constant Current Mode

The main fields of application for this mode are in microelectronics and optoelectronics. It is particularly worth mentioning the work by *Fiege* et al. [78], which demonstrates the ability of the SThM to detect hot spots induced by local degradation within the components of integrated circuits (see Fig. 23). The SThM can reveal highly localised energy dissipation by mapping the surface heating either on the front or the rear side of the thinned substrate of integrated circuits.

Gomès et al. [109] were able to specify where damage had occurred on a MOS structure subjected to strong electrical stress. The DC regime thermal image shown in Fig. 35 reveals highly localised heating with a lateral extent of the order of a few hundred nanometers, not only at the edge of the capacitor (a result already obtained by electrical studies of degradation phenomena), but also in the vicinity of small zones which seem to be the signature of grain boundaries resulting from the fabrication process.

Patino-Lopez et al. [110] used an SThM to characterise the thermal behaviour of PN thermoelectric couples. Figures 36a and b show the amplitude and phase, respectively, of the AC temperature measured along an active PN junction for three modulation frequencies f of the current supplying the junction.

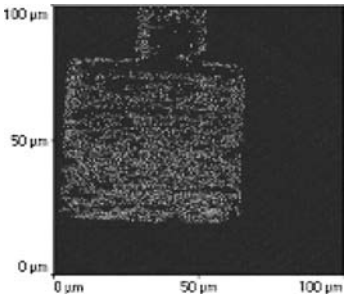

Fig. 35. Thermal image of the gate surface in a MOS structure at breakdown [109]

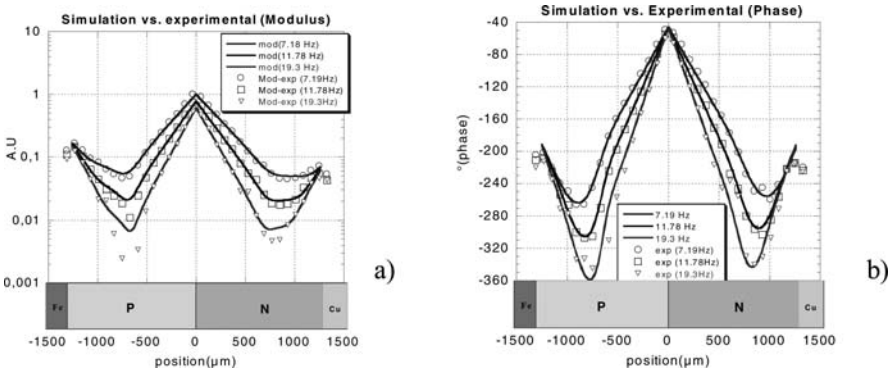

Fig. 36. Temperature profiles. (**a**) Amplitude. (**b**) Phase. Courtesy of *Patino-Lopez* et al. [110]

These profiles show that there are three heat sources within the sample. The largest is located at the PN interface, and the other two at the metal–doped semiconductor interfaces. The PN junction has higher Peltier coefficient than the other junctions and heat is more confined at the PN junction given the low thermal conductivity of the constituents of the structure here. Note also that the thermal diffusivity of the semiconducting materials making up this junction is estimated, successfully in this characterisation, by analysing the variation of the phase of the signal measured at frequency f as a function of the distance between the measurement point and the central source of excitation within the component.

The use of the SThM tip to test the surface temperature of materials is also found in hybrid photothermal setups where attempts are being made to improve the lateral resolution of the probe [111]. These setups will be considered in the Chapter by *Cretin* and *Vairac* in this volume.

3.7 Conclusion

This microscope is now an integral part of the experimental landscape in submicron heat studies. However, much remains to be done to understand the measurements, especially in the constant current mode. The reader is referred to the review [79] for a more complete overview of applications and developments regarding the use of the SThM.

References

[1] W. J. Croft: *Under the Microscope: A Brief History of Microscopy*, Series in Popular Science (World Scientific Publishing Company, Singapore 2004) 184
[2] R. S. Clay, T. H. Court: *The History of the Microscope* (Charles Griffin and Company, London 1932) 184
[3] D. B. Murphy: *Fundamentals of Light Microscopy and Electronic Imaging* (Wiley-Liss 2001) 184
[4] J. W. Goodman: *Introduction to Fourier Optics* (McGraw-Hill, New York 1996) 184
[5] M. Born, E. Wolf: *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge 1999) 184
[6] A. R. Hibbs: *Confocal Microscopy for Biologists* (Springer, Heidelberg 2004) 184
[7] F. S. Crawford: *Waves*, vol. 3, Berkeley Physics Course (1965) 185
[8] L. Reimer: *Scanning Electron Microscopy. Physics of Image Formation and Microanalysis*, 3rd ed. (Springer, Berlin 1993) 185
[9] J. C. H. Spence: *Experimental High-Resolution Electron Microscopy*, 2nd ed. (Oxford University Press, New York and Oxford 1988) 185
[10] B. T. Khuri-Yakub, C. F. Quate: *Selected Papers on Scanning Acoustic Microscopy*, vol. MS 53, SPIE Milestone Series (1992) 185
[11] G. M. Crean, M. Locatelli, J. McGilp: *Acoustic, Thermal Wave and Optical Characterization of Materials*, European Materials Research Society Symposia Proceedings 11 (North-Holland 1990) 185
[12] G. A. Briggs: *Advances in Acoustic Microscopy* (Plenum Press 1995) 185
[13] G. Busse, A. Rosencwaig: Subsurface imaging with photoacoustics, Appl. Phys. Lett. **36**, 815–816 (1980) 185
[14] G. Gaussorgues: *La thermographie infrarouge. Technique et documentation* (Lavoisier 1984) 185
[15] A. Mandelis: *Non-Destructive Evaluation* (Prentice Hall 1994) 185
[16] D. Maillet, J. C. Batsale, A. Bendada, A. Degiovanni: Méthodes intégrales et contrôle non-destructif par thermographie infrarouge stimulée, Revue Générale de Thermique **35**, 14S–27S (1996) 185
[17] S. Grauby, B. C. Forget, S. Holé, D. Fournier: High resolution photothermal imaging of high frequency phenomena using a visible CCD camera associated with a multichannel lock-in scheme, Rev. Sci. Instrum. **70**, 3603–3608 (1999) 185
[18] S. Gomès, N. Trannoy, F. Depasse, P. Grossel: AC scanning thermal microscopy: Tip–sample interaction and buried defect modellings, Int. J. Therm. Sci. Microsc. **39**, 526 (2000) 185

[19] B. Cretin: Recent advances in thermoelastic and acoustic microscopy using a local probe, Condens. Matter News **7**, 13–18 (1999) 185
[20] B. Bonello, B. Perrin, C. Rossignol: Photothermal properties of bulk and layered materials by the picosecond acoustics technique, J. Appl. Phys. **83**, 3081–3088 (1998) 185
[21] G. Binnig, H. Rohrer: Scanning tunneling microscopy, Helvetica Physica Acta **55**, 726–735 (1982) 186
[22] G. Binnig, H. Rohrer, C. Gerber, E. Weibel: Surface studies by scanning tunneling microscopy, Phys. Rev. Lett. **49**, 57–60 (1982) 186
[23] G. Binnig, H. Rohrer: Le microscope à balayage à effet tunnel, Pour la Science (1985) 186
[24] E. Chpolski: *Physique atomique*, 2 volumes (Mir Editions, Moscow 1978) 186, 189
[25] G. Binnig, C. F. Quate, C. Gerber: Atomic force microscopy, Phys. Rev. Lett. **56**, 930–933 (1986) 186
[26] J. K. Zienuk, A. Latusezk: Non-conventional pin scanning ultrasonic microscopy, Acoust. Imag. **17**, 219–224 (1989) 186
[27] K. Takata, T. Hasegawa, S. Hosaka, S. Hosoki, T. Komoda: Tunneling acoustic microscopy, Appl. Phys. Lett. **55**, 1718–1720 (1989) 186
[28] D. Sarid: *Scanning Force Microscopy* (Oxford University Press, Oxford 1994) 191
[29] K. L. Johnson: *Contact Mechanics* (Cambridge University Press, Cambridge 1987) 191
[30] E. A. Ash, G. Nicholls: Super-resolution aperture scanning microscope, Nature **237**, 510–512 (1972) 192
[31] U. C. Fischer, U. T. Dürig, D. W. Pohl: Near-field optical scanning microscopy in reflection, Appl. Phys. Lett. **52**, 249–251 (1988) 192
[32] D. Courjon, K. Sarayeddine, M. Spajer: Scanning tunneling optical microscopy, Opt. Commun. **71**, 23–28 (1989) 192
[33] L. Porte, D. Courjon: Les microscopes en champ proche ou à sonde locale, Spectra 2000, **164**, 40–48 (1992) 192
[34] E. Dieulesaint, D. Royer: *Ondes élastiques dans les solides* (Masson, Paris 1974) 193
[35] P. Günther, V. C. Fischer, K. Dransfeld: Scanning near-field acoustic microscopy, Appl. Phys. B **48**, 89–92 (1989) 193
[36] U. Rabe, K. Janser, W. Arnold: Vibrations of free and surface-coupled atomic force microscope cantilevers: Theory and experiment, Rev. Sci. Instrum. **67**, 3281–3293 (1996) 193
[37] P. E. Mazeran, J. L. Loubet: Force modulation with a scanning force microscope: An analysis, Trib. Lett. **3**, 125 (1996) 193
[38] F. Oulevey, G. Gremaud, A. Sémoroz, A. J. Kulik, N. A. Burnham, E. Dupas, D. Gourdon: Local mechanical spectroscopy with nanometer scale lateral resolution, Rev. Sci. Instrum. **69**, 2085–2094 (1998) 193
[39] U. Rabe, E. Kester, W. Arnold: Probing linear and nonlinear tip–sample interaction forces by atomic force acoustic microscopy, Surf. Interf. Anal. **27**, 386–391 (1999) 193
[40] O. Kolosov, K. Yamanaka: Nonlinear detection of ultrasonic vibrations in an atomic force microscope, Jpn. J. Appl. Phys. Lett. **32**, 22–25 (1993) 193

[41] P. Vairac, B. Cretin: Scanning microdeformation microscopy in reflection mode, Appl. Phys. Lett. **68**, 461–463 (1996) 193
[42] P. Maivald, H. J. Butt, S. A. C. Gould, C. B. Prater, B. Drake, J. A. Gurley, V. B. Elings, P. K. Hansma: Using force modulation to image surface elasticities with the atomic force microscope, Nanotechnol. **2**, 103–106 (1991) 194
[43] L. Robert, B. Cretin: Experimental determination of the observation depth in scanning microdeformation microscopy, Surf. Interf. Anal. **27**, 568–571 (1999) 194
[44] C. C. Williams, K. Wickramasinghe: Scanning thermal profiler, Appl. Phys. Lett. **49–23**, 1587–1589 (1986) 195
[45] C. C. Williams, H. K. Wickramasinghe: *Proceedings of the 5th International Meeting on Photoacoustic and Photothermal Phenomema* (Heidelberg 1987) pp. 364–369 196
[46] C. C. Williams, K. Wickramasinghe: Thermal and photothermal imaging on a sub-100 nanometer scale, SPIE, Scanning Microscopy Technologies and Applications **897**, 129–134 (1988) 196
[47] J. M. R. Weaver, L. M. Walpita, H. K. Wickramasinghe: Optical absorption microscopy and spectroscopy with nanometer resolution, Nature Lett. **342**, 783–785 (1989) 197
[48] C. C. Williams, K. Wickramasinghe: Microscopy of chemical-potential variations on an atomic scale, Nature **344**, 317–319 (1990) 197
[49] M. Stopka, L. Hadjiiski, E. Oesterschulze, R. Kassing: Surface investigations by scanning thermal microscope, J. Vac. Sci. Technol. B **13**, 2153–2156 (1995) 197
[50] E. Oesterschulze, M. Stopka, R. Kassing: Photothermal characterization of solids and thin films by optical and scanning probe techniques, Microelectron. Eng. **24**, 107–112 (1994) 197
[51] E. Oesterschulze, M. Stopka, L. Ackermann, W. Scholz, S. Werner: Thermal imaging of thin films by scanning thermal microscope, J. Vac. Sci. Technol. B **14**, 832–837 (1996) 197
[52] E. Oesterschulze, M. Stopka: Photothermal imaging by scanning thermal microscopy, J. Vac. Sci. Technol. A **14**, 1172–1177 (1996) 197
[53] N. Trannoy: *Contribution à l'étude d'effets thermiques liés à l'excitation laser dans un microscope à effet tunnel: Étude de l'échantillon et de la pointe, rôle du couplage pointe-échantillon*, Ph.D. thesis, University of Reims, France (1997) 198, 199
[54] N. Trannoy, P. Grossel, M. Troyon: Thermal effects induced by laser interaction in scanning tunneling microscopy: Thermal expansion of sample and tip influence, Probe Microsc. **1**, 201–206 (1998) 198
[55] N. Trannoy, P. Grossel: Thermal multilayer with recursive method, in *Proceedings of the EUROTHERM Congress*, Seminar 53: Advanced Concepts and Techniques in Thermal Modelling (Elsevier Science, Mons 1997) 199
[56] P. Grossel, F. Depasse, N. Trannoy: Sample–tip thermal coupling in modulated laser surface excitation, Int. J. Eng. Sci. **35**, 699–709 (1997) 199
[57] D. H. Hoffmann, A. Rettenberger, J. Y. Grand, K. Laüger, P. Leiderer, K. Dransfeld, R. Möller: Thermovoltages in vacuum tunneling investigated by scanning tunneling microscopy, Thin Solid Films **264**, 223–225 (1995) 199

[58] M. J. Hagmann: Simulations of rectification in a laser-illuminated scanning tunneling microscope, Appl. Surf. Sci. **87/88**, 368–372 (1995) 199
[59] M. J. Hagmann: Microwave tunneling current from the resonant interaction of an amplitude modulated laser with a scanning tunneling microscope, J. Vac. Sci. Technol. B **14**, 838–841 (1996) 199
[60] M. Nonnenmacher, K. Wickramasinghe: Scanning probe microscopy of thermal conductivity and subsurface properties, Appl. Phys. Lett. **61**, 168–170 (1992) 199
[61] A. Majumdar, J. P. Carrejo, J. Lai: Thermal imaging using the atomic force microscope, Appl. Phys. Lett. **62**, 2501–2503 (1993) 200
[62] A. Majumdar, M. Chandrachood, J. Lai, O. Nakabeppu, Y. Wu, Z. Shu: Thermal imaging by atomic force microscopy using thermocouple cantilever probes, Rev. Sci. Instrum. **66**, 3584–3592 (1995) 201, 203
[63] R. Pylkki, P. J. Moyer, P. E. West: Scanning near-field optical microscopy and scanning thermal microscopy, J. Appl. Phys. **33**, 3785–3790 (1994) 201, 205, 208
[64] R. B. Dinwiddie, R. J. Pylkki, P. E. West: Thermal conductivity contrast imaging with a scanning thermal microscope, in T. W. Tong (Ed.): *Thermal Conductivity*, vol. 22 (Technomic Publishing, Lancaster 1994) pp. 668–677 201, 205, 208
[65] K. Luo, Z. Shi, J. Lai, A. Majumdar: Nanofabrication of sensors on cantilever probe tips for scanning multiprobe microscopy, Appl. Phys. Lett. **68**, 325–327 (1996) 204
[66] K. Luo, Z. Shi, J. Varesi, A. Majumdar: Sensor nanofabrication, performance and conduction mechanisms in scanning thermal microscopy, J. Vac. Sci. Technol. B **15**, 349–360 (1997) 204
[67] Y. Suzuki: Novel microcantilever for scanning thermal imaging microscopy, Jpn. J. Appl. Phys. **35**, L352–L354 (1996) 204
[68] G. Mills, H. Zhou, A. Midha, L. Donaldson, J. M. R. Weaver: Scanning thermal microscopy using batch fabricated thermocouple probes, Appl. Phys. Lett. **72**, 2900–2902 (1998) 204, 205
[69] O. Nakabeppu, M. Chandrachood, Y. Wu, J. Lai, A. Majumdar: Scanning thermal imaging microscopy by using composite cantilever probes, Appl. Phys. Lett. **66**, 694–696 (1995) 205
[70] M. Igeta, T. Inoue, J. Varesi, A. Majumdar: Thermal expansion and temperature measurement in a microscopic scale by using the atomic force microscope, JSME Int. J. B **42** (1999) 205
[71] M. Lederman, D. Richardson, H. C. Tong: Thermal microscopy of spin-valve and magnetoresistive devices, IEEE Trans. Mag. **33**, 2923–2925 (1997) 206
[72] G. Mills, J. M. R. Weaver, G. Harris, W. Chen, J. Carrejo, L. Johnson, B. Rogers: Detection of subsurface voids using scanning thermal microscopy, Ultramicrosc. **80**, 7–11 (1999) 206
[73] I. W. Rangelow, T. Gotszalk, N. Abedinov, P. Grabiec, K. Edinger: Thermal nano-probe, Microelectron. Eng. **57–58**, 737–748 (2001) 206, 207
[74] T. Gotszalk, P. Grabiec, I. W. Rangelow: Calibration and examination of piezoresistive Wheatstone bridge cantilevers for scanning probe microscopy, Ultramicrosc. **97**, 385–389 (2003) 207

[75] T. Ivanov, T. Gotszalk, P. Grabiec, E. Tomerov, I. W. Rangelow: Thermally driven micromechanical beam with piezoresistive deflection readout, Microelectron. Eng. **67–68**, 550–556 (2003) 207

[76] R. B. Dinwiddie, R. J. Pylkki, P. E. West: Thermal conductivity contrast imaging with a scanning thermal microscope, in T. W. Tong (Ed.): *Thermal Conductivity*, vol. 22 (Technomic Publishing, Lancaster 1994) pp. 668–677 208

[77] S. Gomès: *Contribution théorique et expérimentale à la microscopie thermique à sonde locale: Calibration d'une pointe thermorésistive, analyse des divers couplages thermiques*, Ph.D. thesis, University of Reims, France (1999) 211, 212, 213, 219, 222

[78] G. B. M. Fiege, W. Schade, M. Palaniappan, V. Ng, J. C. H. Phang, L. J. Balk: Front- and backside investigations of thermal and electronic properties of semiconducting devices, Microelectron. Reliab. **39**, 937–940 (1999) 211, 212, 213, 228

[79] H. M. Pollock, A. Hammiche: Micro-thermal analysis: Technique and applications, J. Phys. D: Appl. Phys. **32**, L13–L17 (2001) 211, 230

[80] S. Gomès, N. Trannoy, P. Grossel, F. Depasse, C. Bainier, D. Charraut: D.C. scanning thermal microscopy: Characterisation and interpretation of the measurement, Int. J. Therm. Sci. **40**, 949–958 (2001) 212, 213, 214, 216, 221, 222, 223

[81] K. Luo, Z. Shi, J. Varesi, A. Majumdar: Sensor nanofabrication, performance and conduction mechanisms in scanning thermal microscopy, J. Vac. Sci. Technol. B 15-2, **1**, 349–360 (1997) 213, 215

[82] S. Lefèvre: *Modélisation et élaboration des métrologies de microscopie thermique à sonde locale résistive*, Ph.D. thesis, University of Poitiers, France (2004) 214, 215, 216, 226

[83] S. Gomès, N. Trannoy, P. Grossel: D.C. thermal microscopy: Study of the thermal exchange between a probe and a sample, Meas. Sci. Technol. **10**, 805–811 (1999) 215, 216

[84] L. Shi, A. Majumdar: Thermal transport mechanism at nanoscale point contacts, J. Heat Transfer **124**, 329–337 (2002) 215, 216

[85] A. Hammiche, H. M. Pollock, M. Song, S. Yoshimura: Subsurface imaging by scanning thermal microscopy, Meas. Sci. Technol. **7**, 142–150 (1997) 217

[86] M. Maywald, R. J. Pylkki, L. J. Balk: Imaging of local thermal and electrical conductivity with scanning force microscopy, Scanning Microsc. **8**, 181–188 (1994) 217

[87] L. Zhou, G. Q. Xu, H. T. Ng, S. F. Y. Li: Scanning thermal microscope tip-induced chemical reaction on solid organometallic compound thin films, J. Vac. Sci. Technol. B **15**, 1871 (1997) 217, 227

[88] L. J. Balk, M. Maywald, R. J. Pylkki: Nanoscopic detection of the thermal conductivity of compound semiconductor materials by enhanced thermal microscopy, Inst. Phys. Conf. Ser. **146**, 655–658 (1995) 217

[89] F. Depasse, P. Grossel, S. Gomès: Theory of DC and AC heat diffusion for submicro- and nanoscopies, J. Phys. D: Appl. Phys. **36**, 204–210 (2003) 217, 226

[90] F. Ruiz, W. D. Sun, F. H. Pollak, C. Venkatraman: Determination of the thermal conductivity of diamond-like nanocomposite films using a scanning thermal microscope, Appl. Phys. Lett. **73**, 1802–1804 (1998) 218, 221, 227

[91] J. P. Bardon: La mesure des températures de surface par contact. Erreurs liées aux transferts de chaleur parasites, Rev. Gen. Therm. **170**, 123–135 (1976) 220
[92] V. V. Gorbunov, N. Fuchigami, J. L. Hazel, V. V. Tsukruk: Probing surface microthermal properties by scanning thermal microscopy, Langmuir **15**, 8340–8343 (1999) 221
[93] S. Callard, G. Tallarida, A. Borghesi, L. Zanotti: Thermal conductivity of SiO_2 films by scanning thermal microscopy, J. Non-Cryst. Solids **245**, 203–209 (1999) 221, 222, 227
[94] S. Lefèvre, S. Volz, J. B. Saulnier, C. Fuentes, N. Trannoy: Thermal conductivity calibration for hot wire based DC scanning thermal microscope, Rev. Sci. Instrum. **74**, 2418–2423 (2003) 221, 222, 223, 224
[95] A. Majumdar: Scanning thermal microscopy, Annu. Rev. Sci. **29**, 505–585 (1999) 222, 226
[96] M. Maywald, R. J. Pylkki, F. J. Reineke, L. J. Balk: Modulated thermal profiling on devices, Inst. Phys. Conf. Ser. **146**, 655–658 (1995) 225
[97] D. G. Cahill: Thermal conductivity measurement from 30 to 750 K: The 3ω method, Rev. Sci. Instrum. **61**, 802–806 (1990) 225
[98] S. Lefèvre, J. B. Saulnier, C. Fuentes et, S. Volz: J. Superlat. Microstruct. **35**, 283–288 (2004) 225
[99] G. B. M. Fiege, A. Altes, R. Heiderhoff, L. J. Balk: Quantitative thermal conductivity with nanometre resolution, J. Phys. D: Appl. Phys. **32**, L13–L17 (1999) 225, 226
[100] S. Gomès, F. Depasse, N. Trannoy, P. Grossel: A.C. scanning thermal microscopy: Tip–sample interaction and buried defect modelling, Int J. Therm. Sci. **39**, 526–531 (2000) 226, 227
[101] F. Depasse, P. Grossel, N. Trannoy: Probe temperature and output voltage calculation for the SThM in AC mode, Superlat. Microstruct. **35**, 269–282 (2004) 226
[102] F. A. Guo, N. Trannoy, J. Lu: Analysis of thermal properties by scanning thermal microscopy in nanocrystallized iron surface induced by ultrasonic shot peening, Mater. Sci. Eng. A **369**, 36–42 (2004) 227
[103] F. A. Guo, K. Y. Zhu, N. Trannoy, J. Lu: Examination of thermal properties by scanning thermal microscopy in ultrafine-grained pure titanium surface layer induced by ultrasonic shot peening, Thermochimica Acta **419**, 239–246 (2004) 227
[104] V. M. Asnin, F. H. Pollack, J. Ramer, M. Schurman, I. Ferguson: High spatial resolution thermal conductivity of lateral epitaxial overgrown GaN/sapphire (0001) using a scanning thermal microscope, Appl. Phys. Lett. **75**, 1240–1242 (1999) 227
[105] A. Hammiche, D. J. Hourston, H. M. Pollock, M. Reading, M. Song: Scanning thermal microscopy: Subsurface imaging, thermal mapping of polymer blends and localized calorimetry, J. Vac. Sci. Technol. B **14**, 1486–1491 (1996) 227
[106] A. Hammiche, M. Reading, H. M. Pollock, M. Song, D. J. Hourston: Localized thermal analysis using a miniaturized resistive probe, Rev. Sci. Intrum. **67**, 4268–4274 (1996) 227

[107] H. M. Pollock, A. Hammiche, M. Song, D. J. Hourston, M. Reading: Interfaces in polymeric systems as studied by CASM. A new combination of localised calorimetric analysis with scanning microscopy, J. Adhesion **67**, 217–234 (1996) 227
[108] J. Altet, S. Dilhaire, S. Volz, J.-M. Rampnoux, A. Rubio, S. Grauby, L. D. P. Lopez, W. Claeys, J. B. Saulnier: Four different approaches of IC surface temperature: Application to thermal testing, Microelectron. J. **33**, 689–696 (2002) 228
[109] S. Gomès, D. Ziane: Investigation of the electrical degradation of a metal-oxide-silicon capacitor by scanning thermal microscopy, Solid State Electron. **47**, 919–922 (2003) 228, 229
[110] L. D. Patino-Lopez, S. Grauby, S. Dilhaire, M. A. Salhi, W. Clayes, S. Lefèvre, S. Volz: Characterization of the thermal behaviour of PN thermoelectric couples by scanning thermal microscope, Microelectronics **I.35**, 797–803 (2004) 228, 229
[111] A. Hammiche, H. M. Pollock, M. Reading, M. Claybourn, P. H. Turner, K. Jewkes: Photothermal FT-IR spectroscopy: A step towards FT-IR microscopy at a resolution better than the diffraction limit, Appl. Spectrosc. **53**, 810–815 (1999) 229

Index

Optical Techniques for Local Measurement

Stefan Dilhaire[1], Danièle Fournier[2], and Gilles Tessier[3]

[1] Centre de Physique Moléculaire Optique et Hertzienne, University of Bordeaux
s.dilhaire@cpmoh.u-bordeaux1.fr
[2] Laboratoire d'Optique Physique, Ecole Supérieure de Physique et de Chimie Industrielles de Paris, University of Paris VI
fournier@optique.espci.fr
[3] Laboratoire d'Optique Physique, Ecole Supérieure de Physique et de Chimie Industrielles de Paris
tessier@optique.espci.fr

Abstract. Optical measurement techniques are well suited to many heat transfer problems, insofar as they are non-contact and generally non-invasive. Far-field measurements of infrared thermal emission were for a long time the predominant method. But today these techniques have reached their limits: the spatial resolution is not adequate for micro and nanoheat transfer analysis.

In this Chapter, we shall review the main optical techniques devised recently to overcome these limitations. A great many of these techniques operate in a modulated regime, taking advantage of the excellent signal-to-noise ratio provided by lock-in detection, but also exploiting the spatial confinement of the modulated part of the heat obtained in the alternating regime.

1 Generating Thermal and Thermoelastic Waves

Among all phenomena able to create thermal excitations in a material, it is mainly thermoelectric and optical excitations that are used experimentally. In an electrically functionalised solid, it is thus possible to generate a thermal excitation by passing an electric current through the sample. This excitation can be used as a probe to measure the thermal properties of nearby materials. However, it sometimes constitutes the subject of the investigation itself, particularly in the case of integrated circuits within which it constitutes a potential source of damage that must be characterised.

In materials or systems that are not suitably functionalised for electrical excitation, it is clear that thermal waves cannot be created by the Joule effect. When the material is optically absorbent, an amplitude-modulated light wave, often called a pump, can be used to generate a thermal wave. These so-called photothermal methods have been used with some success since the 1970s, since they are able to probe absolutely passive samples without any contact.

In the next section, we shall successively examine these phenomena and establish their contribution to the source term in the heat equation.

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 239–286 (2007)

1.1 Generating Waves by Thermoelectric Effects

Joule, Peltier, and Thomson Effects

In addition to the energy taken from the electric field by the electron gas and the energy removed by heat conduction, there is a further release or absorption of heat energy, characteristic of the Peltier and Thomson effects. Under these conditions the heat equation becomes [1]

$$k\nabla^2(\Delta T) - \rho C\frac{\mathrm{d}\Delta T}{\mathrm{d}t} = -\frac{\boldsymbol{j}^2}{\sigma} + T\boldsymbol{j}(\Delta S)_T + T\boldsymbol{j}\frac{\partial S}{\partial T}(\nabla T)_x\,, \tag{1}$$

where the first term on the left-hand side is the energy removed by heat conduction and the second is the change in the internal energy, i.e., the energy stored in the form of heat. On the right-hand side, the first term, always negative, represents the energy borrowed from the electric field by electrons. This energy, once given up to the lattice by electron collisions with impurities, produces the Joule effect. It is always present and its sign is independent of the current direction. The heat source thereby produced is delocalised. The second term on the right-hand side is nothing other than the heat exchanged by the Peltier effect between the lattice and the electron gas, with S the thermoelectric power of the material. It is generated by a discontinuity or inhomogeneity in the medium. Its sign depends on the current direction. The Peltier heat source is localised in space, e.g., at the interface of two materials. The third term on the right-hand side of (1) arises from the coupling between the electric current and the temperature gradient it induces. This is the Thomson effect. It behaves as a delocalised source whose sign depends on both the current direction and the direction of heat flow (the temperature gradient).

Note. The Peltier and Thomson effects, whose signs depend on the current direction, are thermodynamically irreversible even though the sign of the heat source term changes when the direction of the current is reversed.

All these terms can be identified experimentally by the type of nonlinearity they produce. Indeed, the Joule effect produces a signal at twice the excitation frequency. The Peltier effect is linear in the current and the signal it produces is at the same frequency as the current. On the other hand, the expression for the Thomson effect shows us that this type of nonlinearity generates a great many harmonics which cannot necessarily be detected owing to the weak dependence of the thermoelectric power on the temperature.

Temperature Fields and Displacement Fields

The frequency behaviour or temporal dynamics of a thermal system are tightly linked to the boundary conditions or the mode of excitation (heat

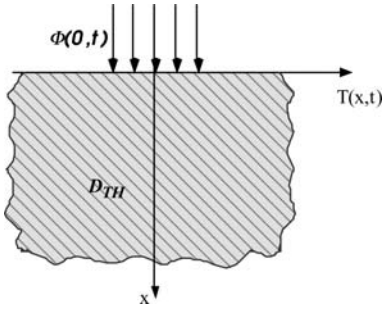

Fig. 1. 1D model showing boundary conditions

or temperature source). Indeed, the size and shape of a system will largely determine its temporal or frequency response. Waves will diffract on obstacles or interfere with one another [2–4].

1D Semi-Infinite System

The system is illustrated in Fig. 1. A heat flux with uniform distribution is injected at the interface and diffuses into the semi-infinite medium. Temporal responses are given by [5]

$$\begin{aligned}\Re\Delta\big[T(x,t)\big] &= \frac{\phi_0}{k}\delta_{\mathrm{TH}}\mathrm{e}^{-x/\delta_{\mathrm{TH}}}\cos\left(\omega t - \frac{x}{\delta_{\mathrm{TH}}} - \frac{\pi}{4}\right)\\ &= \frac{\phi_0}{k}\sqrt{\frac{2\alpha}{\omega}}\exp\left(-x\sqrt{\frac{\omega}{2\alpha}}\right)\cos\left(\omega t - x\sqrt{\frac{\omega}{2\alpha}} - \frac{\pi}{4}\right), \qquad (2)\\ \Re\big[u(x,t)\big] &= -\frac{\phi_0}{k}\delta_{\mathrm{TH}}^2\mathrm{e}^{-x/\delta_{\mathrm{TH}}}\cos\left(\omega t - \frac{x}{\delta_{\mathrm{TH}}} - \frac{\pi}{2}\right)\\ &= -\frac{\phi_0}{k}\frac{2\alpha}{\omega}\exp\left(-x\sqrt{\frac{\omega}{2\alpha}}\right)\cos\left(\omega t - x\sqrt{\frac{\omega}{2\alpha}} - \frac{\pi}{2}\right). \qquad (3)\end{aligned}$$

The angular frequency ω and phase reference are those of the heat flux imposed at the surface, viz.,

$$\phi(0,t) = \phi_0\cos(\omega t)\,. \qquad (4)$$

Diffusive Propagation

This propagation occurs in the positive x direction with speed $v(\omega)$ given by the phase of the wave:

$$\omega\left(t - \frac{x}{v}\right) = \omega t - x\sqrt{\frac{\omega}{2\alpha}}\,, \quad \text{hence} \quad v(\omega) = \sqrt{2\alpha\omega}\,. \qquad (5)$$

This speed is seriously affected by dispersion since it depends on the frequency. We thus note the following points:

- Low frequency signals propagate very slowly. On the other hand, high frequency signals decay rapidly but propagate at very high speeds.
- The attenuation depends on the depth and the frequency.
- There is a phase difference of $\pi/4$ between the exciting flux (4) and the temperature response at $x = 0$ given by (2). This reflects the existence of a skin effect, perfectly analogous to the electromagnetic effect occurring in conducting materials.
- The spectral composition of a signal varies as a function of the position x. Figure 2 shows the temperature response to a heat pulse at different depths in the material. These results were simulated using the model described by *Carslaw* and *Jaeger* [5].

Equations (2) and (3) show the rather special behaviour of thermal waves near a surface. Indeed, the temperature and displacement of the surface can be written in the form

$$\Delta T(0,t) = \frac{\phi_0}{k}\delta_{\mathrm{TH}}\cos\left(\omega t - \frac{\pi}{4}\right) = \frac{\phi_0}{k}\sqrt{\frac{2\alpha}{\omega}}\cos\left(\omega t - \frac{\pi}{4}\right), \tag{6}$$

$$u(0,t) = -\frac{\phi_0}{k}\delta_{\mathrm{TH}}^2\cos\left(\omega t - \frac{\pi}{2}\right) = -\frac{\phi_0}{k}\frac{2\alpha}{\omega}\cos\left(\omega t - \frac{\pi}{2}\right). \tag{7}$$

Note that the surface temperature behaves to fractional order in $(j\omega)^{-1/2}$ (in the sense of the Fourier transform), whereas the expansion obeys an integer power law in $(j\omega)^{-1}$. This behaviour arises naturally in the step temperature and expansion responses. It can be shown that the step response of the temperature to a constant heat flux ϕ_0 is given by [5]

$$\Delta T(x,t) = \frac{2\phi_0}{k}\left[\left(\frac{\alpha t}{\pi}\right)^{1/2}\mathrm{e}^{-x^2/4\alpha t} - \frac{x}{2}\mathrm{erfc}\left(\frac{x}{2\sqrt{\alpha t}}\right)\right], \tag{8}$$

and the temperature at $x = 0$ is given by

$$\Delta T(x,t) = \frac{2\phi_0}{k}\left(\frac{\alpha t}{\pi}\right)^{1/2}. \tag{9}$$

This step response is typical of a fractional system. Indeed, the step response (8) has Laplace transform

$$L\left[\Delta T(x,t)\right] = \sqrt{\alpha}\frac{\phi_0}{k}\frac{\mathrm{e}^{-x\sqrt{p/\alpha}}}{p^{3/2}}, \tag{10}$$

where p is the Laplace variable. The temperature transfer function for the system can then be deduced:

$$H_T(p,x) = \frac{\sqrt{\alpha}}{k}\frac{\mathrm{e}^{-x\sqrt{p/\alpha}}}{p^{1/2}}. \tag{11}$$

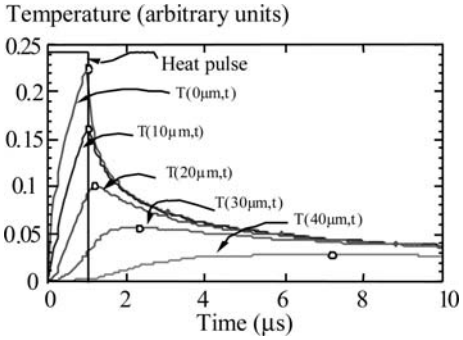

Fig. 2. Spread of a heat pulse at different stages in its propagation

Likewise, the transfer function for the surface displacement is

$$H_u(p,x) = \eta \int_x^\infty H_T(p,\xi)\,\mathrm{d}\xi = -\eta\frac{\alpha}{k}\frac{\mathrm{e}^{-x\sqrt{p/\alpha}}}{p}\,. \tag{12}$$

Only the motion of the surface $x = 0$ is relevant here and we retrieve the integer behaviour of (7), viz.,

$$H_u(p,0) = -\eta\frac{\alpha}{k}\frac{1}{p}\,. \tag{13}$$

Thermoelastic Waves

The temperature field produced in a system subject to thermal excitation can be calculated using analytic or numerical methods of varying degrees of complexity. The thermoelastic problem is much more difficult to formulate since it involves the coupling between heat conduction and elastic wave propagation. It cannot be solved by a simple expansion calculation, which remains too rough an approximation in most cases. The behaviour of the surface motion produced by a thermal excitation can be modelled. The model chosen is depicted in Fig. 3. It represents an isotropic semi-infinite medium. The surface is assumed to be stress free and heat is propagated purely by linear conduction in the semi-infinite medium. We are concerned with the surface displacement. This displacement is due to the contribution of (transverse and longitudinal) elastic terms, together with a contribution of thermal origin.

Calculation shows that, superposed on a Rayleigh wave, there is a surface thermal wave originating from the coupling between thermal and elastic effects. This wave is interpreted in terms of plane evanescent waves. In silicon, the two waves, acoustic and thermal, propagate at speeds 3217 m/s and 1.46 m/s, respectively, for a frequency of 2 kHz. The motion of particles in the medium is illustrated in Figs. 3 and 4. Particles at the interface describe

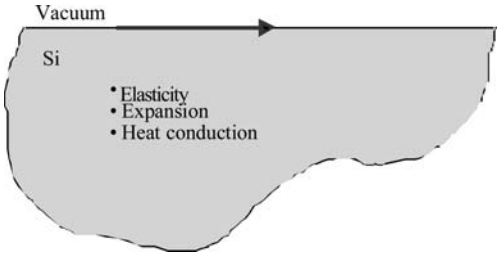

Fig. 3. Thermoelastic model

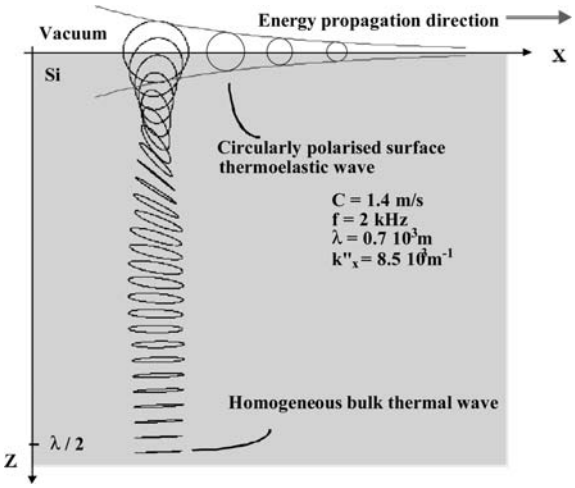

Fig. 4. Representation of thermoelastic waves

circles with radii that fall off exponentially as they propagate. This circular polarisation becomes linear for particles located a half wavelength under the surface. This confers a rather unusual surface aspect on the wave.

1.2 Optical Generation

When the material is optically absorbent, an amplitude-modulated light wave can be used to generate a thermal wave. (These so-called photothermal methods are able to probe absolutely passive samples without any contact.)

Among the non-stationary photothermal methods, we shall distinguish those using modulation, which appeal to the true notion of thermal wave, from the pulsed methods. The latter include in particular the flash method [6], the converging thermal wave method [7, 8], and the transient thermal grating [9], which we shall not discuss in any detail here.

Optical Generation of a Thermal Wave

Modulated photothermal methods are based on work by Bell in the nineteenth century, taken up again by Rosencwaig and Gersho in the 1970s. In the

simplest case, an amplitude-modulated plane wave with angular frequency ω produces a thermal wave of the form $\Delta T(z) = T(z_0)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{e}^{-\sigma(z-z_0)}$ if it is fully absorbed at z_0 (opaque sample), with a propagation constant $\sigma = \sqrt{\mathrm{i}\omega/\alpha}$.

However, this 1D formalism is only rarely used in practice, whenever the heating is obtained by a focused laser beam (confined source with Gaussian distribution). Let us consider the simple case of a semi-infinite medium with thermal conductivity α and heat density ρC, at the surface of which is focused an intensity-modulated laser beam in TEM_{00} mode. The transverse energy distribution in the beam is Gaussian. Moreover, the light energy of the laser is partially absorbed in the bulk according to the Beer–Lambert law and then converted into heat. This heat source can be described by

$$S(r,z) = \frac{1}{2}p\frac{\beta(1-R)}{\pi r_{\mathrm{g}}^2}\mathrm{e}^{-r^2/r_{\mathrm{g}}^2}\mathrm{e}^{-\beta z}(1+\mathrm{e}^{\mathrm{i}\omega t})\,, \tag{14}$$

where P is the incident power of the pump beam, R and β are the coefficients of optical reflection and absorption of the sample at the wavelength of the pump beam, r_{g} is the Gaussian radius at 1/e of the pump beam, ω is the angular frequency of modulation, and r and z are the polar coordinates defining the position in the medium, with $-\boldsymbol{z}$ the outward normal to the sample surface. The aim is to determine the temperature field, which must satisfy the heat conservation equation, viz.,

$$\frac{\partial(\rho C T)}{\partial t} = -\mathrm{div}\boldsymbol{q} + S\,.$$

In a homogeneous and isotropic medium, we assume that the heat flux $\boldsymbol{q}$ is related to the thermal gradient by the Fourier law $\boldsymbol{q} = -k\boldsymbol{\nabla}T$. Moreover, there are boundary conditions at $z = 0$. If we assume that heat losses by conduction, convection, or radiation are negligible, these can be written

$$-\boldsymbol{q}\cdot\boldsymbol{z} = k\frac{\partial T}{\partial z} = 0\,. \tag{15}$$

At $z = +\infty$, the temperature T of the sample is equal to the ambient temperature T_{a}, i.e.,

$$\lim_{z\to+\infty} T = T_{\mathrm{a}}\,. \tag{16}$$

When periodic boundary conditions are imposed on the medium (harmonic flux or bulk heat source), the problem can be solved by decomposing the temperature into three terms, viz.,

$$T(\boldsymbol{M}) = T_{\mathrm{a}} + T_{\mathrm{c}}(\boldsymbol{M}) + \Delta T(\boldsymbol{M})\mathrm{e}^{\mathrm{i}\omega t}\,, \tag{17}$$

where T_{c} is the continuous component of the local temperature rise, ΔT is the complex amplitude of local periodic temperature variations, and $\omega = 2\pi f$

with f the modulation frequency of the periodic heat source. As the problem is linear, ΔT and T_c can be determined independently.

In the periodic regime, the heat conservation equation has the form of a wave equation, viz.,

$$\nabla^2(\Delta T) - \mathrm{i}\frac{\omega}{\alpha}\Delta T = -\frac{\Delta s}{k}\,, \tag{18}$$

where Δs is the periodic component of the heat source given by

$$\Delta s = \frac{1}{2}P\frac{\beta(1-R)}{\pi r_g^2}\exp\left(-\frac{r^2}{r_g^2}\right)\exp(-\beta z)\,. \tag{19}$$

Note that the characteristic wave number of the last equation is complex, reflecting the irreversibility of the heat diffusion phenomenon as well as the evanescent nature of thermal waves.

The problem has cylindrical symmetry. Writing the equation in the system of polar coordinates specified above, we obtain

$$\frac{\partial^2\Delta T}{\partial z^2} + \frac{\partial^2\Delta T}{\partial r^2} + \frac{1}{r}\frac{\partial\Delta T}{\partial r} - \mathrm{i}\frac{\omega}{\alpha}\Delta T = -\frac{\Delta s}{k}\,. \tag{20}$$

Integral transformations can be used to simplify the solution of these equations. The Hankel transformation (of order 0 here) is well suited to this type of problem with cylindrical geometry. The last equation transforms to

$$\frac{\partial^2\Delta\tilde{T}(u,z)}{\partial z^2} - \left(4\pi^2u^2 + \mathrm{i}\frac{\omega}{\alpha}\right)\Delta\tilde{T} = -\frac{\Delta\tilde{s}}{k}\,, \tag{21}$$

where $\Delta\tilde{T}(u,z)$ and $\Delta\tilde{s}(u,z)$ are the zero order Hankel transforms of the functions $\Delta T(u,z)$ and $\Delta s(u,z)$, defined by

$$\Delta\tilde{T}(u,z) = 2\pi\int_0^{+\infty}\Delta T(r,z)J_0(2\pi ru)r\,\mathrm{d}r\,, \tag{22}$$

and

$$\Delta\tilde{s}(u,z) = 2\pi\int_0^{+\infty}\Delta s(r,z)J_0(2\pi ru)r\,\mathrm{d}r\,, \tag{23}$$

with J_0 the order 0 Bessel function of the first kind and u the variable conjugate to r in the Hankel space. Equations (15) and (16) then have the following simple expressions in the Hankel space:

$$\frac{\partial\Delta\tilde{T}}{\partial z} = 0\,, \quad \lim_{z\to+\infty}\Delta\tilde{T} = 0\,. \tag{24}$$

Formulated in this way, it is easy to solve the second order differential equation (21). Taking into account the conditions (24), the solution can now be written

$$\Delta\tilde{T}(u,z) = \frac{\beta(1-R)P}{2k(\beta^2-q_{\mathrm{th}}^2)}\exp\left[-(\pi r_g u)^2\right]\left[\frac{\beta}{q_{\mathrm{th}}}\exp(-q_{\mathrm{th}}z) - \exp(-\beta z)\right]\,,$$

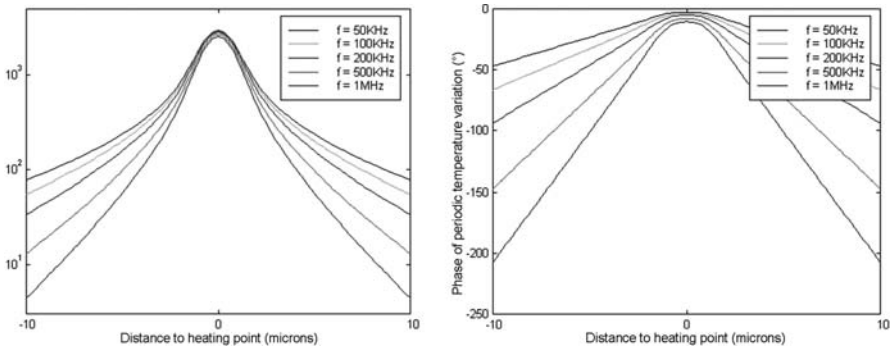

Fig. 5. Amplitude (*left*) and phase (*right*) of the periodic temperature rise at the surface of a sample of nickel heated by an intensity-modulated Gaussian laser of radius $1\,\mu\mathrm{m}$ at e^{-1}

(25)

with

$$q_{\mathrm{th}}^2 = 4\pi^2 u^2 + \mathrm{i}\frac{\omega}{\alpha}\,. \tag{26}$$

Apart from a few simple cases, it is difficult to obtain analytic solutions. To find the thermal profile, the inverse transformations are often carried out numerically:

$$\Delta T(r,z) = 2\pi \int_0^{+\infty} \Delta\tilde{T}(u,z) J_0(2\pi u r) u \,\mathrm{d}u\,. \tag{27}$$

Figure 5 displays the results of such a calculation for a nickel sample, with $\alpha = 0.23\,\mathrm{cm}^2\cdot\mathrm{s}^{-1}$ and $k = 91\,\mathrm{W}\cdot\mathrm{m}^{-1}\cdot\mathrm{K}^{-1}$, showing the effect of an extended heat source. The amplitude of the temperature measured at the center becomes finite, whilst the phase is shifted away from $\phi = 0$. Near the heating point, the phase no longer varies linearly in time. One must move to a distance of the order of $2r_{\mathrm{g}}$ to recover this linear decay. The effect of the extension of the heat source, determined by diffraction phenomena, is a limitation of the spatial resolution in the measurements which is fixed partly by the chosen modulation frequency, and partly by the wavelength of the pump beam.

Semiconducting Materials

More complex phenomena are produced when semiconducting materials are excited photothermally. We shall discuss these briefly here and the reader is referred to the review in [10] for further details. The equilibrium state of the semiconductor is perturbed by light since carriers in the form of electron–hole

pairs are created when a photon is absorbed with energy greater than the band gap. Indeed, such a photon absorption allows an electron to transit from the valence band to the conduction band. The rate at which such carriers are generated has the form

$$G_{\mathrm{L}}(\lambda, z) = \big[1 - R(\lambda)\big]\alpha(\lambda)\Phi_0(\lambda)\mathrm{e}^{-\beta(\lambda)z}\,, \tag{28}$$

where λ is the wavelength, Φ_0 is the light flux, R is the reflection coefficient, and α is the absorption coefficient of the material.

In a direct band gap semiconductor, the absorbed energy is reemitted radiatively by band-to-band recombination. In an indirect band gap semiconductor such as silicon, the carriers relax non-radiatively (phonon), with characteristic time of picosecond order. The free carriers then diffuse in the bulk before recombining, essentially non-radiatively. The diffusion equations in a semiconductor must therefore account for all these phenomena. Taking into account carrier generation by thermal agitation, the temperature rise ΔT and the excess carrier density Δn for a semiconductor subjected to modulated illumination can be written

$$\frac{\mathrm{d}\Delta n}{\mathrm{d}t} = D\nabla^2\Delta n + \frac{\mathrm{d}n_0}{\mathrm{d}t}\frac{\Delta T}{\tau} + G_{\mathrm{L}} - \frac{\Delta n}{\tau}\,, \tag{29}$$

$$\frac{\mathrm{d}\Delta T}{\mathrm{d}t} = \alpha\nabla^2\Delta T + \frac{1}{\rho C_{\mathrm{p}}}\left[\frac{E_{\mathrm{g}}\Delta n}{\tau} + (h\nu - E_{\mathrm{g}})G_{\mathrm{L}}\mathrm{e}^{\mathrm{i}\omega t}\right]\,, \tag{30}$$

where D is the electron diffusivity and τ the recombination lifetime.

Much work is still being carried out on these phenomena, since they provide information about many important properties of semiconductors. However, in the heat transfer context, experimenters commonly seek to eliminate effects due to carriers in order to simplify signal interpretation, either by coating the semiconductor with an opaque layer, or by reducing the light flux sufficiently to make the influence of carriers negligible.

Light Sources and Modulation

In order to obtain a sufficient power, the most widely used sources are lasers, although arc or filament lamps are preferred in certain cases. The main criteria to guide this choice are power, wavelength (to obtain enough absorption in the relevant material), and the required range of modulation frequencies.

Modulation can be obtained by modulating the source itself. This is the case with Q-switch lasers (Nd:YAG up to a few kHz), passive mode-locked lasers (Ti:sapphire, of the order of 100 MHz), or active mode-locked lasers. More simply, some sources have a fast enough response to allow direct modulation of their supply, e.g., arc lamps and some gas lasers (CO_2, a few kHz), but also laser diodes and lasers pumped by laser diodes. Modulation frequencies of a few GHz can then be attained.

Another method involves modulation a posteriori using a continuous source. For frequencies up to a few kHz, the easiest approach is often to use a mechanical modulator or chopper.

Beyond this, if the source is sufficiently collimated (laser), acousto-optical modulators or Pockels cells are generally used. These provide efficiencies of the order of 80 % if the light is polarised, and half this value otherwise. The first type of modulator uses an acoustic wave at a few hundred MHz. In a crystal, this creates a refractive index grating which diffracts the light beam. By modulating the amplitude of this acoustic wave, an alternating deflection is obtained, and hence the beam is modulated at frequencies limited by the carrier frequency to about 30 MHz. Pockels or electro-optical cells use the rotation of the polarisation induced by an electric field in a crystal placed between polariser and analyser. The Pockels effect is fast and frequencies of MHz order can be achieved, although at relatively high cost.

For the record, we should mention modulators using magneto-optical effects or based on liquid crystals.

2 Detecting Thermal and Thermoelastic Waves

A fair number of optical phenomena can be induced by the presence of a thermal wave in a material. Many infrared detection techniques are based on the modification of black body emission. However, with the exception of recent work in the very near infrared[1] and studies in the near field [11], diffraction limits the resolution of these techniques to around ten microns, precluding any possibility of submicron scale measurement.

Raman spectroscopy uses inelastic light scattering, which yields information about vibrations of the crystal lattice, and hence also the temperature. Although it is a relatively slow and complex technique, it can yield absolute temperatures by calculating the ratio between the Stokes and anti-Stokes scattering lines [12]. In direct band gap semiconductors, one can also use photoluminescence, mentioned briefly above, to obtain information about the temperature [13, 14].

Here we shall focus on measurements obtained by analysing an optical wave that has been reflected by the surface under investigation. Indeed, the temperature influences the permittivity ε of a material, and thereby also its complex index and its reflection coefficient. Moreover, changes in temperature lead to surface displacements which modify the phase of the optical wave. To account for these effects, it is useful to introduce the idea of a complex reflection coefficient for an electric field:

$$\mathcal{R}(t) = \left[r_0 + \Delta r(t)\right] \mathrm{e}^{\mathrm{i}[\phi_0 + \Delta\phi(t)]} \,. \tag{31}$$

[1] B. Serio and B. Cretin: presented at the SFT Congress, 2003

For the very small variations that we are measuring, interferometric and reflectometric signals $s_{\mathrm{int}}(t)$ and $s_{\mathrm{refl}}(t)$, respectively, are proportional to variations in the phase and amplitude of the complex reflection coefficient. In the following sections, we shall make a distinction between techniques based on the measurement of $s_{\mathrm{refl}}(t)$ (reflectometry) and those based on measurement of $s_{\mathrm{int}}(t)$ (interferometry):

$$s_{\mathrm{int}}(t) = s_{\mathrm{i0}} \Delta\phi(t)\,, \tag{32}$$

$$s_{\mathrm{refl}}(t) = s_{\mathrm{r0}} \frac{\Delta R(t)}{R_0} = 2 s_{\mathrm{r0}} \frac{\Delta r(t)}{r_0}\,, \tag{33}$$

where s_{i0} and s_{r0} are the interfringe and average reflectance signals, respectively, and $\Delta R(t)/R_0$ is the relative variation of the reflection coefficient of the sample.

In the experimental situations we are concerned with here, variations in the amplitude and phase of the reflection coefficient are mainly due to temperature variations $\Delta T(t)$ and normal surface displacements $u(t)$, respectively. The quantities most directly accessible to an external probe are the surface displacement and the temperature. Under these conditions, for small signals such that $u(t) \ll \lambda$, the signals can be interpreted according to

$$s_{\mathrm{int}}(t) = s_{\mathrm{i0}} 2ku(t)\,, \tag{34}$$

$$s_{\mathrm{refl}}(t) = s_{\mathrm{r0}} \kappa \Delta T(t)\,, \tag{35}$$

where k and κ are the wave vector of the laser and the coefficient of thermoreflectance of the tested component, respectively. It thus transpires that measurements of $s_{\mathrm{refl}}(t)$ only provide temperature information after calibration, whereas displacement measurements using $s_{\mathrm{int}}(t)$ are absolute, but only lead to information about the temperature after an analysis that sometimes proves complex.

2.1 Reflectometry

Temperature Dependence of R

As we have just explained, reflectometry measurements depend on the relationship between permittivity and temperature. This dependence is phenomenologically common to all materials, but can involve very different physical processes. Unfortunately, apart from those materials commonly used in optics (see the volume entitled *Thermodynamic Coefficients* by *Ghosh* in [15]) and some semiconductors, it is difficult to find values of κ in the literature. In some transparent dielectrics, this dependence can be largely explained using a simple electrostriction model to calculate $\varepsilon(T)$, and hence $R(T)$ and κ, taking expansion into account [16].

In metals, several phenomena come into play to modify optical properties with temperature [17, 18]:

- Expansion reduces the plasma frequency and shifts the energy levels, leading to a shift in the Fermi energy. This can be affected by certain constraints, in particular related to the substrate. An increase in the Fermi level with temperature is also possible, but should have little effect for moderate temperature variations.
- The jump in the Fermi distribution increases, affecting certain interband transitions.
- The increase in the phonon population reduces the electron relaxation times and shifts energy bands because of electron–phonon interactions.

In metals correctly described by a Drude model, e.g., aluminium, a reasonable value of the coefficient κ can be obtained.

Finally, in semiconductors, it is essentially the temperature dependence of the band gap that causes variations in $R(T)$, and hence in κ.

In addition to this, just as the permittivity ε, the refractive index n, and therefore the reflection coefficient R all depend strongly on the wavelength λ of the illuminating light, the coefficient κ also varies significantly with λ. In many cases, one must carry out a genuine spectroscopic investigation to optimise the sensitivity of thermoreflectance measurements. This can be obtained sequentially by changing the illumination wavelength by means of a monochromator, or in a parallel way with a CCD setup like the one described below (see Fig. 6) [19].

Finally, the presence of a transparent layer on the surface can significantly alter the measured signal. This situation is commonplace with materials used in microelectronics. These are usually coated with a passivation or encapsulation layer, generally of SiO_2 or Si_3N_4, which is transparent in the visible. Interference then occurs, inducing fringes in the spectra of $R(\lambda)$ and $\kappa(\lambda)$. These phenomena have been measured [19–22] and modelled [19, 23].

As a consequence it is generally impossible to use the values of κ provided in the literature for this type of stratified structure, where tiny variations in the refractive index and thickness can significantly modify κ. Moreover, it is crucial to choose λ so that $\kappa(\lambda)$ is not too small, or even zero. Generally, a calibration has to be carried out on each material in order to obtain κ and hence the temperature.

Calibration

Three methods are available to obtain a value of ΔT which can be correlated with reflectometry measurements to obtain κ:

- Temperature values can sometimes be obtained by modelling, but the temporal sampling required for numerical calculation of ΔT distinctly complicates the calculation. Moreover, the result depends sensitively on how accurately one knows the thermophysical properties of the medium. On length scales shorter than the micron, these properties are often poorly understood and an accurate calibration is rarely possible [24].

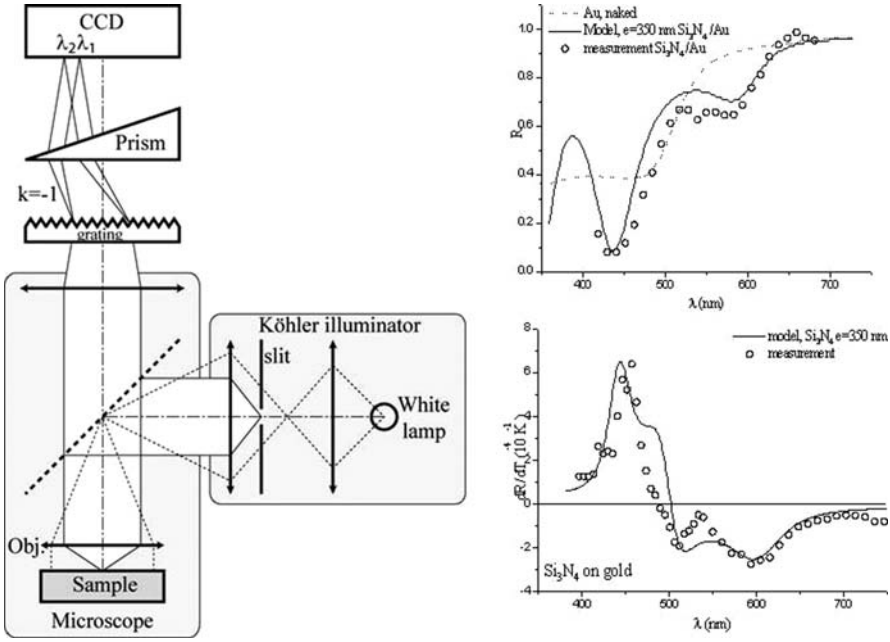

Fig. 6. *Left*: Spectroscopic setup for obtaining spectra $R(\lambda)$ and $\kappa(\lambda)$. *Right*: $R(\lambda)$ (*top*) and $\kappa(\lambda)$ (*bottom*) measured for gold coated with a layer of Si_3N_4

- A thermometer standard can be deposited on the sample. For passive samples, in particular, this may be a microthermocouple deposited on the surface by means of a micromanipulator [22]. If this thermocouple is small enough not to perturb the temperature value, and if the thermal contact is good, a local temperature value can be obtained on any material. In electronic circuits, many elementary components such as transistors, diodes, and resistors, have stongly temperature-dependent electrical characteristics. Their characteristic can then be calibrated in an oven so that they can subsequently be used as temperature sensors. In particular, interconnects behave as resistive elements in which the temperature induces resistance variations, and these in turn react back on the electrical current to generate harmonics that can be analysed to obtain a temperature value [24].
- Finally, when a sample does not lend itself to the above methods, the temperature has to be controlled and measured by external means. To do so, the sample is placed on a Peltier module and a temperature sensor fixed as close as possible to it. The reflectance and temperature are measured simultaneously by varying the temperature more slowly than the thermal time constant of the system (typically 10^{-2} Hz). At such low frequencies, lock-in detection is not very effective ($1/F$ noise) and a simple averaging filter can be used in the same way. To compensate for motion

of the sample caused by expansion of the Peltier module, which defocuses the optical system, the position can be servo controlled [24, 25].

Focused Approach. Single Detector

The most widely used technique for measuring thermoreflectance [26] consists in collecting on a photodiode the intensity I reflected from a laser beam (called the probe beam) with continuous incident intensity I_0. For small temperature variations ΔT, corresponding to small variations in $R(T)$, we may write

$$R(T + \Delta T) = R(T) + \kappa \Delta T \,. \tag{36}$$

Assuming that $R(T) \approx R(T_\mathrm{a}) = R_0$ and that $\partial R/\partial T$ is constant as a function of T, the periodic component of the reflected light intensity I can be written to first order as

$$\mathrm{i} = I_0 R(T) \approx I_0 R_0 \left(1 + \frac{1}{R_0} \kappa \Delta T\right) . \tag{37}$$

The harmonic signal measured in the photodiode is then directly proportional to the periodic temperature rise ΔT at the point where the probe beam reflects off the surface. The amplitude of the measured signal thus depends via the coefficient κ on the optical properties of the surface. The phase of the signal relative to the thermal excitation depends only on the thermal diffusivity of the sample.

From this idea, and as described earlier, two types of setup can be used, depending on the origin of the heat source used to generate the periodic temperature rise ΔT:

- For most 'passive' samples, the simplest way of producing a temperature modulation is to use a sufficiently powerful heating laser focused on the surface of the material.
- Some 'active' samples already include a heat source. This is the case for integrated circuits, electronic components, and laser diodes.

This microscope [27] comprises three main subsystems:

- an optical device for focusing and positioning the detection (and heating) beams,
- a device for measuring the reflected flux,
- the electronics required for the measurement.

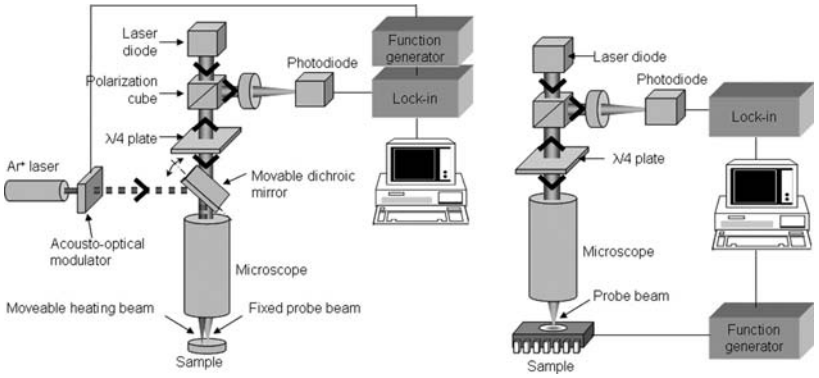

Fig. 7. Single-beam photothermal microscope. *Left*: Setup used to measure local thermal diffusivities in a passive sample. *Right*: Setup used to measure temperatures in an active sample

System for Measuring Local Properties (Passive Samples)

The probe beam comes from a continuous wave laser with power of the order of a few tens of mW, e.g., a diode at $\lambda = 670\,\text{nm}$. It is focused in the same way as the pump at the sample surface. After reflection on the sample surface, the probe passes once more through the dichroic mirror, a quarter wave plate and a polarisation splitter which directs it fully toward the photodiode. An interference filter centered on the probe is used to prevent any part of the pump beam from returning to the photodetector. The signal from the photodiode is filtered by lock-in detection to extract the amplitude and phase of the periodic signal from background noise.

The pump beam comes from a continuous wave laser, e.g., ionised argon Ar^+, $\lambda = 514\,\text{nm}$, with adjustable power typically between 10 mW and 1 W. An acousto-optical crystal modulates the beam intensity at a frequency fixed by a generator. The pump beam, after reflecting off a dichroic mirror, is focused on the sample surface by a metallographic microscope. To make the measurement, the pump beam has to be moved relative to the probe beam. When the sample is homogeneous, it makes strictly no difference which beam is moved. Since the problem is then symmetrical, the only relevant factor is the separation between the heating point and the detection point. In this setup, it is generally the probe that is fixed [27]. This guarantees the stability of the measurement by removing the effects of optical defects on the surface, such as scratches or spikes. By moving the pump beam to different points on the sample surface, one obtains a map of the periodic temperature rise. To facilitate these movements, a computer controls the motion of two step motors which act on the orientation of the dichroic mirror, thereby regulating the relative positions of the pump and probe beams. The tasks carried out by the various pieces of apparatus are piloted by a programme which automates the measurements and stores experimental data.

In order to detect local inhomogeneities in the sample such as thermal barriers, one can use a method derived from the one described here which consists in setting the two beams in coincidence and shifting the sample.

Systems for Measuring Heating (Active Samples)

When making point measurements on an active sample such as an integrated circuit, the sample must be moved to reveal heating, or conversely, the probe beam must be moved. The latter configuration can save scanning time provided that the scanning mirror is light enough. In the end, the measurement speed is limited by the time constant required by the lock-in detection to obtain a satisfactory signal-to-noise ratio. In some cases, a system can be appended to correct defocusing, undesirable interference effects, or lateral movements, in order to compensate for motion due to expansion of the sample [24].

For these two types of setup, in optimal operating conditions, i.e., for a signal limited by photon noise, for a typical reflective sample such as a metal [28], relative reflectance variations of the order of 10^{-8} can be measured. For a coefficient κ of the order of $10^{-4}\,\mathrm{K}^{-1}$, this is equivalent to a temperature variation of $10^{-4}\,\mathrm{K}$.

Detector Arrays and Imaging

These experiments are rather lengthy if one aims to obtain images with several hundred to several thousand measurement points per side. It may then be preferable to use a multiplexed setup, i.e., an array of detectors which can directly obtain a full-field image of the temperature rise at the sample surface without scanning. In the following, we shall limit the discussion to active samples, although it obviously transposes to passive samples by replacing the electrical modulation by external heating.

Multiplexed setups have the advantage of significantly reducing the acquisition time when high definition images are required. Furthermore, they do not involve the use of a coherent light source, i.e., a laser. If one uses a light-emitting diode (LED) or a bandpass interference filter, it is easy to adjust the illumination wavelength so as to obtain an optimal coefficient $\kappa(\lambda)$ for the material under investigation. Finally, the risks of perturbing the operation of an integrated circuit by carrier creation are significantly reduced since the illumination of the sample is much weaker than with a focused laser (up to $1\,\mathrm{kW/mm^2}$).

This reduction in the light flux is imposed by the maximal number of photons that can be accumulated in the CCD sensors per unit time before they become saturated. This quantity is thus related to the depth of the CCD wells (of the order of $N = 250\,000$ for the best cameras, and generally inversely proportional to the pixel size), and to the camera frame rate (a few tens of frames per second). If the setup is limited by photon noise,

Table 1. Characteristics of a CCD camera (Dalsa 1M30)

Well depth N	Readout noise	Definition	Frame rate (frame/s)	Conversion dynamics
250 000	1.5	1024^2	30	4096 (12 bits)

Table 2. Maximal performance (temperature resolution) as a function of integration time for a GaAs sample, with $\mathrm{d}R/\mathrm{d}T = 1.7 \times 10^{-4}\,\mathrm{K}^{-1}$. Note that this is a simulation that only takes into account photon noise and the readout noise of the camera. The true performance is necessarily less good due to the presence of other noise sources

Accumulation time	0.13 s	1 s	1 min	1 h
Temperature resolution (K)	2.39	0.86	0.11	0.01

a signal-to-noise ratio of $\sqrt{N}$ is achieved for one frame. Summing over n frames, one thus achieves a signal-to-noise ratio of $\sqrt{nN}$. For a camera of maximal frame rate F, the signal-to-noise ratio is thus $\sqrt{NF}\sqrt{t}$, where t is the integration time. So in the end, if the system is limited by photon noise, it is the quantity $\sqrt{NF}$ (well depth times frame rate) which determines the sensitivity.

With a camera operating at 30 frames per second, up to $30 \times 250\,000 = 7.5 \times 10^6$ photons can thus be accumulated per second and per pixel, which makes around 3 pW per pixel at $\lambda = 550$ nm. This type of power density limits the risks of perturbing the measurement by carrier creation. However, there is a distinct deterioration in the signal-to-noise ratio which can be partly compensated by accumulating a large number of frames and optimising the wavelength of the source. Depending on the integration time, the signal-to-noise ratio remains about two orders of magnitude lower than what can be obtained with a single detector. However, it should be borne in mind that the measurement is made simultaneously over 10^6 pixels. When an image is required with more than a few tens of pixels per side, the multidetector array approach is considerably faster. In all other cases, the single detector approach is probably preferable.

Homodyne Operation

Since most CCD cameras operate at frequencies of at most a few tens of Hz, the setup has to be adapted to suit the chosen modulation frequency. For frequencies below 10 Hz, the circuit is illuminated by a continuous source and the camera is synchronised directly with the modulated phenomenon.

Sampling is based on multiplexed lock-in detection. A detector functions by successive integrations to carry out the sampling in such a way that one can ascertain the amplitude and phase of a modulated phenomenon of period T [29], as shown in Fig. 9. This provides information similar to the data

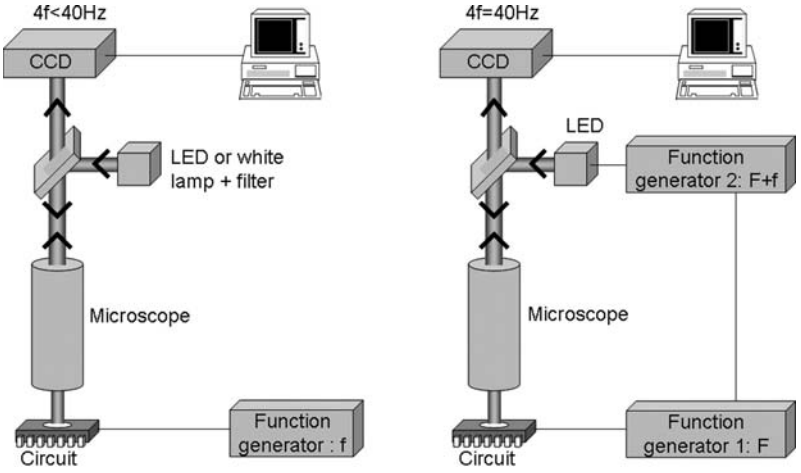

Fig. 8. Thermoreflectance microscope using a detector array. *Left*: Homodyne operation. The circuit operates at frequency f and the camera is triggered at $4f$. *Right*: Heterodyne operation by double amplitude modulation. The circuit operates at frequency F and the illumination is modulated at frequency $F + f$. The camera is triggered at four times the beat frequency f

provided by a lock-in amplifier, but adapted to the possibilities of a CCD camera. Shannon's sampling theorem requires the acquisition of at least two images in order to obtain the modulation amplitude. A classic arrangement for multiplexed lock-in detection consists in using four images I_1, I_2, I_3 and I_4 per period rather than just two, in order to obtain information about the phase as well [29]. The amplitude and phase then represent the modulus and argument of the complex number $I_1 - I_3 + \mathrm{i}(I_4 - I_2)$.

Heterodyning by Double Amplitude Modulation

For circuit operating frequencies greater than one quarter of the maximal frequency of the camera (a few tens of Hz), it becomes impossible to carry out adequate sampling of reflectance variations. One can then modulate the flux of the light-emitting diode at a slightly different frequency to that of the phenomenon under investigation (heterodyne operation, illustrated in Fig. 8). Information about the amplitude and phase of ΔR is then accessible at the beat frequency between the modulation frequency of illumination and that of the heating. If a low enough beat frequency is chosen, the camera can then be used to measure the alternating temperature variations, even for supply frequencies in the circuit as a high as GHz [30].

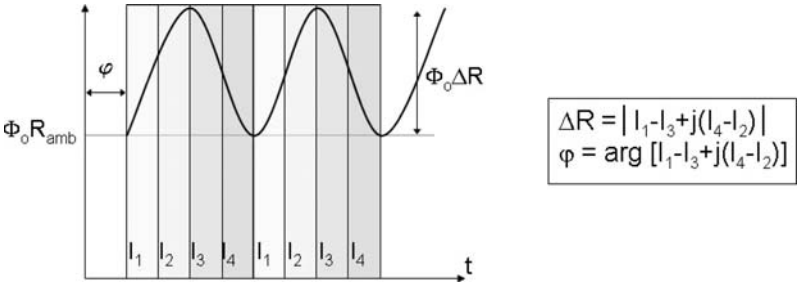

Fig. 9. Multiplexed lock-in detection

2.2 Interferometric Probes

These probes measure any normal displacement of the surface, of whatever origin, i.e., acoustic in the form of Rayleigh, Lamb or bulk waves [31–33], or thermal [34–37], either in permanent or transient regime.

The phase kx of a laser beam reflected from a vibrating surface is modulated by the displacement $u(t)$ of that surface. The reflected electric field E_p of the probe beam is given by

$$E_\mathrm{p} = E_0 \exp\Big\{\mathrm{i}\big[\omega t + \phi_\mathrm{p} + 2ku(t)\big]\Big\}\,, \quad k = \frac{2\pi}{\lambda_\mathrm{opt}}\,, \tag{38}$$

where ϕ_p contains the phase delay due to propagation as well as the phase shift induced by reflection. If $ku(t) \ll 1$, i.e., the amplitude of the surface displacement is small compared with the optical wavelength, the electric field E_p can be considered to be constituted by a low phase modulation. Under these conditions, the spectrum of E_p comprises a central line at the optical carrier frequency and two side bands. Let us consider the standard problem of a sinusoidal displacement

$$u(t) = u\cos(\omega_u t + \phi)\,. \tag{39}$$

The two side bands consist of two lines shifted by $\pm\omega_u$ from the carrier (see Fig. 10). The amplitude of the two lines is given by $J_1\big(2ku(t)\big)$. For low index modulations, the Bessel function can be approximated by

$$J_1\big(2ku(t)\big) \approx ku(t)\,. \tag{40}$$

The electric field can thus be expressed in the form

$$E_\mathrm{p} = E_{\mathrm{p}0}\Bigg\{\exp[\mathrm{i}(\omega t + \phi_\mathrm{p})] + ku\exp\Big[\mathrm{i}\Big((\omega+\omega_u)t + \phi_\mathrm{p} + \frac{\pi}{2}\Big)\Big] - ku\exp\Big[\mathrm{i}\Big((\omega-\omega_u)t - \phi_\mathrm{p} - \frac{\pi}{2}\Big)\Big]\Bigg\}\,. \tag{41}$$

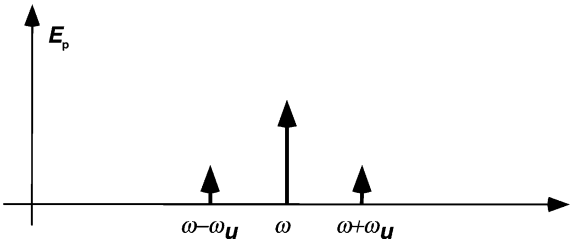

Fig. 10. Spectrum of the magnitude of the phase-modulated electric field

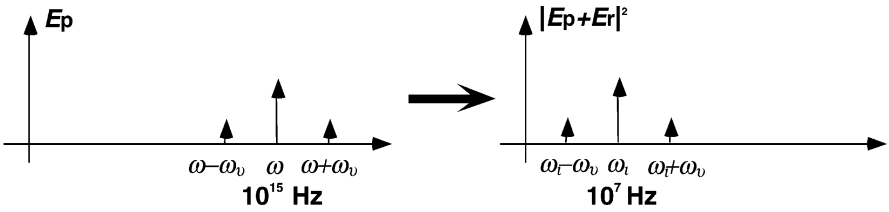

Fig. 11. Transposition of the optical frequency spectrum to radiofrequencies

Information concerning the amplitude of the surface motion u is contained in the side bands. When the modulation frequency is high enough, the side bands can be detected using a Fabry–Perot cavity as a frequency discriminator. These cavities behave like resonant circuits with very high quality factors. It is thus possible to achieve a frequency resolution of a few tens of MHz.

Another method consists in mixing the probe beam E_p with a reference beam E_r which is coherent with the first in such a way that they interfere. The beams are mixed in a photodetector sensitive to the square of the electric field intercepting it. It thus delivers an electric current proportional to the mixture of the two waves.

There are two coherent detection methods: a heterodyne method which mixes the coherent beams with a frequency shift between them, and a homodyne method in which there is no shift between the beams. These two techniques are discussed in the next two sections.

Heterodyne Interferometry

In this approach, the frequency of one of the two beams, the reference or the probe beam, is shifted by an amount $f_\mathrm{i} = \omega_\mathrm{i}/2\pi$ called the intermediate frequency [31, 32, 34, 38, 39]. The latter is the frequency at which the signal is processed. The ratio between the carrier frequency and its side bands is thereby reduced. One can then use an electronic filter with lower quality factor than the Fabry–Perot optical cavity. This transposition is illustrated in Fig. 11.

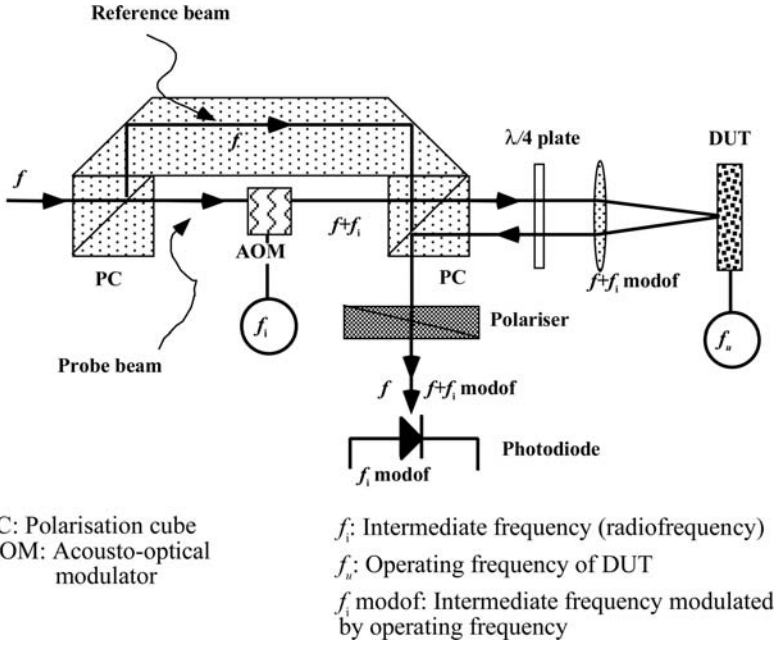

Fig. 12. Device developed by Dieulesaint and Royer [33]

Figure 12 shows a Mach–Zender heterodyne interferometer developed by *Dieulesaint* and *Royer* [32]. A source beam is separated into two parts by a first polarisation splitter. The frequency of the probe beam is shifted by an acousto-optical modulator with collinear input and output. The shift frequency is 70 MHz.

It then crosses another polarisation cube beam splitter and reflects off the device under test (DUT). After crossing the quarter wave plate twice, it is deflected towards the photodetector. The reference beam, deflected by the first cube, goes through the prism to the photodetector. The two beams, probe and reference, interfere after the output polariser.

The electric fields E_p and E_r of the probe and reference waves are

$$E_\mathrm{p} = E_0 \exp\left\{\mathrm{i}\left[(\omega+\omega_\mathrm{i})t + \phi_\mathrm{p} + 2ku(t)\right]\right\}, \tag{42}$$

$$E_\mathrm{r} = E_0 \exp\left[\mathrm{i}(\omega t + \phi_\mathrm{r}\right]. \tag{43}$$

The current $I(t)$ supplied by the photodiode is proportional to the complex square of the resultant electric field:

$$I(t) = S(E_\mathrm{p}+E_\mathrm{r})(E_\mathrm{p}+E_\mathrm{r})^* = \frac{I_0}{2}\left\{1+\cos\left[\omega_\mathrm{i}t + 2ku(t) + \phi_\mathrm{p} - \phi_\mathrm{r}\right]\right\}, \tag{44}$$

where $I_0 = 2SE_0^2$ and S is the sensitivity of the photodetector (A/W).

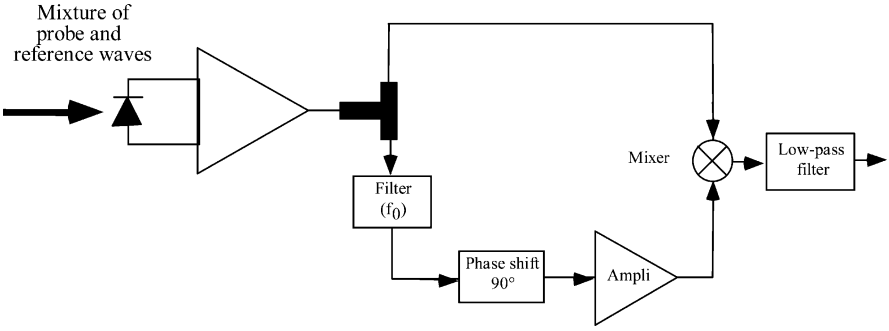

Fig. 13. Broad band coherent electronic detection

The current is thus the image of the probe beam transposed to radiofrequencies. We then deduce the new expression

$$I(t) = I_0 \Big\{ \exp\big[\mathrm{i}(\omega_\mathrm{i} t + \phi_\mathrm{p} - \phi_\mathrm{r})\big] + ku \exp\Big[\mathrm{i}(\omega_\mathrm{i} + \omega_u)t + \frac{\pi}{2} + \phi_\mathrm{p} - \phi_\mathrm{r}\Big] - ku \exp\Big[\mathrm{i}(\omega_\mathrm{i} - \omega_u)t - \frac{\pi}{2} - \phi_\mathrm{p} + \phi_\mathrm{r}\Big] \Big\}. \quad (45)$$

We deduce that the ratio of the amplitude of a side band to its carrier directly yields the amplitude of the surface displacement. This is only valid in the permanent sinusoidal regime. To detect the transient regimes, broad band detection with phase demodulation is required (see Fig. 13).

Part of the output signal is filtered to extract the carrier. Then, after shifting the phase by $\pi/2$, it is mixed with the signal itself. After filtering, this mixing yields the vibration of the object. In order to eliminate random phase fluctuations of mechanical or thermal origin $(\phi_\mathrm{p} - \phi_\mathrm{r})$, present in both signals, the passband of the selective filter must be greater than the phase noise spectrum. Figure 14 shows the passband of the selective filter.

On the one hand, the passband of such a device is bounded by a low-frequency cutoff determined by the quality factor of the selective filter. The latter depends on the required signal-to-noise ratio. In every case, the low-frequency cutoff is bounded by the maximal value of the quality factor of the selective filter. Displacements of frequency lower than a few tens of kHz cannot be detected with this technique. In addition, the passband is also bounded by a high-frequency cutoff imposed by the carrier frequency f_i.

Cretin and *Hauden* [34] have used a fractional phase-locked loop to detect oscillations at frequencies below kHz.

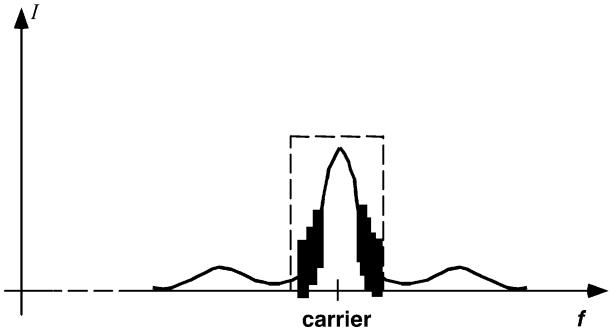

Fig. 14. Passband of selective filter

Stabilised Homodyne Michelson Interferometer

Homodyne interferometry can be used to overcome two difficulties inherent in heterodyne techniques:

- The lower bound of the passband which precludes the possibility of detecting slow surface displacements (at frequencies below kHz). This bound is imposed by the quality factors of selective filters used in the detection setup.
- The need to work with low modulation index, which limits the maximal amplitude of surface displacements to 20 nm.

The homodyne technique consists in mixing the probe wave with a reference wave at the same frequency. Under these conditions, the frequency is shifted towards the zero frequency. We shall show that this type of interferometry compares favourably with heterodyne interferometry, even in experimental situations where the latter can be implemented under optimal conditions.

In the Michelson interferometer shown in Fig. 15, a laser beam is split into two beams by a beam splitter. The probe beam is reflected off the object and the reference beam off a mirror. The intensities of the return beams are assumed equal. These beams are mixed at the device output where they interfere and are picked up by a photodetector. The latter delivers an electric current $I(t)$ proportional to the resultant intensity after wave mixing:

$$E_\mathrm{p} = E_0 \Big\{ \mathrm{i}\big[\omega t + \phi_\mathrm{p} + 2ku(t)\big] \Big\}, \tag{46}$$

$$E_\mathrm{r} = E_0 \big[\mathrm{i}(\omega t + \phi_\mathrm{r})\big], \tag{47}$$

$$I(t) = S(E_\mathrm{p} + E_\mathrm{r})(E_\mathrm{p} + E_\mathrm{r})^* = \frac{I_0}{2} \Big\{ 1 + \cos\big[2ku(t) + \phi_\mathrm{p} - \phi_\mathrm{r}\big] \Big\}, \tag{48}$$

where $I = 2SE_0^2$ and S is the sensitivity of the photodetector (A/W).

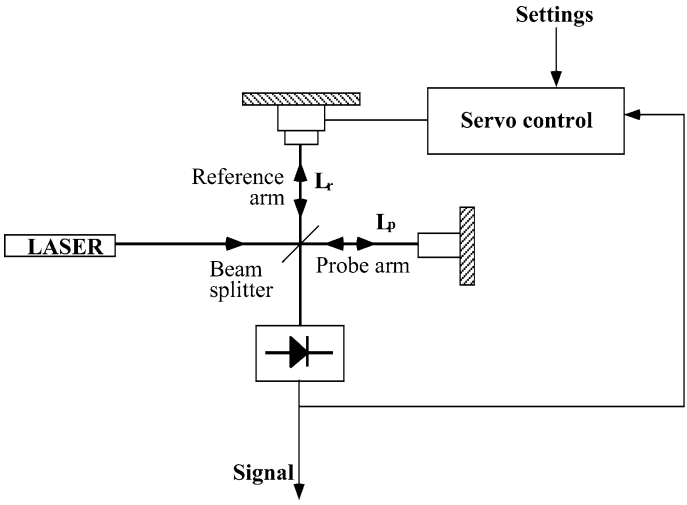

Fig. 15. Michelson interferometer with active stabilisation [1]

The photocurrent thus varies sinusoidally with the difference in length $L_\mathrm{p} - L_\mathrm{r}$ between the probe arm and the reference arm via the 'static' phase difference $\phi_\mathrm{p} - \phi_\mathrm{r}$ between the two arms of the interferometer:

$$\phi_\mathrm{p} - \phi_\mathrm{r} = 2\frac{2\pi}{\lambda} n_\mathrm{air}(L_\mathrm{p} - L_\mathrm{r})\,. \tag{49}$$

The sensitivity of the interferometer is maximal when

$$\cos\bigl[2ku(t) + \phi_\mathrm{p} - \phi_\mathrm{r}\bigr] = 0\,.$$

For small amplitude displacements $2ku(t) \ll 1$, this happens when

$$\phi_\mathrm{p} - \phi_\mathrm{r} = \frac{\pi}{2}(2n+1)\,,$$

where n is an integer, i.e., for an optical path difference equal to

$$n_\mathrm{air}(L_\mathrm{p} - L_\mathrm{r}) = (2n+1)\frac{\lambda}{8}\,. \tag{50}$$

To hold the sensitivity constant at small surface displacements, an operating point is defined for which the two waves are in phase quadrature. This operating point (P in Fig. 16) is held mid-fringe by servo controlling the position of the reference mirror using a feedback loop. The phase quadrature is thus guaranteed independently of the random variations in the optical paths. To achieve this, the reference mirror is carried by a piezoelectric block commanded by the low-frequency part of the interferometric signal. The low-frequency part of the noise contains most of the spectrum of phase variations of thermal and mechanical origin.

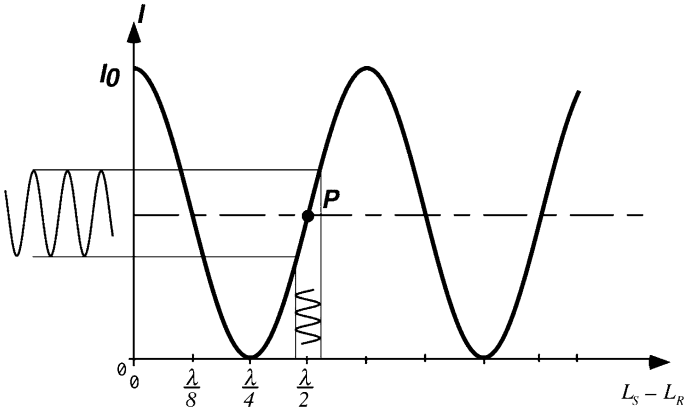

Fig. 16. Output intensity of the interferometer as a function of the length difference $L_\mathrm{p} - L_\mathrm{r}$ between the two arms

When the interferometer is stabilised at its operating point, the instrumental profile I can be linearised about the point of inflection for small displacement amplitudes. Now

$$I(t) = \frac{I_0}{2}\Big\{1 + \cos\big[2ku(t) + \phi_\mathrm{p} - \phi_\mathrm{r}\big]\Big\}\,, \tag{51}$$

$$\phi_\mathrm{p} - \phi_\mathrm{r} = (2n+1)\frac{\pi}{2}\,, \tag{52}$$

and for $ku(t) \ll 1$,

$$\cos\Big[2ku(t) + (2n+1)\frac{\pi}{2}\Big] = (-1)^n \sin\big[2ku(t)\big] \approx (-1)^n 2ku(t)\,,$$

whence

$$I(t) \approx \frac{I_0}{2}\big[1 + (-1)^n 2ku(t)\big]\,. \tag{53}$$

Then for the dynamic component of the signal, we have

$$I(t) = (-1)^n I_0 ku(t)\,. \tag{54}$$

This type of instrument therefore delivers a signal proportional to the normal surface displacement $u(t)$. The sign of the signal can be changed by varying n by one unit.

Focused Approach. Single Detector

Figure 17 shows an experimental device integrating surface reflectance and displacement measurements with a laser beam focused at the diffraction limit.

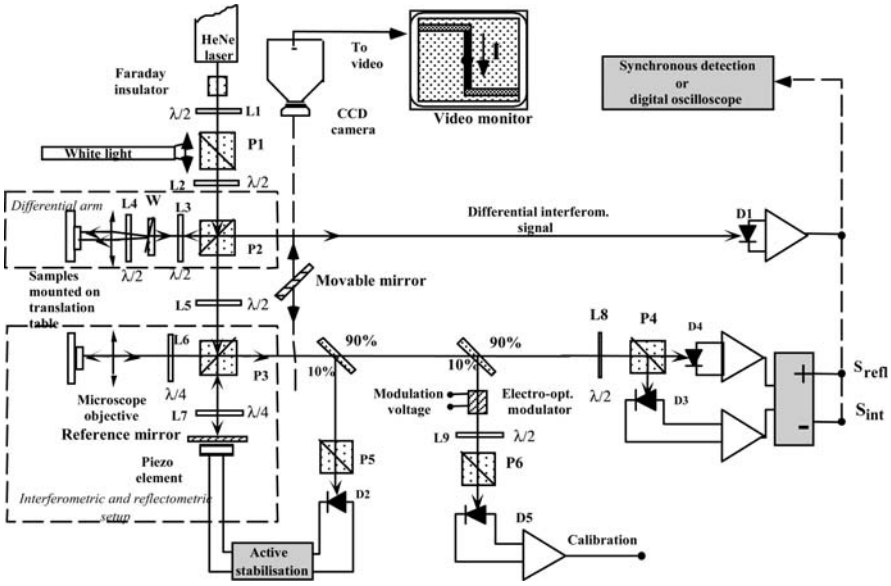

Fig. 17. Setup for combined interferometric, polarimetric and reflectometric measurements

All operating modes are indicated. It can be used to carry out interferometric, reflectometric, or polarimetric measurements to analyse variations in the reflection coefficient by amplitude, phase or depolarisation measurements. It integrates a visualisation system and the signal is detected by lock-in detection or using a digital oscilloscope.

Interferometry

The setup is similar to a homodyne Michelson interferometer with active stabilisation in which the beam splitter is the polarisation cube P3 associated with the quarter wave plates L6 and L7, as shown in Fig. 17 [40, 41]. The performance of this setup depends on how well the position of the reference mirror has been servo controlled. By taking the sum and difference of the interferometric signals obtained for the positive and negative interfringes, this servo control allows one to measure variations in the amplitude and phase of the reflection coefficient for the electric field:

$$\mathcal{R}(t) = \left[r_0 + \Delta r(t)\right] \mathrm{e}^{\mathrm{i}[\phi_0 + \Delta\phi(t)]} \,. \tag{55}$$

For the very small variations measured here, the interferometric and reflectometric signals $S_{\mathrm{int}}(t)$ and $S_{\mathrm{refl}}(t)$ are proportional to the variations in the phase and amplitude of the complex reflection coefficient:

$$S_{\mathrm{int}}(t) = S_{\mathrm{i}0}\Delta\phi(t)\,, \tag{56}$$

$$S_{\mathrm{refl}}(t) = S_{\mathrm{r}0}\frac{\Delta R(t)}{R_0} = 2S_{\mathrm{r}0}\frac{\Delta r(t)}{r_0}\,, \tag{57}$$

where $S_{\mathrm{i}0}$, $S_{\mathrm{r}0}$, and $\Delta R(t)/R_0$ are the interfringe signal, the average reflectance signal, and the relative variation of the reflection coefficient of the sample.

Calibrating Measurements

Variations in the amplitude and phase of the reflection coefficient are essentially due to temperature variations $\Delta T(t)$ and normal surface displacements $u(t)$, respectively. The quantities most directly accessible to an external probe are the surface displacement and the temperature. Under these conditions, and for small signals such that $u(t) \ll \lambda$,

$$S_{\mathrm{int}}(t) = S_{\mathrm{i}0}2ku(t)\,, \tag{58}$$

$$S_{\mathrm{refl}}(t) = S_{\mathrm{r}0}\kappa\Delta T(t)\,, \tag{59}$$

where k and κ are the wave vector of the laser and the thermoreflectance coefficient, respectively, of the device under test. It transpires that displacement measurements are intrinsically absolute, whilst temperature measurements can only be made after calibrating the test bed described previously.

Interferometry

The phase variation

$$\phi_0 + 2k\Delta U(t) + \left[\Delta\phi_{\mathrm{p}}(t) - \Delta\phi_{\mathrm{r}}(t)\right]$$

contains three contributions:

- The first is a constant phase shift, perfectly controlled by the position of the Wollaston.
- The second corresponds to a difference in level $\Delta u(t)$ between the two laser spots.
- The third is produced by a variation in the anisotropy of the refractive index of the material $\left[\Delta\phi_{\mathrm{p}}(t) - \Delta\phi_{\mathrm{r}}(t)\right]$. This signal is maximal when the beam polarisations correspond to the principal axes of anisotropy, and disappears when they are perpendicular to them.

All these contributions can be measured independently from one another by examining the sums and differences and by changing the distance between the two laser spots. This instrument can operate in four different modes: differential interferometry, profilometry, polarimetry, and goniometry.

Performance

The performance of the devices discussed above has been achieved through a concerted effort with regard to each element of the test bed: laser stability, careful construction of mechanical parts, optimised active stabilisation for each feature, the passband and detector noise, have all contributed to achieving a level of performance that remains unequalled to our knowledge in the field of electronic device characterisation. Measurements can be made at any frequency from DC to 125 MHz. The experimentally determined sensitivity is $10^{-16}\,\mathrm{m/Hz^{-1/2}}$ for a 1 mW laser, limited only by photon noise.

Phase Imaging. Detector Arrays

As we shall see later, single point measurement techniques can be used to map displacement or temperature fields. To do this, the whole sample surface must be scanned and the image reconstructed, which is very time-consuming. To remedy this, we have considered imaging techniques that use not just one or a pair of detectors, but CCD arrays. These methods involve illuminating a whole region of the device under test and analysing the phase of the reflected wave front. It is then possible to measure the surface displacement field.

One of the major differences between this type of approach and the point measurement techniques concerns the passband. Indeed, the latter is imposed by the video acquisition rate, generally 50 Hz. For signals with characteristic frequencies less than 50 Hz, the video rate is adequate. However, this imposes a severe restriction on applications. To extend the available frequency domain, we have used heterodyne and stroboscopic techniques.

Interferometric Imaging

The first approach uses the device described above [42, 43]. To this end, a lens is placed upstream of the cube beam splitter in the Michelson interferometer described earlier to disperse the beam. This serves to illuminate a broad region of the sample which, by interfering with the reference beam, produces fringes on a CCD array. This device was used to measure optical path variations in biological cells. It is difficult to process the interferograms obtained, the treatment being poorly suited to deformable objects owing to the fact that the reference wave front is planar.

Thermoplastic Holography

Holographic techniques are ideal for analysing conformational changes in the wave front. One records a hologram of the object at rest that will produce the reference wave. When the object deforms, its image is made to interfere with the reference image on a CCD array. Interference fringes then appear superposed on the image of the object. These fringes constitute isodisplacement contours for the surface.

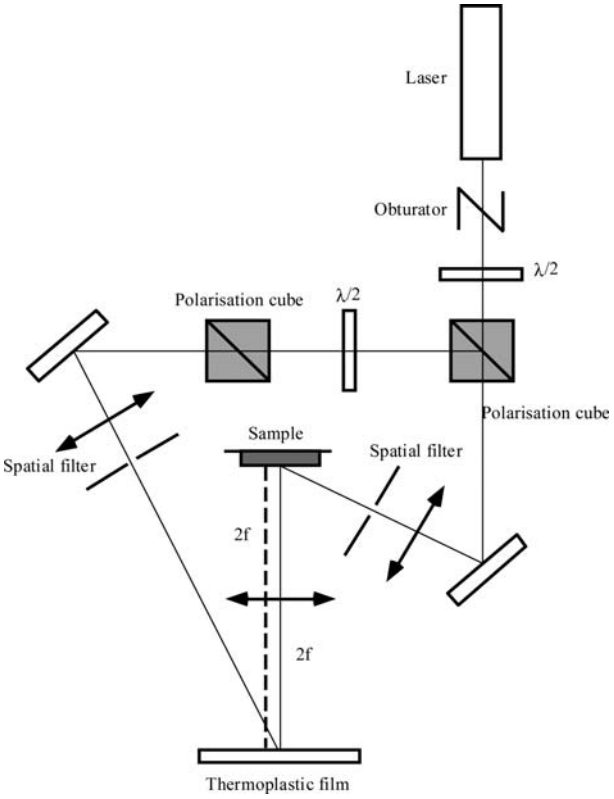

Fig. 18. Holographic test bed

In contrast to interferometric imaging as described earlier, the fringes are only present when the object deforms. As long as the wave front remains unchanged, the fringes do not appear.

The hologram is recorded in a conventional manner, as shown in Fig. 18. In contrast, our choice of technique using films, then thermoplastic plates, was guided by development methods. Wet chemical development processes are not well suited to the observation of small objects, since the film has to be moved to be developed in this way. Moreover, once the film has been developed, bleached, then fixed, the separation of the fringes in the recorded interference pattern will have evolved from its initial state, so that a shift is required in the wavelength to restore the image as it was when recorded. It then becomes difficult to reposition it perfectly in order to superpose the 'object' and 'reference' waves.

The test bed is illustrated schematically in Fig. 18. In this approach, the hologram is developed thermally. The idea is to heat the thermoplastic film above its glass transition temperature so that the change in its thickness modulated by the light intensity distribution is permanently fixed. The diffraction efficiency obtained with this kind of holographic material can be as

much as 20 %. The recording is obtained at a temperature of around 80 °C. To reconstitute the hologram correctly, the film must be kept at a temperature close to the recording temperature, otherwise the recorded grating may contract and one obtains the same wavelength shift for the image restitution. At 80 °C, it is difficult to eliminate convection, especially over areas of several square centimeters. Moreover, the thermoplastic films are costly and consumable. These are the two main reasons for moving towards a digital technique, viz., speckle interferometry.

Electronic Speckle Pattern Interferometry (ESPI)

Speckle interferometry techniques are a natural development of the holographic methods described above. The difficulties arising in the development of the thermoplastic film led us to replace it by a CCD sensor on which the 'object' and 'reference' waves interfere. The spatial resolution of the CCD sensor (20 µm or 50 lines/mm) is much lower than can be obtained with a thermoplastic film (1 µm or 1 000 lines/mm). This is why the angle between the 'object' and 'reference' waves is significantly smaller.

Two approaches can be distinguished in ESPI, one for measuring surface displacements in the plane, and the other for measuring them out of the plane:

– **Measuring Out-of-Plane Displacements.** The idea here is to measure the phase change of each speckle of the moving surface [44]. This is achieved by interfering a reference wave with the object wave, as shown in Fig. 19. The setup is similar to the Michelson interferometer in which the images are made to interfere. Superposing the light scattered by the object on that arriving from the reference arm produces a new speckle pattern on the CCD camera. Everything happens as though the rough sample surface were made up of tiny mirrors with sizes related to the speckle sizes. In fact, the average speckle size on the CCD is directly related to the magnification of the imaging system and its numerical aperture. A diaphragm can be used to adjust the size of the grains to the pixel size in the camera in order to obtain good spatial sampling of the image. The displacement of each mirror is then measured by interferometry on each grain formed on the camera.
The acquisition and processing procedure has four stages. The phase ϕ_{sp} of each grain relative to the reference wave is first determined for the object at rest. The intensity of each pixel then recorded by the camera can be written in the form

$$I_{\mathrm{rest}} = \frac{I_0}{2}\left(1 + V\cos\phi_{\mathrm{sp}}\right), \tag{60}$$

where I_0 is the average intensity, and V and ϕ_{sp} are the contrast and phase of the speckle. A four-image phase shift technique is used. The idea is to record a set of four images obtained by carrying out controlled

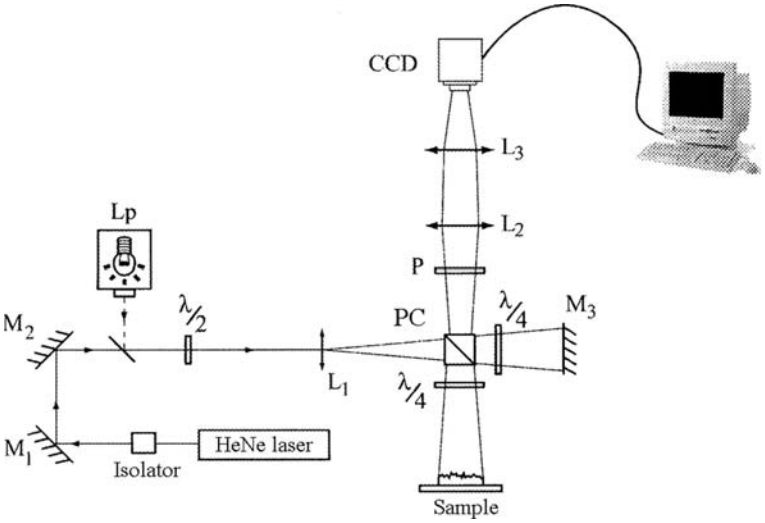

Fig. 19. ESPI setup for measuring out-of-plane displacements

phase steps. A combination of these four images can be used to determine in an independent manner the intensity, the contrast and the phase of the speckle image of the sample at rest.
Expansion of the sample modifies the phase of speckle grains and the intensity detected by the camera can be written in the form

$$I_{\text{rest}} = \frac{I_0}{2}\left(1 + V\cos\phi_{\text{sp}} + \Delta\right), \tag{61}$$

where Δ is the phase associated with the normal displacement of the sample surface.
A new set of four images is recorded. The same procedure is then applied as for the sample at rest. One thus obtains the same contrast image, the same intensity image, and the phase image of the deformed device. At this stage, none of the images yet obtained are in a usable form, because overwhelmed by noise due to the random nature of the speckle distribution. To establish the surface displacement undergone by the sample, one subtracts the two phase images obtained for the sample at rest and the operating sample. The result constitutes a phase map which can be used to measure the surface displacements modulo 2π. A phase unwrapping algorithm is then applied to reconstruct the deformed surface. As long as the displacements are small enough, the decorrelation of the speckle grains does not affect the image quality. For larger surface displacements, the images must be filtered, which affects lateral resolution.

– **Measuring In-Plane Displacements.** In-plane displacement measurements are made by illuminating the sample surface by a sinusoidal grating [44]. To do this, the surface is illuminated by two plane waves that

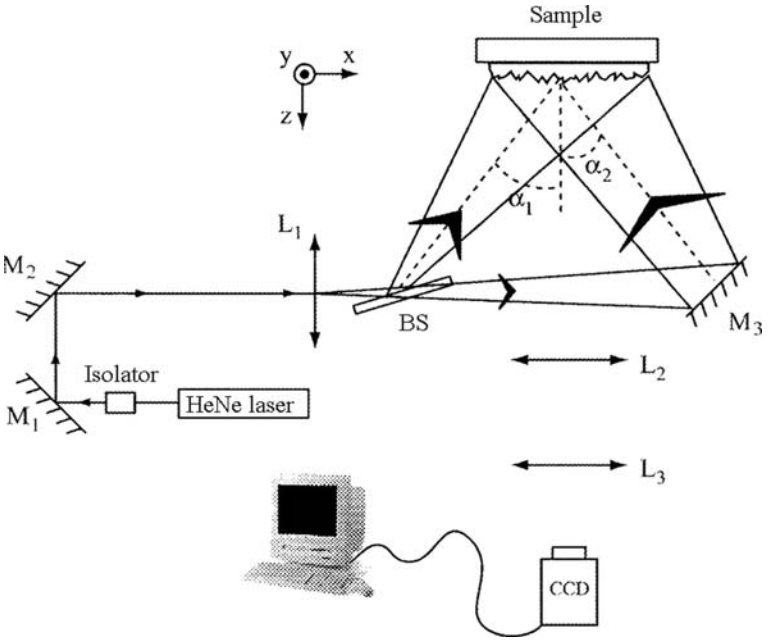

Fig. 20. ESPI setup for measuring in-plane displacements

interfere with an angle $\alpha_1 + \alpha_2$, as shown in Fig. 20. The process is the same as the one used for out-of-plane displacement measurements.

3 Applications

3.1 Temperature and Displacement Fields. Orders of Magnitude

Local Thermal Diffusivity and Thermal Resistance

To illustrate the prospects for photoreflectance as a way of characterising complex materials, we consider the example of studies carried out on aluminium nitride ceramics [45]. An approach combining macroscopic and submicron measurements has been used to establish the properties of different phases in a composite ceramic and relate them to the macroscopic conductivity of the ensemble. Figure 21 shows the isophases, obtained using the single-beam photothermal microscope described above, around the heating point located at the grain center. The separation between these isophases reveals the thermal conductivity of the material. The more widely spaced they are, the greater is the thermal conductivity. The AlN grains and secondary

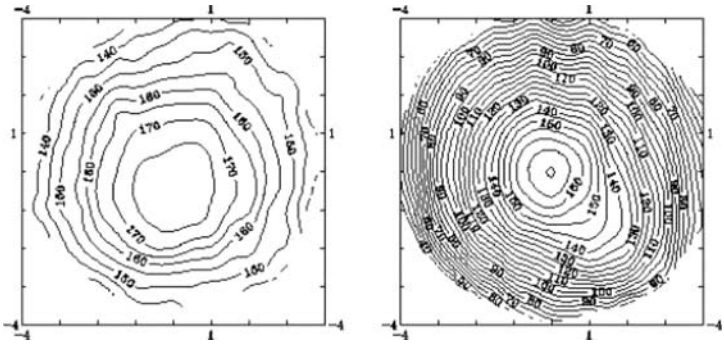

Fig. 21. Isophases obtained in a grain of AlN (*left*) and in the secondary phase (*right*) at a modulation frequency of 1 MHz. The length scale is in microns

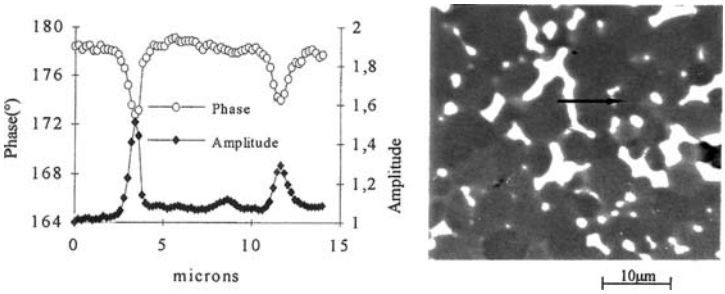

Fig. 22. Thermal barriers in moving from one grain to another. Experiments are carried out with the two beams superposed, at a modulation frequency of 1 MHz

phase grains (YAG or YAP) exhibit thermal conductivity values differing by almost an order of magnitude.

In order to deduce a macroscopic conductivity from the properties of these two phases, one must also be able to measure the interface resistance associated with grain boundaries [46]. This type of study can be carried out using the same microscope in a slightly different mode from the last. The pump and probe beams are now focused on the same point. A thermal barrier perpendicular to the surface will cause an asymmetry in the hot spot associated with a point source close to the thermal barrier. Indeed, as one moves towards the obstacle, the heat can no longer spread out isotropically and one observes an anisotropic local temperature rise. One thus scans the sample in search of phase and amplitude variations in moving from one grain to the next, as shown in Fig. 22.

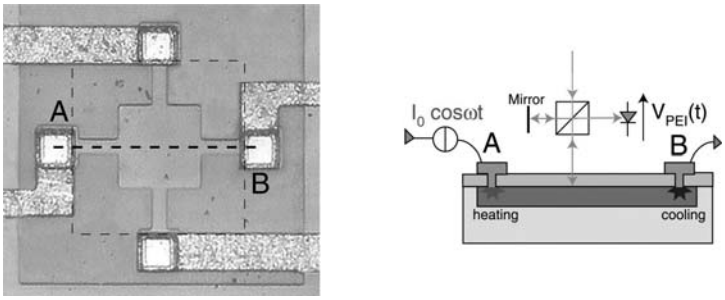

Fig. 23. *Left*: Diffused resistance on an integrated circuit. *Right*: Schematic of interferometric measurement

Detecting Interference of Thermoelastic Waves in Operating Integrated Circuits. Optical Ammeter

One technique for characterising a material involves generating thermoelastic waves inside it and examining the way they propagate. When devices are running, they can generate elastic waves which can be used to probe the semiconductor on which the device is implanted. For example, thermoelastic waves are emitted by ohmic contacts when a current passes. Consider the ohmic contacts represented by points A and B on the device in Fig. 23 (left).

When an electric current goes through these contacts, heat is both deposited and absorbed by the Peltier effect due to the potential barrier crossed by the electrons when they move from the metal to the semiconductor. The two facing contacts are Peltier heat sources in phase opposition. These sources generate thermoelastic waves that propagate in the surrounding material.

The thermoelastic deformations of the contacts can be detected by interferometry and followed as they propagate in the silicon. To do this, the resistor is supplied by a sine wave current of amplitude 50 mA at a frequency of 10 kHz. Figure 24 shows how the amplitude of the deformation of contact A varies with the electrical current through it. The Joule and Peltier contributions are measured there, discriminating between them by the choice of harmonic analysed in the lock-in detector. Indeed, the Joule effect, which is quadratic in the current, occurs at 20 kHz, whereas the Peltier effect, being linear, occurs at the same frequency as the current. Measuring the slopes of the two curves plotted on a log–log graph, the linear and quadratic dependencies of the these two phenomena can be identified.

Given the two curves in Fig. 24, the interferometer can be calibrated for use as an optical ammeter. Indeed, by measuring the amplitude of the surface motion, the electric current through the device can be unambiguously determined.

The amplitude and phase of the surface displacement field can be mapped by scanning the device. The latter is supplied by a sine wave current at 10 kHz between points A and B in Fig. 23. The resulting phase and amplitude im-

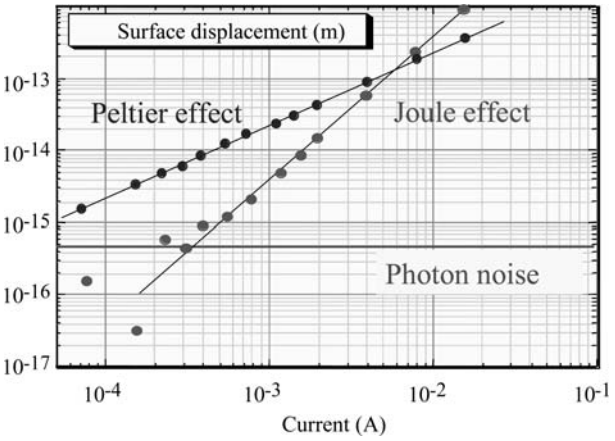

Fig. 24. Surface displacement of an ohmic contact for different currents. Discriminating between Joule and Peltier effects

ages are shown in Fig. 25. Destructive interference between the thermoelastic waves originating in contacts A and B is clearly visible. Figure 25 (left) shows a map of the amplitude of the surface displacement. The two heat sources can be perfectly located at the two ohmic contacts. There is very rapid decay as one moves away from the sources and the disappearance of the signal on the median line between them attests to the effects of destructive interference. The two sources have roughly the same amplitude, but are completely out of phase. Figure 25 (right) shows the phase map of the surface motion. The phase opposition of the two sources is clearly visible. Likewise, one can see a region near the sources where the phase varies linearly with the position, reflecting the wavelike nature of the heat transfer, whence the name thermoelastic waves. Destructive interference is revealed by an abrupt jump of 180° in the phase. The planar aspect of this phase change reflects the homogeneous nature of the medium explored by the thermoelastic waves.

This example provides a good illustration of how the exploration of devices by optical probes can take advantage of phenomena arising from the operation of the device itself. These naturally occurring phenomena are often ignored because they cannot be detected by standard electronic test equipment.

3.2 Locating Hot Spots and Mapping Temperature

Locating Defects. Qualitative Imaging

A rapid CCD-based imaging technique is particularly well suited to the problem of locating defects. In a first stage, approximate localisation can be achieved for circuits measuring up to a few millimeters across. This can then be followed up by high-resolution imaging of a more restricted region centered on the detected defect. We shall illustrate this process for devices made

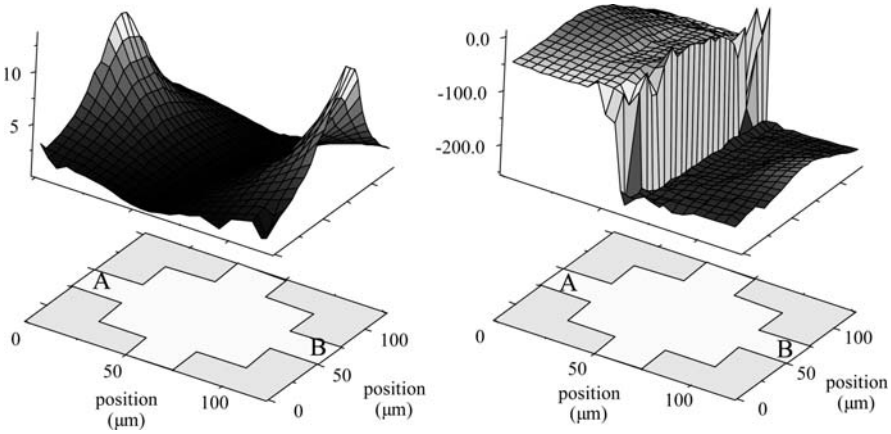

Fig. 25. Map of the sinusoidal surface displacement generated by the Peltier effect at the ohmic contacts A and B. *Left*: Amplitude (pm). *Right*: Phase (degree)

by ST Microelectronics with an etch resolution of 180 nm. These circuits include several ensembles each comprising 10 000 transistors, among which one must be able to detect a single transistor with a source–drain leakage current generating anomalous heating. Even at low frequencies, this hot spot may be extremely localised as a result of thermal insulation due to the fact that the transistor is embedded in a thermally insulating dielectric matrix. This stacking of dielectric layers used as electrical insulation between the different metal layers, or as a final protection of the circuit, can measure as much as 6 µm thick. This precludes the use of a surface probe, e.g., a scanning probe, since it would be impossible to achieve the required resolution once the heat has diffused and spread through such a thickness. Optical techniques are therefore particularly well suited to this task.

The transistors shown in Fig. 26 have been biased at $V_{\text{source}} = 0 = V_{\text{substrate}}$ and $V_{\text{drain}} = 1.4\,\text{V}$, whilst the gate voltage is modulated between 0 and 1.26 V at a frequency of 1 Hz. Two seconds of data acquisition suffice to locate the hot spot at low magnification. The higher resolution images shown here result from three minutes of accumulated data. This type of image can be used to study the effects of the transistor geometry on the appearance of crystal defects which lie at the origin of these source–drain leakages [47].

Quantitative Imaging

Laser Diodes

Optoelectronic components such as laser diodes are also subject to significant thermal strain [48,49]. For laser diodes using a ridge waveguide on the surface, confinement of the light generates significant heating. The laser diodes shown in Fig. 27 displayed an abnormally low damage threshold, suggesting the

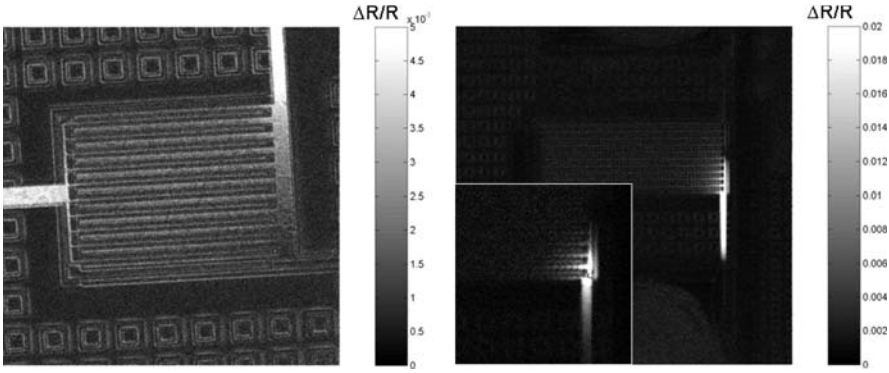

Fig. 26. *Left*: Thermoreflectance image of an array of 10 000 transistors (ST Microelectronics) observed using a ×50 objective ($50 \times 50\,\mu m^2$). When the ensemble is non-defective, only the supply tracks manifest a moderate temperature rise. *Right*: Similar structure with localised defect (image $135 \times 135\,\mu m^2$). *Insert*: Zoom ($42 \times 42\,\mu m^2$) on an image obtained on the same sample 20 min later. The appearance of the hot spot has changed due to a significant source–drain leakage. The spatial resolution is evaluated here at 350 nm (FWHM of the smallest imaged structures). The active regions visualised here are located 6 μm below the sample surface (Si_3N_4)

existence of a localised hot spot that was the source of the rupture. The small size of these structures obviously precludes the use of infrared imaging as a quantitative method. On the other hand, thermoreflectance imaging has been carried out on both the upper face of laser diodes (gold coated to provide electrical contact) and on the emission facets (GaSb substrate and active layers). The wavelength was optimised and calibration was carried out on each material in turn.

These measurements and calibrations led to concordant temperature images in a quite independent manner. The continuity of the temperatures across adjacent facets was obtained to an accuracy of the order of a few percent.

It was thus possible to measure heating effects in several of these diodes for different supply voltages, thereby revealing the anomalous heating of one of the emission facets. This facet then led to the destruction of the ensemble by prematurely reaching the melting temperature of one of the materials.

MEMS Study

The sample here is a microelectromechanical system (MEMS) [50] made from Si_3N_4 and a polysilicon heating resistor deposited on a thin SiO_2 membrane. An image of this resistor is shown in Fig. 28. The size of the membrane is $800 \times 800\,\mu m^2$ and the active zone covers a square measuring $200 \times 200\,\mu m^2$. Two metallic lines end the resistor to allow a four-point measurement. The heating element continuously dissipates 2 mW by the Joule effect.

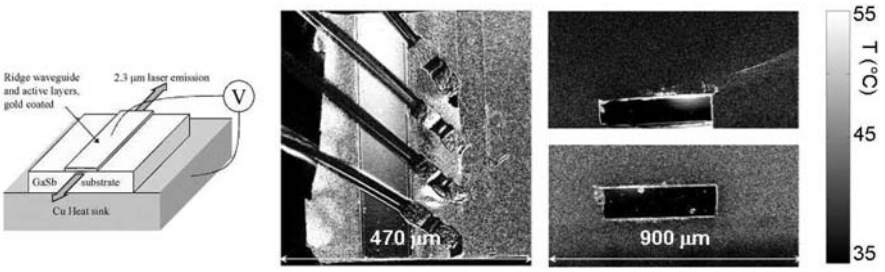

Fig. 27. *Left*: Schematic of a GaInAlAsSb laser diode used with the P face upwards. *Centre*: Temperature image on the upper surface, coated with gold. The laser diode is supplied with a square modulation between 0 and 312 mA, at 1.78 V and 1 Hz. Light wavelength $\lambda = 473\,\text{nm}$. Calibration coefficient $(1/R)\,\mathrm{d}R/\mathrm{d}T = 2.6 \times 10^{-4}\,\text{K}^{-1}$. *Right*: Temperature images obtained on the lateral emission facets, with $\lambda = 670\,\text{nm}$ and $(1/R)\,\mathrm{d}R/\mathrm{d}T = 2.6 \times 10^{-4}\,\text{K}^{-1}$

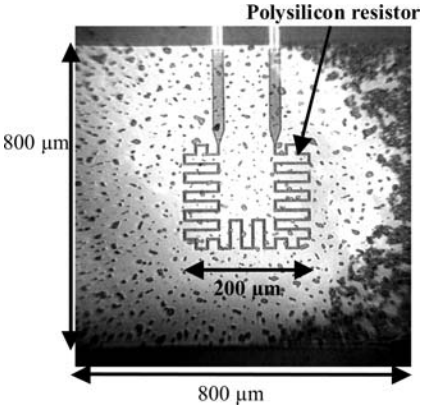

Fig. 28. MEMS made from Si_3N_4 and a polysilicon heating resistor deposited on a thin SiO_2 membrane

Thermoreflectance imaging can be used to map regions of higher temperature, as shown in Fig. 29. Quantitative measurement of the temperature landscape is obtained by calibrating on a sample in an oven. To begin with, the electrical resistance of the sample is measured as a function of the temperature in the oven. The optical measurement is understood and calibrated by simultaneously measuring the reflectance variation and the electrical resistance. It is then possible to acquire the reflectance map calibrated as a function of temperature. A temperature variation of 45 K with respect to the ambient temperature can be measured on the image in Fig. 29, from which the thermal resistance of the sample can be deduced to be $22.5 \times 10^3\,\text{K/W}$.

Metallic Interconnects of Components in an Integrated Circuit

The structures studied here are known as SWEATs (standard wafer level electromigration accelerated test). Two structures of this kind are shown in

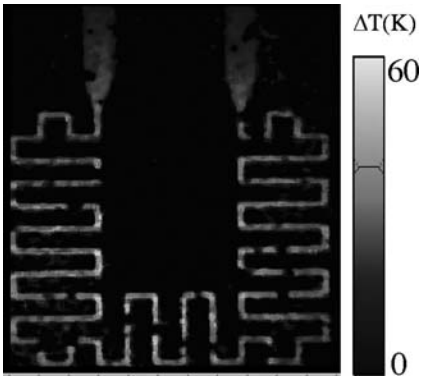

Fig. 29. Temperature map of the MEMS in Fig. 28

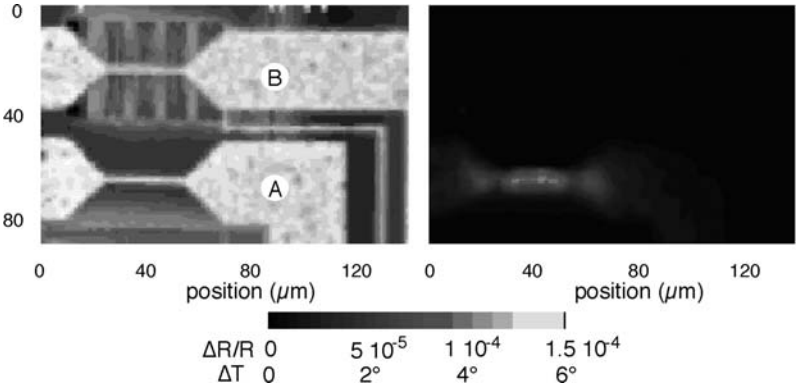

Fig. 30. *Left*: Map of average reflectance. *Right*: Map of relative reflectance variation converted to temperature

Fig. 30 (left). The image shows an aluminium track with irregular geometry deposited on a dielectric layer, itself deposited on a silicon substrate.

Measurements are made pointwise with a step of 2 µm over an area of $90 \times 140\,\mu m^2$ [see Fig. 30 (left)]. Bright and dark zones are visible, corresponding to large and small temperature increases, respectively, and a highly non-uniform temperature distribution appears on the right-hand part of structure A, which is the only part to be supplied electrically. The sensitivity of the measurement is sufficient to be able to map temperature fields with amplitudes as low as a few degrees.

3.3 Measuring Thermophysical Properties

Measuring Diffusivities in Composite Structures

As discussed in Sect. 1.1, heat diffusion in the alternating regime depends on the modulation frequency. Using this frequency dependence of the ther-

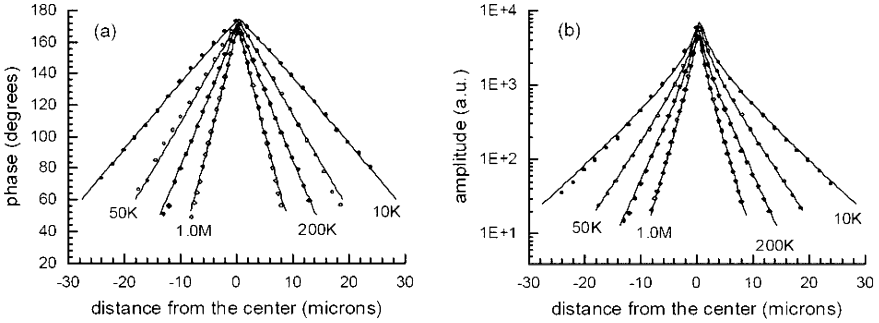

Fig. 31. Phase and amplitude of the thermoreflectance signal for a gold-coated ZrO_2 ceramic. The simulation (*continuous curves*) provides a good description of the experiment for modulation frequencies between 10 kHz and 1 MHz using the same parameters: $\alpha_{ZrO_2} = 0.01\,\mathrm{cm}^2/\mathrm{s}$ and $\alpha_{Au} = 1\,\mathrm{cm}^2/\mathrm{s}$

mal diffusion ($\mu = \sqrt{D/\pi F}$), different depths of the same sample can be preferentially probed. This possibility is particularly well suited to the study of multilayer structures, for which the thermal parameters can be extracted from individual layers.

Figure 31 shows measurements made on a simple system comprising a ZrO_2 substrate coated with gold which absorbs the beam at the surface. Hence, at high frequencies, heating due to the pump laser remains essentially confined within the surface gold layer. Using a model similar to the one described in Sect. 1.2, fitting is mainly sensitive to the thermal parameters of this gold layer. Lower frequencies can then be used to probe the substrate and hence to obtain its thermal properties. The fact that a satisfactory fit can be obtained with the same parameters over a wide frequency range is generally a good indicator of the reliability of the diffusivity values.

Identifying Thermophysical Properties in Integrated Circuits

The method uses a metal interconnect track on an integrated circuit as heating element [51,52]. Its operating temperature is closely linked to the thermal properties of the substrate, but also those of the dielectric layer on which it is deposited. Figure 32 is a schematic cross-section of the test structure.

The reflectometer must first be calibrated. This means determining the thermoreflectance coefficient of the sample at the point being probed. It can be shown that this coefficient depends heavily on experimental conditions, such as the numerical aperture of the microscope objective, or the presence of a dielectric (passivation) layer on the samples, should there be one. The repeatability of measurements is ensured by servo controlling the sample position in the neck of the beam.

After calibration, the reflectometer can be used to carry out absolute measurements of surface temperature. We have been able to determine a

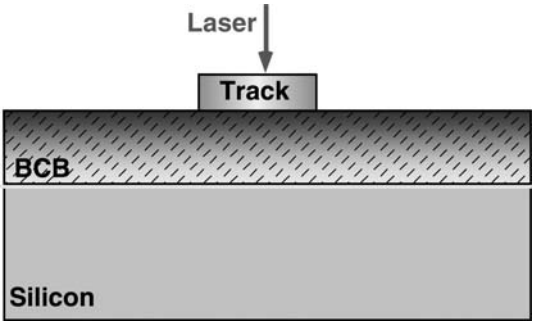

Fig. 32. Cross-section of the device used for identifying thermophysical properties. BCB = benzocyclobutene

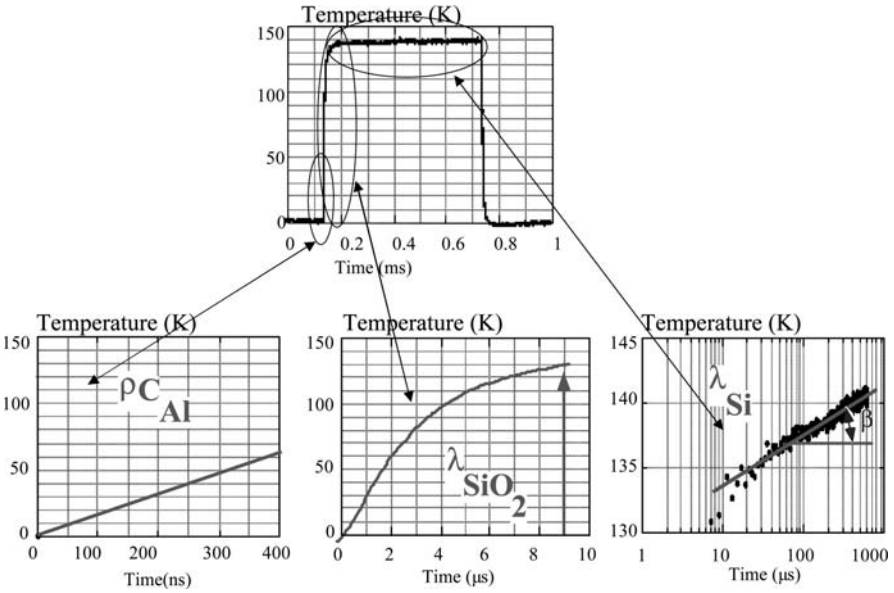

Fig. 33. Role played by time scales in the identification of thermophysical properties

thermoreflectance coefficient of $-2.5 \times 10^{-5}\,\mathrm{K}^{-1}$ with 14 % uncertainty for gold at a wavelength of 633 nm.

Thermophysical properties are identified by comparing measurements with a model for the diffusion of heat in a multilayer medium. The model uses the thermal quadrupole method. For a given layer, each incoming (temperature flux) vector is related to an outgoing (temperature flux) vector by a 2×2 matrix. The problem is simplified by a double Fourier or Laplace–Fourier transform depending on whether the analysis is harmonic or transient. We

then compare the reflectometric measurements with the model predictions. The properties of the various layers can subsequently be estimated by an optimisation method, a linear regression based on a least squares analysis. These values minimise the mean squared deviation between the optically measured and calculated temperatures. The thermal conductivities and heat capacities of thin films can then be determined directly in integrated circuits. The values obtained can subsequently be used in more sophisticated numerical simulations. Figure 33 shows how thermal conductivities and heat capacities can be determined, when uncorrelated, simply by analysing on different time scales. Short times (the first few hundred nanoseconds) provide information about the properties of the heat source. Heat propagation at intermediate times (microseconds) can be used to identify the properties of the insulating layer. Finally, temperature variations over long times are governed by the thermal conductivity of the silicon substrate.

References

[1] S. Dilhaire: *Développement d'un interféromètre laser très haute résolution pour la caractérisation de composants microélectroniques*, Ph.D. thesis, University of Bordeaux, France (1994) 240, 263

[2] A. Mandelis: Theory of photothermal wave diffraction and interference in condensed media, J. Opt. Soc. Am. A: Optics Image Science and Vision **6**, 298–308 (1989) 241

[3] J. Opsal, A. Rosencwaig: Thermal wave depth profiling: Theory, J. Appl. Phys. **53**, 4240–4246 (1982) 241

[4] S. K. Lau, D. P. Almond, P. M. Patel: Transient thermal wave techniques for the evaluation of surface coatings, J. Phys. D: Appl. Phys. **24**, 428–436 (1991) 241

[5] H. S. Carslaw, J. C. Jaeger: *Conduction of Heat in Solids* (Oxford University Press, Oxford 1993) 241, 242

[6] W. J. Parker, R. J. Jenkins, G. L. Abbott, C. P. Butler: Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity, J. Appl. Phys. **32**, 1679 (1961) 244

[7] P. Cielo, L. A. Utracki, M. Lamontagne: Thermal diffusivity measurements by the converging-thermal-wave technique, Canadian J. Phys. **64**, 1172–1177 (1986) 244

[8] F. Enguehard, D. Boscher, A. Deom, D. Balageas: Measurement of the thermal radial diffusivity of anisotropic materials by the converging thermal wave technique, Mater. Sci. Eng. B: Solid State Materials for Advanced Technology **5**, 127–134 (1990) 244

[9] O. W. Kading, H. Skurk, K. E. Goodson: Thermal conduction in metallized silicon dioxide layers on silicon, Appl. Phys. Lett. **65**, 1629–1631 (1994) 244

[10] A. Mandelis, P. Hess: Progress in photothermal and photoacoustic science and technology. Vol. 6, in *Semiconductors and Electronic Materials*, vol. PM74 (SPIE Press 1994) 247

[11] F. Formanek, Y. De Wilde, L. Aigouy, W. K. Kwok, L. Paulius, Y. Chen: Nanometer-scale probing of optical and thermal near-fields with an aperture-less NSOM, Superlatt. Microstruct. **35**, 315–323 (2004) 249
[12] G. Abstreiter: Micro-Raman spectroscopy for characterization of semiconductor devices, Appl. Surf. Sci. **50**, 73–78 (1991) 249
[13] L. Pavesi, M. Guzzi: Photoluminescence of $Al_xGa_{1-x}As$ alloys, J. Appl. Phys. **75**, 4779–4842 (1994) 249
[14] M. Cazzanelli, C. Vinegoni, L. Pavesi: Temperature dependence of the photoluminescence of all-porous-silicon optical microcavities, J. Appl. Phys. **85**, 1760–1764 (1999) 249
[15] E. D. Palik: *Handbook of Optical Constants*, J. Opt. Soc. Am. A: Optics Image Science and Vision **1** (1984) pp. 1297–1297 250
[16] G. Tessier, S. Hole, D. Fournier: Ultraviolet illumination thermoreflectance for the temperature mapping of integrated circuits, Opt. Lett. **28**, 875–877 (2003) 250
[17] W. J. Scouler: Temperature-modulated reflectance of gold from 2 to 10 eV, Phys. Rev. Lett. **18**, 445 (1967) 250
[18] R. Rosei, D. W. Lynch: Thermomodulation spectra of Al, Au, and Cu, Phys. Rev. B **5**, 3883 (1972) 250
[19] G. Tessier, G. Jerosolimski, S. Hole, D. Fournier, C. Filloy: Measuring and predicting the thermoreflectance sensitivity as a function of wavelength on encapsulated materials, Rev. Sci. Instrum. **74**, 495–499 (2003) 251
[20] R. Abid, F. Miserey, F. Z. Mezroua: Temperature effect on oxidized silicon reflectivity: Experimental determination of the relative sensitivity; Application to temperature non-contact measurements on the surface of a GTO thyristor in commutation, J. de Phys. III **6**, 279–300 (1996) 251
[21] V. Quintard, G. Deboy, S. Dilhaire, D. Lewis, T. Phan, W. Claeys: Laser beam thermography of circuits in the particular case of passivated semiconductors, Microelectron. Eng. **31**, 291–298 (1996) 251
[22] G. Tessier, S. Hole, D. Fournier: Quantitative thermal imaging by synchronous thermoreflectance with optimized illumination wavelengths, Appl. Phys. Lett. **78**, 2267–2269 (2001) 251, 252
[23] O. B. Wright, R. L. Voti, O. Matsuda, M. C. Larciprete, C. Sibilia, M. Bertolotti: Photothermal probing of inhomogeneously modulated transparent thin films, J. Appl. Phys. **91**, 5002–5009 (2002) 251
[24] S. Dilhaire, S. Grauby, W. Claeys: Calibration procedure for temperature measurements by thermoreflectance under high magnification conditions, Appl. Phys. Lett. **84**, 822–824 (2004) 251, 252, 253, 255
[25] G. Tessier, A. Salhi, Y. Rouillard, F. Genty, J.-P. Roger, D. Fournier: Quantitative thermal imaging of GaInAsSb/AlGaAsSb laser diodes by thermoreflectance, presented at the ICPPP 13, Rio de Janeiro (2004) 253
[26] A. Rosencwaig, J. Opsal, W. L. Smith, D. L. Willenborg: Detection of thermal waves through optical reflectance, Appl. Phys. Lett. **46**, 1013–1015 (1985) 253
[27] L. Pottier: Micrometer scale visualization of thermal waves by photoreflectance microscopy, Appl. Phys. Lett. **64**, 1618–1619 (1994) 253, 254
[28] T. Velinov: On analysis of signals from a photothermal microscope, Meas. Sci. Technol. **6**, 28–32 (1995) 255

[29] P. Gleyzes, F. Guernet, A. C. Boccara: Picometric profilometry 2. Multidetector approach and multiplexed lock-in detection, J. Opt. **26**, 251–265 (1995) 256, 257
[30] C. Filloy, G. Tessier, S. Hole, G. Jerosolimski, D. Fournier: The contribution of thermoreflectance to high resolution thermal mapping, Sensor Rev. **23**, 35–39 (2003) 257
[31] J. P. Monchalin: Heterodyne interferometric laser probe to measure continuous ultrasonic displacements, Rev. Sci. Instrum. **56**, 543–546 (1985) 258, 259
[32] D. Royer, E. Dieulesaint: Optical probing of the mechanical impulse response of a transducer, Appl. Phys. Lett. **49**, 1056–1058 (1986) 258, 259, 260
[33] D. Royer, E. Dieulesaint, P. Leclaire: Remote sensing of the thickness of hollow cylinders from optical excitation and detection of lamb waves, in (IEEE Ultrasonics symposium, Montreal 1989) 258, 260
[34] B. Cretin, D. Hauden: Joule displacement and thermoelastic microscopies of electronic components with optical probing, presented at the ESREF, Bordeaux (1989) 258, 259, 261
[35] Y. Martin, E. A. Ash: Photo-displacement microscopy using a semiconductor laser, Electron. Lett. **18**, 763–764 (1982) 258
[36] J. E. Rothenberg: Observation of the transient expansion of heated surfaces by picosecond photothermal deflection spectroscopy, Opt. Lett. **13**, 713–715 (1988) 258
[37] H. K. Wickramasinghe, Y. Martin, S. Ball, E. A. Ash: Thermodisplacement imaging of current in thin-film circuits, Electron. Lett. **18**, 700–701 (1982) 258
[38] B. Cretin, D. Hauden: Thermoacoustic scanning microscope using a laser probe, in (Ultrasonic Symposium 1984) 259
[39] B. Cretin, W. X. Xie, S. Wang, D. Hauden: Practical limitations and improvements, Opt. Commun. **65**, 157–162 (1988) 259
[40] W. Claeys, S. Dilhaire, E. Schaub: Laser probing techniques and methods for the thermal characterization of microelectronic components, in *Thermal Management of Electronic Systems* (Elsevier, Paris 1998) pp. 227–237 265
[41] W. Claeys, S. Dilhaire, S. Jorez, L. D. Patino-Lopez: Laser probes for the thermal and thermomechanical characterisation of microelectronic devices, Microelectron. J. **32**, 891–898 (2001) 265
[42] E. Laffon, D. Dulon, C. Aurousseau, S. Dilhaire, W. Claeys: Organic material concentration in auditory outer hair cells measured by laser interferometry, Cytometry **20**, 1–6 (1995) 267
[43] E. Laffon, S. Dilhaire, J. L. Leveque, P. Corcuff: An improved technique for optical interferometric imaging of isolated cells, Cytometry **24**, 93–96 (1996) 267
[44] K. Nassim, L. Joannes, A. Cornet, S. Dilhaire, E. Schaub, W. Claeys: Thermomechanical deformation imaging of power devices by electronic speckle pattern interferometry (ESPI), Microelectron. Reliab. **38**, 1341–1345 (1998) 269, 270
[45] B. C. Li, L. Pottier, J. P. Roger, D. Fournier, K. Watari, K. Hirao: Measuring the anisotropic thermal diffusivity of silicon nitride grains by thermoreflectance microscopy, J. European Ceramic Soc. **19**, 1631–1639 (1999) 271
[46] C. Pelissonnier-Grosjean, D. Fournier, A. Thorel: Thermal resistance of grain boundary interfaces in polycrystalline aluminum nitride, J. de Phys. IV **9**, 201–206 (1999) 272

[47] I. Mica, M. L. Polignano, G. Carnevale, P. Ghezzi, M. Brambilla, F. Cazzaniga, M. Martinelli, G. Pavia, E. Bonera: Crystal defects and junction properties in the evolution of device fabrication technology, J. Phys.: Condens. Matter **14**, 13403–13410 (2002) 275
[48] S. Dilhaire, S. Grauby, S. Jorez, L. D. Patino-Lopez, E. Schaub, W. Claeys: Laser diode COFD analysis by thermoreflectance microscopy, Microelectron. Reliab. **41**, 1597–1601 (2001) 275
[49] S. Dilhaire, S. Jorez, L. D. Patino-Lopez, W. Claeys, E. Schaub: Laser diode light efficiency determination by thermoreflectance microscopy, Microelectr. J. **32**, 899–901 (2001) 275
[50] J. Laconte, C. Dupont, A. Akheyar, J.-P. Raskin, D. Flandre: *Symposium on Designe, Test, Integration and Packaging of MEMS and MOEMS* (Cannes-Mandelieu 2002) 276
[51] T. Phan, S. Dilhaire, V. Quintard, W. Claeys, J. C. Batsale: Thermoreflectance measurements of transient temperature upon integrated circuits: Application to thermal conductivity identification, Microelectron. J. **29**, 181–190 (1998) 279
[52] T. Phan, S. Dilhaire, J. C. Batsale, V. Quintard, W. Claeys: Laser probing determination of the thermal conductivity of integrated circuits dielectric layers, High Temp. High Press. **29**, 81–88 (1997) 279
[53] M. Deschamps, O. Poncelet, S. Dilhaire, W. Claeys: Surface acoustic waves at the vacuum–thermoviscoelastic solid interface, Ultrasonics **37**, 677–680 (2000)

Index

Hybrid Techniques and Multipurpose Microscopes

Bernard Cretin and Pascal Vairac

FEMTO-ST Institute/Department LPMO, 32 Avenue de l'Observatoire,
25044 Besançon Cedex
{bcretin,pvairac}@femto-st.fr

Abstract. In this Chapter we discuss microscopes able to achieve submicron resolution using thermoelastic effects. Section 1 reviews the physical effects that can be exploited, while Sect. 2 uses a 3D model to illustrate the main features of this kind of microscopy, describing the phenomenon of super-resolution common to all near-field imaging techniques. Section 3 then discusses several hybrid microscopes and Sect. 4 describes the prospects for a technique that is still in its infancy.

1 Physics of Microscopes Combining Thermal and Thermoelastic Effects

As we saw in Sect. 1.1 of the Chapter by *Cretin* et al. in this volume, heat transfer phenomena, which are diffusion phenomena, are naturally comparable with the evanescent fields to be found in the various forms of near-field microscopy. Several phenomena are often coupled together and it is sometimes difficult to determine the origin of the various contributions making up the results. For instance, a SNOM heats the surface and thereby behaves as a scanning thermal microscope (SThM).

For a phenomenological approach, we consider the basic far-field configuration [1]. We assume that the sample is heated by a modulated laser. (This could be another kind of beam, such as an electron beam, but this type of excitation makes it impossible to work at ambient pressure.) The physical processes coming into play and the main sensors used to detect their effects are shown schematically in Fig. 1.

The thermal wave produced (dynamical solution of the heat diffusion equation) can be detected directly by a microthermocouple [2] or a pyroelectric sensor [3], but the latter form of detection only works correctly in transmission, i.e., coupling is only possible with reflected radiation from the source. To obtain an image, the surface of the object has to be scanned, usually electromechanically. The following techniques have been used successfully:

– Detection of the thermoacoustic signal in the surrounding gas using a microphone. This method follows directly from the use of the photoacoustic effect discovered by Bell in 1881. It is very sensitive provided that one uses a photoacoustic cell [4–6] leading to a resonance at the working

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 287–307 (2007)

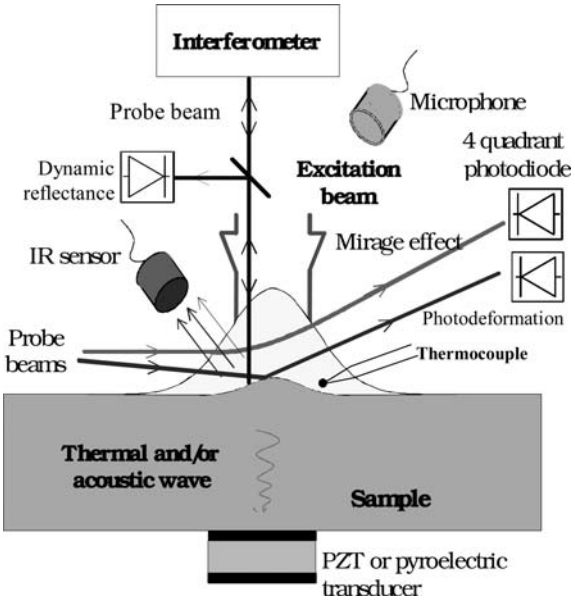

Fig. 1. Photothermal effect, secondary effects, and detection method

frequency. The microphone nevertheless limits the frequency band to a maximum of the order of 50 kHz.

– Radiometric detection [7–9]. Modulation of the infrared emission from the material reflects the dynamical evolution of the temperature since, according to Stefan's law for a black body (a correction has to be applied in practice)

$$\Delta W = 4\sigma T^3 \Delta T\,.$$

An infrared (IR) detector allows one to determine the temperature (dynamical component) and also the thermal conduction properties.

– Use of the mirage effect [10–13]. This method uses the fact that a light beam is deflected when it passes through a region where the gas is heated, due to variations in the refractive index. In this case, a second beam, parallel to the sample surface, is required to carry out the detection. The method is limited physically to low modulation frequencies since it is sensitive to the temperature gradient near the heated surface (probe beam at grazing incidence).

– Measurement of reflectance variations [14–16]. The optical reflection coefficient of a material depends on the surface temperature of the sample. The relative reflectance variation is related to the temperature by

$$\frac{\Delta R}{R} = \frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}T}\Delta T\,,$$

where $\mathrm{d}R/\mathrm{d}T$ is a temperature coefficient that depends on the material. This method uses a light beam that reflects off the sample surface near the heated point. Normal incidence is required for fine-tuning purposes.

– Use of thermodeformation [17]. Under heating, the sample surface expands and deforms with an amplitude that depends on the thermal properties and expansion. Let α_T be the expansion coefficient of the material and $T(z,t)$ the temperature of the sample in a 1D model. Then the motion of the sample surface can be expressed in the form

$$u(0,t) = \int_0^\infty \alpha_\mathrm{T} T(z,t)\,\mathrm{d}z\,.$$

In this 1D model, the displacement varies as $1/f$, favouring low frequency operation. The information obtained is more complete (including the thermomechanical effect), but at the same time more difficult to interpret. The probe beam can be simply deflected, detecting its deviation by a quadrant photodiode, or its phase can be detected using an interferometric technique. In the first case, the device is sensitive to the local slope of the surface (gradient technique), whereas in the second it is the normal displacement at the surface that is measured. Note that only interferometry can provide an absolute displacement measurement that is reliable over time, since a variation in the laser power or local reflectance modifies the result obtained with an optical deflectometer.

– Thermoelastic or thermoacoustic measurements [18,19]. Dynamical heating involves high frequency components which, through expansion, produce elastic waves that propagate through the sample. The waves can be directly converted into electrical signals using a piezoelectric sensor, or detected without contact using an interferometer. Note that the thermoacoustic method favours acoustic information at high frequencies. The microscope then behaves as an acoustic microscope with thermal excitation. Frequencies in the range 10–200 MHz are then used.

Table 1 summarises the different possibilities for detection and their characteristics in the far field. The lateral resolution is not specified in the table because it depends on the excitation beam, e.g., using a laser focused by a microscope objective, a resolution of 500 nm is easily achieved. In transmission, detection is made on the opposite side of the sample. This kind of detection is not ideal because it requires the sample to have a specific type of geometry.

Combined microscopes using several of these detection methods at the same time, e.g., simultaneous thermal and thermoelastic detection, have only been developed since the mid-1990s, due to their complexity. In going from the far field to the near field, several options become possible:

– Reduction of source dimensions, e.g., replacing the focused laser beam by a tapered optical fibre with very small radius of curvature at the apex (SNOM type).

Table 1. Possible means of detection. To give an order of magnitude, a displacement resolution of $10^{-15}\,\mathrm{m/Hz^{1/2}}$ corresponds to a typical temperature resolution of $10^{-5}\,\mathrm{K/Hz^{1/2}}$

Detection method	Detected contributions	Mode of operation	Characteristics (pass band, resolution)
Piezoelectric	Thermal, thermomechanical	Transmission	$\sim 100\,\mathrm{MHz}$ $10^{-17}\,\mathrm{m/Hz^{1/2}}$ (at resonance)
Photoacoustic (microphone)	Thermal, thermomechanical	Reflection (closed chamber)	$\sim 10\,\mathrm{kHz}$ $10^{-5}\,\mathrm{K/Hz^{1/2}}$
Thermo-reflectance	Optical, thermal	Reflection	$\sim 100\,\mathrm{MHz}$ $\sim 1\,\mu\mathrm{m}$, $10^{-3}\,\mathrm{K/Hz^{1/2}}$
Mirage effect	Optical, thermal	Reflection	$\sim 10\,\mathrm{kHz}$ $10^{-7}\,\mathrm{K/Hz^{1/2}}$
Thermo-deformation	Optical, thermal, thermomechanical	Reflection	$\sim 10\,\mathrm{MHz}$ $10^{-14}\,\mathrm{m/Hz^{1/2}}$
Pyroelectric	Thermal	Transmission	$\sim 100\,\mathrm{kHz}$ Variable resolution
Thermocouple	Thermal	Reflection	$\sim 100\,\mathrm{Hz}$ $10^{-4}\,\mathrm{K/Hz^{1/2}}$
Radiometric	Optical, thermal	Reflection, transmission	$\sim 1\,\mathrm{GHz}$ $10^{-4}\,\mathrm{K/Hz^{1/2}}$
Interferometric	Optical, thermal, thermomechanical	Reflection, transmission	$\sim 100\,\mathrm{MHz}$ $10^{-15}\,\mathrm{m/Hz^{1/2}}$

– Reduction of the sensor dimensions. A local probe of the kind used in STM or AFM can detect thermomechanically generated deformations, or a thermal microsensor can be placed on an AFM tip.

These microscopes will be discussed in the next section.

2 Microscopes and Their Resolutions

As we have seen, the thermal and mechanical aspects of detection are two important features to be taken into account (the optical resolution is obvious here). We shall thus be concerned with these two features. To estimate the resolution, the thermoelastic conversion has to be modelled. Indeed, many models have been devised to do this, both in 1D [20,21] and in 3D [22,23]. We shall simplify the thermoelastic model, complex by its very nature, as far as possible, by assuming cylindrical symmetry and using integral transforms [24, 25] which reduce the complexity of the calculation.

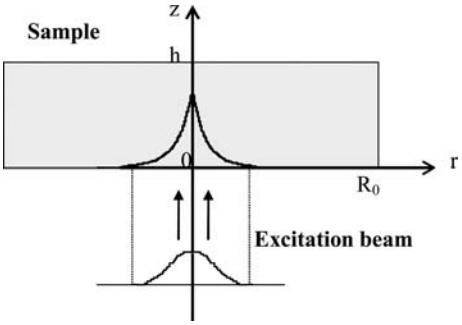

Fig. 2. Geometry of the photothermal excitation model

2.1 3D Model with Cylindrical Symmetry

In order to interpret the mechanisms of photothermoacoustic generation in thermoelastic microscopy, a 3D model is required [26]. Indeed, the heat source is extremely small compared with the dimensions of the sample. To analyse the behaviour of the material, we shall make the following simplifying assumptions:

- The material is homogeneous and isotropic, with linear behaviour and radius $R_0 \gg h$ for thickness h.
- The excitation has Gaussian profile, a rough approximation in the near field, but one that makes it possible to use cylindrically symmetric integral transforms.
- The excitation occurs at the centre of the object at a fixed frequency f.
- Inertial effects are neglected, which is perfectly realistic at low frequencies.

The Hankel transformation, which is so useful in cylindrical geometry (see Fig. 2), is given by

$$F(s) = \int_0^\infty rJ_0(sr)f_r(r)\,\mathrm{d}r = TH_0\big[f_r(r)\big]\,.$$

To carry out the temperature calculation, we separate the space and time variables, viz., $\Theta(r,\theta,z,t) = T(r,\theta,z)\mathrm{e}^{\mathrm{i}\omega t}$ and calculate T. The heat diffusion equation is

$$\nabla^2\Theta - a\frac{\partial\Theta}{\partial t} = -P_\mathrm{c}\alpha\exp\left(-\frac{r^2}{w^2}\right)\exp(-\alpha z)\exp(\mathrm{i}\omega t)\,,$$

where w is the radius of the excitation at $1/\mathrm{e}$, α is the reciprocal of the penetration depth of the excitation beam, and P_c is a constant to be specified later. This is solved using the zero order Hankel transform with boundary conditions

$$\frac{\partial T}{\partial z} = 0 \quad \text{at} \quad z = 0 \quad \text{and} \quad z = h\,,$$

i.e., zero dynamic flux on the upper and lower faces. The full expression for the temperature field is given by the sum of the particular and the general solutions:

$$T = K' \int_0^\infty sJ_0(sr)\,\mathrm{d}s \frac{\mathrm{e}^{-w^2s^2/4}}{\alpha^2 - s^2 - \nu^2} \times \left\{ \mathrm{e}^{-\alpha z} - \frac{\alpha \cosh\left[(h-z)\sqrt{s^2+\nu^2}\,\right]}{\sqrt{s^2+\nu^2}\sinh\left(h\sqrt{s^2+\nu^2}\right)} \right\},$$

where

$$K' = -\frac{P_\mathrm{c}\alpha w^2}{2}, \quad P_\mathrm{c} = \frac{4P_\mathrm{t}(1-R)}{\pi^2\kappa w^2}.$$

The calculation of the displacement is considerably simplified by using the Youngdahl stress functions which lead to the thermoelastic equations [25]

$$\Delta\Omega = \frac{1}{2(1-\nu)}\left[-\frac{\partial^2\Psi}{\partial z^2} + 2(1+\nu)\alpha_\mathrm{T}\Theta\right], \quad \Delta\Psi = 0,$$

where ν is the Poisson ratio. The constants appearing in the homogeneous solutions of the last system are found from the boundary conditions

$$T_{zz}^{(\mathrm{D})} = 0, \quad T_{rz}^{(\mathrm{D})} = 0 \quad \text{at} \quad z = 0 \quad \text{and} \quad z = h,$$

expressing the fact that the dynamic strains $T_{zz}^{(\mathrm{D})}$ and $T_{rz}^{(\mathrm{D})}$ are zero at the surface. The details of the calculation are given in [26]. The final result for the Hankel transform of the displacement of the sample surface at $z = 0$ is

$$\left.\frac{\tilde{u}_z^{(\mathrm{D})}}{C_0}\right|_{z=0} = -s\frac{3-4\nu}{4(1-\nu)}P(s) + s\frac{3-4\nu}{4(1-\nu)}Q(s) - sG(s) + sH(s) - \frac{\alpha \mathrm{e}^{-w^2s^2/4}}{\alpha^2 - s^2 - \nu^2}\left(\frac{1}{\alpha^2 - s^2} - \frac{1}{\nu^2}\right),$$

where the polynomials $P(s)$, $Q(s)$, $G(s)$, and $H(s)$ are calculated from the boundary conditions, and C_0 is a constant. The interesting solution to the problem, i.e., normal surface displacement, is calculated numerically from the relation

$$u_z^{(\mathrm{D})} = C_0 \int_0^\infty sJ_0(sr)\frac{\tilde{u}_z^{(\mathrm{D})}}{C_0}\,\mathrm{d}s.$$

We have used this model to calculate the temperature and displacement fields in a metal, viz., aluminium, which is weakly anisotropic. The normalised temperature is plotted in dimensionless coordinates (variable r/w, where r is the radial coordinate) in Fig. 3. The excitation frequency f has been fixed

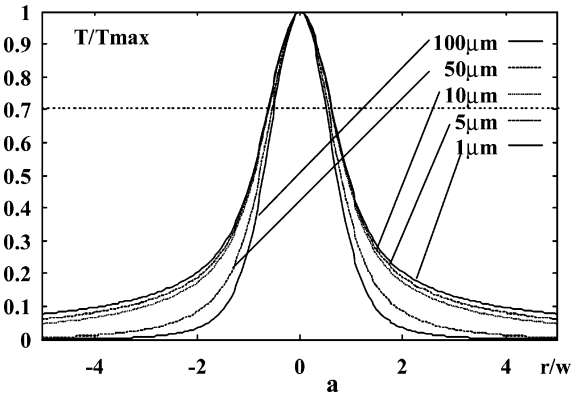

Fig. 3. Radial temperature field in dimensionless coordinates r/w for several values of the source radius w. Excitation frequency 1 kHz. Material Al

at 1 kHz, guaranteeing operation in super-resolution mode [27], which is of near-field type. At this frequency, the thermal diffusivity μ_{T} in aluminium is around 150 µm. The model was applied for various values of the source radius in order to bring out the scaling effect. It is clear from the curves that the classic behaviour of exponential type becomes less marked as the source radius decreases below the value of μ_{T}.

The second important feature of the analysis concerns the resolution. By analogy with the classic resolution criteria of Rayleigh or Sparrow, for example, we define the resolution from a −3 dB criterion. This value may seem somewhat arbitrary, but it leads to a physically acceptable resolution, roughly equal to the radius of the excitation source, as can be seen in the figure. (Note that this criterion is defined here relative to the field amplitude. In optics, the square of the field is usually used.) Clearly, for very small radii compared with the thermal diffusivity, energy dispersion dominates and, for this reason, the resolution is constant in dimensionless variables. This is a fundamental point, justifying the reduction of the source dimensions in order to obtain a high resolution. Note that the thermoelastic resolution for its part is strongly dependent on the frequency. This point is not discussed here but can be found in [27].

The model can also be used to calculate the temperature distribution inside the sample in order to determine another fundamental imaging parameter, namely the investigation depth. We define this depth from the conventional definition (depth at 1/e) which results from the 1D analysis and many experiments carried out with thermal images [28–30]. It is interesting to note the differences between the levels chosen for the investigation depth and resolution criteria. These values result from the physical model and the history of microscopy. Figure 4 shows the isolevel curves at 1/e for temperature fields and displacement fields, for the same source radius (1 µm in this

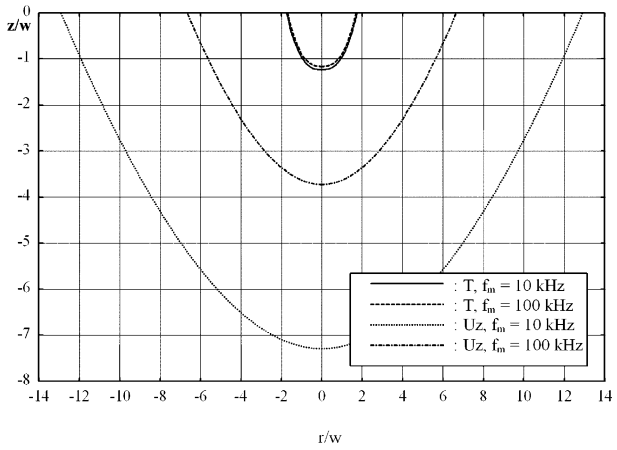

Fig. 4. Profiles at 1/e of the temperature and normal displacement fields as a function of r and z for a source with Gaussian profile of radius 1 µm heating an aluminium sample at different modulation frequencies f_{m}

case, well below the value of μ_{T}) and for different excitation frequencies. As can be seen from Fig. 3, the isothermal curves are superposed, whereas the isodisplacement curves are very different and hence depend on the excitation frequency. The latter must therefore be chosen high enough to guarantee a surface effect. (The curve at 1 kHz comes out of the figure and is not therefore shown here.) The difference between these two types of behaviour is explained by the effect of integrating the expansion over the whole dynamically heated region. (The investigation depth clearly involves the value of μ_{T}.)

We have seen that the thermal resolution R_{T} is approximately equal to w, the source dimension, in super-resolution mode. Moreover, R_{T} is independent of the frequency as long as the source dimension is small enough compared with the diffusion length. If we consider a submicron source (the model is not appropriate for near-field optical excitation, but the thermal model is valid down to the nanoscale), we may expect the same behaviour.

Super-resolution has also been demonstrated experimentally. Figure 5 shows images obtained with an argon (Ar^+) laser focused on a polycrystalline nickel sample. The region scanned is $256\,\mu\mathrm{m} \times 256\,\mu\mathrm{m}^2$, with each pixel measuring 1 µm. The excitation frequency was fixed at 80 kHz. (Any other frequency could have been chosen. At this frequency, the diffusion length in nickel is around 10 µm and super-resolution is achieved.) The normal displacement was measured with an optical interferometer [31]. Amplitude images are shown for two optical magnifications, viz., ×20 and ×100. In the image obtained with the ×100 objective, the contrast is rather limited. As we have seen, this follows from the smaller investigation depth. For this magnification, no contrast can be seen in the phase image (not shown here), as predicted by the model when the spot is much smaller than the thermal diffusion length.

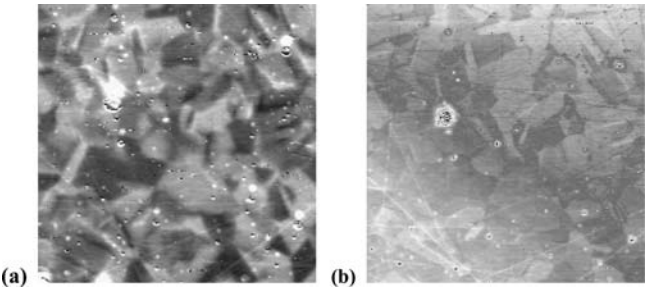

Fig. 5. Super-resolution in thermoelastic amplitude images of a polycrystalline nickel sample. (**a**) ×20 objective. (**b**) ×100 objective

When focused with the ×20 microscope objective, the excitation beam has radius approximately 4 μm (compared with an estimated 0.7 μm for the ×100 objective). The amplitude of the displacement normal to the surface is of the order of 1 nm. This value is easily detected with any laser probe that has a reasonable signal-to-noise ratio. The high resolution interferometer used, with a typical vertical resolution of 3 fm/$Hz^{1/2}$, was able to obtain very high quality images. Metal samples are particularly interesting to observe because thermomechanical anisotropy is a source of contrast within grains [32, 33].

3 Combined Photothermoelastic Microscopy

With the advent of the scanning thermal microscope, many other microscopes were then developed to tackle the problem of non-destructive evaluation (NDE), or for combined metrology. These photothermal and photothermoelastic microscopes of near-field type derive from the three main families of near-field microscope: STM, AFM, and SNOM. However, a more global classification can be based on the physical features of the local probe used in the microscope, as we shall now describe.

3.1 Microscopes Based on a Thermoelectric Probe

Techniques deriving from SNOM [34,35] are particularly simple to implement since tapered optical fibres (which may be metallised at the end) receive a high power density at their apex and thereby implicitly constitute a source of heat [36]. Thermal effects can thus be detected in all near-field optical microscopes using a fibre with metallised tip. Several variants have been investigated. For example, *Goodson* et al. [37] called their method by the name of near-field optical thermometry (NFOT). A laser beam is injected into the optical fibre. At the other end of the fibre, the tip is held at a constant distance from the sample by the standard shear-force method (frictional motion

produced by a piezoelectric tuning fork). Light leaving the fibre and interacting with the sample is collected by a microscope objective and detected by a photomultiplier tube (PMT). The amplitude and phase of the signal are extracted using a lock-in amplifier. In this experiment, the sample is heated by the Joule effect and the method is thus limited to conducting samples like STM. At least four configurations can be applied in the near-field thermo-optical techniques:

- Detection in transmission mode (through the sample), with light collection in the far field.
- Measurement of light reflected from the sample, with light collection in the far field.
- Detection of light going back into the fibre after interaction with the sample.
- Measurement of the infrared wave emitted by the sample following heating (emission of black body type) and collected by a fibre that is transparent to infrared wavelengths.

The microscope developed at LPMO [38] is shown in Fig. 6. This incorporates a highly sensitive photodetector to pick up reflectance variations in the near field (by transformation of evanescent waves into propagative waves as a result of the near-field optical coupling). In parallel, a piezoelectric sensor placed under the sample (PZT tube) for the purpose of topographic servo control is also used to obtain thermoacoustic information in the near field. To do this, the piezoelectric signal is acquired at the modulation frequency of the high power laser beam. This microscope is called the multi-acquisition scanning thermoacoustic microscope (M-SThAM). Six images can be acquired simultaneously:

- an optical near-field image,
- a topographic near-field image,
- two photothermal near-field images (amplitude and phase),
- two photothermoacoustic near-field images (amplitude and phase).

Figure 6 shows the setup schematically.

In this device, the near-field photothermoelastic source is an intensity-modulated laser whose beam, injected into a tapered optical fibre, heats the sample locally. In a similar way to the far-field device, the excitation laser beam is produced by a high power laser (Ar^+). The intensity is modulated by an acousto-optical modulator with control electronics piloted by a low frequency generator which delivers a voltage at the frequency $f_{\text{modulation}}$. The excitation laser beam is thus reflected by the dichroic mirror and then injected into the optical fibre after being focused by a microscope objective. The optical energy injected into the fibre propagates to the end where it becomes the near-field photothermoelastic source. Local dynamical heating of the sample spreads, as we have seen theoretically, over a region with dimensions of the same order as the terminal diameter of the tapered fibre, i.e.,

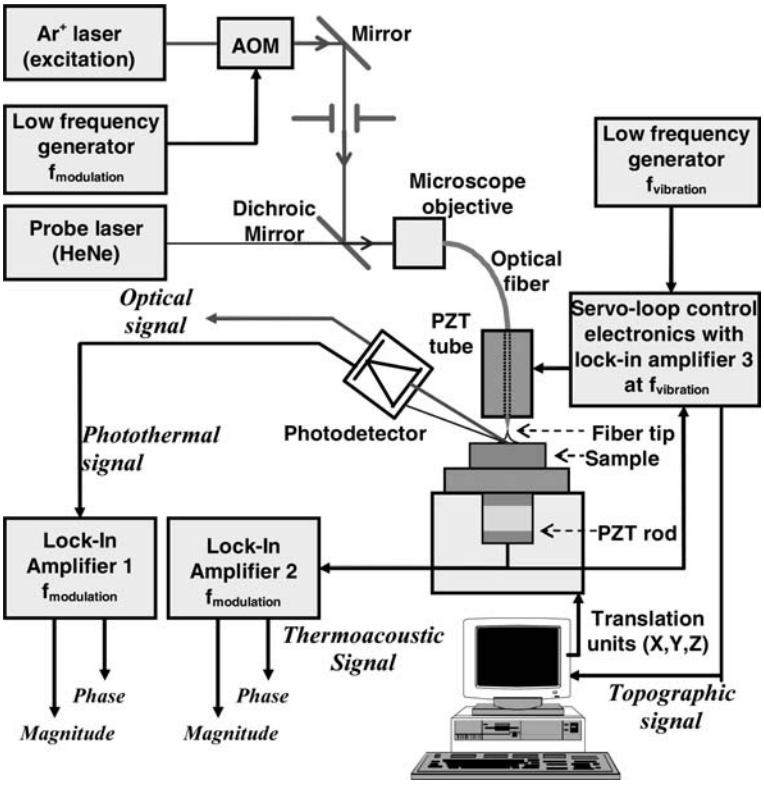

Fig. 6. Configuration of the M-SThAM

50–200 nm. This local heating modulates the local reflectance of the sample and generates thermoacoustic waves which are detected in transmission.

Thermoacoustic vibrations at the modulation frequency $f_{\text{modulation}}$ (LF generator) are detected by a piezoelectric sensor placed below the sample. The signal obtained in this way is sent to the lock-in amplifier 2 in order to extract the amplitude (energy part) and phase (propagative part) of the thermoacoustic signal. Photothermal detection is carried out by a sensitive (far-field) photodetector equipped with an interference filter with central wavelength equal to the wavelength of the HeNe laser beam. This probe beam is also injected into the optical fibre and used to detect both the optical signal (static reflectance) and the photothermal signal by transforming evanescent waves into propagative waves via the near-field coupling. Photothermal information is provided by the photodetector, its associated electronics, and the lock-in amplifier 1.

In order to separate thermoacoustic or photothermal information from topographic information, we have designed a topographic servo control device for SNOM-type fibres. (In near-field thermoelastic microscopy, control of the

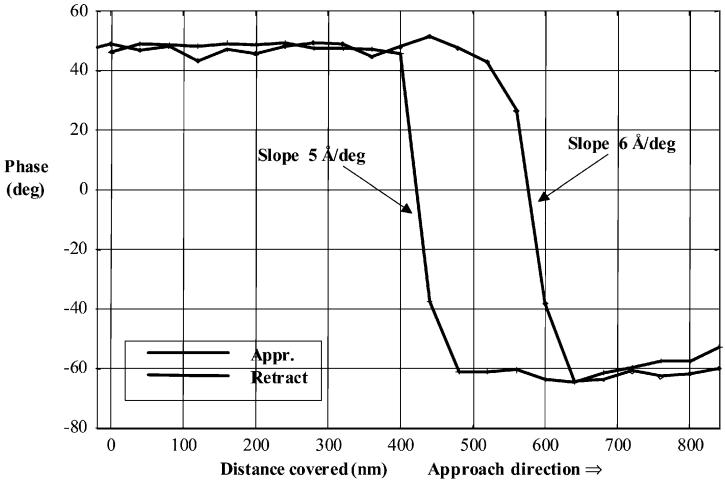

Fig. 7. Phase of the signal detected by acoustic transmission beneath the sample, and hysteresis. Frequency 10.9 kHz. Vibration amplitude 2 nm peak-to-peak. Tip dimensions 100 nm (optical fibre)

tip–sample separation is used to eliminate the effects of interference fringes occurring between the end of the fibre and the sample surface.) Now most near-field optical microscopes use the shear-force technique which is based on a transverse vibration of the fibre and which therefore, by friction, leads to a further elastic effect. A new method has been devised, inspired directly by work on near-field acoustic transmission microscopy undertaken at LPMO: the perturbation generated by a tip with vertical vibrational motion in quasi-contact conditions with a plane surface is transmitted to the lower face of the sample. Figure 7 shows the phase of the acoustic transfer function for the vibrating fibre–sample–piezoelectric transducer system.

As the vibrating tip approaches the sample, several phenomena arise in the signal detected beneath the sample. With vertical vibration, the tip induces:

- elastic (contact) forces, mainly Hertz and adhesive forces,
- viscoelastic (non-contact) forces due to coupling with the air.

In the harmonic regime, this is expressed theoretically by a phase rotation of 90° in the quasi-contact region. The servo control uses this phase sensitivity of the acoustic interaction, the phase being detected by the lock-in amplifier 3 in Fig. 6. It allows one to obtain a vertical resolution of around 2 nm over a range of 2 µm. The microscope has obtained sets of images with resolutions of the order of 50 nm. The images shown in Fig. 8 were obtained on a porous silicon sample coated with an aluminium layer of thickness about 1 µm. The image size is $2 \times 2\,\mu m^2$. The amplitude of the vibrations at the end of the fibre is 10 nm and the excitation frequency is 7 kHz. The working distance is around 50 nm and the estimated resolution about 80 nm. Note that the

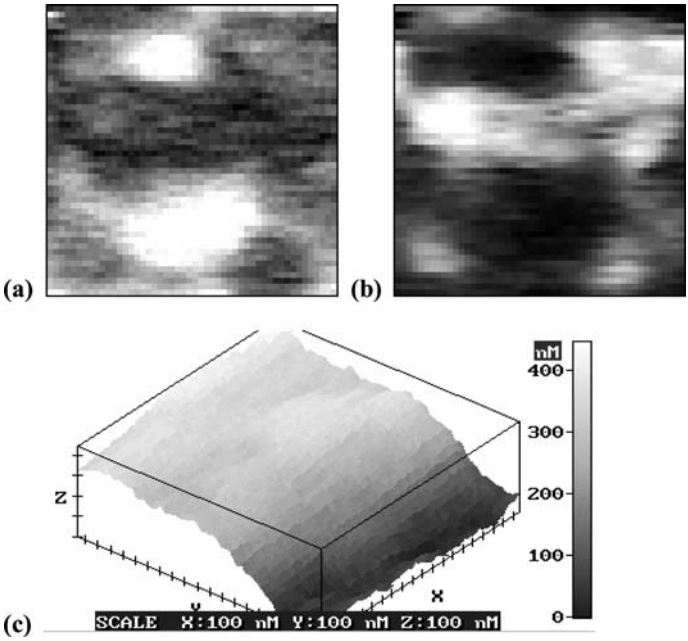

Fig. 8. Images of a porous silicon sample coated with an aluminium layer. Image size $2 \times 2\,\mu m^2$. (**a**) Photothermal amplitude image (7 kHz). (**b**) Photothermal phase image. (**c**) Topographic image

topographical profile does not correspond to the amplitude and phase images which are essentially related to thermal features of the sample.

3.2 Microscopes Based on Detection of Expansion

As we saw in the Chapter by *Cretin* et al. in this volume, AFM and STM microscopes can provide precise measurements of sample topography. If the sample is heated by the Joule effect or by a light beam (other methods can also be used, such as an electron beam), the thermoelastic deformation or vibration can be detected by these near-field microscopes with a lateral resolution that may reach or go below the nanometer. Thermal expansion can be related to the temperature gradient but this technique is not calibrated for absolute temperature (expansion of the tip which is also heated). Various acronyms have been used for this type of microscope. For example, microscopes measuring the expansion due to the Joule effect, called scanning Joule expansion microscopes (SJEM) [39], derive from AFM. The sample is dynamically heated by a current and the normal component of the displacement due to thermal expansion is detected by servo controlling the AFM tip separation. Temperature calibration requires a good understanding of the physical properties of the sample [40]. The thermal expansion of the tip can also be

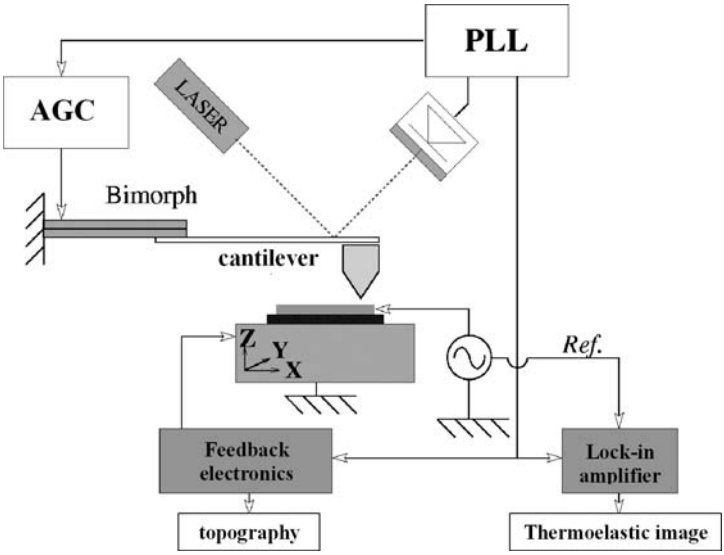

Fig. 9. Scanning thermoelastic microscope (SThEM)

detected because it is coupled to that of the sample [41] and several interdisciplinary methods have been developed [42]. As we saw in the theoretical discussion, the resolution is essentially related to the dimensions of the local probe which behaves as a heat sink. The STM is extremely sensitive to thermal expansion because its vertical resolution is around 1 pm. This interesting behaviour has been successfully used to study local thermal expansion [43]. The hybrid microscope based on the AFM technique (SThEM) [44] measures the local expansion of the sample. The setup is shown schematically in Fig. 9.

The probe is topographically servo controlled at the sample surface by an acoustic method. It is placed at the end of a vibrating beam, excited at its resonance frequency by an oscillator. The resonance frequency of the beam is modified by the interaction between tip and sample. It is this frequency, measured by a phase-locked loop, which constitutes the servo signal. Working distances of the order of 100 nm are then ensured, with a resolution of the order of 10 nm. The system can be used to detect thermomechanical effects by demodulating the servo signal at the excitation frequency of the sample. The resulting images (see Fig. 10) provide local thermoelastic information.

These images include thermal properties, expansion coefficients and the mechanical anisotropy of the structure. The lateral thermoelasticity resolution can reach the same level as the lateral topographic resolution. The images shown in Fig. 10 were obtained with an AFM tip (radius around 20 nm). The sample is an integrated circuit of the type used in an operational amplifier, which has been opened to give access to the microchip. The topographic image was obtained with a cantilever vibrating at 65 kHz. The circuit was

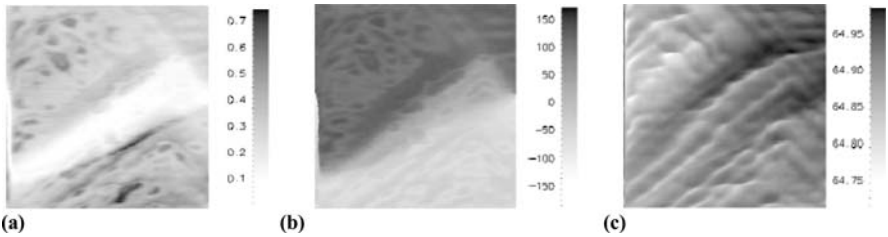

Fig. 10. Scanning thermoelastic microscope (SThEM) images of an AD549 circuit supplied at 1.3 kHz. Image size $100 \times 100\,\mu m^2$. Maximum height difference 200 nm. (**a**) Amplitude (mV). (**b**) Phase (degrees). (**c**) Topography (frequency in kHz)

used electrically at a frequency of 1.3 kHz in operating conditions close to maximum power (integrated current limited to 20 mA in the chip). The thermoelastic image shows the same region after synchronous detection at the frequency of the current. The absence of frequency doubling (Joule effect) is justified by the presence of a continuous component.

4 Prospects

Combined microscopes do of course remain open to improvement. Furthermore, the underlying physics needs to be much better understood, e.g., in order to be sure that measured data uniquely reflects some specific physical phenomenon. On the other hand, all the preliminary stages in the development of these microscopes have now been accomplished and they could soon be made commercially available, like the SThM, which combines the AFM with heat measurements. The most probable applications concern the three fields of nanometrology, nanofabrication and data storage.

The widespread industrial demand, e.g., in the semiconductor industry, nanobiotechnology, and development of materials, for accurate measurements on the nanoscale (submicron spatial resolution in 3 space directions) and simultaneous measurement of the physical properties of materials will inevitably lead to the design of new measurement instruments based on near-field microscopy. An illustration of the trend in the fabrication of ever smaller devices by high technology industries is provided by Moore's law, shown graphically in Fig. 11 [45]. Very high density circuits are likely in the near future. This trend will have repercussions in other areas such as micromechanics, and more generally, in microfabrication, opening the way to nanotechnology.

To satisfy this demand, major research programmes have been set in motion and they have already produced the first tools in this field [46, 47], still specialising in one type of measurement. The main obstacles to the development of these tools are of a physical nature: geometry of the tip and the way it evolves in time, and the many forces (capillary, adhesive, frictional, etc.)

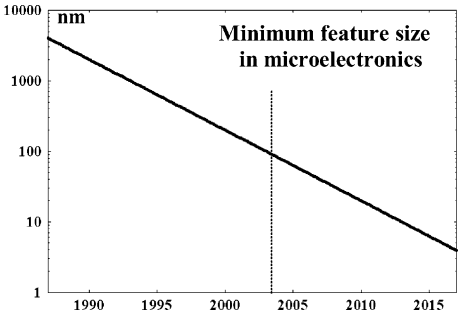

Fig. 11. Moore's law. The trend in microelectronic component sizes

that can come into play to influence measurement results. Another practical problem is the (re)positioning of the sample (3 rotations and 3 translations) independently of drift effects. Finally, one of the limitations of scanning techniques is the unavoidably long aqcuisition time (acquisition frequency less than 1000 sample/s). Regarding local thermal measurements, the stakes are high, because it is one of the few techniques able to access the third dimension, and it proves to be particularly useful for characterising new components such as nanotransistors, carbon nanotubes, data storage devices, and so on, or thin films.

A major challenge for nanotechnology is nanofabrication, naturally carried out with the help of scanning probe microscopes. Once again, a multipurpose microscope is required for manipulation and testing during fabrication. Many techniques have been published, but nanofabrication systems including testing based on SPM are still at the laboratory research stage [47–54].

One application which looks interesting, but which has not yet led to any form of combined microscopy, is data storage. Several demonstrators have been made to validate the principle of thermomechanical read and write methods [46, 47, 55–57]. A specific illustration is the high density AFM data storage technique developed by IBM, Zurich. It uses an array of 32×32 tiny cantilevers, each of which can read and write on a polymer base. When the tips are heated, the surface takes on an uneven aspect which can then be read by AFM. In the latest version of this system, data is stored at a much higher density ($200\,\mathrm{Gbit/in^2}$) than on today's magnetic hard disks. This system, known as Millipede, can now achieve data flow rates of 32 kbit/s, which is still rather slow for computer applications but already acceptable for onboard data systems. It is not clear whether this technique will ever be widely adopted, but it offers good prospects for the development of combined microscopes.

References

[1] D. P. Almond, P. M. Patel: *Photothermal Science and Techniques* (Chapman & Hall, London 1996) 287

[2] L. Thiery, N. Marini, J. P. Prenel, M. Spajer, C. Bainier, D. Courjon: Temperature profile measurements of near-field optical microscopy fiber tips by means of sub-micronic thermocouple, Int. J. Therm. Sci. **39**, 519–525 (2000) 287

[3] D. Royer, E. Dieulesaint, T. Valero: Pyroelectric non-destructive testing of BAW transducers, Ultrasonics Symposium Proceedings pp. 908–911 (1984) 287

[4] A. Rosencwaig: Photoacoustic spectroscopy of solids, Opt. Commun. **7**, 305–308 (1973) 287

[5] A. Rosencwaig, A. Gersho: Theory of the photoacoustic effect with solids, J. Appl. Phys. **47**, 64–69 (1976) 287

[6] G. Busse, A. Rosencwaig: Subsurface imaging with photoacoustics, Appl. Phys. Lett. **36**, 815–816 (1980) 287

[7] R. Santos, L. C. Miranda: Theory of photothermal radiometry with solids, J. Appl. Phys. **52**, 4194–4198 (1981) 288

[8] S. J. Sheard, R. K. Appel, M. G. Somekh: Photothermal radiometric imaging of semiconductors, Electron. Lett. **23**, 227–228 (1987) 288

[9] D. Maillet, J. C. Batsale, A. Bendada, A. Degiovanni: Méthodes intégrales et contrôle non-destructif par thermographie infrarouge stimulée, Revue Générale de Thermique **35**, 14S–27S (1996) 288

[10] A. C. Boccara, D. Fournier, J. Badoz: Thermo-optical spectroscopy: Detection by the mirage effect, Appl. Phys. Lett. **36**, 130–132 (1980) 288

[11] J. C. Murphy, L. C. Aadmot: Photothermal spectroscopy using optical beam probing: Mirage effect, J. Appl. Phys. **51**, 4580–4588 (1980) 288

[12] A. Salazar, A. Sanchez-Lavega, J. Fernandez: Thermal diffusivity measurements on solids using collinear mirage detection, J. Appl. Phys. **74**, 1539–1547 (1993) 288

[13] F. Lepoutre, D. Fournier, A. C. Boccara: Nondestructive control of weldings using mirage detection, J. Appl. Phys. **57**, 1009–1015 (1985) 288

[14] J. Opsal, M. W. Taylor, W. L. Smith, A. Rosencwaig: Temporal behavior of modulated optical reflectance in silicon, J. Appl. Phys. **61**, 240–248 (1987) 288

[15] A. Mandelis, J. F. Power: Frequency-modulated impulse response photothermal detection through optical reflectance. 1: Theory, Appl. Opt. **27**, 3397–3417 (1988) 288

[16] F. Lepoutre, P. Forge, F. C. Chen, D. Balageas: Micronic thermal characterizations of cracks and interfaces in composite materials by photoreflectance, La Recherche Aérospatiale **1**, 39–52 (1994) 288

[17] G. Rousset, L. Bertrand, P. Cielo: A pulsed thermoelastic analysis of photothermal surface displacements in layered materials, J. Appl. Phys. **57**, 4396–4405 (1985) 289

[18] J. C. Murphy, J. W. Maclachlan, L. C. Aamodt: Imaging contrast processes in thermal and thermoacoustic imaging, IEEE Trans. Ultrason., Ferroelec., Freq. Cont. **33**, 529–541 (1986) 289

[19] B. Cretin, D. Hauden: Thermoacoustic microscopy using optical excitation and detection, SPIE Proc. **809**, 64–69 (1987) 289

[20] R. M. White: Generation of elastic waves by transient surface heating, J. Appl. Phys. **34**, 3559–3567 (1963) 290
[21] A. Rosencwaig, A. Gersho: Theory of the photoacoustic effect with solids, J. Appl. Phys. **47**, 64–69 (1976) 290
[22] G. C. Wetsel: Photothermal generation of thermoelastic waves in composite media, IEEE Trans. Ultrason., Ferroelec., Freq. Cont. **33**, 450–461 (1986) 290
[23] W. Jackson, N. M. Amer: Piezoelectric photoacoustic detection: Theory and experiment, J. Appl. Phys. **51**, 3343–3353 (1980) 290
[24] H. Delavault: Transformation de Hankel en coordonnées cylindriques, Publications scientifiques et techniques **71** (1957) 290
[25] C. K. Youngdahl: On the completeness of a set of stress functions appropriate to the solution of elasticity problems, Int. J. Eng. Sci. **7**, 61–79 (1969) 290, 292
[26] B. Cretin, N. Daher, B. Cavallier: Thermoelastic modeling: Application to super-resolution in photothermal and thermoelastic microscopy, in *Proceedings of SPIE Optical Inspection and Micromeasurements II*, SPIE 3098 (1997) pp. 466–475 291, 292
[27] B. Cretin: Super-resolution in photothermal and thermoelastic microscopies: Extension of the near-field concept, Revue Générale de Thermique **37**, 556–564 (1998) 293
[28] A. Rosencwaig: High resolution photoacoustic thermal-wave microscopy, Appl. Phys. Lett. **36**, 725–727 (1980) 293
[29] F. A. McDonald, G. C. Wetsel, C. G. Clark: Effects of frequency on definition and resolution in photothermal imaging of subsurface structure, Ultrasonics Symposium Proc. pp. 672–676 (1983) 293
[30] L. J. Inglehart, K. R. Grice, L. D. Favro, P. K. Kuo: Spatial resolution of thermal wave microscopes, Appl. Phys. Lett. **43**, 446–448 (1983) 293
[31] P. Vairac, B. Cretin: Heterodyne laser probe using a double pass, Opt. Commun. **132**, 19–23 (1996) 294
[32] J. C. Murphy, J. W. Maclachlan, L. C. Aamodt: Imaging contrast processes in thermal and thermoacoustic imaging, IEEE Trans. Ultrason., Ferroelec., Freq. Cont. **33**, 529–541 (1986) 295
[33] B. Cretin, J. Takadoum, A. Mahmoud, D. Hauden: Metallurgical applications of the thermoelastic microscope, Thin Solid Films **209**, 127–131 (1992) 295
[34] U. Durig, D. W. Pohl, F. Rohner: Near-field optical-scanning microscopy, J. Appl. Phys. **59**, 3318–3327 (1986) 295
[35] D. Courjon, K. Sarayeddine, M. Spajer: Scanning tunneling optical microscopy, Opt. Commun. **71**, 23–28 (1989) 295
[36] D. I. Kavaldjiev, R. Toledo-Crow, M. Vaez-Iravani: On the heating of the fiber tip in a near-field scanning optical microscope, Appl. Phys. Lett. **67**, 2771–2773 (1995) 295
[37] K. E. Goodson, M. Asheghi: Near-field optical thermometry, Microscale Thermophys. Eng. **1**, 225–235 (1997) 295
[38] B. Cavallier: *Microscopies photothermiques et thermoélastique conventionnelles et à sonde locale: Théorie et expérimentation*, Ph.D. thesis, University of Franche-Comté, France (2000) 296
[39] J. Varesi, A. Majumdar: Scanning Joule expansion microscopy at nanometer scales, Appl. Phys. Lett. **72**, 37–39 (1998) 299

[40] M. Cannaerts, D. Buntinx, A. Volodin, C. Van Haesendonck: Calibration of a scanning Joule expansion microscope (SJEM), Appl. Phys. A **72**, 67–70 (2001) 299
[41] J. Pelzl, J. Bolte, F. Niebisch, D. Dietzel, H. H. Althaus: New developments in thermal wave microscopy, Anal. Sci. **17**, s53–s56 (2001) Special Issue 300
[42] A. Hammiche, D. M. Price, E. Dupas, G. Mills, A. Kulik, M. Reading, J. M. R. Weaver, H. M. Pollock: Two adaptations of thermomechanical modulation for microscopy: SThEM and dynamic L-TMA (scanning thermal expansion microscopy and dynamic localised thermomechanical analysis), J. Microsc. **199**, 180–190 (2000) 300
[43] N. Trannoy, P. Grossel: Photothermal effects induced by laser excitation in scanning tunneling microscope, Int. J. Thermal Sci. **39**, 532–536 (2000) 300
[44] R. Patois: *Méthodes optiques et acoustiques pour les microscopies thermiques et thermo-élastiques aux échelles micro- et nanométriques*, Ph.D. thesis, University of Franche-Comté, France (2003) 300
[45] G. E. Moore: Cramming more components onto integrated circuits, Electron. **38**, 114–117 (1965) 301
[46] H. J. Mamin, D. Rugar: Thermomechanical writing with an atomic force microscope tip, Appl. Phys. Lett. **61**, 1003–1005 (1992) 301, 302
[47] B. W. Chui, T. D. Stowe, Y. S. Ju, K. E. Goodson, T. W. Kenny, H. J. Mamin, B. D. Terris, R. P. Ried, D. Rugar: Low-stiffness silicon cantilevers with integrated heaters and piezoresistive sensors for high-density afm thermomechanical data storage, IEEE J. Microelectromech. Syst. **7**, 69–78 (1998) 301, 302
[48] R. Held, T. Heinzel, A. P. Studerus, K. Ensslin, M. Holland: Semiconductor quantum point contact fabricated by lithography with an atomic force microscope, Appl. Phys. Lett. **71**, 2689–2691 (1997) 302
[49] D. M. Eigler, E. K. Schweizer: Positioning single atoms with a scanning tunnelling microscope, Nature **344**, 524–526 (1990) 302
[50] E. S. Snow, P. M. Campbell, F. K. Perkins: Nanofabrication with proximal probes, Proc. of the IEEE **85**, 601–611 (1997) 302
[51] C. F. Quate: Manipulation and modification of nanometer scale objects with the STM, in L. Esaki (Ed.): *Highlights in Condensed Matter Physics and Future Prospects* (Plenum, New York 1991) pp. 573–630 302
[52] R. Leach, J. Haycocks, K. Jackson, A. Lewis, S. Oldfield, A. Yacoot: Advances in traceable nanometrology at the National Physical Laboratory, Nanotechnol. **12**, R1–R6 (2001) 302
[53] O. Jusko, X. Zhao, H. Wolf, G. Wilkening: Design and three-dimensional calibration of a measuring scanning tunneling microscope for metrological applications, Rev. Sci. Instrum. **65**, 2514–2518 (1994) 302
[54] R. Berger, C. Gerber, H. P. Lang, J. K. Gimzewski: Micromechanics: A toolbox for femtoscale science: Towards a laboratory on a tip, Microelectron. Eng. **35**, 373–379 (1996) (International Conference on Micro- and Nanofabrication, Glasgow, 1996) 302
[55] E. Betzig, J. K. Trautman, R. Wolfe, E. M. Gyorgy, P. L. Finn, M. H. Kryder, C. H. Chang: Near-field magneto-optics and high density data storage, Appl. Phys. Lett. **61**, 142–144 (1992) 302

[56] M. Lutwyche, C. Andreoli, G. Binnig, J. Brugger, U. Drechsler, W. Haeberle, H. Rohrer, H. Rothuizer, P. Vettiger: Microfabrication and parallel operation of 5 × 5 AFM cantilever arrays for data storage and imaging, presented at MEMS 98, Heidelberg (1998) 302
[57] B. W. Chui, H. J. Mamin, B. D. Terris, T. D. Stowe, D. Rugar, T. W. Kenny: Low-stiffness silicon cantilevers for thermal writing and piezoresistive readback with the atomic force microscope, Appl. Phys. Lett. **69**, 2767–2769 (1996) 302

Index

Energy Exchange at Short Time Scales: Electron–Phonon Interactions in Metals and Metallic Nanostructures

Fabrice Vallée

CPMOH, University of Bordeaux I,
351 Cours de la Libération, 33405 Talence Cedex, France
f.vallee@cpmoh.u-bordeaux1.fr

1 Introduction

Electron interactions and in particular the coupling between electrons and phonons are all-important parameters in many electronic properties of metals, such as resistivity or superconductivity. The development of femtosecond laser sources and time-resolved optical techniques has given selective access to various interaction mechanisms in bulk solid metals and metallic nanoparticles [1–4]. These studies are based on selective excitation of electrons by a first pulse (pump) and real-time monitoring of the energy redistribution in a metallic object and towards its surroundings by a second pulse (probe). Information is extracted via the strong correlation between electronic and vibrational kinetics and changes in the measured optical properties as a function of the delay between the pump and probe pulses [1]. After the electrons have thermalised, i.e., a few hundred femtoseconds [5–7], changes in the optical properties essentially reflect changes in the electron temperature. Its dynamical evolution is thus correlated with energy exchange mechanisms: interactions between electrons and vibrational modes of the lattice, energy transfer from electrons or the metallic lattice to vibrational modes of the environment, and so on [8, 9].

Following a general discussion of electronic and vibrational structures in noble metals in Sect. 2, we go on to consider several important aspects of their optical properties, i.e., of the dielectric constant, and in particular, their dependence on the electron distribution in bulk metals and in confined systems such as nanoparticles (Sect. 3.1). We then describe the excitation of conduction electrons by femtosecond laser pulse (Sect. 3.2). The temporal evolution of the electron distribution and its return to equilibrium will be modelled in Sect. 4 using the Boltzmann equation for an electron gas and its limit for a thermalised electron distribution, viz., the two-temperature model. Several experimental results obtained for films and metallic nanoparticles in various time or intensity regimes are also outlined in Sect. 4. We end with a brief introduction to the detection in the time domain of acoustic vibrations in metallic nanospheres (Sect. 5).

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 309–332 (2007)

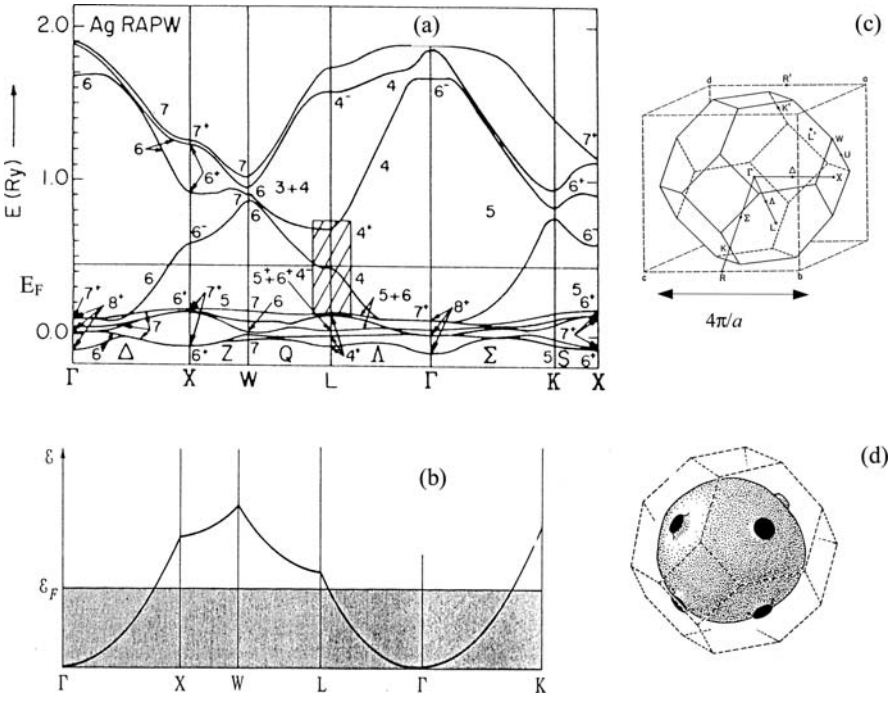

Fig. 1. Silver. (**a**) Calculated band structure [10]. (**b**) Band structure for free electrons [11]. (**c**) First Brillouin zone (bcc). (**d**) Fermi surface [10]

2 Electronic and Vibrational Structures in Metallic Systems

2.1 Electronic Structure of Noble Metals

Gold, silver, and copper are noble metals made of atoms with the same electronic structure: their *d* orbitals are full and they each have one electron on an *s*-type orbital. The latter are delocalised in the crystal lattice and will constitute the conduction electrons (see Table 1). Their crystal lattice is face-centered cubic (fcc). The first Brillouin zone of the reciprocal lattice (bcc) is shown in Fig. 1 [10].

Electronic Band Structure in Bulk Metals

The noble metals have similar band structures, including a set of five valence bands with low dispersion, called the *d* bands (because mainly composed of electron levels originating in the *d* orbitals of the atoms), and a half-filled *sp* band called the conduction band (see Fig. 1) [10]. All the electron bands are therefore either filled or empty, except for the conduction band

Table 1. Noble metals. Electron configuration in the atomic structure. Lattice constant a. Electron density in the conduction band n_e. Effective mass for the conduction band m_e normalised by the free electron mass m_0. Fermi energy E_F. Interband transition threshold $\hbar\Omega_{ib}$ [10, 11]

	Atomic structure	a (Å)	n_e (10^{22} cm^{-3})	m_e/m_0	E_F (eV)	$\hbar\Omega_{ib}$ (eV)
Ag	[Kr] $4d^{10}5s^1$	4.08	5.86	1	5.49	3.9
Au	[Xe] $4f^{14}5d^{10}6s^1$	4.07	5.90	1	5.53	2.4
Cu	[Ar] $3d^{10}4s^1$	3.61	8.47	1.5	4.67	2.1

which is filled up to the Fermi level E_F. (This is only strictly true at zero electron temperature T_e, but remains an excellent approximation at higher temperatures in metals for which $k_B T_e \ll E_F$.) The conduction band is quasi-parabolic, which justifies the commonly found description of the conduction electrons as being quasi-free. Their dispersion relation is thus

$$E(\boldsymbol{k}) = \frac{\hbar^2 k^2}{2m_e}. \tag{1}$$

This expression represents the energy E of a state of the conduction band with wave vector $\boldsymbol{k}$, taking the zero energy at the bottom of the band. The corresponding band structure gives a fairly good representation of the true conduction band (see Fig. 1), justifying the approximation. The effective electron masses m_e, determined from the infrared dielectric constant [11, 12], are given in Table 1. The Fermi surface separating occupied and free states at zero temperature is then spherical, giving a fairly good representation of the true Fermi surface except in the vicinity of the L points (see Fig. 1d).

Most electron properties only involve a small number of the electron bands, and often even involve only the conduction band states near the Fermi energy. This is the case, at least qualitatively, for the interactions of the conduction electrons, such as electron–electron and electron–phonon interactions, and for the optical response in the infrared to visible range of the electromagnetic spectrum. These properties can to a large extent be described using a simplified model of the band structure. The conduction band is then described by an isotropic parabolic model with dispersion relation given by (1) and the d bands by dispersionless states (see Fig. 2).

The model is well-suited to interpreting studies of electron properties based on optical techniques, provided that the photon energies are significantly lower than the interband transition threshold. For optical frequencies close to this threshold, the detailed band structure and anisotropy of the material must be taken into account. Absorption is then dominated by electron states near the L point of the Brillouin zone, whose dispersion has been modelled by Rosei et al. [13, 14].

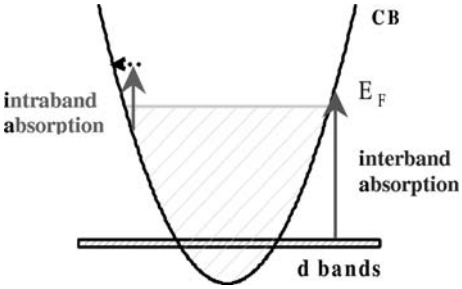

Fig. 2. Simplified model of the electronic band structure in noble metals. Isotropic parabolic conduction band and dispersionless d bands

Electronic Band Structure in Confined Metals

Although quantum confinement effects are very important in nanoscale semiconductor particles (confinement of excitons whose Bohr radius is comparable with the particle dimensions and confinement of electron states with low quantum number, close to the centre of the Brillouin zone), these effects are considerably less relevant to metallic particles. For the latter, since the electron density is much higher, electrical and optical properties arising from the response of electrons near the Fermi level involve states with high quantum number [15, 16]. It is then possible to introduce a continuum of states and use an approach similar to the one for the bulk medium, at least for sizes greater than 2 nm [17]. Size effects then arise as corrections, especially for optical properties, and to a large extent, the electronic band structure described above can be used down to dimensions of a few nanometers.

Electron Distribution

An electron state with wave vector $\boldsymbol{k}$ and given spin can only be occupied by one electron (fermion). For a thermalised distribution at temperature T_e, its occupation probability is given by a Fermi–Dirac distribution (see Fig. 5)

$$f\big(E(\boldsymbol{k})\big) = \frac{1}{\exp \dfrac{E(\boldsymbol{k}) - E_\mathrm{F}}{k_\mathrm{B} T_\mathrm{e}} + 1} \,. \tag{2}$$

This description of the electrons as independent quasi-particles (Fermi fluid model) corresponds to the bulk medium (states defined by $\boldsymbol{k}$) [11]. Given the fact that confinement has little effect on the band structure, it can be extended to confined media.

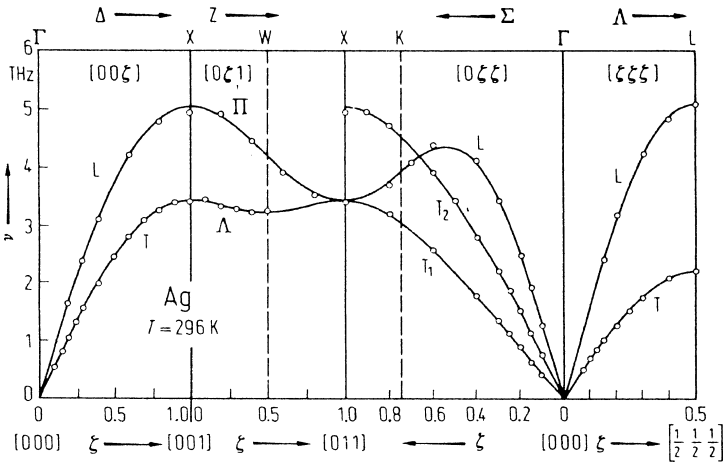

Fig. 3. Dispersion curves of acoustic phonons measured by neutron diffraction in silver [10]

2.2 Lattice Vibrations

Phonon Dispersion Relation

The unit cell of noble metals comprises one atom. The only lattice vibrations are thus the acoustic modes, which can be longitudinal (L) or transverse (T_1 and T_2) [10, 11]. For each of the three branches, there are N_i modes in the first Brillouin zone, where N_i is the number of ions forming the lattice. The dispersion relation $\omega = \omega(\boldsymbol{q})$ of acoustic phonons is shown for silver in Fig. 3 [10]. For small wave vectors, the relations are linear and a speed of sound can be defined by $\omega(\boldsymbol{q}) = v_\mathrm{s} q$.

The true dispersion relation is often replaced by an isotropic linear relation, viz., $\omega(\boldsymbol{q}) = \omega_q = v_\mathrm{s} q$ (Debye model). The first Brillouin zone must then be replaced by a sphere of radius q_D which contains N_i modes:

$$q_\mathrm{D} = (6\pi^2 N_i)^{1/3} = \left(\frac{2}{Z}\right)^{1/3} k_\mathrm{F}\,,$$

where $Z = 1$ is the valence of the metal. The phonon dispersion relation is better reproduced using a sinusoidal model, at least along axes of high symmetry (ΓL and ΓX):

$$\omega_q = \omega_\mathrm{M} \sin \frac{\pi q}{2 q_{\Gamma X}}\,,$$

with $\hbar\omega_\mathrm{M} = 21\,\mathrm{meV}$ in Ag. If we assume an isotropic dispersion relation, we must once again restrict to wave vectors shorter than q_D (close to $q_{\Gamma X}$).

The occupation number of the vibrational mode $\omega_{\boldsymbol{q}}$ at the lattice temperature T_{L} is given by the Bose–Einstein distribution

$$N(\boldsymbol{q}) = \frac{1}{\exp \dfrac{\hbar\omega_{\boldsymbol{q}}}{k_{\mathrm{B}}T_{\mathrm{L}}} - 1} \,. \tag{3}$$

Lattice Vibrations. Vibrational Modes of Nanoparticles

The phonon description of vibrational modes in a crystal lattice is based on the translation invariance of the lattice structure. For a finite medium, this invariance is spoilt by the presence of interfaces which impose new boundary conditions and hence a new quantisation of the vibrations. Using the solid state physics approach, only phonons associated with a wavelength λ_{ph} greater than or comparable with the size of the object are significantly modified. For acoustic vibrations, with frequency $\omega_{\mathrm{ph}} = 2\pi v_{\mathrm{s}}/\lambda_{\mathrm{ph}}$, where v_{s} is the speed of sound, the low frequency modes will therefore be replaced by confined vibrational modes.

For large clusters comprising several hundred atoms, these modes can be described using a macroscopic approach. They then correspond to the deformation eigenmodes of a finite homogeneous elastic object (see Sect. 5). Their properties, frequency and damping, are correlated with the size, shape and environment of the object and can therefore be used to analyse nanomaterials. In this model, only low energy modes are altered and this leads to only small changes in properties associated with lattice vibrations, except at low temperatures. A more accurate description must be based on an atomic approach which takes into account the different environment of the surface atoms and hence the different forces to which they are subject. A reduction in the density of states of high frequency vibrational modes has thus been discovered for nanospheres, as compared with the bulk medium [18]. This model can only be carried out numerically, however, and it is difficult to integrate into an electronic relaxation model. The bulk medium model will be used as a first approximation.

3 Optical Properties of Metals

3.1 Optical Response at Equilibrium

Bulk Noble Metals

In a metal, light–matter interactions involve two electronic mechanisms which explain the optical response. The first is due to quasi-free electrons of density n_{e} in the conduction band and is associated with optical transitions

within the same band, or intraband transitions (see Fig. 2). Its contribution to the dielectric constant is described by the Drude model [9]

$$\varepsilon_{\text{Drude}} = 1 - \frac{\omega_{\text{p}}^2}{\omega(\omega + \text{i}\gamma_0)}\,, \tag{4}$$

where $\omega_{\text{p}} = n_{\text{e}}e^2/\varepsilon_0 m_{\text{e}}$ is the plasma frequency of the metal. Absorption, described by the imaginary part of the dielectric constant, is proportional to the average optical collision rate of the electrons, denoted by $\gamma_0(\omega)$. This term corresponds to collisions between three quasiparticles, where an electron–electron or electron–phonon collision occurs with simultaneous absorption of a photon of frequency ω. This reflects the fact that direct absorption of a photon by an electron cannot occur because the wave vector of the photon is very small compared with the wave vector of the electrons, and such an absorption would correspond to a vertical transition with constant electron wave vector (see Fig. 2). Absorption must therefore be assisted by a third particle, either another electron or a phonon, to ensure conservation of both energy and momentum. At room temperature, electron–phonon interactions dominate.

The second mechanism is related to optical transitions between two electronic bands, or interband transitions, from filled bands to states in the conduction band, or from the latter to empty bands of higher energy. In the alkali metals, these transitions occur at high frequencies and only lead to a small correction to the real part of ε in the optical domain. In the noble metals, it is much more important and is essentially due to transitions between the d bands and the sp conduction band (see Fig. 2). A d-band electron can only be excited into an unoccupied state of the conduction band (Pauli exclusion principle), i.e., above the Fermi energy E_{F}. There is therefore a threshold frequency Ω_{ib} for interband photon absorption (see Table 1) which gives the characteristic colour of these metals in their bulk solid phase.

When there are bound electrons, the dielectric constant of the metal is written as the sum of two terms:

$$\varepsilon = \varepsilon_{\text{Drude}} + \delta\varepsilon_1^{\text{ib}}\,. \tag{5}$$

The interband term dominates the optical response in the visible and ultraviolet, whereas the intraband term is the more important in the red and infrared. It leads to a negative real part for the dielectric constant in this spectral region.

Metallic Nanoparticles

For clusters with dimensions greater than 2 nm, absorption mechanisms are very similar and involve the same basic concepts. Size effects appear as corrections and in the optical properties are manifested mainly by an increase

in the optical collision rate of electrons arising in the Drude model [19]. From a classical standpoint, this corresponds to the fact that electron–surface collisions are no longer negligible compared with other interaction processes and must therefore be taken into account in the optical collision rate of the electrons. In contrast, the interband dielectric constant of the metal is only slightly modified down to sizes of the order of 2 nm [20].

The most important effect on the optical properties is due to the dielectric confinement described in the Mie theory [21, 22]. For metallic nanospheres that are much smaller than the relevant optical wavelengths, absorption is enhanced near the frequency Ω_{R} such that

$$\varepsilon_1(\Omega_{\mathrm{R}}) + 2\varepsilon_{\mathrm{m}} = 0\,. \tag{6}$$

At the frequency Ω_{R}, a resonance is observed in the absorption spectrum. This is the so-called surface plasmon resonance (SPR). This corresponds to a frequency for which the electromagnetic field seen by a nanoparticle is greatly increased compared with that of the incident wave, due to the resonant polarisation caused by charge displacement. From a classical standpoint, the surface plasmon resonance can be associated with the resonant excitation by the external electromagnetic field of a collective electron oscillation within the nanoparticle. This oscillation can be compared with the collective oscillation of an electron gas in a bulk solid (plasmon mode at ω_{p}), modified by the presence of interfaces.

3.2 Femtosecond Pump–Probe Method

Basic Principle

Ultrafast electron kinetics and, in particular, interactions with the lattice can be investigated using a femtosecond pump–probe technique. The basic idea is illustrated schematically in Fig. 4. A first pulse (pump) perturbs the medium, inducing a change in its optical properties (transmission T or reflection R). Their dynamic evolution and hence also the return of the system to equilibrium are measured by a second, time-delayed pulse (probe). The change ΔT or ΔR in the transmission and reflection of the probe for a delay t between the pump and probe is given by the convolution of the impulse response of the system with the pump–probe cross-correlation. The time resolution is thus limited by the duration of the intensity correlation between the two pulses and is measured by replacing the sample by a nonlinear crystal.

Femtosecond Optical Excitation

In most experiments, the electron distribution is first driven out of equilibrium by intraband absorption of a femtosecond laser pulse with central frequency ω_{pp} (see Fig. 5). For a purely intraband excitation ($\omega_{\mathrm{pp}} < \Omega_{\mathrm{ib}}$),

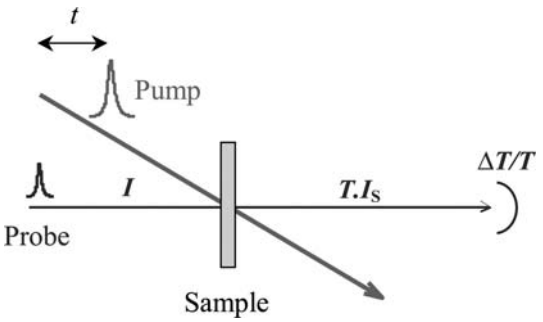

Fig. 4. Transmission pump–probe experiment

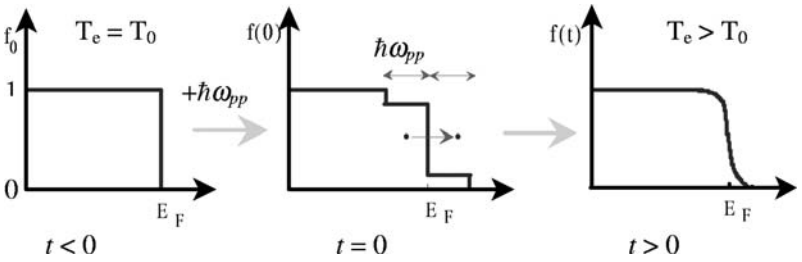

Fig. 5. Creation of a non-equilibrium electron distribution in a metal by intraband absorption of a laser pulse of frequency ω_{pp} at time $t = 0$

an electron in a state of energy E below the Fermi level absorbs a photon and is excited into a state of energy $E + \hbar\omega_{\mathrm{pp}}$. More precisely, the change in the occupation number for a state of energy E is expressed as the sum of two terms: one expresses the departure of electrons which absorb a photon and are excited into a state of higher energy, and one expresses the arrival of electrons from states of lower energy. Taking into account the density of states (parabolic band), the change in occupation number is given by

$$\mathrm{d}f_{\mathrm{exc}}(E) = A_{\mathrm{exc}}\Big\{\sqrt{E-\hbar\omega_{\mathrm{pp}}}f(E-\hbar\omega_{\mathrm{pp}})\big[1-f(E)\big] - \sqrt{E+\hbar\omega_{\mathrm{pp}}}f(E)\big[1-f(E+\hbar\omega_{\mathrm{pp}})\big]\Big\}, \quad (7)$$

where the constant of proportionality A_{exc} describes the energy injected into the electron gas. For a finite width laser pulse, the excitation is assumed to be proportional to the temporal shape $I(t)$ of the pump pulse.

Femtosecond Optical Probe

For a sufficiently weak perturbation, changes ΔT or ΔR in the transmission or reflection can be written as linear combinations of the changes in the dielectric constant $\varepsilon = \varepsilon_1 + \mathrm{i}\varepsilon_2$ of the medium with, for example,

$$\frac{\Delta T}{T}(\omega_\mathrm{s}) = \frac{\partial \ln T}{\partial \varepsilon_1}(\omega_\mathrm{s})\Delta\varepsilon_1(\omega_\mathrm{s}) + \frac{\partial \ln T}{\partial \varepsilon_2}(\omega_\mathrm{s})\Delta\varepsilon_2(\omega_\mathrm{s}) \tag{8}$$

for the change ΔT in the transmission. The information obtained will depend on the probe frequency ω_s. Out of resonance, i.e., when ω_s is far from the interband transition threshold Ω_ib (see Fig. 2), the change in the imaginary part of the interband dielectric constant is simply

$$\Delta\varepsilon_2^\mathrm{ib}(\hbar\omega_\mathrm{s}) \propto \frac{\sqrt{\hbar\Omega_\mathrm{d} + \hbar\omega_\mathrm{s}}}{(\hbar\omega_\mathrm{s})^2}\Delta f(\hbar\Omega_\mathrm{d} + \hbar\omega_\mathrm{s})\,, \tag{9}$$

using a simplified model which neglects bending of the d bands and assumes an isotropic parabolic conduction band as shown in Fig. 2.

For a probe frequency ω_s such that the occupation number of the final states is not modified by the pump, i.e., such that $\hbar\Omega_\mathrm{d} + \hbar\omega_\mathrm{s} < E_\mathrm{F} - \hbar\omega_\mathrm{pp}$, equivalent to $\omega_\mathrm{s} + \omega_\mathrm{pp} < \Omega_\mathrm{ib}$, the change in $\varepsilon_2^\mathrm{ib}(\omega_\mathrm{s})$ will be zero. However, the real part of the interband dielectric constant, which depends on the integrated absorption spectrum, will be changed. $\Delta\varepsilon_1^\mathrm{ib}$ can be calculated from (7) using the Kramers–Krönig relation:

$$\Delta\varepsilon_1^\mathrm{ib}(\hbar\omega_\mathrm{s}) \propto -\int \frac{\sqrt{E'}\big[f(E') - f_0(E')\big]}{(E' - E_\mathrm{F} + \hbar\Omega_\mathrm{ib})\big[(E' - E_\mathrm{F} + \hbar\Omega_\mathrm{ib})^2 - (\hbar\omega_\mathrm{s})^2\big]}\,\mathrm{d}E'\,, \tag{10}$$

where $E' = \hbar\Omega_\mathrm{d} + \hbar\omega'$. The integral is restricted to energies for which the distribution f is out of equilibrium. The states with energy E' furthest away from the Fermi level, initially populated or depopulated by the pump, relax very quickly on a time scale of a few femtoseconds. The integral will thus be limited to states close to E_F and, away from resonance ($\hbar\omega_\mathrm{s} < \hbar\Omega_\mathrm{ib}$), can be simplified by expanding in terms of the parameters $(E' - E_\mathrm{F})/\hbar\Omega_\mathrm{ib}$ and $\big[\hbar\omega_\mathrm{s}/(E' - E_\mathrm{F} + \hbar\Omega_\mathrm{ib})\big]^2$, viz.,

$$\Delta\varepsilon_1^\mathrm{ib}(\hbar\omega_\mathrm{s}) \propto \int E'\sqrt{E'}\big[f(E') - f_0(E')\big]\,\mathrm{d}E' \equiv \Delta u_\mathrm{e}\,. \tag{11}$$

The change in $\varepsilon_1^\mathrm{ib}$ out of resonance is thus proportional to the excess energy Δu_e of the electron gas compared with its initial energy. The same result is obtained from numerical simulations, using the calculated electron distributions and the Rosei band structure model or a parabolic model.

In addition, the change in the intraband contribution to ε_1 is negligible compared with that of the interband term. Out of resonance, the time evolution of $\Delta\varepsilon_1$ thus reflects the time evolution of the excess energy of the electron gas. This provides a selective way to study the evolution of electron–lattice energy transfers as a function of the electron distribution.

4 Electron–Lattice Interactions. Energy Exchange

Electron interactions and in particular energy exchanges between electrons and the lattice can be studied by driving the two systems out of equilibrium by selectively exciting electrons, then monitoring their thermalisation in the time domain. The system, initially at equilibrium ($T_\mathrm{e} = T_\mathrm{L} = T_0$), is excited by absorbing a laser pulse, whereupon the electron distribution is modified. If short enough pulses are used ($< 100\,\mathrm{fs}$), the distribution is then out of equilibrium and can no longer be described by an internal temperature. The return to equilibrium occurs via two interaction mechanisms. Collisions between electrons conserve the total energy of the system and redistribute the energy in the electron gas (internal thermalisation at an electron temperature $T_\mathrm{e} > T_0$). This thermalisation takes place on a time scale of a few hundred femtoseconds [6, 7]. Collisions with phonons transfer the energy injected into the electron gas to the lattice within a few picoseconds [8], effecting a return to equilibrium at a temperature T_L. Since the lattice has a much bigger heat capacity than the electron gas (typically, 100 times greater), it is only slightly perturbed ($T_\mathrm{L} \approx T_0$).

4.1 Kinetic Model. Boltzmann Equation

Except at very short times, less than or of the order of 10 fs, electron relaxation kinetics can be described using a one-particle approach, neglecting coherence effects. The electron gas is then fully characterised by its distribution function $f(\boldsymbol{k})$. Energy exchanges, and hence the evolution of f, are described using the Boltzmann equation. For a homogeneous system, this reads [6]

$$\frac{\mathrm{d}f(\boldsymbol{k})}{\mathrm{d}t} = \left.\frac{\mathrm{d}f(\boldsymbol{k})}{\mathrm{d}t}\right|_{\mathrm{e-e}} + \left.\frac{\mathrm{d}f(\boldsymbol{k})}{\mathrm{d}t}\right|_{\mathrm{e-ph}} + L(\boldsymbol{k}, t)\,. \tag{12}$$

The first two terms represent the change in occupation number of the state $\boldsymbol{k}$ due to electron–electron and electron–phonon interactions, respectively, while the last term $L(\boldsymbol{k}, t)$ describes its perturbation by the pump pulse at time t. For a conduction band with isotropic dispersion, this equation simplifies enormously because the occupation numbers only depend on the magnitude of the wave vector $\boldsymbol{k}$. The same goes for the scattering rate and the excitation function. In the isotropic parabolic band model, the Boltzmann equation can

thus be uniquely expressed in terms of the energy of the electron states: $f(\boldsymbol{k}) = f(E)$.

For an intraband excitation, the perturbation function $L(E,t)$ for the pump pulse is proportional to (7):

$$L(E,t) = \frac{I(t)}{\int I(\tau)\,\mathrm{d}\tau}\,\mathrm{d}f_{\mathrm{exc}}(E)\,. \tag{13}$$

We shall only be concerned here with electron–phonon interactions, i.e., energy exchanges between conduction electrons and the lattice.

4.2 Electron–Phonon Interaction. Bulk Metals

Electron–phonon interactions correspond to processes in which one electron emits or absorbs a phonon with energy $E_{\mathrm{ph}} = \hbar\omega_{\boldsymbol{q}}$ and wave vector $\boldsymbol{q}$:

$E-E_{\mathrm{ph}}(q)$	em	E	abs	$E+E_{\mathrm{ph}}(q)$
$\bullet$	$\leftrightarrows$	$\bullet$	$\leftrightarrows$	$\bullet$
$\boldsymbol{k}-\boldsymbol{q}$	abs	$\boldsymbol{k}$	em	$\boldsymbol{k}+\boldsymbol{q}$

The scattering rate is given by

$$\begin{aligned}\left.\frac{\mathrm{d}f(\boldsymbol{k})}{\mathrm{d}t}\right|_{\mathrm{e-ph}} &= \frac{2\pi}{\hbar}\sum_{\boldsymbol{q}}|M|^2F^-(\boldsymbol{k},\boldsymbol{q})\delta\big(E(\boldsymbol{k})-E(\boldsymbol{k}-\boldsymbol{q})-\hbar\omega_{\boldsymbol{q}}\big)\\ &\quad+\frac{2\pi}{\hbar}\sum_{\boldsymbol{q}}|M|^2F^+(\boldsymbol{k},\boldsymbol{q})\delta\big(E(\boldsymbol{k})-E(\boldsymbol{k}+\boldsymbol{q})+\hbar\omega_{\boldsymbol{q}}\big)\,,\end{aligned} \tag{14}$$

where

$$F^-(\boldsymbol{k},\boldsymbol{q}) = -f(\boldsymbol{k})\big[1-f(\boldsymbol{k}-\boldsymbol{q})\big]\big[1+N(\boldsymbol{q})\big] + \big[1-f(\boldsymbol{k})\big]f(\boldsymbol{k}-\boldsymbol{q})N(\boldsymbol{q})\,, \tag{15}$$

$$\begin{aligned}F^+(\boldsymbol{k},\boldsymbol{q}) &= -f(\boldsymbol{k})\big[1-f(\boldsymbol{k}+\boldsymbol{q})\big]N(\boldsymbol{q}) + \big[1-f(\boldsymbol{k})\big]f(\boldsymbol{k}+\boldsymbol{q})\big[1+N(\boldsymbol{q})\big]\\ &= -F^-(\boldsymbol{k}+\boldsymbol{q},\boldsymbol{q})\,.\end{aligned} \tag{16}$$

The phonon emission process depends on the phonon occupation number through the factor $1+N$, which includes a spontaneous term and a stimulated term, whereas the absorption rate only involves a stimulated term, proportional to N.

Interactions between electrons and acoustic phonons in noble metals can be described to a first approximation by a deformation potential mechanism, i.e., lattice vibrations modulate the crystal potential and hence the energy of the electronic bands. The interaction matrix element is obtained by replacing the electron–phonon interaction Hamiltonian by the first order expansion

with respect to q of the energy shift of the conduction band due to a vibrational mode with wave vector $\boldsymbol{q}$ [23, 24]. The scattering amplitude is then given by

$$\left|M(q)\right|^2 = \frac{\hbar^2 \Xi^2 q^2}{2\rho V \hbar\omega_q}\,, \tag{17}$$

where ρ is the density of the material, and Ξ is the deformation potential defined as the change in energy of the minimum of the conduction band for a unit expansion of the lattice. This expression is only valid for small q, i.e., $q \ll k_\mathrm{F}$, a condition which is not satisfied for all electron–phonon collisions. Actually, for the conditions we have assumed, viz., $\hbar\omega_\mathrm{pp}$, $k_\mathrm{B}T_0 > \hbar\omega_\mathrm{D}$, where ω_D is the Debye frequency, and for isotropic bands, the dependence of the coupling constant on q has very little influence on the results and only leads to a modification of the value of Ξ, which will be treated as an adjustable parameter. In the following, we will write the scattering rate in the form

$$\left.\frac{\mathrm{d}f(E)}{\mathrm{d}t}\right|_\mathrm{e-ph} = \Gamma^-(E) + \Gamma^+(E)\,. \tag{18}$$

The expression for $\Gamma^\pm$ depends on the model used to describe the phonon dispersion relation:

– **Debye Model.** In this model,

$$\Gamma^\pm(E) = \frac{G_\mathrm{D}}{\sqrt{E}} \int_0^{E_{\mathrm{ph\,max}}} E_\mathrm{ph}^2 F^\pm(E, E_\mathrm{ph})\,\mathrm{d}E_\mathrm{ph}\,, \tag{19}$$

where

$$G_\mathrm{D} = \frac{\Xi^2\sqrt{m}}{4\sqrt{2}\pi\hbar^4\rho v_\mathrm{s}^4}\,, \quad E_\mathrm{ph} = \hbar v_\mathrm{s} q\,,$$
$$E_{\mathrm{ph\,max}} = \min\left(\hbar v_\mathrm{s} q_\mathrm{D}, 2v_\mathrm{s}\sqrt{2mE}\right).$$

For isotropic dispersions for the conduction band and vibrational modes, the functions $F^\pm$ depend only on the energies, viz.,

$$F^\pm(E, E_\mathrm{ph}) = F^\pm\left(E(\boldsymbol{k}), \hbar\omega(\boldsymbol{q})\right) = F^\pm(\boldsymbol{k}, \boldsymbol{q})\,.$$

– **Sinusoidal Dispersion Relation.** In this model,

$$\Gamma^\pm(E) = \frac{G_\mathrm{S}}{\sqrt{E}} \int_0^{q_\mathrm{max}} \frac{q^3}{\omega_\mathrm{M}\sin(\pi q/2q_\mathrm{D})} F^\pm(E, q)\,\mathrm{d}q\,, \tag{20}$$

where

$$G_\mathrm{S} = \frac{\Xi^2\sqrt{m}}{4\sqrt{2}\pi\hbar\rho}\,, \quad q_\mathrm{max} \approx q_\mathrm{D} \approx q_{\Gamma X}\,, \quad \hbar\omega_\mathrm{m} \approx 21\,\mathrm{meV} \text{ for silver}\,.$$

The results of numerical calculations made with the two models, and also for the Einstein model ($\omega_{\boldsymbol{q}} = \text{constant}$), are in fact identical, up to a renormalisation of the effective coupling constant G (or Ξ). This is due to the fact that, since the initial energy given to the electrons is much greater than the phonon energy and the temperature is greater than the Debye temperature, the dynamics of energy loss from the electron gas to the lattice is insensitive to the actual dispersion of the phonons.

4.3 Energy Exchange in the Thermal Regime. Two-Temperature Model

Several hundred femtoseconds after the optical excitation, the electrons are thermalised and energy transfer from the electrons to the lattice can be described by the two-temperature model. This considers the energy balance between two subsystems, viz., the conduction electrons and the lattice, each described by its own temperature. It can be deduced from the Boltzmann equation for the electrons where only the electron–phonon collision term is included. This amounts to assuming that the electron–electron collisions are efficient enough to maintain the electronic system at all times in internal thermal equilibrium at temperature T_e. The dynamical evolution of the energy density (or the temperature) in each system is then given by a system of two coupled differential equations [25, 26]:

$$\begin{cases} C_\mathrm{e}\dfrac{\partial T_\mathrm{e}}{\partial t} = -H(T_\mathrm{e}, T_\mathrm{L}) + p(t)\,, \\ C_\mathrm{L}\dfrac{\partial T_\mathrm{L}}{\partial t} = +H(T_\mathrm{e}, T_\mathrm{L})\,, \end{cases} \tag{21}$$

where H represents the energy exchanges and $p(t)$ the selective excitation of the electron gas by the pump pulse. [$p(t)$ must in fact be replaced by the initial condition $T_\mathrm{e}(t = 0) = T_\mathrm{exc.}$] The quantities C_e and C_L are the electron and lattice specific heat capacities, respectively. This expression is only valid for a homogeneous system, i.e., where the temperatures have no spatial dependence, a hypothesis also previously used implicitly for the electron distribution, assumed uniform in the system studied in Sect. 4.1. It can be generalised to the case of inhomogeneous electron or phonon temperatures by adding a heat diffusion term to (21). This generalisation is necessary when optically thick systems are involved, since longitudinal diffusion then plays an important role (the optical absorption length in a metallic film is typically of the order of 15 nm), or when time scales are long enough for transverse scattering out of the excited region to become significant.

For high temperatures T_e and T_L compared with the Debye temperature T_D, the electron temperature evolves according to

$$\frac{\mathrm{d}u_\mathrm{e}}{\mathrm{d}t} = C_\mathrm{e}\frac{\mathrm{d}T_\mathrm{e}}{\mathrm{d}t} - H(T_\mathrm{e}, T_\mathrm{L}) = -g(T_\mathrm{e} - T_\mathrm{L})\,, \tag{22}$$

where g, independent of the lattice temperature, is the electron–phonon coupling constant usually used in femtosecond studies [1–4, 27]. For an electron–phonon interaction described by a deformation potential and the Debye model, g is given by

$$g = \Xi^2 \frac{k_B m^2 q_D^4}{16\rho\pi^3\hbar^3} . \tag{23}$$

The system (21) associated with (22) is nonlinear, because the electron heat capacity depends on the temperature. For free electrons,

$$C_e(T_e) = \frac{\partial u_e}{\partial T} = \frac{\pi^2 n_e k_B}{2} \frac{T_e}{T_F} = \gamma_0 T_e . \tag{24}$$

The system (21) can be solved analytically to give

$$T_{eq} \ln \frac{T_e - T_{eq}}{T_{exc} - T_{eq}} + \tilde{T} \ln \frac{T_e + \tilde{T}}{T_{exc} + \tilde{T}} = -G \frac{\tilde{T} + T_{eq}}{2C_L} t , \tag{25}$$

where

$$\begin{cases} T_{eq} & = -\dfrac{C_L}{\gamma_0} + \left[\left(\dfrac{C_L}{\gamma_0} \right)^2 + T_{exc}^2 + 2\dfrac{C_L}{\gamma_0} T_0 \right]^{1/2} \approx T_0 + \dfrac{\gamma_0}{2C_L} (T_{exc}^2 - T_0^2) , \\ \tilde{T} & = T_{eq} + 2\dfrac{C_L}{\gamma_0} \approx 2\dfrac{C_L}{\gamma_0} . \end{cases} \tag{26}$$

Here T_{eq} is the temperature of the thermalised electron–lattice system and T_{exc} the excitation temperature of the electrons, i.e., the temperature of the conduction electrons after excitation by the pump pulse at $t = 0$, assuming that the system is thermalised at all times and neglecting the pulse width.

However, when the perturbation is weak ($T_e - T_L \ll T_L$), we may assume that C_e is constant, i.e., $C_e \approx \gamma_0 T_0$, which leads to a simpler expression where T_e falls off exponentially:

$$T_e(t) - T_L(t) = (T_{exc} - T_0) e^{-t/\tau_{e-ph}} , \tag{27}$$

with time constant $\tau_{e-ph} = C_e C_L / g(C_e + C_L)$. The electron gas then has excess energy $\Delta u_e(t) = C_e \big[T_e(t) - T_L(t) \big]$ and evolves as ΔT_e. Since the heat capacity of the lattice is large compared with that of the electrons, we have $\tau_{e-ph} \approx C_e/g = C_0 T_0/g$, and $T_L(t) \approx T_0$. This exponential decrease predicted by the two-temperature model agrees well with experimental results in the weak excitation regime (see Fig. 6). For strong excitations, the energy transfer is slower and has no exponential behaviour. In agreement with the two-temperature model, the exponential regime is reached after a few picoseconds, when the electron distribution is cold enough to justify neglecting the temperature dependence of the electron heat capacity C_e (see Fig. 6).

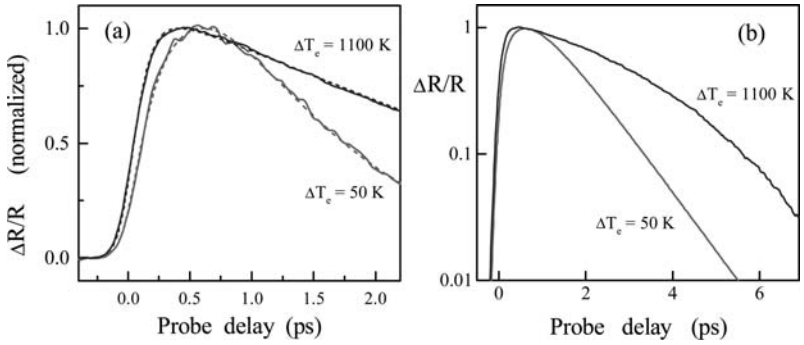

Fig. 6. (**a**) Change in the reflectance of a gold film of thickness 20 nm for weak and strong excitation pulses (maximum electron temperature rises 50 K and 1100 K, respectively). *Dotted curves* are calculated using the Boltzmann model. (**b**) As in (**a**), but plotted on a logarithmic scale

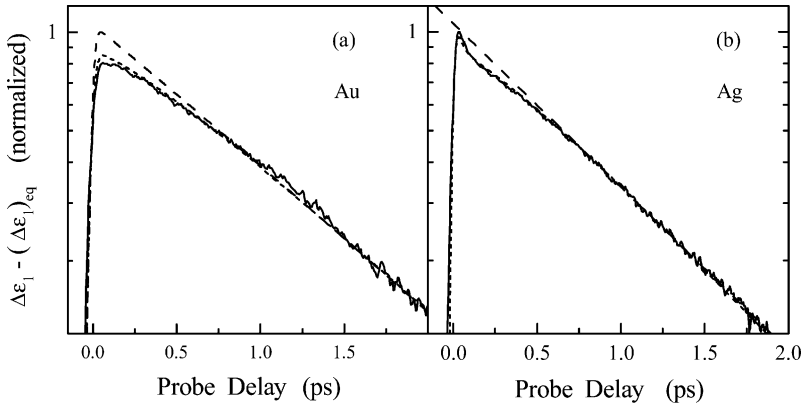

Fig. 7. (**a**) Change in the real part of the dielectric constant of a gold film of thickness 20 nm in the case of a weak excitation. *Dashed* and *dotted curves* are calculated assuming instantaneous thermalisation of the electrons (two-temperature model) and using the Boltzmann model, respectively. (**b**) As in (**a**), but for a silver film. The *dashed line* corresponds to exponential decay

As stressed above, this regime is of course only reached when the electron temperature has been established. Prior to this, a non-exponential regime is predicted which reflects the slower energy exchanges between an athermal electron distribution and the lattice. The results obtained by numerical simulation of the electron kinetics (Boltzmann equation) agree quantitatively with experimental results (see Fig. 7). The exponential decay predicted by the two-temperature model is only observed when the electron temperature has been established, i.e., after a few hundred femtoseconds.

4.4 Electron–Lattice Interactions in Metallic Nanoparticles

Many studies have been carried out on clusters of different shapes and sizes. The intrinsic effect of size reduction on electron–phonon interactions is still the subject of debate. In particular, energy transfer to the surrounding medium, which becomes more efficient as the cluster size decreases, can play a role. Indeed, in contrast to metallic films, when a medium comprising metallic nanoparticles dispersed in a matrix is excited by a laser pulse, a set of hot spots is created, viz., the nanoparticles. For a uniform temperature to be established, energy must diffuse out from the nanoparticles into the matrix. The two-temperature model must therefore be generalised by introducing the matrix temperature T_m. Assuming that the nanoparticles are independent, we can write [28]

$$\begin{cases} C_\mathrm{e}\dfrac{\partial T_\mathrm{e}}{\partial t} = -g(T_\mathrm{e} - T_\mathrm{L}) + p(t)\,, \\ C_\mathrm{L}\dfrac{\partial T_\mathrm{L}}{\partial t} = g(T_\mathrm{e} - T_\mathrm{L}) + S_\mathrm{Lm}\,, \\ C_\mathrm{m}\dfrac{\partial T_\mathrm{m}}{\partial t} = \boldsymbol{\nabla}\cdot(\kappa_\mathrm{m}\boldsymbol{\nabla}T_\mathrm{m})\,, \end{cases} \tag{28}$$

where C_m is the heat capacity of the matrix and κ_m is its thermal conductivity. T_m varies spatially with the distance r from the relevant nanoparticle. The electron and lattice temperatures are assumed uniform in the particle. Assuming also that the temperature is continuous at the particle–matrix interface, energy exchange between a nanosphere of diameter D and the matrix is given by

$$S_\mathrm{Lm} = 6\frac{\kappa_\mathrm{m}}{D}\left.\frac{\partial T_\mathrm{m}}{\partial r}\right|_{r=D/2}. \tag{29}$$

For short pulses, the system is thus much more difficult to describe. The description then depends on the matrix and the quality of the interface. However, if the relevant time scale is short compared with the typical time for transfer of energy from the particle to the matrix, this energy transfer can be neglected and the problem is the same as for a bulk medium.

Recent results obtained for weak excitation in noble metals have shown that the electron–lattice energy transfer time is independent of the matrix and also of the way the nanoparticles were elaborated, indicating a weak contribution from the environment [9]. In complete contrast, in the strong excitation regime, measured relaxation times were observed to depend significantly on the matrix [2, 29].

In the weak excitation regime, a reduction in the energy transfer time $\tau_\mathrm{e\text{–}ph}$ was shown for silver and gold as the cluster size diminished (see Fig. 8). This variation is similar to the one demonstrated for the electron–electron coupling [7] and is probably due to a similar surface effect.

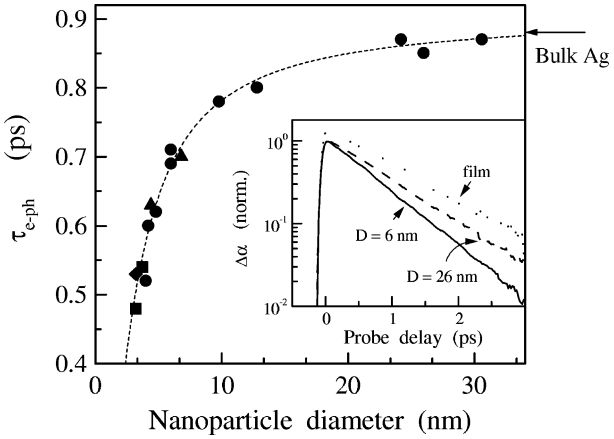

Fig. 8. Electron–lattice energy transfer time τ_{e-ph} as a function of size for silver nanospheres dispersed in $BaO-P_2O_5$ (*black dots*), Al_2O_3 (*black squares*), MgF_2 (*black diamonds*), and water (*black triangles*). The *dotted curve* interpolates between experimental points. The *insert* shows changes in the measured absorption on a logarithmic scale as a function of the pump–probe delay for two average diameters (26 and 6 nm) and in a silver film of thickness 23 nm

5 Acoustic Vibrational Modes of Nanospheres

The acoustic vibrational modes of lowest frequency in a material are associated with atomic displacements in the lattice on a much longer scale than the interatomic distance ($\omega_{ph} = 2\pi v_s/\lambda_{ph}$, where v_s is the speed of sound). For a finite material, they can be described using a macroscopic approach that neglects the periodic and discontinuous features of the crystal lattice. They then correspond to deformational eigenmodes of a continuous and homogeneous elastic object. Their properties depend only on the speed of sound, the density ρ of the material, and the boundary conditions imposed by the shape of the object and the surrounding material.

5.1 Vibrational Modes

The vibrational eigenmodes of an isotropic elastic sphere of diameter D were studied by *Lamb* in the nineteenth century [30–33]. This work was generalised to the case of a sphere in an infinite and isotropic elastic matrix, notably in geophysics to model inclusions in the Earth's crust [34, 35]. Up to a scale factor, the problem is the same for nanospheres. It consists in solving the Navier equation for displacements inside and outside the sphere which satisfy the requirement that the displacement and stress should be continuous at the interface $r = D/2$. There are two types of vibrational eigenmode: torsional and spheroidal modes. The first have zero radial displacement component and thus correspond to a constant volume oscillation [32,33]. The direction of the

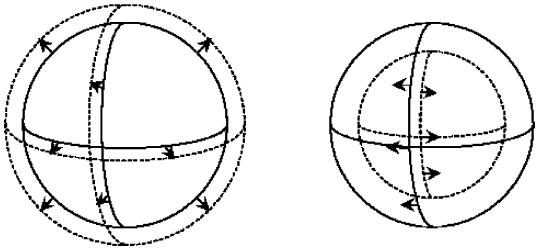

Fig. 9. Vibrational modes of a homogeneous elastic sphere [30]. *Left*: Radial spheroidal mode. The expanding and contracting motion of the sphere is like breathing. *Right*: Torsional mode without expansion. *Continuous* and *dotted curves* show the surface displacement and internal motion, respectively

displacement vector is parallel to the surface of the sphere, which corresponds to the twisting of an elastic body. The second type of mode have a nonzero radial displacement component and is therefore associated with a change in volume (see Fig. 9).

Only the second type of mode is observed in time-resolved experiments for which the excitation is isotropic. For the same reason, among the spheroidal modes, those corresponding to a purely radial displacement are preferentially observed. The fundamental radial mode ($n = 1$) of a sphere is its breathing mode (expansion and contraction). The frequencies of the radial modes are inversely proportional to D, i.e.,

$$\omega_n \approx \frac{2\pi n v_{\mathrm{L}}^{(\mathrm{s})}}{D}\,, \tag{30}$$

where $v_{\mathrm{L}}^{(\mathrm{s})}$ is the longitudinal speed of sound in the sphere and n is the order of the mode. The vibration frequency is largely independent of the environment and very close to the frequency of the acoustic phonon with wavelength $\lambda_{\mathrm{ph}} = D/n$.

The motion of a sphere placed in a matrix is accompanied by the propagation of a spherical wave in the matrix. The mechanical contact between sphere and matrix induces an energy loss and hence a damping of the radial modes. The damping rate γ_n is proportional to the ratio between the energy flux through the surface ($\propto D^2$) and the energy stored in the sphere ($\propto D^3$), hence proportional to D^{-1}. More precisely, it is in most cases determined by the ratio of the acoustic impedances ρv_{L} of the materials making up the matrix (m) and the sphere (s):

$$\gamma_n \approx \frac{\rho^{(\mathrm{m})} v_{\mathrm{L}}^{(\mathrm{m})}}{\rho^{(\mathrm{s})} D}\,. \tag{31}$$

It is only nonzero when there is a matrix, and it is very sensitive to its nature and in particular its elastic properties, as well as the quality and type of interface.

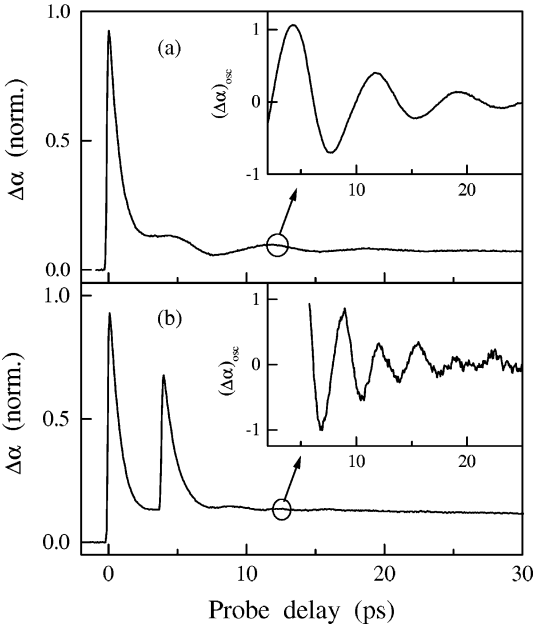

Fig. 10. (**a**) Change in the absorption detected by a probe pulse as a function of its delay with respect to the pump pulse for silver nanospheres of diameter $D = 24\,\text{nm}$ dispersed in glass. The oscillations of period $T_1 = 7.6\,\text{ps}$ are due to the fundamental radial vibrational mode $n = 1$ (breathing mode). (**b**) Signal measured under the same conditions for two pump pulses separated by a time $T_1/2$. The oscillations of period $T_2 = 3.5\,\text{ps}$ are due to the harmonic mode $n = 2$. Transient peaks correspond to the electron response

5.2 Time-Resolved Studies

The period of these modes lies somewhere in the picosecond region. The associated periodic motion can thus be monitored by optical methods with subpicosecond resolution. Using a femtosecond pump–probe technique, the excitation and detection of the fundamental breathing mode ($n = 1$) of the spherical nanoparticles have been demonstrated in the case of metallic and semiconducting materials [4, 36, 37]. In this type of study, several thousand clusters are observed at the same time. Their motion, triggered by the pump pulse, induces a modulation in the interatomic separation of atoms making up the clusters, and hence also a modulation in their electronic properties, i.e., density of conduction electrons and energies of the electron bands. This is reflected in a modulation of the absorption by the medium (a modulation of the spectral position of the surface plasmon resonance in the case of silver) which is detected by the probe pulse (see Fig. 10).

Such a response can only be observed if all the clusters oscillate in phase. For metals, the energy given to the electrons by the pump pulse is quickly

transferred to the lattice. It is thus heated in a time $\tau_{\text{e–ph}} \approx 1\,\text{ps}$ which is shorter than the expansion time of a cluster, i.e., the period T_n of the order n breathing mode ($T_1 \approx 3\,\text{ps}$ for $D = 10\,\text{nm}$). This rapid heating triggers a simultaneous increase in size in all the excited nanoparticles, followed by an oscillation around their new equilibrium size corresponding to their new temperature. The fundamental breathing mode is preferentially launched because the associated displacement is very close to an isotropic expansion, and the other modes make smaller contributions.

This process can be simply described using the image of an oscillator whose equilibrium position is displaced more quickly than its own period. The oscillation begins with an extremal displacement amplitude, in agreement with the phase of oscillations observed experimentally (see Fig. 10). In fact, the highly excited electrons exert a pressure which induces an expansion of the lattice, and this also contributes. This adds to the expansion due to the anharmonicity of the crystal potential and modifies the phase of the motion for strong excitations or very small nanoparticles.

The radial modes have also been studied by Raman spectroscopy, but it was only possible to measure their frequencies. Time-resolved techniques also provide information about the frequency $\omega_1 = 2\pi/T_1$ of the fundamental mode and its intrinsic damping γ_1. The damping observed in the time domain is due partly to the size dispersion of the nanoparticles. They oscillate at slightly different frequencies, and after their initial in-phase excitation, slowly drift out of phase (inhomogeneous damping). It is also caused by energy transfer to the environment. The latter depends on its nature, e.g., glass, polymer, liquid, etc., and on the interface, e.g., the presence of defects, adsorbed molecules, and so on. Once the size dispersion has been determined by electron microscopy, this intrinsic contribution γ_1 can be measured. We have shown that it varies as $1/D$ and can be used to obtain information about the material, such as the quality of the nanoparticle–matrix interface or the presence of adsorbed molecules.

By demonstrating the coherent excitation of cluster motion by a femtosecond pulse, one can envisage its optical control by means of a sequence of pulses (see Fig. 10). This has been achieved in the simplest case, with a pair of pulses separated by a half-period of the fundamental breathing mode. The second pulse then serves to stop the fundamental mode. After a half-period, the size of the clusters is maximal. If the intensity of the second pulse is chosen correctly, the energy injected carries the cluster temperature to the value corresponding to their maximal radius. The system is therefore in equilibrium once again. However, the motion of the first harmonic mode ($n = 2$), with almost double the frequency, is strengthened. As for the fundamental mode, its properties (frequency and damping) can then be determined.

References

[1] F. Vallée: C.R. Acad. Sci. **2**, 1469 (2001) 309, 323
[2] S. Link, M. A. El-Sayed: J. Phys. Chem. B **103**, 8410 (1999) 309, 323, 325
[3] C. Voisin, N. Del Fatti, D. Christofilos, F. Vallée: J. Phys. Chem. B **105**, 2264 (2001) 309, 323
[4] N. Del Fatti, F. Vallée: C.R. Acad. Sci. **3**, 365 (2002) 309, 323, 328
[5] M. Aeschlimann, M. Bauer, S. Pawlik, W. Weber, R. Burgermeister, D. Oberli, H. C. Siegmann: Phys. Rev. Lett. **79**, 5158 (1997) 309
[6] N. Del Fatti, C. Voisin, M. Achermann, S. Tzortzakis, D. Christofilos, F. Vallée: Phys. Rev. B **61**, 16956 (2000) 309, 319
[7] C. Voisin, D. Christofilos, N. Del Fatti, F. Vallée, B. Prével, E. Cottancin, J. Lermé, M. Pellarin, M. Broyer: Phys. Rev. Lett. **85**, 2200 (2000) 309, 319, 325
[8] S. D. Brorson, A. Kazeroonian, J. S. Modera, D. W. Face, T. K. Cheng, E. P. Ippen, M. S. Dresselhaus, G. Dresselhaus: Phys. Rev. Lett. **64**, 2172 (1990) 309, 319
[9] A. Arbouet, C. Voisin, D. Christofilos, P. Langot, N. Del Fatti, F. Vallée, J. Lermé, G. Celep, E. Cottancin, M. Gaudry, M. Pellarin, M. Broyer, M. Maillard, M. P. Pileni, M. Treguer: Phys. Rev. Lett. **90**, 177401 (2003) 309, 315, 325
[10] Landolt-Börnstein: *New Series, Group III: Crystal and Solid State Physics*, vol. 13a and 13c (Springer, Berlin 1981) 310, 311, 313
[11] N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Holt-Saunders, Tokyo 1981) 310, 311, 312, 313
[12] P. B. Johnson, R. W. Christy: Phys. Rev. B **6**, 4370 (1972) 311
[13] R. Rosei: Phys. Rev. B **10**, 474 (1974) 311
[14] R. Rosei, C. H. Culp, J. H. Weaver: Phys. Rev. B **10**, 484 (1974) 311
[15] W. P. Halperin: Rev. Mod. Phys. **58**, 533 (1986) 312
[16] J. A. A. Perenboom, P. Wyder, F. Maier: Phys. Rep. **78**, 173 (1981) 312
[17] C. Flytzanis, F. Hache, M. C. Klein, D. Ricard, P. Roussignol: *Progress in Optics*, vol. XXIX (North Holland, Amsterdam 1991) p. 321 312
[18] A. Eiguren, B. Hellsing, F. Reinert, G. Nicolay, E. V. Chulkov, V. M. Silkin, S. Hüfner, P. M. Echenique: Phys. Rev. Lett. **88**, 066805 (2002) 314
[19] U. Kreibig, M. Vollmer: *Optical Properties of Metal Clusters* (Springer, Berlin 1995) 316
[20] H. Hovel, S. Fritz, A. Hilger, U. Kreibig, M. Vollmer: Phys. Rev. B **48**, 18178 (1993) 316
[21] G. Mie: Am. Phys. (Leipzig) **25**, 377 (1908) 316
[22] C. F. Bohren, D. R. Huffman: *Absorption and Scattering of Light by Small Particles* (John Wiley, New York 1998) 316
[23] E. M. Conwell: *High Field Transport in Semiconductors* (Academic Press, New York 1967) 321
[24] J. M. Ziman: *Electrons and Phonons* (Clarendon Press, Oxford 1960) 321
[25] S. I. Anisimov, B. L. Kapeliovitch, T. L. Perel'man: Sov. Phys. JETP **39**, 375 (1974) 322
[26] M. I. Kaganov, I. M. Lifshitz, L. V. Tanatarov: Sov. Phys. JETP **4**, 173 (1957) 322
[27] R. Groeneveld, R. Sprik, A. Lagendijk: Phys. Rev. B **51**, 11433 (1995) 323

[28] P. Grua, J. P. Morreeuw, H. Bercegol, G. Jonusauskas, F. Vallée: Phys. Rev. B **68**, 035424 (2003) 325
[29] V. Halté, J. Y. Bigot, B. Palpant, M. Broyer, B. Prével, A. Pérez: Appl. Phys. Lett. **75**, 3799 (1999) 325
[30] N. Nishiguchi, T. Sakuma: Solid State Commun. **38**, 1073 (1981) 326, 327
[31] H. Lamb: Proc. Lond. Math. Soc. **13**, 189 (1882) 326
[32] Y. Satô, T. Usami: Geophys. Magn. **31**, 15 (1962) 326
[33] A. C. Eringen, E. S. Suhubi: *Elastodynamics* (Academic Press, New York 1975) 326
[34] V. A. Dubrovskiy, V. S. Morochnik: Earth Phys. **17**, 494 (1981) 326
[35] A. Tamura, K. Higeta, T. Ichinokawa: J. Phys. C **15**, 4975 (1982) 326
[36] M. Nisoli, S. D. Silvestri, A. Cavalleri, A. M. Malvezzi, A. Stella, G. Lanzani, P. Cheyssac, R. Kofman: Phys. Rev. B **55**, R13424 (1997) 328
[37] T. D. Krauss, F. W. Wise: Phys. Rev. Lett. **79**, 5102 (1997) 328

Index

Investigation of Short-Time Heat Transfer Effects by an Optical Pump–Probe Method

Bernard Perrin

Institut de Nanosciences de Paris, UMR 7588 CNRS and University of Pierre and Marie Curie, 140 rue du Lourmel, 75015 Paris
bper@ccr.jussieu.fr

Abstract. The study of heat transfer properties on micro- and nanoscales generally requires one to work on time scales ranging between a few picoseconds and a few nanoseconds, whether one is concerned with diffusion over short length scales or heat exchanges involving small volumes of matter. Optical methods are particularly well-suited to the study of small systems for various reasons:

- Non-contact measurements are possible.
- The optical penetration length can be extremely short, e.g., 10–30 nm in metals.
- Measurements can be made with a broad temporal dynamic range, whether one is working in the frequency domain, e.g., photothermal measurements with modulated optical sources, or in the time domain.

In this context, the development of reliable and easy-to-use femtosecond laser sources has considerably enhanced the potential of optical methods. In the present Chapter, we describe how femtosecond lasers are used, and in particular, the pump–probe method for studying heat transfer in micro- and nanoscopic systems. In this technique, an ultrashort laser pulse is split into two parts: one (the pump), very intense, excites the medium under investigation, whilst the other (the probe), weaker and slightly delayed with respect to the first by being made to follow an optical path of variable length, is used to detect the physical effects induced by the first. By controlling the path of the probe beam to within 1 μm, a time resolution well below the picosecond is possible, more than adequate to study heat transfer phenomena.

Although thermal and elastic effects may be quite well decoupled on long time scales, this is certainly not the case for the time scales considered here (picosecond to nanosecond). We therefore begin by describing acoustic effects caused by absorption of a femtosecond pulse in a simple geometry, namely, an opaque film, neglecting heat diffusion (Sect. 1.1). Having understood the way acoustic effects manifest themselves in the response of a system heated by a first pump pulse, we shall examine three simple situations (Sect. 1.2):

- heat diffusion in an opaque film,
- heat diffusion in a substrate in the presence of an interfacial thermal resistance,
- cooling of a nanoparticle in solution or embedded in a matrix.

We then tackle the mechanisms relevant to optical detection of heat or acoustic gradients, and discuss the difficulties involved in determining the heat diffusion (Sect. 2). Finally, in Sect. 3, we describe the main features of a pump–probe experiment, including several detection devices (reflectometry and interferometry). We also discuss effects induced by the repeat rate of pulsed lasers.

S. Volz (Ed.): Microscale and Nanoscale Heat Transfer, Topics Appl. Physics **107**, 333–359 (2007)

1 Acoustic and Thermal Generation by Ultrashort Laser Pulse

1.1 Acoustic Generation in the Absence of Heat Diffusion

Consider an opaque film in which a laser pulse is absorbed over a distance κ_ν^{-1}, where κ_ν is the optical absorption coefficient. The following assumptions are made:

- The width of the laser pulse is much shorter than the time required for acoustic transients to appear.
- The gradient of the energy deposited by the laser pulse is not modified by either heat or electron diffusion while the acoustic fields are being created.
- The size of the laser spot is much greater than the distance over which the acoustic fields propagate. The problem can then be treated as one-dimensional, in the direction normal to the absorbing medium.
- The medium is isotropic.

The energy $W(z,t)\,\mathrm{d}t$ deposited in the sample by absorption of the laser pulse is given by

$$W(z,t) = \frac{\kappa_\nu(1-R)Q}{S}\mathrm{e}^{-\kappa_\nu z}\Gamma(t)\,, \tag{1}$$

where Q is the energy of one laser pulse, S is the cross-sectional area of the laser beam at the focal point, and R is the reflectance of the film. To simplify the problem, the intensity of the laser spot is assumed to be uniform over the area S. Finally, $\Gamma(t)$ is the Heaviside step function. This energy deposit induces a temperature rise in the layer given by

$$\Delta T(z,t) = \frac{W(z,t)}{\rho C_P}\,, \tag{2}$$

where C_P is the specific heat capacity of the film and ρ its density. A thermal stress

$$\sigma_{ij}^{\mathrm{th}}(z,t) = -\sum_{kl} C_{ijkl}\beta_{kl}\Delta T(z,t) = -\beta\Delta T(z,t)\sum_{kl} C_{ijkl}\delta_{kl} \tag{3}$$

is produced by this temperature rise, where C_{ijkl} is the elasticity tensor and β the thermal expansion coefficient. We consider only the z dependence of the temperature gradient. In this direction, the component σ_{33} of the stress tensor is given by

$$\begin{aligned}\sigma_{33}^{\mathrm{th}}(z,t) &= \sigma_{\mathrm{th}}(z,t) = -\beta\Delta T(z,t)(C_{11}+2C_{12}) = -3\beta B\Delta T(z,t)\\ &= -\frac{3\beta B\kappa_\nu(1-R)Q}{\rho C_P S}\mathrm{e}^{-\kappa_\nu z}\Gamma(t) = -\rho v_{\mathrm{s}}^2\eta_0\mathrm{e}^{-\kappa_\nu z}\Gamma(t)\,,\end{aligned} \tag{4}$$

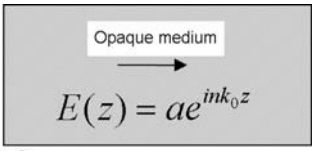

Fig. 1. Semi-infinite, opaque metallic film

where we have introduced the dimensionless parameter

$$\eta_0 = \frac{3\beta B \kappa_\nu (1-R) Q}{\rho^2 v_\mathrm{s}^2 C_P S}\,, \tag{5}$$

which determines the amplitude of the strain induced by the laser pulse. Apart from this, v_s is the speed of longitudinal sound waves in the film and $B = (C_{11} + 2C_{12})/3$ is the incompressibility modulus. The speed of sound is related to the elastic constant C_{11} by $\rho v_\mathrm{s}^2 = C_{11}$. Table 1 gives the values of η_0 for various metals, produced by a pulse of 1 nJ focused on a spot of radius 30 µm. The total stress σ, summing over all thermal stresses σ_th and mechanical stresses $C_{11}\partial u/\partial z$, is given by

$$\sigma(z,t) = \rho v_\mathrm{s}^2 \frac{\partial u}{\partial z} - \rho v_\mathrm{s}^2 \eta_0 \mathrm{e}^{-\kappa_\nu z} \Gamma(t)\,, \tag{6}$$

where $u(z,t)$ is the displacement of the sample surface induced by the acoustic field at time t and at distance z. The equation of propagation for the displacement is thus given by

$$\frac{\partial^2 u}{\partial z^2} - \frac{1}{v_\mathrm{s}^2}\frac{\partial^2 u}{\partial t^2} = -\eta_0 \kappa_\nu \mathrm{e}^{-\kappa_\nu z} \Gamma(t)\,, \tag{7}$$

whilst the equation of propagation for the strain $\eta(z,t) = \partial u/\partial z$ is

$$\frac{\partial^2 \eta}{\partial z^2} - \frac{1}{v_\mathrm{s}^2}\frac{\partial^2 \eta}{\partial t^2} = \eta_0 \kappa_\nu^2 \mathrm{e}^{-\kappa_\nu z} \Gamma(t)\,. \tag{8}$$

The boundary conditions are:

- the sample surface is stress free,
- the film is assumed to be infinitely thick and the acoustic wave thus propagates in the positive z direction.

The solution to this equation is then

$$\eta(z,t) = \frac{\eta_0}{2}\left[(2 - \mathrm{e}^{-\kappa_\nu v_\mathrm{s} t})\mathrm{e}^{-\kappa_\nu z} - \mathrm{sgn}(z - v_\mathrm{s} t)\mathrm{e}^{-\kappa_\nu |z - v_\mathrm{s} t|}\right]\,, \tag{9}$$

and the spatial evolution of the acoustic signal is shown in Fig. 2.

In this expression, there appears a time $\tau_\mathrm{ac} = (\kappa_\nu v_\mathrm{s})^{-1}$ which characterises the scale of acoustic transients and which corresponds to the time taken by

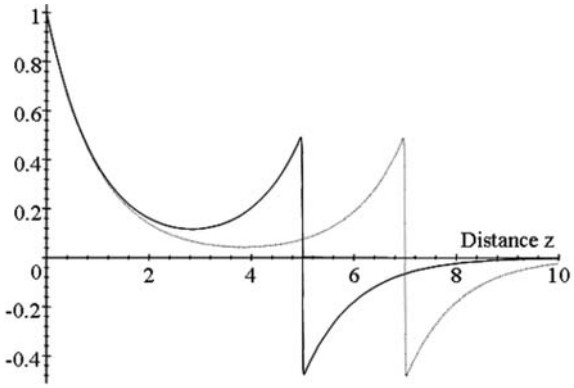

Fig. 2. Strain $\eta(z,t)$ as a function of z for two values of t

a sound wave to travel the optical penetration length κ_ν. As soon as $t \gg \tau_{\rm ac}$, the first term in the expression for $\eta(z,t)$ depends only on z and represents the static expansion induced in the sample by the temperature gradient (since heat diffusion has been neglected). The second term represents an acoustic pulse of width $\tau_{\rm ac}$ propagating in the opaque layer. These acoustic pulses with very high frequency content (several hundred GHz) can only propagate a few hundred nanometers in polycrystalline films at room temperature and over distances of a few hundred μm in extremely pure crystals at very low temperatures, without being attenuated. In these circumstances, the pulses can give rise to one or more acoustic echoes at the surface of the film, after reflecting on interfaces. Translating the time origin so that $t = 0$ corresponds to the arrival of the middle of the acoustic echo at the surface, the strain associated with the first acoustic echo can be written in the form

$$\eta(z,t) = \frac{\eta_0 r_{\rm ac}}{2}\Big[\exp(-\kappa_\nu|z - v_{\rm s}t|)\mathrm{sgn}(z - v_{\rm s}t) + \exp(-\kappa_\nu|z + v_{\rm s}t|)\mathrm{sgn}(z + v_{\rm s}t)\Big]\,, \quad (10)$$

where $r_{\rm ac}$ is the acoustic reflection coefficient for the interface causing the echo. This echo gives rise to a surface displacement given by

$$u(0,t) = \int_{\infty}^{0} \mathrm{d}z\eta(z,t) = -\frac{\eta_0 r_{\rm ac}}{\kappa_\nu}\exp(-\kappa_\nu v_{\rm s}|t|)\,, \quad (11)$$

as shown in Fig. 3.

According to this simple model in which the effects of electron and heat diffusion are neglected, the temporal width of the echoes, directly related to $\tau_{\rm ac} = 1/\kappa_\nu v_{\rm s}$, would be a few picoseconds for many metals, and their spatial extent would be a few tens of nanometers since the speed of sound $v_{\rm s}$ is usually a few nm/ps and the optical penetration depth κ_ν^{-1} in the range 10–30 nm. These acoustic pulses thus have a very high axial resolution (in the direction

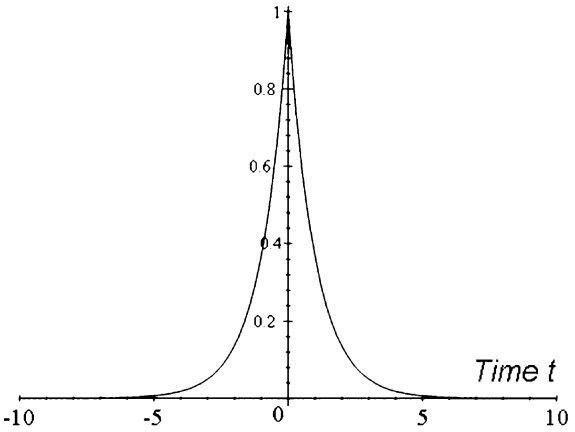

Fig. 3. Surface displacement due to the arrival of an acoustic echo

of propagation) and they give rise to a new field of investigation, viz., picosecond acoustics [1–4]. Apart from its relevance to fundamental research [5–8], this nanoscale SONAR technique has been used for non-destructive assessment of deposited thin films in the semiconductor industry [9–11].

Acoustic phenomena are present whenever a system is heated. They occur in films and bulk samples in the first few picoseconds when a pulse leaves the surface, then in the form of echoes when these pulses return to the surface after a few tens of picoseconds, or a few nanoseconds. For ultrathin films or nanosystems such as nanoparticles, these acoustic phenomena excite resonant modes. Acoustic fields affect the optical properties of systems in two ways: first by altering the geometry of the sample, e.g., strain and displacement of the surface of thin films, or changing the diameter and shape of nanoparticles; and second by perturbing the dielectric properties through a photoelastic coefficient. As an example, the surface displacement induced by the static, non-diffusive expansion and by the acoustic pulse moving away within the film is given by

$$u(0,t) = \int_{\infty}^{0} \mathrm{d}z\eta(z,t) = -\frac{\eta_0}{\kappa_\nu}\left(1 - \mathrm{e}^{-\kappa_\nu v_\mathrm{s} t}\right) . \tag{12}$$

1.2 Taking Heat Diffusion into Account

Opaque Film

Temperature Gradient in the Film

In this section, finite heat conduction is taken into account in the opaque film. For the time domain considered here (from a few picoseconds to a few

Table 1. Predicted surface displacement and strain amplitudes and temperature rise induced by absorption of a laser pulse of energy 1 nJ at wavelength 750 nm, focused on a spot of radius 32 μm. Diffusion has been neglected

	$\eta_0 \times 10^6$	μ_0 (pm)	ΔT (K)		$\eta_0 \times 10^6$	μ_0 (pm)	ΔT (K)
Ag	9.5	0.12	0.2	Hg	1 960	19.5	3.6
Al	100	0.69	2.0	Mg	59	3.51	1.9
Au	9.5	0.12	0.3	Mo	29	0.50	3.1
Co	36	0.46	1.8	Ni	42	0.59	1.8
Cr	25	0.34	2.5	Pt	54	0.68	2.6
Cu	9.2	0.12	0.3	Ti	54	0.99	3.2
Fe	46	0.77	2.2	W	22	0.48	2.8

nanoseconds), the effects of radial diffusion can be neglected and the problem treated as one-dimensional. The temperature field under the surface of the thin film obeys the diffusion equation

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial z^2} + \frac{(1-R)\kappa_\nu Q}{\rho C_P S} \delta(t) \mathrm{e}^{-\kappa_\nu z} , \tag{13}$$

with a source term expressing absorption of the laser pulse. In this equation, $\alpha = k/\rho C_P$ is the diffusivity of the film, with k the thermal conductivity of the film, and $\delta(t)$ the Dirac distribution. The boundary conditions are:

- zero heat flux at the surface, i.e., $\partial T/\partial z|_{z=0} = 0$,
- zero flux from the lower region of the sample.

The Laplace transformed solution $\tilde{T}(p)$ for this equation is thus given by

$$\tilde{T}(p) = \frac{\Delta T}{\sqrt{p}} \frac{\sqrt{p}\mathrm{e}^{-\kappa_\nu z} - \kappa_\nu \sqrt{\alpha}\mathrm{e}^{-z\sqrt{p/\alpha}}}{p - \kappa_\nu^2 \alpha} , \tag{14}$$

where

$$\Delta T = \frac{(1-R)Q\kappa_\nu}{\rho C_P S} .$$

The temperature rises ΔT induced by a pump pulse of 1 nJ focused on a spot of radius 30 μm are shown in Table 1 for the main metals. By inverting the Laplace transform, we retrieve the standard solution to this problem, viz.,

$$T(z,t) = \frac{\Delta T}{\sqrt{4\pi\alpha t}} \int_{-\infty}^{+\infty} \mathrm{d}z' \mathrm{e}^{-(z-z')^2/4\alpha t} \mathrm{e}^{-\kappa_\nu |z'|} , \tag{15}$$

with surface temperature

$$T(0,t) = \Delta T \mathrm{e}^{\kappa_\nu^2 \alpha t} \mathrm{erfc}\left(\sqrt{\kappa_\nu^2 \alpha t}\right) . \tag{16}$$

The characteristic time $\tau_{\rm th}$ in this problem is defined by $\tau_{\rm th} = 1/\kappa_\nu^2\alpha$, equal to several tens of picoseconds in metals. At long times $t \gg \tau_{\rm th}$, $T(0,t) \sim 1/\sqrt{\alpha t}$, as expected for 1D diffusion.

Surface Displacement

As in the last section, this temperature gradient gives rise to an elastic displacement described by

$$\rho\frac{\partial^2 u}{\partial t^2} = \frac{\partial \sigma}{\partial z} = \rho v_{\rm s}^2\frac{\partial^2 u}{\partial z^2} - 3\beta B\frac{\partial T}{\partial z}\,. \tag{17}$$

The Laplace transform $\tilde{u}(p,z)$ of the displacement must therefore satisfy

$$p^2\tilde{u} = v_{\rm s}^2\frac{\partial^2 u}{\partial z^2} + \kappa_\nu v_{\rm s}^2\eta_0\frac{{\rm e}^{-\kappa_\nu z} - {\rm e}^{-z\sqrt{p/\alpha}}}{p-\kappa_\nu^2\alpha}\,, \tag{18}$$

with the same boundary conditions as above (zero stress on the surface and the acoustic wave only propagates in the positive z direction). To keep things simple, we shall only give the expression for the surface displacement:

$$\tilde{u}_0(p) = \frac{\eta_0 v_{\rm s}}{p-\kappa_\nu^2\alpha}\left(\frac{\sqrt{\alpha}}{p}\frac{\kappa_\nu}{\sqrt{p}+v_{\rm s}/\sqrt{\alpha}} - \frac{1}{p+\kappa_\nu v_{\rm s}}\right)\,. \tag{19}$$

By inverse Laplace transform, this leads to the solution

$$u_0(t) = -\frac{\eta_0}{\kappa_\nu}\left[1 - \frac{{\rm e}^{-t/\tau_{\rm ac}}}{\xi+1} - \frac{\xi{\rm e}^{t/\tau_{\rm th}}}{1-\xi^2}{\rm erfc}\left(\sqrt{\frac{t}{\tau_{\rm th}}}\right) + \frac{\xi^2{\rm e}^{\tau_{\rm th}t/\tau_{\rm ac}^2}}{1-\xi^2}{\rm erfc}\left(\sqrt{\frac{\tau_{\rm ph}t}{\tau_{\rm ac}^2}}\right)\right]\,, \tag{20}$$

where

$$\xi = \frac{\tau_{\rm ac}}{\tau_{\rm th}} = \frac{\kappa_\nu\alpha}{v_{\rm s}}\,.$$

The behaviour at long times is given by

$$u_0(t) \sim -\frac{\eta_0}{\kappa_\nu}\left(1 - \sqrt{\frac{\tau_{\rm th}}{\pi t}}\right)\,.$$

The dynamical evolution of the surface displacement is shown in Fig. 4 when heat diffusion is taken into account and compared with the case where it is not.

The same treatment could be used to calculate the strain associated with the acoustic pulse that propagates in the opaque film. Taking into account heat diffusion [2,12], the acoustic pulse is broadened in a way that is no longer antisymmetric. The surface displacement due to echoes is also extended and is no longer symmetric.

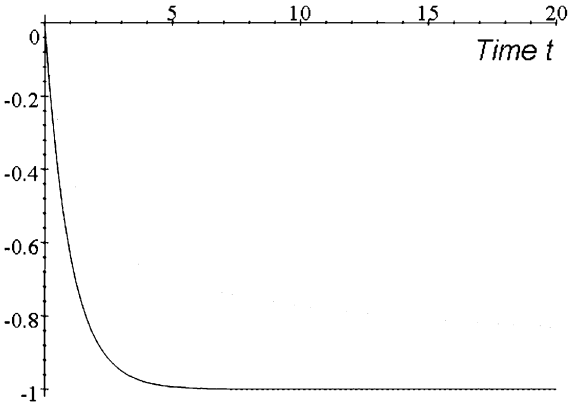

Fig. 4. Surface displacement without diffusion (*continuous curve*) and with (*dotted curve*) for a semi-infinite medium (no acoustic echo)

Experimental Difficulties to Determine the Diffusivity of an Opaque Thin Film

The long-time behaviour going as $1/\sqrt{\alpha t}$ makes it difficult to obtain an accurate determination of the thermal diffusivity because the constant of proportionality $1/\sqrt{\alpha}$ must be multiplied by a prefactor which depends critically on a great many experimental parameters that are hard to control, e.g., geometry of the pump and probe beams, optical absorption coefficient, photo- and acousto-optical parameters, gain of the detection chain, and so on.

The Problem of Electron Diffusion

This example constitutes one of the simplest problems of heat diffusion that can be studied on short time scales by optical techniques. However, this type of study is only useful for materials with a short absorption length, and hence especially for metals. In these systems, electron diffusion can greatly modify the energy gradient deposited by the pump laser, particularly in the noble metals (gold, silver, copper, etc.) where the mean free path of conduction electrons can be much greater than the optical absorption length. To give a rigorous treatment of this problem in metals, one must take into account the thermalisation of both the electron gas and the phonon gas. The temperatures of these two systems satisfy the system of coupled equations

$$\rho c_{\mathrm{e}} \frac{\partial T_{\mathrm{e}}}{\partial t} = k \frac{\partial^2 T_{\mathrm{e}}}{\partial z^2} - g(T_{\mathrm{e}} - T_{\mathrm{l}}) + \frac{(1-R)\kappa_\nu Q \delta(t)}{S} \mathrm{e}^{-\kappa_\nu z} \,, \tag{21}$$

$$\rho c_{\mathrm{p}} \frac{\partial T_{\mathrm{l}}}{\partial t} = g(T_{\mathrm{e}} - T_{\mathrm{l}}) \,, \tag{22}$$

where c_{e} is the specific heat of the electron gas, k the thermal conductivity, g the electron–phonon coupling coefficient, and T_{e} and T_{l} the temperatures of

the electron and phonon gases, respectively. Since the thermal conductivity of the lattice is small in metals, compared with that of the electrons, k will be assumed equal to the macroscopic thermal conductivity. This investigation will not be discussed here but can be found in [13–15].

Opaque Thin Film on a Substrate with Interface Resistance

To extend the range of application of these optical methods to the case of transparent materials, the system under investigation can be coated with a thin metal film which serves both to heat the underlying material and to detect the cooling. Moreover, since the electrons of the metal are confined within the metal film, the problems of electron diffusion are more easily controlled.

In this section, we shall therefore consider a metal film with thickness d subjected at time $t = 0$ to uniform heating ΔT by absorption of a laser pulse. We examine the subsequent decrease of $T_\mathrm{f}(t)$, which characterises the initial temperature difference of this film, by heat diffusion through a thermal resistance R_th into the underlying substrate, assuming that the film is a perfect heat conductor with specific heat c_f and density ρ_f. The substrate has thermal diffusivity $\alpha_\mathrm{s} = k_\mathrm{s}/\rho_\mathrm{s}c_\mathrm{s}$, where k_s is the thermal conductivity of the substrate. The increase in the temperature T_s of the substrate satisfies the diffusion equation

$$\frac{\partial^2 T_\mathrm{s}}{\partial z^2} = \frac{1}{\alpha_\mathrm{s}} \frac{\partial T_\mathrm{s}}{\partial t} \,. \tag{23}$$

The Laplace transform $\tilde{T}_\mathrm{s}(p)$ is thus given by

$$\tilde{T}_\mathrm{s}(p, z) = \tilde{T}_\mathrm{s}(0) \mathrm{e}^{-z\sqrt{p/\alpha_\mathrm{s}}} \,. \tag{24}$$

The cooling $\rho_\mathrm{f} c_\mathrm{f} dS \partial T_\mathrm{f}/\partial t$ is equal to the heat flux in the substrate at $z = 0$, viz.,

$$\rho_\mathrm{f} c_\mathrm{f} dS \frac{\partial T_\mathrm{f}}{\partial t} = S\kappa_\mathrm{s} \left.\frac{\partial T_\mathrm{s}}{\partial z}\right|_{z=0} \,, \tag{25}$$

which leads to the relation between Laplace transforms

$$p\tilde{T}_\mathrm{f} - \Delta T = -\frac{\kappa_\mathrm{s}}{d\rho_\mathrm{f} c_\mathrm{f}} \sqrt{\frac{p}{\alpha_\mathrm{s}}} \tilde{T}_\mathrm{s}(0) \,, \tag{26}$$

where c_f and ρ_f are the specific heat and density of the metal film, respectively. The thermal resistance at the film–substrate interface induces a discontinuity in the temperature given by

$$T_\mathrm{f} - T_\mathrm{s} = -\kappa_\mathrm{s} R_\mathrm{th} \left.\frac{\partial T_\mathrm{s}}{\partial z}\right|_{z=0} \,, \tag{27}$$

whereupon

$$\tilde{T}_\mathrm{f} - \tilde{T}_\mathrm{s}(0) = \kappa_\mathrm{s} R_\mathrm{th} \sqrt{\frac{p}{\alpha_\mathrm{s}}} \tilde{T}_\mathrm{s}(0)\,. \tag{28}$$

Hence,

$$\tilde{T}_\mathrm{s}(0) = \frac{1}{\kappa_\mathrm{s} R_\mathrm{th} \sqrt{\dfrac{p}{\alpha_\mathrm{s}}} + 1} \tilde{T}_\mathrm{f} = \frac{\Delta T - p\tilde{T}_\mathrm{f}}{\dfrac{\kappa_\mathrm{s}}{d\rho_\mathrm{f} C_\mathrm{f}} \sqrt{\dfrac{p}{\alpha_\mathrm{s}}}}\,, \tag{29}$$

and

$$\tilde{T}_\mathrm{f} = \tilde{T}_\mathrm{f} = \frac{\tau_\mathrm{th}}{\sqrt{p\tau_\mathrm{th}}} \frac{\chi\sqrt{p\tau_\mathrm{th}} + B_i}{B_i + \sqrt{p\tau_\mathrm{th}}\left(\chi\sqrt{p\tau_\mathrm{th}} + B_i\right)} \Delta T\,, \tag{30}$$

where $\chi = \rho_\mathrm{s} c_\mathrm{s} / \rho_\mathrm{f} c_\mathrm{f}$. The thermal relaxation time is

$$\tau_\mathrm{th} = \frac{d^2}{\alpha_\mathrm{s} \chi^2} = \frac{d^2 \rho_\mathrm{f}^2 c_\mathrm{f}^2}{\alpha_\mathrm{s} \rho_\mathrm{s}^2 c_\mathrm{s}^2}\,,$$

and

$$B_i = \frac{d}{k_\mathrm{s} R_\mathrm{th}}$$

can be identified as the Biot number characterising exchange between the thin layer and the substrate. By inverting the Laplace transform, we obtain finally

$$T_\mathrm{f}(t) = \frac{1}{2\mathrm{i}\pi} \int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} \mathrm{d}p \tilde{T}_\mathrm{f} \mathrm{e}^{pt} = \frac{2\Delta T\sqrt{\tau_\mathrm{th}}}{\pi} \int_0^{+\infty} \mathrm{d}u \frac{\mathrm{e}^{-u^2 t}}{(1-\tau_\mathrm{R} u^2)^2 + \tau_\mathrm{th} u^2}\,, \tag{31}$$

integrating around a contour that avoids the negative real axis, and defining the resistive relaxation time

$$\tau_\mathrm{R} = \frac{\tau_\mathrm{th}\chi}{B_i} = d\rho_\mathrm{f} c_\mathrm{f} R_\mathrm{th}\,.$$

For a very large Biot number (low interface resistance or low thermal conductivity in the substrate), the relaxation time for diffusion in the substrate dominates and we retrieve the well known result

$$T_\mathrm{f}(t) = \Delta T \exp\left(\frac{t}{\tau_\mathrm{th}}\right) \mathrm{erfc}\left(\sqrt{\frac{t}{\tau_\mathrm{th}}}\right)\,. \tag{32}$$

When the thermal conductivity of the substrate becomes very large ($\tau_\mathrm{th} \to 0$), the resistive time predominates and the decay becomes exponential, viz.,

$$T_\mathrm{f}(t) = \Delta T \exp\left(-\frac{t}{\tau_\mathrm{R}}\right)\,. \tag{33}$$

Table 2. Physical quantities used for the simulation in Fig. 5

	k_s $(W \cdot m^{-1} \cdot K^{-1})$	$\rho_s c_s$ $(J \cdot m^{-3} \cdot K^{-1})$	R_{th} $(W \cdot m^{-2} \cdot K^{-1})$	τ_{th} (ps)	τ_R (ns)	B_i
Silicon	150	1.65×10^6	10^{-8}	240	2.43	0.066
Diamond	2 000	1.76×10^6	2×10^{-8}	16.8	4.86	0.002 5

Two examples of the behaviour of $T_f(T)$ are shown in Fig. 5 for silicon and diamond. The heating metallic film is assumed to be aluminium and to have a thickness of 100 nm. The physical quantities used in this simulation are provided in Table 2.

When the values of τ_{th} and τ_R do not differ too greatly, it is in principle possible to determine both the diffusivity of the investigated substrate and the thermal resistance of the interface. The pump–probe technique has been used with a source of femtosecond laser pulses to measure the Kapitza resistance between a film and its substrate [16–19], the thermal conductivity of thin films as a function of the temperature or composition [20–24], and the thermal conductivity of semiconducting superlattices [25].

Metallic Particle

As in the last example, we consider the cooling of a metallic nanoparticle in contact with its environment (a liquid or transparent solid matrix) via an interfacial thermal resistance R_{th}. This problem is very close to the one considered in the last section. The nanoparticle, assumed to be spherical with radius a, is subjected at time $t = 0$ to uniform heating ΔT due to the absorption of a laser pulse. We study the subsequent decrease in the temperature $T_p(t)$ of this sphere by heat diffusion into the liquid or matrix through the thermal resistance R_{th}, assuming that the particle is a perfect heat conductor with specific heat c_p and density ρ_p. The surroundings have thermal diffusivity $\alpha_s = k_s/\rho_s c_s$, where k_s is the thermal conductivity of the solution or matrix. The increase in the temperature T_s of the solution satisfies the diffusion equation, given in spherical coordinates by

$$\frac{\partial^2 (rT_s)}{\partial r^2} = \frac{1}{\alpha_s} \frac{\partial (rT_s)}{\partial t} \,. \tag{34}$$

The Laplace transform $\tilde{T}_s(p)$ is thus given by

$$\tilde{T}_s(p, r) = \tilde{T}_s(a) \frac{a e^{-w(r/a-1)}}{r} \,, \tag{35}$$

where $w = a\sqrt{p/\alpha_s}$. The cooling

$$\frac{4}{3} \pi a^3 \rho_p c_p \frac{\partial T_p}{\partial t}$$

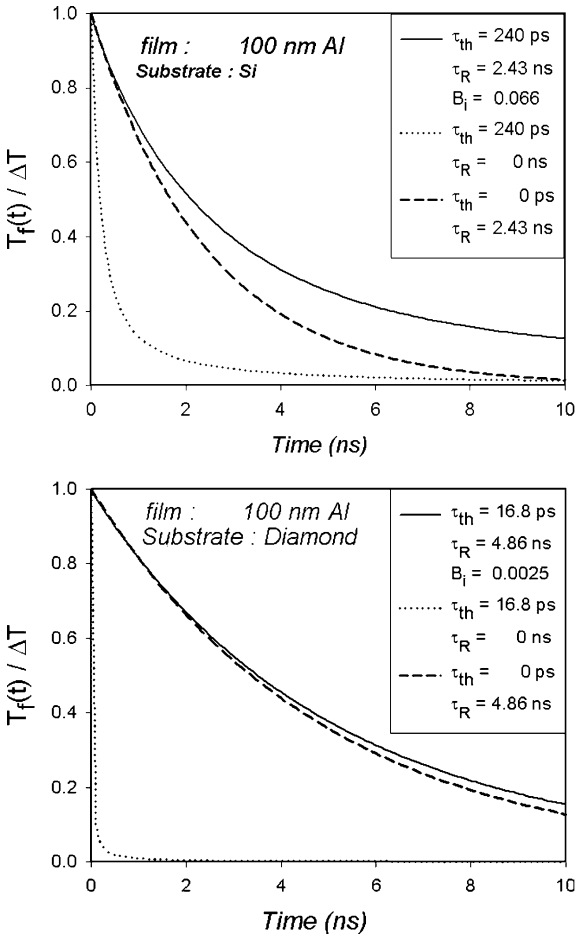

Fig. 5. Simulation of cooling in a metallic film deposited on silicon (*top*) and on diamond (*bottom*)

is equal to the heat flux in the solution at $r = a$, i.e.,

$$\frac{4}{3}\pi a^3 \rho_\mathrm{p} c_\mathrm{p} \frac{\partial T_\mathrm{p}}{\partial t} = 4\pi a^2 k_\mathrm{s} \left.\frac{\partial T_\mathrm{s}}{\partial r}\right|_{r=a} , \tag{36}$$

leading to the following relation between Laplace transforms:

$$p\tilde{T}_\mathrm{p} - \Delta T = -\chi \alpha_\mathrm{s} \frac{w+1}{a^2} \tilde{T}_\mathrm{s}(a) , \tag{37}$$

where

$$\chi = \frac{3\rho_\mathrm{s} c_\mathrm{s}}{\rho_\mathrm{p} c_\mathrm{p}} .$$

At the particle–solution interface, the thermal resistance induces a discontinuity in the temperature given by

$$T_\mathrm{p} - T_\mathrm{s} = -k_\mathrm{s} R_\mathrm{th} \left.\frac{\partial T_\mathrm{s}}{\partial r}\right|_{r=a} , \tag{38}$$

whence,

$$\tilde{T}_\mathrm{p} - \tilde{T}_\mathrm{s}(a) = \frac{w+1}{B_i}\tilde{T}_\mathrm{s}(a) , \tag{39}$$

where $B_i = a/k_\mathrm{s}R_\mathrm{th}$ is the Biot number. Hence,

$$\tilde{T}_\mathrm{s}(a) = \frac{B_i}{aw + B_i + 1}\tilde{T}_\mathrm{p} , \tag{40}$$

and

$$\tilde{T}_\mathrm{p} = \frac{a^2\Delta T}{\alpha_\mathrm{s}} \frac{aw + B_i + 1}{w^2(w + B_i + 1) + \chi B_i(w+1)} . \tag{41}$$

Inverting the Laplace transform, we obtain

$$\begin{aligned} T_\mathrm{p}(t) &= \frac{1}{2\mathrm{i}\pi}\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} \mathrm{d}p\tilde{T}_\mathrm{p}\mathrm{e}^{pt} \\ &= \frac{2\chi B_i^2\Delta T}{\pi}\int_0^\infty \frac{u^2\mathrm{e}^{-u^2t/\tau_\mathrm{c}}\,\mathrm{d}u}{\left[(B_i+1)u^2 - \chi B_i\right]^2 + u^2(u^2-\chi B_i)^2} , \end{aligned} \tag{42}$$

where $\tau_\mathrm{c} = a^2/\alpha_\mathrm{s}$. The same asymptotic behaviour is obtained as in the last case for small and large Biot numbers. After long times, the diffusion goes as $1/t\sqrt{t}$, as expected for a 3D diffusion law, and the thermal relaxation time is $\tau_\mathrm{th} = \tau_\mathrm{c}/\sqrt[3]{4\chi^2}$. As an indication, characteristic cooling times τ_th for gold nanoparticles of diameter 10 nm are 37 ps in an aqueous solution, 80 ps in glycerol, and 10 ps in a silica matrix.

2 Optical Detection of Thermal and Acoustic Transients

Thermal and acoustic effects affect the optical response of a system in two ways:

– The dielectric constant, and hence the refractive index, of the material is modified by a temperature rise or an elastic strain. The reflection and transmission coefficients, or the scattering cross-section thus depend on the temperature and the strain.

– The geometry of the object also changes. The surface of a film is curved and moves when an acoustic echo reaches the surface. The diameter and shape of a nanoparticle vary under the effect of a resonant vibration. These perturbations are likely to modify the amplitude, direction of propagation, phase, and polarisation of a reflected, transmitted or scattered beam. They can also modify the frequency of this electromagnetic wave, but the spectral shifts remain much smaller than the spectral width of the probe pulse.

The various characteristics of a probe beam can thus be used for the purposes of detection. The simplest technique is to measure the intensity of the probe beam and thus to carry out transient reflectance or transmittance measurements. Deflection measurements have also been proposed to study thermal [26] and acoustic transients [27]. Interferometric detection can be used for simultaneous measurement of the amplitude and phase of the electromagnetic field of the probe beam [28] and it is has been proposed to determine acoustic effects [4, 7, 29, 30].

Although the optical penetration length is small in metals, to be rigorous, one ought to take into account the gradient of the refractive index induced by a temperature or strain gradient to be sure to determine the thermal and acoustic transients correctly. We shall discuss a simple case, namely, the change in reflectance of an opaque thin film.

Change in Reflectance of an Opaque Thin Film

At any point of the sample, the presence of a thermal or acoustic gradient induces a gradient in the dielectric constant

$$\varepsilon(z,t) = \varepsilon + \Delta\varepsilon(z,t) = \left[n + \Delta n(z,t)\right]^2 , \quad \Delta n(z,t) = \frac{\partial n}{\partial X} X(z,t) ,$$

where X represents the temperature or the strain, n is the complex refractive index of the medium, and $\partial n/\partial X$ is a complex thermo-optical or acousto-optical constant. We must then solve the equation of propagation of the electromagnetic field E, viz.,

$$\frac{\partial^2 E}{\partial z^2} = \left[\varepsilon + \Delta\varepsilon(z,t)\right] \frac{\partial^2 E}{\partial t^2} . \tag{43}$$

The solution to this equation is

$$E_t(z,t) = t_{01} \int_{-\infty}^{+\infty} \mathrm{d}\Omega E_i(\Omega) \mathrm{e}^{\mathrm{i}(\Omega t - n k_0 z)} ,$$

when there is no gradient, where Ω is the angular frequency of the electromagnetic field, k_0 is the wave vector in vacuum, and $t_{01} = 2/(1+n)$ is the transmission coefficient of the air–sample interface. Since the frequency of the light wave is much greater than the spectrum of $\Delta\varepsilon(z,t)$, the equation

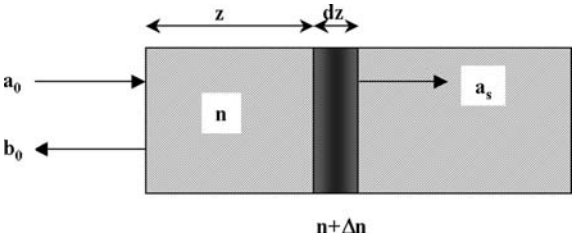

Fig. 6. Scattering of an electromagnetic field by an acoustic perturbation localised at z

of propagation can be solved for a Fourier component of the electromagnetic field, thus treating the perturbation induced by the gradient as quasi-static.

Assume to begin with that only a slice of the material of thickness $\mathrm{d}z$ has its refractive index n modified by an amount Δn at a distance z from the surface (see Fig. 6). We seek an expression for the reflection coefficient of the material in the presence of this perturbed layer.

Using a transfer matrix technique, the amplitudes of the incident, reflected and transmitted fields can be related by

$$\begin{pmatrix} a_0 \\ b_0 \end{pmatrix} = \frac{1}{t_{01}t_{12}t_{21}} \begin{pmatrix} 1 & r_{01} \\ r_{01} & 1 \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}nk_0z} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}nk_0z} \end{pmatrix} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \times \begin{pmatrix} \mathrm{e}^{-\mathrm{i}(n+\Delta n)k_0\,\mathrm{d}z} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}(n+\Delta n)k_0\,\mathrm{d}z} \end{pmatrix} \begin{pmatrix} 1 & r_{21} \\ r_{21} & 1 \end{pmatrix} \begin{pmatrix} a_\mathrm{s} \\ 0 \end{pmatrix}, \quad (44)$$

where

$$r_{01} = \frac{1}{1+n}, \quad r_{21} = \frac{\Delta n}{2n+\Delta n}, \quad \Delta n = \frac{\Delta\varepsilon}{2n}.$$

The reflection coefficient $r(z,t)$ can then be deduced from a truncated expansion in powers of Δn and $\mathrm{d}z$:

$$r(z,t) = \frac{b_0}{a_0} = r_{01} + \mathrm{i}k_0(1-r_{01}^2)\mathrm{e}^{2\mathrm{i}k_0nz}\Delta n(z,t)\,\mathrm{d}z \quad (45)$$

and

$$\frac{\Delta r}{r} = \frac{r(z,t)-r_{01}}{r_{01}} = \mathrm{i}k_0\frac{1-r_{01}^2}{r_{01}}\mathrm{e}^{2\mathrm{i}k_0nz}\Delta n(z,t)\,\mathrm{d}z\,. \quad (46)$$

To obtain the total contribution of the gradient to the change in reflectance, one must integrate over the whole thickness of the material. Hence,

$$\left.\frac{\Delta r(t)}{r}\right|_{\mathrm{grad}} = \mathrm{i}k_0\frac{\partial n}{\partial X}\frac{4n}{1-n^2}\int_0^{+\infty}\mathrm{d}zX(z,t)\mathrm{e}^{2\mathrm{i}k_0nz}\,. \quad (47)$$

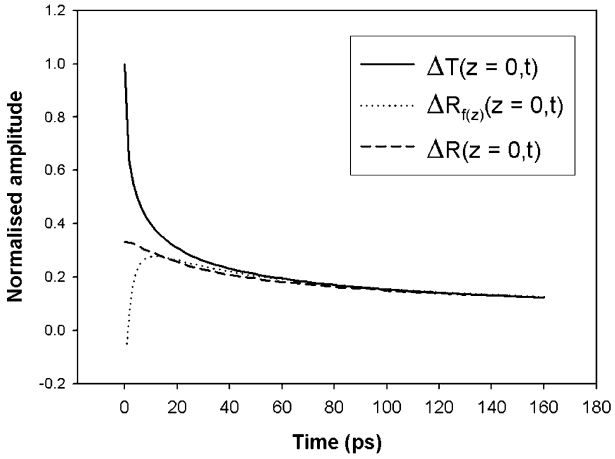

Fig. 7. Time dependence of the reflectance of a bulk tungsten layer. *Continuous curve*: Surface temperature. *Dashed curve*: Experimental curve. *Dotted curve*: Simulation taking into account detection over the optical penetration depth. It has been assumed that $\Im(\partial n/\partial T) = 0$

The surface displacement u_0 also contributes to the imaginary part of $\Delta r/r$, viz.,

$$\left.\frac{\Delta r(t)}{r}\right|_{\text{geom}} = 2\mathrm{i}k_0 u_0(t)\,. \tag{48}$$

The total contribution from acoustic and thermal effects is therefore

$$\frac{\Delta r(t)}{r} = \mathrm{i}k_0 \left\{ 2u_0(t) + \frac{4n}{1-n^2} \times \int_0^{+\infty} \mathrm{d}z \left[\frac{\partial n}{\partial T}\Delta T(z,t) + \frac{\partial n}{\partial \eta}\Delta\eta(z,t) \right] \mathrm{e}^{2\mathrm{i}k_0 n z} \right\}. \tag{49}$$

Thermal and acoustic effects contribute simultaneously to $\Delta r/r$ in the first few picoseconds after absorption of the pump laser pulse and there is no hope of using the first few picoseconds of a pump–probe experiment to measure the thermal diffusivity of a film without taking into account the acoustic pulse that leaves the surface. The second consequence of (49) is that the thermal gradient beneath the surface must also be taken into account in order to give an adequate treatment of the first few moments of diffusion. After the time required for the heat to diffuse beyond the optical penetration length ($\tau_{\text{th}} = 1/\kappa_\nu^2\alpha \sim 10\,\text{ps}$), one can justifiably consider that only the temperature at $z = 0$ is probed [31]. The consequences for the change in reflectance of taking into account the thermal gradient are illustrated in Fig. 7.

One of the difficulties raised by using (49) relates to the fact that the coefficients $\partial n/\partial X$ are not well known, or not known at all. As far as the coeffi-

cients $\partial n/\partial T$ are concerned, there are differential measurement techniques for determining them over a broad spectral range [32,33]. The coefficients $\partial n/\partial X$ can be determined by simultaneous analysis of the two forms of acoustic echo seen in the real and imaginary parts of $\Delta r/r$. Fitting these signals is realistic enough, since a significant part of the contribution to the imaginary part of $\Delta r/r$ comes from the surface displacement, which can therefore be used to calibrate [4, 12, 30]. As for the photothermal coefficients, the photoelastic coefficients also exhibit a great sensitivity to the optical wavelength [34]. This is particularly true for nanoparticles when the probe frequency is close to the plasmon resonance [35, 36].

3 Experimental Setups

In this section, we shall describe a standard coupled pump–probe setup, either with transient transmittance/reflectance detection, or with interferometric detection. As mentioned in the introduction, titanium:sapphire oscillators are the most commonly used today, delivering laser pulses with width less than 100 fs in the wavelength range 700–1000 nm, with a repeat rate close to 100 MHz and power output around 1 W. The energy per pulse is thus around 10 nJ.

As shown in Fig. 8, as soon as it leaves the oscillator, the beam is split into two: a pump beam which serves to heat the system and a probe beam which measures the subsequent cooling. There are various ways to carry out this separation of the beam: a beam splitting plate or cube, a polarising cube, or a frequency-doubling crystal. The most effective way to discriminate between the pump and probe in the detector is to use a frequency-doubling crystal together with dichroic plates or filters, since the pump and probe beams have different colours in this configuration, which is said to be non-degenerate. For a degenerate configuration, where the pump and probe have the same wavelength, a polarising cube can be used to obtain orthogonal polarisations in the pump and probe, whereupon they may once again be discriminated at the detector.

One of the two beams is then delayed (or advanced) with respect to the other. If the detection scheme is complicated, it may be preferable to advance the pump. For delays exceeding a few nanoseconds (1 ns corresponds to a change of 30 cm in the optical path), one must take into account the spread of the laser beam because, at the point where the pump and probe beams overlap, the diameter of the focal spot of the retarded (or advanced) beam can vary significantly. Furthermore, it is difficult to maintain a perfect overlap of the two beams during the delay. A correcting device has been proposed by *Capinski* [37]. The beam is injected into an optical fibre at the output of the delay line. When it emerges from the fibre, the diameter and direction of the beam are then independent of the pump–probe delay. The intensity may vary but a servo system can be used to hold it constant, or the intensity

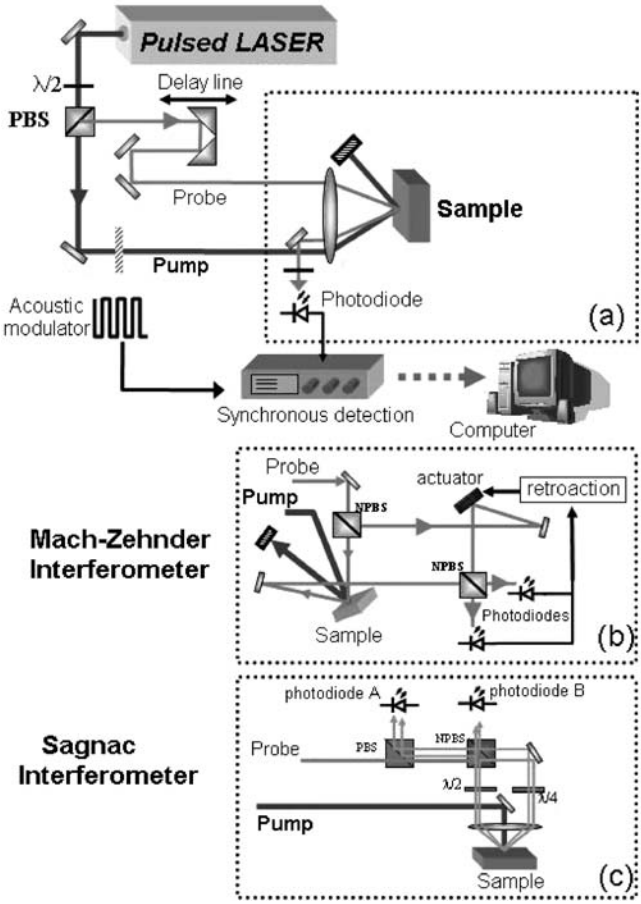

Fig. 8. Experimental setups. Detection by (**a**) reflectance, (**b**) Mach–Zender interferometry, (**c**) Sagnac interferometry. The three forms of detection are interchangeable. Beam splitting cubes can be non-polarising (NPBS) or polarising (PBS)

can be measured and the signal normalised at the output of the synchronous detector. This correction would appear to be necessary if one is to maintain an acceptable dynamic measurement range. The disadvantage in using a fibre is a deterioration in the pulse width and spectrum, although this may be compensated at the outset, and the possible appearance of nonlinear effects in the fibre.

Reflectance variations expected as a result of thermal effects are typically in the range 10^{-6}–10^{-4}. Since intensity fluctuations at the oscillator output are of the order of 1 %, synchronous detection must be used. The pump beam is modulated with an angular frequency Ω that is then sought in the signal emerging from the detectors. Several types of modulator can be used:

– Mechanical choppers up to a few tens of kHz. The advantage here is that they do not broaden the laser pulses.
– Acousto-optical or electro-optical modulators which can modulate up to a few MHz, but which significantly broaden the laser pulses.

After interaction with the system under investigation, the reflected, transmitted or scattered probe beam is detected using a simple photodiode whose output is amplified by the synchronous detector. It is not generally necessary to use ultrasensitive or ultrafast detectors, because the average power of the probe beam is of the order of the mW and the modulation frequency does not exceed a few MHz. The sensitivity expected for this type of device is around 10^{-7}.

3.1 Interferometric Detection

It is sometimes useful to carry out a simultaneous measurement of the real and imaginary parts of the change in reflectance (or transmittance) induced by the pump beam. When measuring the reflectance of an opaque film, the surface displacement contributes to the imaginary part of this coefficient. We set

$$\frac{\widehat{\Delta r}}{r} = a + \mathrm{i}\phi\,. \tag{50}$$

Since this quantity is very small, a can be considered as the change in amplitude of the reflected probe beam and ϕ as the change in phase. In any interferometer, the beam carrying the measurement $(a + \mathrm{i}\phi)$, with intensity $I_\mathrm{s} = E_\mathrm{s}^2$, is made to interfere on a detector with a reference beam of intensity $I_\mathrm{r} = E_\mathrm{r}^2$. The detected intensity is thus

$$\begin{aligned} I(\theta) &= \left|E_\mathrm{s}(1 + a + \mathrm{i}\phi) + E_\mathrm{r}\mathrm{e}^{\mathrm{i}\theta}\right|^2 \\ &= E_\mathrm{s}^2(1 + 2a) + E_\mathrm{r}^2 + 2E_\mathrm{s}E_\mathrm{r}\left[(1 + a)\cos\theta + \phi\sin\theta\right], \end{aligned} \tag{51}$$

where θ is the optical path difference between probe and reference. By choosing an operating point for the interferometer such that $\theta = \pm\pi/2$, we obtain the intensity variation

$$\frac{\Delta I}{I_0}\left(\pm\frac{\pi}{2}\right) = \frac{2I_\mathrm{s}}{I_0}(a \pm \xi\phi)\,, \quad \text{where} \quad \xi = \sqrt{\frac{I_\mathrm{r}}{I_\mathrm{s}}}\,.$$

Therefore, using a single detector and two experiments carried out successively for $\theta = +\pi/2$ and $\theta = -\pi/2$, the quantities a and ϕ can be obtained independently. Moreover, in most interferometric detection devices, two detectors can be used simultaneously to detect complementary signals $I(\theta)$ and $I(\theta+\pi)$, so that a and ϕ can be obtained in a single measurement by addition and subtraction, i.e., $I(\theta)\pm I(\theta+\pi)$. Two types of interferometer are depicted in Figs. 8b and c.

Sagnac Interferometer. Differential Detection

The Sagnac interferometer is particularly useful when pulsed lasers are used [4, 7]. In this type of setup, the probe beam travels around a loop in two opposite directions and the sample is placed at a non-symmetrical position on this loop. One can then compare two probe beams that have been reflected (or transmitted) by the same point of the sample at two different times t and $t+\tau_s$ relative to the pump pulse. In addition, a half-wave plate is inserted to rotate the polarisation of the probe through 90° along its path. A quarter-wave plate is also introduced and oriented in such a way that the two polarisations are out of phase by $\pm\pi/2$. In this way, the two beams propagating in opposite directions cross the quarter-wave plate polarised in different directions and are retarded or advanced by $\pi/2$ relative to one another. Assuming to simplify that the beam splitting cubes (whether polarising or not) are perfect (splitting into two exact halves), it is easy to show that the photodiodes A and B detect signals given by

$$\begin{aligned} S_A &= \left[a(t)+a(t+\tau_s)\right] \pm \left[\phi(t+\tau_s)-\phi(t)\right], \\ S_B &= \left[a(t)+a(t+\tau_s)\right] \mp \left[\phi(t+\tau_s)-\phi(t)\right]. \end{aligned} \tag{52}$$

By adding and subtracting the signals from the two diodes, $a(t+\tau_s)+a(t)$ and $\phi(t+\tau_s)-\phi(t)$ can thus be measured simultaneously. This device thus provides a natural way of subtracting, in the imaginary part ϕ of $\Delta r/r$, the stationary signal arising from cumulative effects due to the train of pump pulses to be described in the next section. The other great advantage with the Sagnac interferometer is its high mechanical stability, since any vibrations will affect both arms in the same way. However, for slowly decaying phenomena, the deconvolution of effects arising at times t and $t+\tau_s$ is sometimes problematic.

Sensitivity of Interferometric Measurements

The sensitivity of an interferometric measurement is equivalent to the sensitivity of a reflectance measurement (10^{-7}). If it is assumed that the change in phase ϕ is only due to a surface displacement, this sensitivity leads to a displacement $4\pi u/\lambda = 10^{-7}$, which gives, for a wavelength of 750 nm, a resolution of 6 fm in the measurement of the surface displacement induced by the pump pulse. Moreover, for weakly reflecting systems, this kind of setup allows the possibility of heterodyning the reflected beam with a more intense reference beam (except in the Sagnac interferometer, where probe and reference are reflected from the same point of the sample).

3.2 Cumulative Effects Due to the Pump Pulse Train

The repeat period of the most commonly used titanium:sapphire oscillators is around ten nanoseconds, and thermal effects induced by a first pump pulse

may not have time to fully relax before the next pulse arrives to heat the sample [25, 37]. Moreover, in order to obtain a suitable dynamic detection range, the pump pulse train must be modulated and variations in the probe beam recorded (whether it is reflected, scattered or transmitted) using synchronous detection. The modulation frequency can be anywhere between a few kHz and a few MHz. Three time scales are thus relevant in this pump–probe experiment:

- the laser pulse width (100 fs),
- the time interval between two consecutive pump or probe pulses (10 ns),
- the modulation period (from 1 ms to 1 μs).

Since the laser pulses are very short compared with the duration of the thermal phenomena under investigation, the pump and probe pulse train can be expressed in the form of Dirac combs:

$$\begin{aligned} I_{\mathrm{P}}(t) &= I_{\mathrm{P}} \sum_{m=-\infty}^{m=+\infty} M(mT)\delta(t - m\check{T})\,, \\ I_{\mathrm{S}}(t) &= I_{\mathrm{S}} \sum_{n=-\infty}^{n=+\infty} \delta(t - \tau_{\mathrm{ps}} - n\check{T})\,, \end{aligned} \tag{53}$$

where $M(t)$ represents the modulation of the pump beam, $\check{T}$ the repeat period, τ_{ps} the delay of the probe with respect to the pump, and I_{P} and I_{S} the intensities of the pump and probe beams, respectively. The change in reflectance (or transmittance) induced by the pump pulse train can thus be written

$$\Delta r(t) = I_{\mathrm{P}} \sum_{n=-\infty}^{n=+\infty} M(m\check{T})\widehat{\Delta r}(t - mT)\,, \tag{54}$$

where $\widehat{\Delta r}(t)$ is the change induced by a single pulse. The signal detected at the photodiode is therefore

$$S(t) \propto I_{\mathrm{S}}(t)\Delta r(t) = S_0 \sum_{m,n} M(m\check{T})\widehat{\Delta r}(t - m\check{T})\delta(t - \tau_{\mathrm{ps}} - n\check{T})\,. \tag{55}$$

The Fourier transform $S(\omega)$ of $S(t)$ is given by

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{+\infty} \mathrm{d}t S(t)\mathrm{e}^{-\mathrm{i}\omega t} \\ &= S_0 \sum_{m,n} M(m\check{T})\widehat{\Delta r}(\tau_{\mathrm{ps}} + n\check{T} - m\check{T})\mathrm{e}^{-\mathrm{i}\omega(\tau_{\mathrm{ps}} + n\check{T})} \\ &= S_0 \sum_{m,p} M(m\check{T})\widehat{\Delta r}(\tau_{\mathrm{ps}} + p\check{T})\mathrm{e}^{-\mathrm{i}\omega\left[\tau_{\mathrm{ps}} + (m+p)\check{T}\right]}\,. \end{aligned} \tag{56}$$

The modulation function can be expanded in a Fourier series

$$M(m\check{T}) = \sum_{k=0}^{\infty} M_k \mathrm{e}^{\mathrm{i}mk\Omega\check{T}} ,$$

where $\Omega/2\pi$ is the modulation frequency. Hence,

$$\begin{aligned} S(\omega) &= S_0 \sum_{m,p,k} M_\mathrm{k} \widehat{\Delta r}(\tau_\mathrm{ps} + p\check{T}) \mathrm{e}^{\mathrm{i}m\check{T}(\Omega k-\omega)} \mathrm{e}^{-\mathrm{i}\omega(\tau_\mathrm{ps}+p\check{T})} \\ &= \frac{2\pi S_0}{\check{T}} \sum_{p,r,k} M_\mathrm{k} \widehat{\Delta r}(\tau_\mathrm{ps} + p\check{T}) \mathrm{e}^{-\mathrm{i}\omega(\tau_\mathrm{ps}+p\check{T})} \delta(\Omega k - \omega) . \end{aligned} \tag{57}$$

Assuming that the synchronous detection gives access to the line $k = 1$ in the spectrum of $S(\omega)$, the signal output from this detection is

$$S_\Omega(\tau_\mathrm{ps}) = \overline{S} \sum_{p=0}^{+\infty} \widehat{\Delta r}(\tau_\mathrm{ps} + p\check{T}) \mathrm{e}^{-\mathrm{i}\Omega(\tau_\mathrm{ps}+p\check{T})} , \tag{58}$$

where only the positive values of p are taken into account, given the causality of the response $\widehat{\Delta r}(t)$. We also assume that the reference phase of the synchronous detector is defined so that the prefactor $\overline{S}$ is real. Note that

$$S_\Omega(\check{T}) = \overline{S} \sum_{p=1}^{+\infty} \widehat{\Delta r}(p\check{T}) \mathrm{e}^{-\mathrm{i}p\Omega\check{T}}$$

and that the discontinuity at $\tau_\mathrm{ps} = 0$ is given by

$$S_\Omega(0) - S_\Omega(\check{T}) = \overline{S}\widehat{\Delta r}(0) .$$

To assess the difference between the behaviour of $S_\Omega(\tau_\mathrm{ps})$ and $\widehat{\Delta r}(\tau_\mathrm{ps})$, it is best to consider a concrete example that is not too difficult to calculate, e.g., $\widehat{\Delta r}(t) = \mathrm{e}^{-t/\tau_\mathrm{c}}$. We then have

$$S_\Omega(\tau_\mathrm{ps}) = \overline{S} \frac{\mathrm{e}^{-\tau_\mathrm{ps}/\tau_\mathrm{c}} \mathrm{e}^{-\mathrm{i}\Omega\tau_\mathrm{ps}}}{1 - \mathrm{e}^{-\check{T}/\tau_\mathrm{c}} \mathrm{e}^{-\mathrm{i}\Omega\check{T}}} . \tag{59}$$

Several conclusions can be drawn from this relation:

– The time τ_c is generally shorter than $\check{T}$ and much shorter than $2\pi/\Omega$ (otherwise one could not justify using short pulses). For values $\check{T}/\tau_\mathrm{c} = 2$ and $\Omega\check{T} = 0.15$,

$$S_\Omega(\tau_\mathrm{ps}) \approx 1.15\overline{S}\mathrm{e}^{-\tau_\mathrm{ps}/\tau_\mathrm{c}} .$$

Only the 'in phase' output of the synchronous detector [the real part of $S_\Omega(\tau_\mathrm{ps})$] contains a significant contribution. The dependencies of $S_\Omega(\tau_\mathrm{ps})$ and $\widehat{\Delta r}(\tau_\mathrm{ps})$ are roughly the same but the prefactor is increased by 15 %.

– At long times, new thermal diffusion phenomena occur: lateral diffusion from the initial laser spot, diffusion into the underlying substrate, etc. Consequently, $S_\Omega(\tau_{\rm ps})$ always contains a 'continuous' component that can be defined by

$$\overline{S}_\Omega = \overline{S}\sum_{p=0}^{+\infty}\widehat{\Delta r}(p\check{T})\mathrm{e}^{\mathrm{i}\Omega p\check{T}} \approx \frac{\overline{S}}{\overline{T}}\int_0^\infty \widehat{\Delta r}(t)\mathrm{e}^{\mathrm{i}\Omega t}\,.$$

The short-time dependence can thus be neglected and the 'continuous' component $\overline{S}_\Omega$ represents the photothermal response of the material heated periodically by the pump beam at the modulation frequency $2\pi/\Omega$. The impulsive aspect of the pump and probe beams disappears totally in the analysis of this stationary contribution, whose component in quadrature is no longer negligible. A phase and amplitude analysis of the kind ordinarily carried out in photothermal experiments is then preferable. This stationary contribution, whose amplitude falls off with the modulation frequency, can be much greater than the time-dependent part $S_\Omega(\tau_{\rm ps}) - \overline{S}_\Omega$. To reduce it, one can either increase the modulation frequency or reduce the modulation width. A Sagnac interferometer can also eliminate this continuous component in the detection of the imaginary part $\Im(\widehat{\Delta r}(\tau_{\rm ps}))$. To quantify this continuous component, we may once again assume an exponential decrease for $\widehat{\Delta r}$, but with a much longer characteristic time $\tau_{\rm l}$ than considered previously ($\tau_{\rm l} \gg \tau_{\rm c}$) and longer than the modulation period $\check{T}$. This gives

$$\overline{S}_\Omega = \overline{S}\frac{\tau_{\rm l}}{\overline{T}}\frac{1-\mathrm{i}\Omega\tau_{\rm l}}{1+(\Omega\tau_{\rm l})^2}\,.$$

For long enough times $\tau_{\rm l}$ (or high enough modulation frequencies), the imaginary part (in phase quadrature) of the continuous component $\overline{S}_\Omega$ is no longer negligible compared with the real part (in phase). Moreover, the amplitude of this component falls off as Ω increases.

The conclusions from this calculation do not really contain any surprises: at short times, the time dependence $\widehat{\Delta r}$ only appears in the real part of the synchronous detector, provided that the reference phase of the detector has been suitably adjusted. The in-phase signal is proportional to the real part of the reflectance but an accumulation effect due to successive pump pulses must nevertheless be taken into acount when analysing the absolute amplitude of the signal. This time dependence comes along with a continuous background component which appears in both the output in phase and the output in quadrature from the synchronous detection. The amplitude of the continuous component tends to zero at high modulation frequencies. The pulse-like aspect of the pump beam does not have to be taken into account to analyse this stationary signal.

4 Conclusion

Femtosecond laser sources can thus be used with a pump–probe setup to investigate the thermal properties of thin films, multilayers, and nano-objects, and also to determine contact resistances. They provide the very high temporal resolution required to study small spatial dimensions, and a good dynamic range, combined with the fact that they can be implemented without alteration of the subject under investigation. Several problems have been raised here:

- Thermal effects on short time scales ($< 100\,\mathrm{ps}$) are always associated with acoustic phenomena which must be correctly understood if the determination of thermal properties is to remain unbiased.
- Behaviour on very short time scales cannot be described without taking into account the electron gas.
- A measurement of transient reflectance does not detect the surface temperature in the first few picoseconds but is sensitive to the thermal gradient over the whole optical penetration depth.
- Since the asymptotic behaviour on long time scales ($> 10\,\mathrm{ns}$) goes as $t^{-D/2}$, where D is the size of the relevant system, it serves little purpose in the determination of the thermal diffusivity.

One must therefore work in the time domain from 50 ps to 10 ns. A configuration involving a perfectly conducting medium (a film deposited on a substrate or nanoparticle) in contact with its surroundings would seem to be the most favourable both for measuring the diffusivity of the substrate or matrix, and for determining the interfacial thermal resistance. It is easy enough to measure the transient reflectance (or transmittance), even if certain precautions must be taken for long optical delays.

References

[1] C. Thomsen, J. Strait, Z. Vardeny, H. J. Maris, J. Tauc, J. J. Hauser: Phys. Rev. Lett. **53**, 989 (1984) 337
[2] C. Thomsen, H. T. Grahn, H. J. Maris, J. Tauc: Phys. Rev. B **34**, 4129 (1986) 337, 339
[3] H. J. Maris: Pour la Science **245**, 88 (1998) 337
[4] B. Perrin: *Systèmes Femtosecondes* (Publications de l'Université de Saint Etienne 2001) pp. 65–89 337, 346, 349, 352
[5] T. C. Zhu, H. J. Maris, J. Tauc: Phys. Rev. B **44**, 4281 (1991) 337
[6] H.-Y. Hao, H. J. Maris: Phys. Rev. B **64**, 064302 (2001) 337
[7] J.-Y. Duquesne, B. Perrin: Phys. Rev. B **68**, 134205 (2003) 337, 346, 352
[8] C. Rossignol, B. Perrin, B. Bonello, P. Djémia, P. Moch, H. Hurdequint: Phys. Rev. B **70**, 094102 (2004) 337
[9] H. N. Lin, R. J. Stoner, H. J. Maris: J. Nondestructive Evaluation **9**, 239 (1990) 337

[10] C. Rossignol, B. Perrin, S. Laborde, L. Vandenbulcke, M. I. Barros, P. Djémia: J. Appl. Phys. **95**, 4157 (2004) 337
[11] C. Rossignol, B. Perrin: IEEE Trans. Ultrason., Ferroelec., Freq. Cont. **52**, 1354–1360 (2005) 337
[12] C. Rossignol: *Etude théorique et numérique d'expériences d'acoustique picoseconde*, Ph.D. thesis, University of Pierre & Marie Curie, Paris (2000) 339, 349
[13] G. Tas, H. J. Maris: Phys. Rev. B **49**, 15046 (1994) 341
[14] O. B. Wright, V. E. Gusev: IEEE Trans. Ultrason., Ferroelec., Freq. Cont. **42**, 331 (1995) 341
[15] N. del Fatti, C. Voisin, D. Christofilos, F. Vallee, C. Flytzanis: J. Phys. Chem. A **104**, 4321 (2000) 341
[16] R. J. Stoner, H. J. Maris, T. R. Anthony, W. F. Banholzer: Phys. Rev. Lett. **68**, 1563 (1992) 343
[17] R. J. Stoner, H. J. Maris: Phys. Rev. B **48**, 16373 (1993) 343
[18] B. Bonello, B. Perrin, C. Rossignol: J. Appl. Phys. **83**, 3081 (1998) 343
[19] A. N. Smith, J. L. Hostetler, P. M. Norris: vol. 57, EUROTHERM Seminar (1998) p. 277 343
[20] C. J. Morath, H. J. Maris, J. J. Cuomo, D. L. Pappas, A. Grill, V. V. Pattel, J. P. Doyle, K. L. Saenger: J. Appl. Phys. **76**, 2636 (1994) 343
[21] W. S. Capinski, H. J. Maris, E. Bauser, I. Silier, M. Asen-Palmer, T. Ruf, M. Cardona, E. Gmelin: Appl. Phys. Lett. **71**, 2109 (1997) 343
[22] T. Baba, N. Taketoshi: *Microscale Heat Transfer*, vol. 57, EUROTHERM Seminar (1998) p. 285 343
[23] B. C. Daly, H. J. Maris, A. V. Nurmiko, M. Kuball, J. Han: J. Appl. Phys. **92**, 3820 (2002) 343
[24] B. C. Daly, H. J. Maris, W. K. Ford, G. A. Antonelli, L. Wong, E. Andideh: J. Appl. Phys. **92**, 6005 (2002) 343
[25] W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, D. S. Katzer: Phys. Rev. B **59**, 8105 (1999) 343, 353
[26] J. E. Rothenberg: Opt. Lett. **13**, 713 (1988) 346
[27] O. B. Wright, K. Kawashima: Phys. Rev. Lett. **69**, 1668 (1992) 346
[28] L. Sarger, P. Segonds, L. Canioni, F. Adamietz, A. Ducasse, C. Duchesne, E. Fargin, R. Olazcuaga, G. L. Flem: J. Opt. Soc. Am. B **11**, 995 (1994) 346
[29] B. Perrin, B. Bonello, J.-C. Jeannet, E. Romatet: Progr. Nat. Sci. **S6**, 444 (1996) 346
[30] B. Perrin, C. Rossignol, B. Bonello, J.-C. Jeannet: Physica B **263–264**, 571 (1999) 346, 349
[31] E. Romatet: *Réalisation et applications d'un dispositif d'acoustique picoseconde*, Ph.D. thesis, University of Pierre & Marie Curie, Paris (1996) 348
[32] R. Abid, F. Miserey: C.R. Acad. Sci. Paris **319**, 631 (1994) 349
[33] G. Tessier, G. Jerosolimski, S. Holé, D. Fournier, C. Filloy: Rev. Sci. Instrum. **74**, 495 (2003) 349
[34] A. Devos, C. Lerouge: Phys. Rev. Lett. **86**, 2669 (2001) 349
[35] N. del Fatti, S. Tzortzakis, C. Voisin, C. Flytzanis, F. Vallee: Physica B **263–264**, 54 (1999) 349
[36] N. del Fatti: *Dynamique electronique femtoseconde dans les systèmes métalliques massifs et confinés*, Ph.D. thesis, Ecole Polytechnique, Palaiseau, France (1999) 349
[37] W. S. Capinski, H. J. Maris: Rev. Sci. Instrum. **67**, 2720 (1996) 349, 353

Index

Index