Encyclopedia of Nanoscience and Nanotechnology

Volume 9 Number 1 2004

view Polymeric Nanoparticles for Drug and Gene Delivery 1
M. N. V. Ravi Kumar; M. Sameti; C. Kneuer; A. Lamprecht; C.-M. Lehr

view Porphyrin-Based Chemical Sensors 21
Roberto Paolesse; Federica Mandoj; Alessia Marini; Corrado Di Natale

view Preparation of Vesicles (Liposomes) 43
Peter Walde

view Protein-Doped Nanoporous Silica Gels 81
Stefano Bettati; Barbara Pioselli; Barbara Campanini; Cristiano Viappiani; Andrea Mozzarelli

view Quantum Dot Atoms, Molecules, and Superlattices 105
Hiroyuki Tamura; Hideaki Takayanagi; Kenji Shiraishi

view Quantum Dot Infrared Photodetector 131
Jamie Phillips; Adrienne Stiff-Roberts; Pallab Bhattacharya

view Quantum Dots: Artificial Atoms and Molecules 143
Philippe Matagne; Jean-Pierre Leburton

view Quantum Well Infrared Detectors 179
W. Lu; Y. Fu

view Quantum Well Infrared Photodetectors: Theoretical Aspects 199
A. F. M. Anwar; Kevin R. Lefebvre

view Raman Scattering in Nanostructures 225
C. E. Bottani; C. Castiglioni; G. Zerbi

view Raman Spectroscopy of Quantum Wires and Quantum Dots 273
V. Wagner; J. Geurts; W. Kiefer

view Raman Spectroscopy in Carbon Nanotubes 307
M. S. Dresselhaus; A. M. Rao; G. Dresselhaus

view Resists for Nanolithography 339
P. Argitis

view Resonant Tunneling Devices 357
A. F. M. Anwar; Mirza M. Jahan

view Room-Temperature Ballistic Nanodevices 371
Aimin M. Song

view Scanning Probe Techniques for Semiconductor Nanostructures 391
Thomas Ihn

view Scanning Tunneling Microscopy of Carbon Nanotubes 415
László P. Biró; Philippe Lambin

view Self-Assembled Monolayers on Semiconductor Surfaces 427
D. Zerulla

view Self-Assembled Nanobiomaterials 459
Steve S. Santoso; Shuguang Zhang

view Self-Assembled Nanofibers 473
Hirotaka Ihara; Makoto Takafuji; Toshihiko Sakurai

view Self-Assembled Nanostructures at Silicon Surfaces 497
D. Y. Petrovykh; F. J. Himpsel

view Self-Assembled Organic/Inorganic Nanocomposites 529
Byron McCaughey; J. Eric Hampsey; Donghai Wang; Yunfeng Lu

view Self-Assembled Porphyrin Arrays 561
Kazuya Ogawa; Yoshiaki Kobuke

view Self-Assembled Porphyrinic Nanoarchitectures 593
Xin Chen; Charles Michael Drain

view Self-Organization of Colloidal Nanoparticles 617
Joydeep Dutta; Heinrich Hofmann

view Self-Organized Nanostructure Formation on Surfaces 641
Andrew T. S. Wee

view Semiconductor Nanodevice Modeling 653
Eric A. B. Cole

view Semiconductor Nanoparticles for Photocatalysis 669
W. Li; S. Ismat Shah

view Semiconductor Nanotransistors 697
Y. Fu

view Semiconductor Quantum Dots 719
Lucjan Jacak; Arkadiusz Wójs; Paweł Machnikowski

view III/V Semiconductor Quantum Dots 735
M. Guzzi; S. Sanguinetti; M. Gurioli

view Semiconductor Quantum Wires 763
Xing-Quan Liu; Xue-Lun Wang; C. Jagadish; M. Ogura

view SiGe/Si Heterostructures 775
C. W. Liu; L. J. Chen

view Silicon Nanocrystals in SiO_2 Thin Layers 793
A. G. Nassiopoulou

view Silicon Nanowires 815
Klaus Sattler

view Silicon Quantum Dots 837
Anri Nakajima

view Silicon Surface Nanooxidation 859
D. Stiévenard; B. Legrand

view Single-Electron Devices 865
Konstantin Likharev

view Single-Electron Dynamics 885
Toshimasa Fujisawa

view Single-Electron Transistors 903
Jia Grace Lu

view Single Wall Carbon Nanotubes 923
Ákos Kukovecz; Zoltán Kónya; Imre Kiricsi

view Sliding, Rotation, and Rolling of Nanoparticles 947
K. Miura

Copyright © 2004 American Scientific Publishers

Polymeric Nanoparticles for Drug and Gene Delivery

M. N. V. Ravi Kumar, M. Sameti, C. Kneuer, A. Lamprecht, C.-M. Lehr

Saarland University, Saarbarucken, Germany

CONTENTS

1. Introduction
2. Features of Polymeric Particles
3. Preparation and Characterization of Nanoparticles
4. Recent Developments in Nanoparticle Technology
5. Nanoparticles for Some Specific Applications
Glossary
References

1. INTRODUCTION

There is a plethora of reasons why the simple capsule containing pure active drug is not an acceptable delivery system for a drug at any stage of its life (Fig. 1). Conventional formulations are usually thought of as oral (tablets and capsules), topical (ointments, creams, etc.), and injections. It is often desirable for a drug to be delivered at a steady, uniform rate; zero order is ideal. This implies that the concentration of the drug in blood or tissue is independent of the concentration of the drug at the site of absorption. None of the conventional formulations meet these demands. Controlled drug delivery occurs when a polymer, whether natural or synthetic, is judiciously combined with a drug or other active agent in such a way that the active agent is released from the material in a predesigned manner. The release of the active agent may be constant over a long period, it might be cyclic over a long period, or it may also be triggered by the environment or other external events. Over the past few decades, there has been considerable interest in developing biodegradable nanoparticles as potential candidates for controlled drug delivery. Various polymers [1] have been used in drug delivery research as they are expected to be capable of delivering drugs to target site and thus increase the therapeutic benefit, while minimizing side effects [2].

Nanoparticles were first developed by Speiser and co-workers [3] around 1970 and are defined as solid colloidal particles, less then 1 μm in size, that consist of macromolecular compounds. Since then, a considerable amount of work on nanoparticles is being carried out around the world in the field of drug/gene delivery (Fig. 2). They were initially devised as carriers for vaccines and anticancer drugs [4]. The use of nanoparticles for ophthalmic and oral delivery was also investigated [5]. Drugs or other biologically active molecules are dissolved, entrapped, or encapsulated in the nanoparticles or are chemically attached to the polymers or adsorbed to their surface. The selection of the appropriate method for preparing drug-loaded nanoparticles depends on the physicochemical properties of the polymer and the drug. On the other hand, the procedure and the formulation conditions will determine the inner structure of these polymeric colloidal systems. Two types of systems with different inner structures are possible: (i) a matrix-type system composed of an entanglement of oligomer or polymer units, defined here as a nanoparticle or nanosphere; (ii) a reservoir-type system, consisting of an oily core surrounded by a polymer wall, defined here as a nanocapsule. Various colloidal nanoparticulate systems in use for drug/gene delivery are as shown in Figure 3. The term nanoparticle shall be used to refer to both systems including nanoparticles as well as nanocapsules. In this chapter we discuss preparation techniques, characterization, and some reported nanoparticulate delivery systems and their applications.

2. FEATURES OF POLYMERIC PARTICLES

2.1. Small Size and Volume

It is practically impossible to fabricate a polymeric particle less than 5 nm in size, considering the fact that polymer molecules usually have a molecular weight higher than 10,000 Da. Particles composed of 1000 units of polymers with molecular weight 10,000 Da would be 30 nm in size.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (1–19)

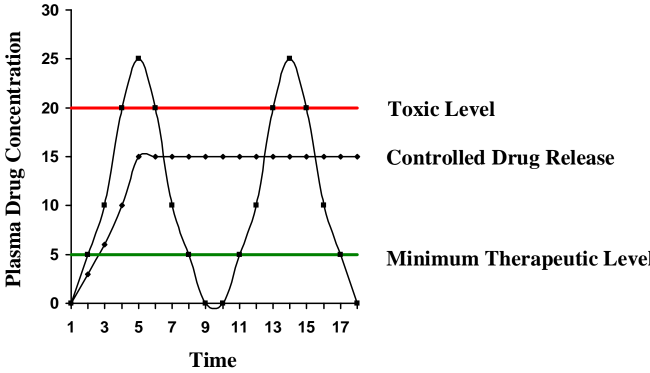


Figure 1. Comparison of typical pharmacokinetic profiles seen for conventional vs controlled release formulations. (*Source:* Science Finder, reports included until July 2002)

Their small volume enables the whole body of particles to respond quickly to stimuli. The relaxation time of volume change of a gel was indicated to be proportional to the square of its radius [6]. Therefore, fine particles can be a microreactor with a high reaction rate.

2.2. Large Specific Surface Area

The large surface area of the particles would be available for sites of absorption and desorption, chemical reactions, light scattering, etc. The total surface area of 1 g of particles having a diameter of 0.1 μm would be about 60 m^2. The total surface area is inversely proportional to the diameter.

2.3. High Diffusibility and Mobility

Polymer colloids have a low viscosity and high fluidity compared to solutions containing the same amount of solid. The viscosity of polymer colloids is a universal function of apparent volume fraction of the particles. The apparent volume fraction of particles can be changed by environmental conditions such as pH and temperature for some polymer colloids [7]. In dispersion the particles can move macroscopically through the medium by gravity, electric field, etc. and microscopically via Brownian motion. These movements keep a fresh interface between the particles and the medium. Particles having soft layers on the surface allow water to penetrate the layer and meet less resistance when they move through water. Such particles occasionally exhibit a fairly high electrophoretic mobility even if they have little charge [8]. This explains the extraordinarily high electrophoretic mobility of cells and other biocolloids.

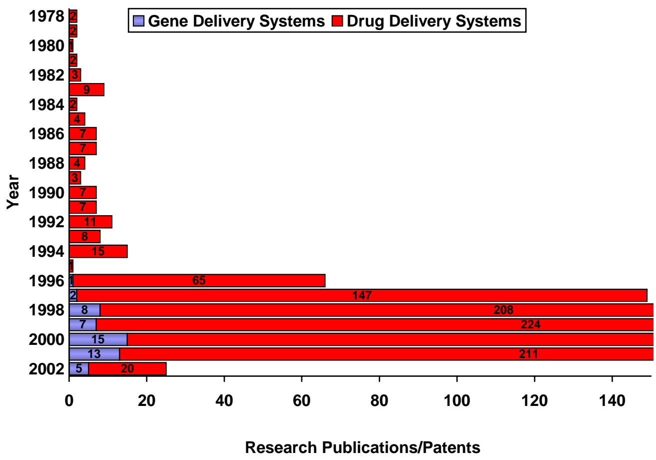


Figure 2. Updates on nanoparticle research (*Source:* Science Finder).

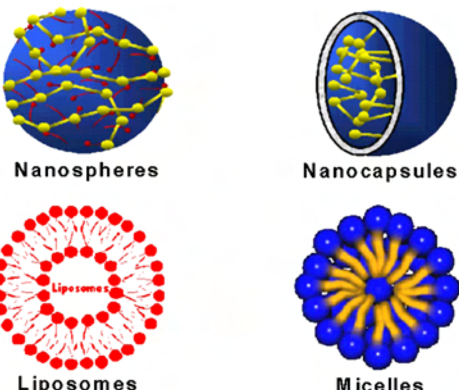


Figure 3. Various colloidal particles in use for numerous pharmaceutical applications.

2.4. Stable Dispersions

The potential energy and stability of the polymer particles in the dispersion are decided collectively by the three factors, viz., electrostatic repulsive forces, van der Waals attractive forces, and steric repulsive forces among the particles. If the particle has a potential energy larger than 15 kT, it has a high stability ratio W and is stable enough to be stored for a relatively long period, say, more than one month [9]. The stability ratio $[W]$ is expressed as $W = K_q/K_s$, where K_q and K_s are the rate constants for rapid and slow flocculation, respectively.

The surface potential energy is affected significantly by environmental factors such as ionic strength and pH. The minimum concentration of a salt to flocculate the particles is termed flocculation concentration [10]. Steric repulsive interactions are a crucial force for microspheres covered with polymer layers protruding in solution. The steric stabilization effect consists of both an enthalpy and an entropy effect. The overlap of the surface layers of two vicinal particles results in a shift from the equilibrium state of the layer (i.e., enthalpy gain) and a decrease in the conformational freedom of solvated polymer chains (i.e., entropy loss) [11]. If the polymer layers on the particles are sensitive to temperature, the stability of the dispersion due to the steric stabilization effect is influenced by temperature. The critical temperature for flocculation, in such systems, is called the critical flocculation temperature. Dissolved, but nanoadsorbable, polymer molecules cause the coagulation of particles. This phenomenon is referred to as the depression effect [12, 13]. This effect may work in many practical uses of polymer colloids, although it has not been considered extensively. The high stabilities of polymer colloids may bring about some difficulties in the separation of particles from the dispersion medium. Centrifugation sometimes fails

for dispersions of very fine particles because the sedimentation rate is proportional to the square of the diameter. Coagulation with salt is one of the most conventional methods to separate particles from water. Magnetic or electric fields are sometimes applied for collecting particles from dispersion.

2.5. Uniformity

Techniques for the preparation of monodisperse polymer colloids received greater attention during 1980s [14, 15]. The use of monodisperse particles makes it possible to give sharp, reliable, and reproducible results for their respective applications. Emphasis should be on their uniformity and size; the chemistry and morphology of the particles also deserve attention. Technologies to satisfy these conditions have been developed with an understanding of the principles for particle nucleation and growth in particle-forming polymerizations.

2.6. Variety

Polymer particles can be prepared by physical methods, such as emulsification, coacervation, and spray-drying, and by chemical methods like heterogeneous polymerization. These preparative methods give a variety of particles in terms of size, surface chemistry, composition, surface texture, and morphology.

3. PREPARATION AND CHARACTERIZATION OF NANOPARTICLES

3.1. Cross-Linking of Amphiphilic Macromolecules

Nanoparticles have been prepared from polysaccharides, proteins, and amphiphilic macromolecules by inducing their aggregation followed by stabilization by either heat denaturation or chemical cross-linking. It can be done by a water-in-oil emulsion system or in aqueous environment. The cross-linking technique was first used by Kramer et al. in 1974 [16]. In this technique, an aqueous solution of protein was emulsified in oil using a high-speed homogenizer/sonicator and the water-in-oil emulsion was then poured in a hot oil having a temperature greater than 100 °C and held for a specific time (to denature the protein), thereby leading to the formation of submicroscopic particles. These particles were finally washed with organic solvents and subsequently collected by centrifugation. The crucial factors in nanoparticle production were emulsification energy and stabilization temperature; however, the limitation of high stabilization temperature was overcome by adding a chemical cross-linking agent (e.g., glutaraldehyde) to the system. To achieve the variable size of the nanosphere, several formulations were adopted and optimized. Protein and polysaccharide nanoparticles can be obtained by a phase separation process in aqueous medium. This can be induced by desolvation of the macromolecule, by change in pH or temperature, or by adding counterions in acid medium [17].

3.2. Polymerization of Acrylic Monomers

Couvreur et al. [18, 19] carried out extensive investigations on nanoparticles and nanospheres composed of polyalkylcyanoacrylate (PACA) polymers, which are bioresorbable and were used as surgical glues for several years. PACA nanoparticles have been prepared by an emulsion polymerization method in which droplets of water-insoluble monomers are emulsified in aqueous/acidic phase containing a stabilizer. The monomers polymerize relatively faster by an anionic mechanism, the rate of polymerization being pH dependent. The system is maintained under magnetic agitation while the polymerization reaction takes place. The duration of polymerization reaction is determined by the length of the alkyl chain varying from 2 to 12 h for ethyl and hexylcyanoacrylate, respectively. Finally, the colloidal suspension is neutralized and lyophilized following incorporation of glucose as a cryoprotectant. Water-soluble drug may be associated with PACA nanosphere either by dissolving the drug in the aqueous polymerization medium or by incubating blank nanospheres in an aqueous solution of the drug. The drug loading efficiency is dependent on various factors, including the pK_a and polarity of the drug, size and surface charge of the nanospheres, and the drug concentration in aqueous medium [18].

In another method of encapsulation of lipophilic drugs into PACA polymers, the monomers and the drug have been dissolved in a mixture of a polar solvent (acetone or methanol), an oil (coconut oil or benzyl benzoate), and a lipophilic surfactant, such as lecithin. The organic phase is added into aqueous phase having a hydrophobic surfactant (e.g., Poloxamer 188) under magnetic agitation. Thus, diffusion of the polar solvent into aqueous phase and the polymerization of the monomer at the oil–water interface take place simultaneously. Polymerization is initiated by the hydroxyl anions and leads to the formation of nanocapsules having an oily core surrounding by a polymer coat. The organic solvent is eliminated completely from the colloidal suspension. The selection of the oil has a great role to play, as it influences the size of the nanocapsules, the molecular weight of the polymer coat, and the stability of the suspension after storage [19].

3.3. Polymer Precipitation

Solvent precipitation techniques have been generally applied for hydrophobic polymers, except for dextran nanospheres. Several techniques described in the literature are based on the mechanism of polymer precipitation.

3.3.1. Solvent Extraction–Evaporation

In this technique, hydrophobic polymer is dissolved in an organic solvent, such as chloroform, ethyl acetate, or methylene chloride, and is emulsified in an aqueous phase having a stabilizer [e.g., polyvinyl alcohol (PVA)]. Just after the formation of nanoemulsion, solvent diffuses to the external phase until saturation. The solvent molecules that reach on the water–air interphase evaporate, which leads to continuous diffusion of the solvent molecules from the inner

droplets of the emulsion to the external phase. Simultaneously, the precipitation of the polymer leads to the formation of nanospheres. The extraction of solvent from the nanodroplets to the external aqueous medium can be induced by adding an alcohol (e.g., isopropanol), thereby increasing the solubility of the organic solvent in the external phase. A purification step is required to assure the elimination of the surfactant in the preparation. This technique is most suitable for the encapsulation of lipophilic drugs, which can be dissolved in the polymer solution.

3.3.2. Solvent Displacement or Nanoprecipitation

In this method, the organic solvent selected is completely dissolved in the external aqueous phase; thus there is no need for evaporation or extraction for polymer precipitation. Polymer and drug are dissolved in acetone, ethanol, or methanol and incorporated under magnetic stirring into an aqueous solution of surfactant. The organic solvent diffuses instantaneously to the external aqueous phase followed by precipitation of the polymer and drug. After the formation of nanoparticles, solvent is eliminated and the suspension concentrated under reduced pressure. This method is surfactant-free. However, this method is limited only to the drugs that are highly soluble in a polar solvent.

3.3.3. Salting Out

A technique based on the precipitation of a hydrophobic polymer is useful for the encapsulation of either hydrophilic or hydrophobic drugs because of a variety of solvents, including polar (e.g., acetone or methanol) and nonpolar (methylene chloride or chloroform) solvents, can be chosen for dissolving the drug. This procedure is just like nanoprecipitation; however, the miscibility of both phases is prevented by the saturation of the external aqueous phase with electrolytes. Precipitation occurs when a sufficient amount of water is added to allow complete diffusion of the acetone in the aqueous phase.

3.4. Nanocapsule Preparation

We considered PACA while discussing the preparation and characterization techniques of nanoparticles for the sake of convenience. Nanocapsules of PACA are obtained via an interfacial polymerization process in emulsion. Nanocapsules are formed by mixing an organic phase with an aqueous phase. The organic phase is generally an ethanolic solution of the monomer mixed together with the oily core material and the lipophilic drug to be encapsulated and occasionally, soya bean lecithin is added as an additional surfactant. Oils, viz., Miglyol, benzylbenzoate, ethyl oleate and lipiodoal, have been frequently used in preparation of nanocapsules. The encapsulation efficiency of a lipophilic drug is dependent on its partition coefficient between the oil and the aqueous phase, so the oil must be chosen accordingly. The aqueous phase is a solution of a nonionic surfactant, usually Synperonic PE-F68 at 0.5% at a pH between 4 and 10. Nanocapsules are formed by adding the organic phase dropwise into the aqueous phase under stirring, through a wide-bore syringe needle or a micropipette tip. The mixture immediately becomes opalescent. After stirring for 15–30 minutes the ethanol is removed by evaporation under reduced pressure. If required, the nanocapsules can be further concentrated by evaporation; this allows them to be rediluted in a physiological buffer for injection. Nanocapsules formed by this way have a mean diameter between 200 and 300 nm with a narrow polydispersity [20]. The speed of the magnetic stirring has no influence on the particle size, which depends solely on the nature and the volume of the oil and on the volume of the diffusing organic phase. The proportion of the monomer must be correctly chosen to avoid simultaneous formation of either flakes of polymer or of a single oily emulsion. The presence of surfactant in the aqueous phase is, in fact, not necessary for the successful formation of nanocapsules but does prevent them from aggregating on storage to a cake, which is difficult to disperse. Nanocapsules formed in this way are physically stable for several years at ambient temperature and may be sterilized by autoclaving at 120 °C for 20 min. However, as a result of their vesicular character, nanocapsules are not easily lyophilized, since they tend to collapse releasing the oily core [20].

The colloids formed by the interfacial polymerization of PACA could be influenced by many factors, viz. the nature of the aqueous phase, the pH, the composition of the organic phase (monomer, oil, ethanol), the ratio of monomer to the aqueous phase, and the emulsification conditions. The degree of polymerization, and therefore the molecular weight, depends on a balance between initiation, propagation, and termination [21]. The number of growing chains depends on the concentration of initiators. For the same quantity of monomer, when the number of live chains is high, the degree of polymerization (DP) is low. However, DP is reduced when the concentration of the terminating agents is high [22, 23]. The concentration of the initiating and the terminating agents available for the monomer depends on the physicochemical nature of the system in which the components are dispersed at the moment of the formation of the colloid.

In the case of a system composed of two immiscible phases, such as an emulsion, the presence and the concentration of the solutes in the different phases is a function of the polarity of the solute molecules and of the dielectric constant of the medium. In contrast, at an interface between an aqueous and an organic medium, there can be large local variations in properties, which can themselves cause changes in the interfacial properties in a dynamic system. Nevertheless, the degree of polymerization depends on the propagation reaction; other characteristics of the colloids, viz. particle size and morphology, depend on interfacial phenomena inducted by the dynamic mixing of an organic phase with an aqueous phase.

Since the preparation of nanocapsules of PACA is rather similar to the preparation of PACA nanospheres, it is necessary to verify that the colloidal suspension does not consist of a simple mixture of an oil-in-water emulsion containing polymeric nanospheres of similar size. Rollot et al. by their various experiments confirmed that the preparation indeed was composed of oil-filled nanocapsules [24].

3.5. Nanosuspensions

The nanosuspension technique is an alternative and promising approach for the production of drug nanoparticles, where the drugs are poorly soluble in both aqueous as well organic medium. The major advantages of this technology are its general applicability to most of the drugs and its easy fabrication/handling. The disintegration principle for obtaining nanosuspensions is the cavitation forces in high pressure homogenizers (e.g. piston-gap homogenizers like APV gaulin types). The preparation method involves the dispersion of the drug powder in an aqueous surfactant solution by high speed stirring. The obtained macrosuspension is then passed through a high pressure homogenizer applying typically 1500 bar and 3 to 10 up to a maximum of 20 passes (=homogenization cycles). The suspensions pass a very small homogenization gap in the homogenizer, typically having a width of 25 μm at 1500 bar. Due to the narrowness of the gap the streaming velocity of the suspension increases tremendously with an increase in fluid pressure. Simultaneously, the static pressure on the fluid decreases below the boiling point of water at room temperature [25, 26]. As a consequence, water starts boiling at room temperature due to the high pressure; gas bubbles are formed which implode cavitation when the fluid leaves the homogenization gap. These cavitation forces are strong enough to break the drug microparticles to drug nanoparticles [25]. The mean particle size in the nanometer range obtained by this procedure depends on the pressure and number of cycles applied; in addition it is affected by the hardness of the drug itself. Mean diameters of 330 and 600 nm are reported for paclitaxel nanosuspensions [26] and clofazemine [27] respectively. Müller and co-workers pioneered the field of nanosuspensions [28–32]. Müller [33] et al. in their recent article reviewed the preparation of nanosuspensions on a laboratory scale, and large scale, physical and chemical properties of nanosuspensions, surface modification of nanosuspensions, biological properties, and perspectives were also discussed.

3.6. Characterization

Nanoparticles as a colloidal carriers mainly depend on the particle size distribution, surface charge, and hydrophilicity. These physicochemical properties affect not only drug loading and release, but also the interaction of these particulate carriers with biological membranes.

3.6.1. Particle Size Analysis

Two main techniques have been used to determine the particle size distribution of colloidal systems: photon correlation spectroscopy (PCS) and electron microscopy including both scanning electron microscopy (SEM) and transmission electron microscopy (TEM). The quasi electron light scattering technique for Brownian moment measurement offers an accurate procedure for measuring the size distribution of nanoparticles. The PCS technique does not require any particular preparation for analysis and is excellent due to its efficiency and accuracy. However, its dependency on Brownian movement of particles in suspended medium may affect the particle size determination.

Electron microscopy provides an image of the particles to be measured. In particular, SEM is used for vacuum dried nanoparticles that are coated with a conductive carbon–gold layer for analysis and TEM is used to determine the size, shape, and inner core structure of the particles. TEM in combination with freeze–fracture procedures differentiates among nanocapsules, nanospheres, and emulsion droplets. Atomic force microscopy (AFM) is an advanced microscopic technique and its images can be obtained in aqueous medium. AFM images nowadays are powerful support for the investigation of nanoparticles in biological media.

3.6.2. Surface Charge and Hydrophobicity

The interaction of nanosphere with biological environment and electrostatic interaction with biological compounds occur due to the charge on the surface (e.g., negative charge promotes the adsorption of positively charged drug molecules such as aminoglucosides as well as enzymes and proteins). The surface charge of colloidal particles can be determined by measuring the particle velocity in an electric field. Nowadays laser light scattering techniques, in particular laser Doppler anemometry, are fast enough to measure the surface charge with high resolution. Hydrophobicity of the nanoparticles can be determined by the methods including adsorption of hydrophobic fluorescent or radiolabeled probes, two phase partitions, hydrophobic interaction chromatography, and contact angle measurements. Recently, X-ray photoelectron spectroscopy has been developed which offers the identification of chemical groups in the 5-Å-thick coat on the external surface of nanospheres. Gref et al. [34] have characterized the poly(ethylene glycol) (PEG)-coated poly(lactide-*co*-glycolide) (PLGA) nanosphere and identified the PEG chemical elements that were concentrated on the nanosphere's surface.

3.6.3. Methods of Changing Particle Size and Surface Characteristics

The fate of colloidal particles inside the body depends on three factors: particle size, particle charge, and surface hydrophobicity. Particles with a very small size (less than 100 nm), low charge, and a hydrophilic surface are not recognized by the mononuclear phagocytic system and, therefore, have a long half-life in the blood circulation. In general, nature and concentration of the surfactant play an important role in determining the particle size, as well as the surface charge (e.g., nanospheres with mean size of less than 50 nm were prepared by increasing concentration of Poloxamer 188 [35]). The approaches for modifying surface charge and hydrophilicity were initially based on the adsorption of hydrophilic surfactants, such as block copolymers of the poloxamer and poloxamine series. The *in vivo* studies of hydrophilic nanospheres limit their usefulness due to their toxicity in intravenous injection. Lately, the idea of using diblock copolymers made of poly(lactic acid) (PLA) and poly(ethylene oxide) (PEO) is widely accepted due to the safety and stability of the hydrophilic coat. For this purpose, the copolymer is dissolved in an organic solvent and then emulsified in an external aqueous phase, thereby orienting the PEO toward the aquous surrounding medium, while in

another method PLA–PEO copolymer is adsorbed onto preformed PLGA nanoparticles. It is found to be efficient in prolonging the nanosphere circulation time following intravenous administration.

3.7. Drug Incorporation and Adsorption

For a nanoparticulate system to be highly successful it should have a high drug loading capacity, which in turn reduces the quantity of carrier required for administration. Two theoretical curves can be proposed to describe the adsorption of drugs onto nanoparticles: Langmuirian-type and constant partitioning-type isotherms. In fact, it was found that nanoparticles can entrap a drug according to a Langmurian adsorption mechanism, because of their large specific area [36]. The drug can either be incorporated into nanospheres during the polymerization process or adsorbed onto the surface of preformed particles. There are reports on vidarabine, an antiviral agent, whose nucleophile N in positions 3 and 7 may play a role of initiator for the anionic polymeric mechanism of the cyanoacrylic monomer [37], which suggests drug–polymer interaction as a covalent linkage. A similar study in this direction was described with peptide compounds such as growth hormone-releasing factor (GRF) [38]. In these studies it was demonstrated that if the drug was added within 5 min of the polymerization process, 50% of the peptide was found to be covalently linked to the polymer. In contrast to these findings, the drug loading capacity was found to be very low when the peptide was added after 60 min during polymerization, but no chemical modification was observed. These findings suggest that there is a narrow window of time for the addition of GRF to the polymerization medium resulting satisfactory drug-loading capacity as well as preservation of chemical structure of the peptides [38]. Generally, the longer the alkyl chain length, the higher the affinity of the drug. Moreover, the Langmurian isotherm states that the percentage of drug adsorption generally decreases with the quantity of drug dissolved in the polymerization medium [39].

Reports were also available on the attempts made for the association of synthetic fragments of deoxy ribonucleic acid (DNA) to PACA nanospheres [39]. The association of antisense oligonucleotides with nanoparticles was achieved only in the presence of a hydrophobic cation, such as triphenylphosphonium or quaternary ammonium salts [40]. The poor yield of oligonucleotide association without cations could be explained by the hydrophilic character of nucleic acid chains that are known to be soluble in water. The adsorption efficiency of oligonucleotide–cation complexes on nanospheres was found to be highly dependent on several parameters: oligonucleotide chain length, nature of the cyanoacrylic polymer, hydrophobicity of the cations used as ion-pairing agents, and ionic concentration of the medium [39].

3.7.1. Biodegradation and Drug Release

The rate of cyano copolymer degradation is dependent on their alkyl chain length. The dominating mechanism of particle degradation was found to be surface erosion [40], resulting via the hydrolysis by the enzymes of the ester side chain of the polymer [41]. This process keeps the polymer chain intact, but it becomes more and more hydrophilic until it is water soluble. The degradation pathway is consistent with the production of alcohol during the bioerosion of PACA nanospheres *in vitro* in the presence of esterases [41]. Indeed, the action of rat liver microsomes and tritosomes on the ester hydrolysis of polyisobutylcyanoacrylate nanospheres was clearly demonstrated [41]. Since the biodegradability of polyalkycyanoacrylate depends on the nature of the alkyl chain, it is possible to choose a monomer whose polymerized form has a biodegradability corresponding to the established program for the drug release [41]. Lenaerts et al. [42] demonstrated that the drug release from nanospheres was a direct consequence of the polymers bioerosion, by using a double radiolabeled technique. Later this was confirmed using GRF, another model drug [42]. No drug release was reported in the absence of esterase with this peptide compound, whereas in the nanosphere suspension turbidity remained unchanged, indicating no polymer bioerosion. Drug release appeared to be esterase dependent. In other studies, dextran sulfate was employed as a stabilizer in the preparation of PACA nanospheres, which slowed down the release rates of drug rose Bengal [42].

4. RECENT DEVELOPMENTS IN NANOPARTICLE TECHNOLOGY

Other than the commonly used synthetic hydrophobic polymers, various other polymers like chitosan, sodium alginate, gelatin, etc. are also being explored in drug delivery [43]. A drug delivery system is most often associated with fine particulate carriers, such as emulsion, liposomes, and nanoparticles, which are designed to localize drugs in the target site. From the clinical point of view, they might have to be biodegradable and/or highly biocompatible. In addition, high drug content is desirable, because in many cases actual drug loading efficiency is often too low to secure an effective dose at the target site. Biodegradable nanoparticles have received considerable attention as potent vehicles for targeting a site and controlled release of drugs/bioactive components [43–47]. Various nanoparticlulate systems composed of different materials are constantly being explored in the areas of drug and gene delivery as shown in Table 1. We hereby discuss some nanoparticulate systems developed by various researchers.

4.1. Doxorubicin Nanoparticle Conjugates

Natural and synthetic polymers have widely been used for conjugating the drugs for improved circulating times in the bloodstream, targeting a specific organ, and sustained release at the injection site [48–50]. The polymer conjugation approach, along with formulations with polymeric micelles and liposomes, is an attractive means to selectively suppress tumor growth [51, 52]. Some of the earlier studies suggested the use of high molecular weight polymers would eventually cause problems [53, 54]. Yoo et al. [55] reported doxorubicin-poly(D, L-lactic-*co*-glycolic acid) (PLGA) conjugates (Fig. 4). Doxorubicin was chemically conjugated to a terminal end group of PLGA by an ester linkage and the doxorubicin–PLGA conjugate was later formulated into nanoparticles. They have observed a 1-month release of

Table 1. Nanoparticle types and their applications.

Nanoparticles	Application
Poly(lactide-*co*-glycolide)	drug/gene delivery
Chitosan	drug/gene delivery
Solid lipids	drug/gene delivery
Liposomes	drug/gene delivery
Block copolymers	drug/gene delivery
Poly(ethylene glycol)	drug/gene delivery
Polycaprolactone	drug delivery
Polycyanoacrylate	drug/gene delivery
Dextran	drug/gene delivery
Poly-L-lysine	drug/gene delivery
Silica	drug/gene delivery
Gelatin	drug/gene delivery
Poly(aspartic anhydride-*co*-ethylene glycol)	gene delivery
Atelocollagen	drug/gene delivery
Alginate	drug/gene delivery

conjugated doxorubicin from the nanoparticles vs a 5 day release of unconjugated doxorubicin. *In vivo* antitumor activity assay also showed that a single injection of the nanoparticles had comparable activity to that of free doxorubicin administered by daily injection [55]. Recently, Mitra et al. [56] and Janes et al. [57] reported dextran–doxorubicin conjugates using chitosan as the nanoparticle carrier.

4.2. Lipid Coated Cisplatin Nanocapsules

Cisplatin is one of the most widely used agents in the treatment of solid tumors, but its clinical utility is limited by toxicity. The poor solubility of the cisplatin results in low

Figure 4. Synthetic route of DOX–PLGA conjugate. Reprinted with permission from [55], H.-S. Yoo et al., *J. Control. Release* 68, 419 (2000), © 2000, Elsevier Science.

encapsulation efficiency [58, 59]. Researchers have tried to use lipophilic derivatives of cisplatin to improve the encapsulation efficiency [60]. Recently, Burger et al. [61] reported a novel procedure to efficiently encapsulate native, nonderivatized cisplatin in lipid formulations. The methodology is based on simple freeze thawing of concentrated solution of cisplatin in the presence of negatively charged phospholipids [61]. The authors claim this procedure as novel with maximum encapsulation efficiency with cisplatin aggregates being covered with a single lipid bilayer. It was found that the *in vitro* cytotoxicity was raised 1000-fold with this new technology [61].

4.3. Poly(DL-lactide-*co*-glycolide) Nanoparticles

The most widely used emulsion solvent evaporation method for preparation of nanoparticles using PLGA requires surfactants to stabilize the dispersed particle [62]. This method often has a problem in that the surfactant remains at the surface of the particles and hence is difficult to remove especially where PVA is used as surfactant. Other surfactants such as span series or tween series, PEO, etc. are also used to stabilize with some disadvantages like removal of solvents, toxicity, low particle yield, consumption of more surfactant, and multisteps.

The most important factor that needs to be considered while using surfactants, is that they are nonbiodegradable and nondigestible and tend to affect humans with allergic reactions. Recently, Jeon et al. [63] proposed a surfactant-free method for preparation of PLGA nanoparticles as an alternative.

The surfactant-free PLGA nanoparticles were prepared by dialysis method using various solvents and their physiochemical properties were investigated against used solvent. Release kinetics of norfloxacin showed that a higher drug content leads to larger particle size and slow release [63].

Kwon et al., [64] reported estrogen loaded PLGA nanoparticles employing emulsification-diffusion method using PVA or didodecyl dimethylammonium bromide (DMAB) as stabilizers. They have studied the influence of process variables on the mean particle size of the nanoparticles. The particle size was less than 100 nm when DMAB was used a stabilizer [64]. Stabilizers like D-α-tocopheryl polyethylene glycol 1000 succinate (vitamin E-TPGS) were also tried by other groups [65].

Santos-Magalhaes et al. [66] reported PLGA nanocapsules/nanoemulsions for benzathine pencillin G. Nanoemulsions were produced by spontaneous emulsification and nanocapsules by interfacial deposition of preformed polymer. They have observed similar release kinetics from both formulations [66].

4.4. Poly(ethylene oxide)-poly(L-lactic acid)/poly(β-benzyl-L-aspartate)

Polymeric micelles are expected to self-assemble, when block copolymers are used for their preparation [67]. Micelles of biocompatible copolymer, viz., PEO with PLA or with poly(β-benzyl-L-aspartate) (PBLA), have been reported in the literature [68, 69]. The synthetic process of

Figure 5. (a) Poly(ethylene oxide)-*co*-b-benzyl-L-aspartate and (b) poly(ethylene oxide)-*co*-L-lactide micelles with aldehyde groups on their surface. Reprinted with permission from [67], J. Jagur-Grodzinski, *React. Funct. Polym.* 39, 99 (1999). © 1999, Elsevier Science.

such nanospheres with functional groups on their surface is shown in Figure 5.

Aldehyde groups on the surface of the PEO–PLA micelles might react with the lysine residues of a cell's proteins and may facilitate attachment of the amino-containing ligands. These hydroxyl groups on the surface of the PEO–PBLA micelles can be further derivatized and conjugated with molecules capable of targeting the modified micelles to specific sites of a living organism. Such nanospheres have been tested as vehicles for delivery of anti-inflammatory and antitumor drugs [70, 71].

4.5. Poly(lactide-*co*-glycolide)–[(propylene oxide)-poly(ethylene oxide)]

Biocompatible and biodegradable poly(lactide-*co*-glycolide) (PLG) nanoparticles (80–150 nm) have been prepared by the following nanoprecipitation technique [72]. The nanoparticles were coated with a 5–10 nm thick layer of polypropylene (PPO)-PEO block copolymer or with tetrafunctional $(PEO\text{-}PPO)_2$–*N*-CH_2-CH_2-*N*–$(PPO\text{-}PEO)_2$ [72]. Such coats are bound to the core of the nanosphere by the hydrophobic interactions of the PPO chains, while PEO chains protrude into the surrounding medium and form a steric barrier, which hinders the adsorption of certain plasma proteins onto the surface of such particles. On the other hand, the PEO coat enhances adsorption of certain other plasma compounds. As a consequence, the PEO-coated nanospheres are not recognized by macrophages as foreign bodies and are not attacked by them [73].

4.6. Polyphosphazene Derivatives

Allock and co-workers developed derivatives of the phosphazene polymers suitable for biomedical applications [74, 75]. Long-circulating in the blood, 100–120 nm in diameter, PEO-coated nanoparticles of the poly(organophospazenes) containing amino acid have been prepared. PEO–polyphosphazene copolymer, or poloxamine 908 (a tetrafunctional PEO copolymer), has been deposited on their surface [76]. Chemical formulae of such polyphosphazene derivatives are shown in Figure 6.

4.7. Poly(ethylene glycol) Coated Nanospheres

Nanospheres of PLA, PLG, or poly(ε-caprolactone) coated with PEG may be used for intravenous drug delivery. PEG and PEO denote essentially identical polymers. The only difference between the respective notations is that methoxy groups in PEO may replace the terminal hydroxyls of PEG. PEG coating of nanospheres provides protection against interaction with the blood components, which induce removal of the foreign particles from the blood. PEG

Figure 6. Structures of polyphosphazenes for medical applications. (a) Poly([(phenylethylalanine ethylester)40%(glycine ethyl ester)60%] phosphazene) PF(GL-PhAL). Reprinted with permission from [75], H. R. Allock et al., *Macromolecules* 10, 824 (1997). © 1997, American Chemical Society.

coated nanospheres may function as circulation depots of the administered drugs [34, 77]. Slow release of the drugs into plasma alters the concentration profiles leading to therapeutical benefits. PEG-coated nanospheres (200 nm), in which PEG is chemically bound to the core, have been prepared, in the presence of monomethoxy PEG, by ring opening polymerization (with stannous octoate as a catalyst) of such monomers as ε-caprolactone, lactide, and/or glycolide [77]. Ring opening polymerization of these monomers in the presence of such multifunctional hydroxy acids as citric or tartaric acid, to which several molecules of the monomethoxy monoamine of PEG (MPEG-NH_2) have been attached, yields multiblock $(PEG)_n$–$(X)_m$ copolymers. PEG–PLA copolymer in which NH_2 terminated methoxy PEG molecules have been attached to tartaric acid is shown in Figure 7. It has been demonstrated that morphology, degradation, and drug encapsulation behavior of copolymers containing PEG blocks strongly depends on their chemical composition and structure. Studies of nanoparticles composed of the diblocks of the PLG with the methoxy terminated PEG (PLG–PEG) or of the branched multiblocks PLA–$(PEG)_3$, in which three methoxy terminated PEG chains are attached through a citric acid residue, suggested that they have a corecorona structure in an aqueous medium. The polyester blocks form the solid inner core. The nanoparticles, prepared using equimolar amounts of the PLLA–PEG and PDLA–PEG stereoisomers, are shaped as discs with PEG chains sticking out from their surface.

Their hydrophobic/hydrophilic content seems to be just right for applications in cancer and gene therapies. Such nanospheres are prepared by dispersing the methylene chloride solution of the copolymer in water and allowing the solvent to evaporate [77]. By attaching biotin to its free hydroxyl groups and complexing it with avidin, cell specific delivery may be attained. Nuclear magnetic resonance studies of such systems [78] revealed that the flexibility and mobility of the thus attached PEG chains are similar to those of the unattached PEG molecules dissolved in water. Recently, PLG microspheres, with the PEG–dextran conjugates attached to their surface, have been investigated as another variant of the above-described approach. Microspheres with a diameter of 400–600 nm have been prepared [79].

Figure 7. Multiblock $(PEG)_n$–$(X)_m$ copolymers. Amino terminated methoxy poly(ethylene glycol) molecules attached to tartaric acid with PLA side chains. Reprinted with permission from [77], R. Gref and A. J. Domb, *Adv. Drug Del. Rev.* 16, 215 (1995). © 1995, Elsevier Science.

4.8. Poly(isobutylcynoacrylate) Nanocapsules

Intragastric administration of insulin-loaded poly(isobutylcyanoacrylate) nanocapsules induced a reduction of the glycemia to normal level in streptozotocin diabetic rats [80] and alloxan induced diabetic dogs [81]. The hypolglycemic effect was characterized by surprising events including a lag time period of 2 days and a prolonged effect over 20 days. Insulin is a very hydrosoluble peptide and should be inactivated by the enzymes of the gastrointestinal tract. Thus, that insulin could be encapsulated with high efficiency in nanocapsules containing an oily core and why these nanocapsules showed so unexpected biological effect remained unexplained. Nanocapsules were prepared by interfacial polymerization of isobutylcyanoacrylate [82]. Any nucleophilic group including those of some of the aminoacids of insulin [83] could initiate the polymerization of such a monomer. In this case, insulin could be found covalently attached to the polymer forming the nanocapsule wall as was recently demonstrated with insulin-loaded nanospheres [84].

Aboubakar et al. [85] studied physicochemical characterization of insulin-loaded poly(isobutyl cyanoacrylate) nanocapsules obtained by interfacial polymerization. They claimed that the large amount of ethanol used in the preparation of the nanocapsules that initiated the polymerization of isobutylcyanoacrylate preserved the peptide from a reaction with monomer resulting in a high encapsulation rate of insulin. From their investigations, it appears that insulin was located inside the core of the nanocapsules and not simply adsorbed onto their surface.

Lambert et al. [86] used poly(isobutyl cyanoacrylate) nanoparticles for the delivery of oligonucleotides. Nanoparticles of size ranging from 20 to 400 nm were prepared. The authors claimed that this technology might offer interesting perspectives for DNA and peptide transport and delivery.

4.9. Folic Acid Conjugated Nanoparticles

Drug targeting has been an issue in the pharmaceutical industries of late and progress in this area of research is noteworthy. Literature suggests the use of nonionic polymers, viz. poly(ethylene glycol), for surface modification would certainly increase the circulating time of the colloidal drug carriers [87–89]. Stella et al. [90] introduced a new concept in drug targeting. The concept involved design of poly(ethylene glycol) coated biodegradable nanoparticles coupled to folic acid to target the folate-binding protein, a soluble form of the folate receptor, which is overexpressed on the surface of many tumor cells. The detailed studies

suggested that the interaction with the folate-binding protein was achieved only when folate conjugated nanoparticles were used, and no interaction resulted with the nonconjugated nanoparticles used as controls [90]. Thus the authors claimed that the folate-linked nanoparticles represent a potential new drug carrier for tumor cell-selective targeting.

4.10. Dextran Hydrogel Nanoparticles

Dextrans are polysaccharides composed of glucose units coupled into long branched chains mainly through a 1–6 and some 1–3 glucosidic linkages. Dextrans are colloidal, hydrophilic, and water-soluble substances, inert in biological systems, and do not affect cell viability. Dextrans have been explored for the delivery of various pharmaceuticals [91–93]. Kim et al. [94] reported hydrogel nanoparticles prepared from glycidyl methacrylate dextran and dimethacrylate poly(ethylene glycol). Drug release studies were performed using clonazepam as a hydrophobic model drug which was found to be dextranase and pH dependent [94].

4.11. Poly(γ-benzyl-L-glutamate)/poly(ethylene oxide)

Hydrophilic–hydrophobic diblock copolymers exhibit amphiphilic behavior and form micelles with core–shell architecture. These polymeric carriers have been used to solubilize hydrophobic drugs, increase blood circulation time, obtain favorable biodistribution, and lower interactions with reticuloendothelial systems [95]. In the same direction, Oh et al. [96] reported the preparation and characterization of polymeric nanoparticles containing adriamycin as a model drug. The nanoparticles are obtained from PBLG/PEO diblock copolymer, which forms a hydrophobic inner core and a hydrophilic outer shell of micellar structure [97, 98], by adopting a dialysis procedure. Their results indicate that only 20% of the entrapped drug was released in 24 h at 37 °C and the release was dependent on the molecular weight of the hydrophobic polymer.

4.12. Chitosan–Poly(ethylene oxide) Nanoparticles

Hydrophilic nanoparticle carriers have important potential applications for the administration of therapeutic molecules [29, 99]. Most of the recently developed hydrophobic–hydrophilic carriers require the use of organic solvents for their preparation and have a limited protein-loading capacity [100, 101]. Calvo et al. [102] reported a new approach for the preparation of nanoparticles, made solely of hydrophilic polymer, to address these limitations. The preparation technique, based on an ionic gelation process, is extremely mild and involves the mixture of two aqueous phases at room temperature.

One phase contains the polysaccharide chitosan (CS) and a diblock copolymer of ethylene oxide and polyanion sodium tripolyphosphate (TPP) (Fig. 8). It was stated that the size (200–1000 nm) and zeta potential (between +20 and +60 mV) of nanoparticles can be conventionally modulated by varying the ratio CS/PEO-PPO. Furthermore,

Figure 8. Schematic representation of chitosan nanoparticle preparation. Reprinted with permission from [102], P. Calvo et al., *J. Appl. Polym. Sci.* 63, 125 (1997). © 1997, Wiley-VCH.

using bovine serun albumin (BSA) as a model protein, it was shown that these new nanoparticles have great protein loading capacity (entrapment efficiency up to 80% of the protein) and provide a continuous release of the entrapped protein for up to 1 week [102].

4.13. Methotrexate-*o*-carboxymethylate Chitosan

Nanoparticles of methotrexate (MTX) were prepared using *o*-carboxymethylate chitosan (*o*-CMC) as wall forming materials and an isoelectric-critical technique under ambient conditions [103]. Drug controlled releases were studied in several media including simulated gastric fluid, intestinal fluid, and 1% fresh mice serum. It was found that acidic media provide a faster release rate than neutral media. The effects of MTX/*o*-CMC ratio and amount of crosslinking agents of drug release in different media were evaluated. The changes of size and effective diameter of *o*-CMC nanoparticles were detected by SEM and a laser light scattering system before and after the drug release. Authors claimed that the *o*-CMC nanoparticles constitute an attractive alternative to other anticancer drugs and enzyme carriers [103].

4.14. Solid Lipid Nanoparticles

Solid lipid nanoparticles (SLNs), one of the colloidal carrier systems, have many advantages such as good biocompatibility, low toxicity, and stability [104]. Schwarz and Mehnert [105] studied the lipophilic model drugs tetracaine and etomidate. The study highlights the maximum drug loading, entrapment efficacy, and effect of drug incorporation on SLN size, zeta potential (charge), and long-term physical stability. Drug loads of up to 10% were achieved with a

good maintenance of physically stable nanoparticle dispersion [105]. They claimed that the incorporation of drugs showed no or little effect on particle size and zeta potential compared to drug-free SLNs [105]. In another study, Kim and Kim [106] studied the effect of drug lipophilicity and surfactant on the drug loading capacity, particle size, and drug release profile. They prepared SLNs by homogenization of melted lipid dispersed in an aqueous surfactant solution. Ketoprofen, ibuprofen, and pranoprofen were used as model drugs and tween and poloxamer surfactants were tested [106]. Mean particle size of prepared SLNs ranged from 100 to 150 nm. It was found that the drug loading capacity was improved with the highly lipophilic drug and low concentration of surfactant [106]. Despite some setbacks, lipid nanoparticles continued to be of great interest in the fascinating area of drug delivery technology [107–109].

4.15. Cholic Acid and Methoxy Poly(ethylene glycol)

Cholic acid (CA) is one of the major bile acids, the main product of cholesterol metabolism, and biologically one of the most detergent-like molecules in the body. Cholic acid acts as a hydrophobic core and represents a potential drug incorporation site. Methoxy poly(ethylene glycol) (MPEG), a modified form of poly(ethylene glycol), is known to be nontoxic, nonimmunogenic, highly biocompatible, and, moreover, very well known to prevent interactions with cells and proteins due to its hydrophilic nature [110]. Kim et al. [111] took advantage of these wonderful properties of CA and MPEG for developing amphiphilic polymeric nanoparticles following their previously reported methods (Fig. 9) [112–114]. The physicochemical characteristics of the polymer conjugates were investigated [111]. Clonazepam was used as a model drug in their studies and the drug loading content was 16.2 wt%. They monitored the drug release and found pseudo-zero-order kinetics [111].

4.16. Poly-L-lysine Nanoparticles

Poly-L-lysine finds application in various biomedical areas, viz., controlled drug delivery [115] and gene delivery [116]. Many hydrophilic polymers have been linked covalently to poly-L-lysine and copolymers such as poly-L-lysine poly(ethylene glycol) block [117, 118], poly-L-lysine poly(ethylene glycol) graft [119], and poly-L-lysine hyaluronic acid graft [120] have been synthesized for biomedical applications. Reports are also available on synthesis of lipid bearing poly-L-lysines, where lipid units are attached to terminal lysine amino units [121]. Recently, Wang et al. [122] reported the covalent attachment of both hydrophilic and hydrophobic units to poly-L-lysine and further fabrication of nanoparticles. The nanoparticles were prepared by sonicating poly-L-lysine graft copolymer amphiphiles in aqueous media [122]. The influence of polymer architecture and molecular weight on the properties of the nanoparticles has also been investigated [122]. The schematic representation of the nanoparticle architecture is shown in Figure 10.

Figure 9. Synthetic scheme of CE conjugate. Reprinted with permission from [111], I.-S. Kim and S.-H. Kim, *Int. J. Pharm.* 226, 23 (2001). © 2001, Elsevier Science.

4.17. Surface Modified Silica Nanoparticles

Surface modified silica nanoparticles were synthesized by modification of commercially available silica particles (IPAST, Nissan Chemical Industries, Tokyo, Japan) with

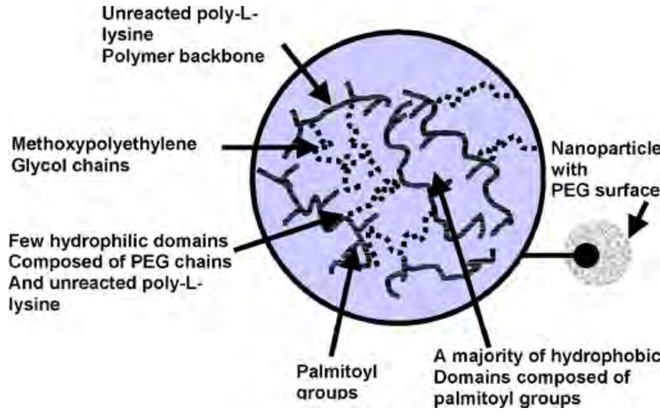


Figure 10. Schematic representation of PLP nanoparticle architecture. Reprinted with permission from [122], W. Wang et al., *Langmuir* 16, 7859 (2000). © 2000, American Chemical Society.

various aminoalkylsilanes, viz., *N*-(2-aminoethyl)-3-aminopropyltrimethoxysilane (AEAPS) and *N*-(6-aminohexyl)-aminopropyltrimethoxysilane (AHAPS). The resulting particles have a hydrodynamic diameter of 26 nm and a zeta potential up to +31 mV [123, 124]. The scheme for modification of the silica particles is shown in Figure 11.

5. NANOPARTICLES FOR SOME SPECIFIC APPLICATIONS

5.1. Gene Delivery Systems

For more than two decades, researchers have been working to alleviate disease through gene therapy. In this type of treatment a gene is delivered to cells, allowing them to produce their own therapeutic proteins. Traditionally, DNA delivery systems have been classified as viral vector mediated systems and nonviral vector mediated systems [125]. Currently, because of their highly evolved and specialized components, viral systems are by far the most effective means of DNA delivery, achieving high efficiencies (>90%) for both delivery and expression [126]. The most promising nonviral gene delivery system thus far, other than the "gene gun," is DNA vaccine applications, comprised of ionic complexes formed between DNA and polycationic liposomes [127, 128]. A major problem in the practical application of gene transfer for therapeutic purposes (gene therapy) is the efficient delivery of the genetic material to the desired place of action, the target tissue. Nanotechnology appears to be one of the potential solutions to take up this technological challenge. In nature, efficient gene delivery is achieved by viruses which have evolved over millions of years into well adapted gene delivery particles. Hence, one approach to improving the chances of successful gene therapy may be the design of fully synthetic virus equivalents with the means of nanotechnology.

The typical components of a virus include (i) the genomic nucleic acid, which is usually condensed by (ii) nucleic acid binding proteins (NBPs) to form a core which is further surrounded by (iii) a capsid formed of proteins and/or a lipid envelope. The overall particle size ranges between 20 and 200 nm. Figure 12 describes the architecture of human adenoviruses which are currently very popular for viral gene therapy trials. The nucleic acid of a virus can be RNA or DNA, single or double stranded, and linear or circular. A common and important property of the nucleic acid binding proteins is their high positive charge excess to achieve efficient binding and condensation of nucleic acids. Each histone complex, probably the best studied nucleic acid binding unit, contains 220 cationic lysine and arginine residues, has a dimension of $10 \times 18 \times 11$ nm, and can bind 146 base pairs of double stranded DNA per unit. These structural characteristics are shared by all other NBPs and were the basis for the design of synthetic equivalents.

5.1.1. Technologies for Production of DNA Nanoparticles

Two major technological approaches to the production of DNA nanocarriers can be distinguished: "conventional" nanoparticle preparation as described in Section 3 and the

(A) Hydrolysis: $R-Si(OR)_3 + nH_2O \longrightarrow R-Si(OR)_{3-n}(OH)_n + nROH$

Condensation: $\sim Si-OH + RO-Si\sim \longrightarrow \sim Si-O-Si\sim + ROH$

(B) Aminoalkylsilanes: n=1 → AEAPS; n=2 → AHAPS

(C) IPAST $\xrightarrow{H^+}$ SURFACE MODIFIED PARTICLES

Figure 11. (A) Hydrolysis and condensation of unmodified particles, (B) alkyaminoalkanes, (C) modification scheme.

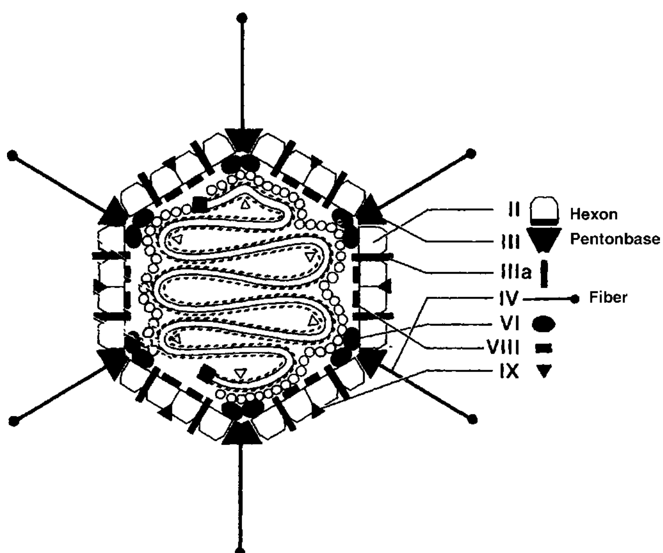


Figure 12. Structure of type 2 adenovirus according to Stewart and Burnett [189]. Adenoviruses consist of a core containing linear double stranded DNA of 40 kilo base pairs which is surrounded by a complex capsid. They do not possess a lipid envelope. The average size of adenovirus particles ranges between 70 and 110 nm. The genomic DNA is covalently capped at its ends with the terminal protein, tp ■. pVII (–), the cationic virus protein VII, forms dimers that fit into the major groove of the DNA double helix and compensate approx. 80% of the negative charge of the DNA in order to pack the genome into dense, nucleosome-like structures. Further components form the capsid surrounding the core, including pIII, the major hexon protein pII, and the fiber protein pIV. pIV und pIII mediate specific receptor mediated uptake of the virus particle by target cells.

self-assembly technique which utilizes interpolyelectrolyte interactions between the polyanionic nucleic acid and polycationic polymers or particles.

Approach 1 Some of the available technologies for production of nano- and microparticles have been adapted for the encapsulation of plasmid DNA. Complex coacervation, as described in Section 3.1, has been used successfully to incorporate the nucleic acid within a matrix of gelatin [129]. Addition of sodium sulphate as desolvating agent to a mixture of DNA and type A gelatin followed by vigorous vortexing produced precipitation of particles that were typically around 400 nm in size. This reaction is favored by slightly acidic pH, suggesting a role for electrostatic interactions, and the formed complexes can be hardened with calcium chloride (0.5 M). DNA encapsulation efficiencies reached >98% at a loading level of 25–30% (w/w). In addition, further bioactive molecules present during the precipitation step can be efficiently incorporated as demonstrated for chloroquine which enhances transfection by inhibition of intracellular inactivation of the transgenic DNA.

Other protocols for the preparation of DNA nanoparticles have used the solvent extraction–evaporation technique as described in Sections 3.3 and 4.3. Due to the low solubility of DNA in many organic solvents, a water–in oil–in water double emulsion is produced and the solvent is removed from the oily phase to form a polymer nano- or microcapsule around the hydrophilic core containing the DNA. Suitable capsule materials include biodegradable polymers like the polyanhydride PLGA. Using this material, Cohen et al. [130] reported the production of DNA nanocapsules with diameters between 400 and 700 nm that did also show significant transfection of mammalian cells *in vitro*. With optimized conditions, the efficiency of DNA encapsulation can exceed 75%. However, the homogenization steps involved cause considerable damage to the plasmid DNA and thereby greatly reduce its activity. To reduce physical stress to the DNA, stabilizing agents like poly-L-lysine [131], EDTA, or carbohydrates [132] can be added and the inner water phase may be frozen before the second homogenization step [132]. The highest reported DNA encapsulation efficiency after utilization of these measures was 90% with a preservation of >90% DNA integrity. However, the overall content of nucleic acid exceeds 1% only occasionally.

Approach 2 Most "conventional" methods for preparation of nanoparticles involve homogenization steps that irreversibly inactivate the relatively fragile nucleic acids. Self-assembly techniques offer the possibility to avoid such damage. The currently most popular self-assembly protocols employ suitable polycationic polymers. Driven by polyelectrolyte interactions, they react with DNA into small, usually toroid structures termed *polyplexes* [133]. Their size can range from 10 to approximately 200 nm and if one of the reactants is used in excess (usually the cationic polymer), the resulting suspension will be stabilized by repulsive forces between the particles. Further stabilization has been achieved by surface modification with hydrophilic polymers as described in Section 4.7. Suitable cationic polymers that have been successfully used in the past include poly-L-lysine [134], polyethylenimine [135], aminoalkylmethacrylate based (co)polymers [136], dendrimers [137], and chitosan [138].

A major limitation of the polymers is the difficulty to include additional bioactive molecules (in order to enhance or modify the DNA activity) during the self-assembly process, unless these are covalently linked to the polymer backbone. This drawback can be overcome if cationic nanoparticles are used instead of polymers. Studies on cationically modified silica nanoparticles have shown that this is, in principle, possible [139], especially for small particles with diameters between 10 and 30 nm that have dimensions similar to naturally occurring NBP like the histone complex produced biological active complexes. A representative example of such a nanoparticle–DNA complex (*nanoplex*) is depicted in Figure 13. *In vitro* gene delivery has also been achieved using nanoparticles consisting of a PLA core and a cationic surface modification with polylysine [140].

A further advantage can be provided by cationic SLNs described in Section 4.14. Nanosuspensions of these are produced by high pressure, hot homogenization of a mixture of matrix lipid, tenside, and cationic modifier. The latter mediates electrostatic binding to DNA and the proximity of multiple SLNs on the DNA macromolecule can induce particle coalescence as observed by AFM (Fig. 9). However, matrix lipid and modifier have to be selected carefully in order to avoid dissociation of the cationic modifier from the particle surface after DNA binding and to retain sufficient fluidity to allow for particle coalescence [141].

5.1.2. Outlook

Further significant improvement of the efficiency of nanoparticular and all other synthetic gene delivery vehicles will be required for their clinical success. This may be achieved by a combination of the basic core containing the

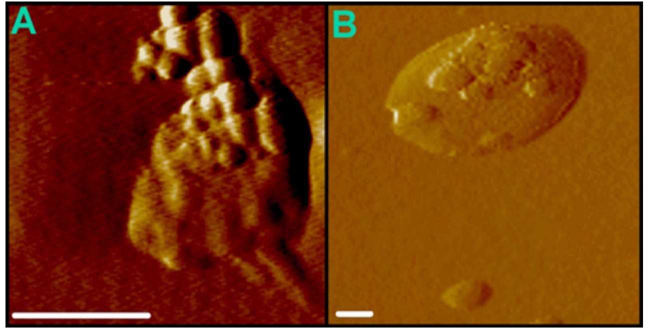


Figure 13. Nanoplexes: Complexes of plasmid DNA and cationic nanoparticles. Imaging was performed by atomic force microscopy (bar = 100 nm). (A) Cationic silica particles ($d = 26$ nm) self-associate with plasmid DNA into submicrometer complexes that protect, release, and transfect the DNA *in vitro*. Reprinted with permission from [139], C. Kneuer et al., *Bioconj. Chem.* 11, 926 (2000). © 2000, American Chemical Society. (B) Solid lipid nanoparticles ($d = 100$ nm) composed of compritol ATO, Tween 80, Span 85, and disteaorylethyl dimethyl ammonium-chloride form discrete submicrometer particles with plasmid DNA (upper half). Coalescence of the individual SLN following interaction with DNA stabilizes the complex. A single SLN can be seen in the lower half of the image. Reprinted with permission from [141], C. Olbrich et al., *J. Control. Release* 77, 345 (2001). © 2001, Elsevier Science.

nucleic acid or the nucleic acid carrier with additional functional units. The latter can include a capsid for protection and target recognition, nucleic acid modifications for intracellular delivery to the cell nucleus, and inhibitors and modulators of "undesired" cellular activities. Such components have been identified in many viruses and are known to be critical for high efficiency. Ogris et al. [142] described a synthetic prototype.

5.2. Peptide Delivery

Recently, colloidal carrier systems, especially nanoparticles, have been receiving much attention in the field of drug targeting because of their high loading capacity for drugs, particularly macromolecules, as well as their unique disposition characteristics in the body [143, 144]. It is well known that the bioavailability of peptide and protein drugs after oral administration is very low because of their instability in the gastrointestinal (GI) tract and low permeability through the intestinal mucosa [145, 146]. Therefore, injectable dosage forms are currently used to obtain therapeutic effects. Due to poor compliance of these administration routes, it is indispensable to develop alternatives such as nasal, buccal, rectal, vaginal, pulmonary, and transdermal routes [147]. Oral administration is the most convenient route for drug delivery, and several approaches such as chemical modification to alter the physicochemical properties of peptide drugs [148], the use of an absorption enhancer to promote drug absorption [149–151], and the use of a protease inhibitor to protect drugs against degradation by enzymes [150] have been investigated in order to achieve oral peptide delivery. Many pharmaceutical scientists have also reported that particulate drug delivery systems such as nanoparticles [152–154], microcapsules [153, 155], liposomes [156], and emulsions [157] are useful in improving the absorption of peptide drugs via the GI tract.

Much of the research has focused on the absorption enhancement of peptide and protein drugs and vaccine antigens. Couvreur and co-workers have been studying whether nanoparticles composed of polyalkylcyanoacrylate derivatives enhance the absorption of orally administered insulin, using animal models [158–161]. They found that insulin encapsulated in polyisobutylcyanoacrylate nanocapsules reduced glycemia by 50–60%, although free insulin did not affect glycemia when administered orally to diabetic rats. Kawashima et al. [162] reported that the absorption of elcatonin, which is a calcitonin of eel, via the GI tract of rats was enhanced by poly(-lactide-glycolide) nanoparticles coated with chitosan, which is a mucoadhesive cationic polymer. They also studied the effect of chitosan-coated liposomes on the intestinal absorption of insulin in rats and reported that the blood glucose concentration decreased significantly after oral administration of insulin-loaded liposomes [163]. Sakuma et al. also demonstrated that nanoparticles composed of novel graft copolymers having a hydrophobic backbone and hydrophilic branches can improve the absorption of salmon calcitonin in rats [164–167].

In general, absorption of a drug via the GI tract is closely related to its solubility. Many drugs with poor absorption characteristics are highly water-soluble and this high polarity causes the low permeability through the intestinal membrane which is mainly composed of hydrophobic lipids. Numerous peptide and protein drugs fall into this category [168–170]. There are two ways to incorporate a hydrophilic drug in a nanoparticle. One is to encapsulate it into the nanoparticle core and the other is to adsorb it onto the nanoparticle surface [170].

5.3. Brain Delivery

The blood–brain barrier (BBB) represents an insurmountable obstacle for a large number of drugs, including antibiotics, antineoplastic agents, and a variety of central nervous system-active drugs, especially neuropeptides. One of the possibilities to overcome this barrier could be drug delivery nanoparticulate systems to the brain. The nanoparticles may be especially helpful for the treatment of the disseminated and very aggressive brain tumors. The mechanism of the nanoparticle-mediated transport of the drugs across the blood–brain barrier at present is not fully elucidated. The most likely mechanism is endocytosis by the endothelial cells lining the brain blood capillaries [171]. The brain blood vessel endothelial cells are characterized by having tight continuous circumferential junctions between them thus abolishing any aqueous paracellular pathways between these cells [172]. Drugs may be bound in the form of a solid solution or dispersion or adsorbed to the surface or chemically attached. Alyautdin et al. claimed that the poly(butyl cyanoacrylate) nanoparticles are the only nanoparticles to date that were so far successfully used for the *in vivo* delivery of drugs to the brain [171]. This polymer has the advantage of rapid biodegradability [173, 174]. Kreuter and co-workers reported the delivery of hexapeptide dalargin (Tyr-Ala-Gly-Phe-Leu-Arg), a Leu-enkephalin analog with opioid activity, the first drug that was delivered to the brain using nanoparticles. The antinociceptive activity after intravenous injection of dalargin-loaded poly(butyl cyanoacrylate) nanoparticles of a size around 250 nm overcoated with polysorbate 80 was demonstrated by the tail-flick test [174] as well as by the hot-plate test [175, 176]. Both tests showed the same tendencies. A circadian-phase dependency of the reaction time in the hot-plate test was also observed by Ramge et al. [176] in Balb/c and DBA/2 mice. The pharmacokinetics of doxorubicin bound to nanoparticles and/or coated with polysorbate 80 after intravenous injection to rats was investigated by Gulyaev et al. [177]. Schroeder et al. [178] also showed that polysorbate 85 coated poly(butyl cyanoacrylate) nanoparticles enable a brain transport after oral administration. The pharmacokinetics of another drug, amitriptyline, normally a tricyclic antidepressant, however, is able to penetrate the BBB [179]. Results demonstrated that intravenous injection of polysorbate 80 coated nanoparticles showed an improvement in brain AUC, whereas their serum AUC was reduced [179].

In vitro experiments using brain blood vessel endothelial cells were conducted to gain insight in the quantitation and possible mechanism of the nanoparticle-mediated transport of drugs into the brain. The first *in vitro* experiments employed poly(methyl-[2-14C]-methacrylate) nanoparticles [180]. More recently, rat brain endothelial cells of the RBE4 cell line and poly(butyl cyanoacrylate) nanoparticles were used to study the nanoparticle uptake by these

cells [181]. As mentioned before, poly(butyl cyanoacrylate) particles are very rapidly biodegradable and are slightly more hydrophilic than the poly(methyl methacrylate) particles reported by Borchard et al. [180]. The poly(butyl cyanoacrylate) nanoparticles in the RBE4 cell experiments were produced using fluoresceine isothiocyanate dextran 70000 instead of the normal dextran as in most of the *in vivo* experiments reported. The toxicity of poly(butyl cyanoacrylate) nanoparticles was determined in the MTT test [182] using RBE4, bovine, and human brain microvessel endothelial cells. The highest toxicity was observed in the RBE4 cells at a nanoparticle concentration of 10 μg/ml and no decrease in viability was observed. Fenart et al. [183] used the Cecchelli model to investigate the effect of charge and lipid coating on the ability of maltodextrin 60 nm nanoparticles to cross the brain capillary endothelial cell layer of the model.

Nanoparticles represent a tool to transport essential drugs across the BBB that normally are unable to cross this barrier. Drugs that have successfully been transported into the brain using this carrier include the hexapeptide dalargin, the dipeptide kytorphin, loperamide, tubocurarine, the NMDA receptor antagonist MRZ 2/576, and doxorubicin [171]. Nanoparticle-mediated drug transport to the brain depends on the overcoating of the particles with polysorbates, especially polysorbate 80. This material seems to act as an anchor for apolipoprotein E (apo E) or other substances following injection into the blood stream. The adsorption of apo E from blood plasma onto the nanoparticle surface has been detected *in vitro* [171].

5.4. Colon Targeting

In several studies the involvement of macrophages and dendritic cells in active inflammatory bowel disease (IBD) has been indicated. An efficient manipulation of those cells might be considered as an interesting potential therapeutic pathway for IBD treatment, since it was reported that biodegradable microspheres can be sufficiently taken by macrophages and M cells [184]. Based on these results, the direct uptake of anti-inflammatory agents by macrophages, achieved with the use of microspheres, appears to have a superior immunosuppressive effect and to be more useful for the treatment of patients with IBD [185, 186]. Nanoparticles showed a size dependent accumulation in the inflamed tissue and were, therefore, thought to be even more promising in the treatment of this disease. They proved the ability to target the macrophages and ruptures (Fig. 14) inside the ulcerated tissue in IBD for a specific local drug delivery [187, 188]. This type of new strategy aims to selectively accumulate the drug delivery system at the site of action. Polymeric nanoparticulate carrier systems were expected to target the inflamed tissue in inflammatory bowel diseases. Since no sedimentation occurs with colloidal drug carriers, they might be less or not at all affected by the streaming due to their diffusing properties. With the anti-inflammatory model drug rolipram, the entire drug loaded nanoparticle formulations proved to be as efficient as the drug in solution in mitigating the experimental colitis. First, the inflammation activity score and myeloperoxidase activity decreased significantly after the oral administration of rolipram nanoparticles or solution. When animals were kept without drug treatment the drug solution group displayed a strong relapse, whereas the nanoparticle groups maintained reduced inflammation levels. Moreover, when free drug was administered, the rolipram solution group exhibited a high adverse effect index, whereas the rolipram nanoparticle groups proved their potential to retain the drug from systemic absorption as evidenced by a significantly reduced index. An essential advantage of this strategy seems to be the direct contact of the carriers with the inflammation site which allows a much higher local drug concentration. Moreover, nanoparticles were found not only to accumulate in the ulceration but also to adhere distinctly to the mucus which allows an increased specificity to nonulcerated inflamed tissue since mucus production is highly increased all over the inflamed tissue.

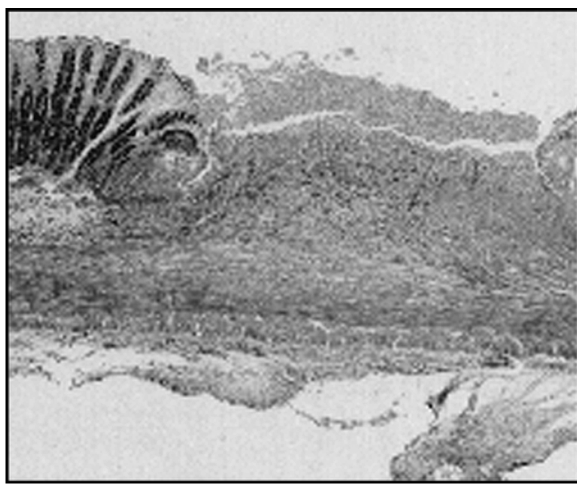


Figure 14. Histological section across inflamed colonic tissue. The ulceration exhibits distinct immune related cell extravasion while the mucus production is highly increased in intact mucosa (black areas on the left). Reprinted with permission from [187], A. Lamprecht et al., *J. Pharmacol. Exp. Ther.* 299, 775 (2001). © 2001, American Society for Pharmacology and Experimental Therapeutics.

These new delivery systems allow the desired drug to accumulate in the inflamed tissue with a relatively high efficiency including two major advantages. The drug is concentrated at its site of action, which reduces possible adverse effects and enhances the effect of the administered dose. Moreover, the sustained drug release allows pharmacological effects to be extended due to the prolonged residence time of the carrier system at the targeted inflamed area. This deposition of polymeric carrier systems in the inflamed tissues might be promising in the design of new carrier systems for the treatment of inflammatory bowel disease.

GLOSSARY

Atomic force microscopy (AFM) A technique for analyzing the surface of a rigid material all the way down to the level of the atom. AFM uses a mechanical probe to magnify surface features up to 100,000,000 times, and it produces 3D images of the surface.

Biodegradable The ability of a substance to be broken down into simpler, smaller parts by a biological process. Many plastics are not biodegradable.

Blood-brain barrier (BBB) The walls of blood vessels in the brain have a special property that prevents large molecules from moving out of the bloodstream and into the brain. This property, called the blood-brain barrier, serves to protect the brain by keeping potentially harmful substances from entering the brain.

Central nervous system (CNS) The central nervous system is that part of the nervous system that consists of the brain and spinal cord.

Colloids These are finely divided dispersions of one material in a second continuous phase. The size of colloidal particles is between that of individual small molecules and macroscopic objects. It is convenient to take a limit as; at least one dimension should be less than or of the order of a micrometer (10^{-6} m). The colloid consists of a dispersed phase in a continuous dispersion medium.

Critical aggregation concentration In mixed solutions of surfactants and polymers there is often association at a surfactant concentration below the normal critical micelle concentration, (CMC). Binding of surfactant to the polymer molecules usually causes this. There will be a discontinuity in the slope of a plot of surface tension versus concentration at this concentration.

Critical coagulation concentration When salt is added to a charge stabilized colloid, the screening length is reduced and at a particular concentration the dispersion is no longer stable (such that there is no secondary minimum in the DLVO potential). The concentration of salt at which the particles are no longer stable and begin to aggregate is the critical coagulation concentration (CCC).

Critical micelle concentration (CMC) This is the concentration above which surfactants form micelles in solution. The change in properties that occur as micelles form is marked by sharp transitions in many physical quantities such as the surface tension of the solution, conductivity, turbidity, and nuclear magnetic resonance chemical shifts.

Deoxy ribonucleic acid (DNA) The molecule that carries the genetic information for most living systems. The DNA molecule consists of four bases (adenine, cytosine, guanine, and thymine) and a sugar phosphate backbone arranged in two connected strands.

Emulsion A colloidal dispersion of one liquid in another liquid (usually an oil and water). It is often "stabilized" with a surfactant or with polymers. Emulsions are not usually truly stable but may be metastable. This contrasts with microemulsions.

Emulsion polymer A polymer that is prepared by polymerization of a dispersion of monomer in a liquid. It is often stabilized with a surfactant or with polymers. This preparation involves an emulsion but the product is a colloidal dispersion of a polymer that is usually solid in the liquid. The product is known as latex. Typically many vinyl polymers are prepared with radical initiators in this way.

Flocculation Colloids are described as flocculated when the individual particles have aggregated together to form clusters. This occurs when the repulsive potential that provides stability in dispersion becomes less than the attraction. Critical flocculation concentration and temperature are the parameters above which the individual particles aggregate to form clusters.

Fluoresceine isothiocyanate (FITC) Green fluorescent dye used to 'tag' antibodies for use in immuno-fluorescence.

Gastrointestinal tract (GIT) Organs that are part of the digestive tract, which include the esophagus, stomach, small intestine, and large intestine (rectum and colon).

Inflammatory bowel disease (IBD) A chronic inflammatory condition of the digestive tract. Ulcerative colitis and Crohn's disease are the most common forms of IBD.

Mononuclear phagocytic system (MPS) A system comprising blood monocytes and tissue macrophages.

Nanoparticles A material with dimensions less than 1000 nanometers and often below 100 nanometers.

Photon correlation spectroscopy (PCS) The correlation of scattered photons is measured. This can be used to determine the dynamic behaviour of particles or polymer molecules at length scales determined by the reciprocal of the scattering vector, $Q = (4\pi/\lambda)\sin(\theta/2)$ and if there is a simple exponential decay characterised by a relaxation time, t, the diffusion coefficient, D and hence hydrodynamic radius, can be determined as follows: $D = 1/\tau Q^2$.

Poly(lactide-*co*-glycolide) (PLGA) Copolyester of lactic acid and glycolic acid.

Reticuloendothelial system (RES) A rather old term for the network of phagocytes and endothelial cells throughout the body.

Ribonucleic acid (RNA) A nucleic acid molecule similar to DNA but containing ribose rather than deoxyribose. RNA is formed upon a DNA template. There are several classes of RNA molecules. They play crucial roles in protein synthesis and other cell activities.

Scanning electron microscopy (SEM) Scanning electron microscopy is performed by scanning a focused probe across the surface of the sample to be studied. Secondary electrons emitted from the sample are typically detected by a photomultiplier system, the output of which is used to modulate the brightness of a TV monitor that is rastered in synchronization with the electron beam scan. The more electrons a particular region emits, the brighter the image at that point. SEM images typically contain a good deal of topographical detail.

Transmission electron microscopy (TEM) The electron beam is passed through a thin film sample (typically ~1–200 nm thick). Bright field diffraction contrast images are formed with the direct (undiffracted) beam. Dark field images are formed with a selected diffracted beam.

ACKNOWLEDGMENT

M. N. V. R. K. is grateful to the Alexander Von Humboldt Foundation, Germany, for providing him with a personal fellowship.

REFERENCES

1. R. Chandra and R. Rustgi, *Progr. Polym. Sci.* 23, 1273 (1998).
2. J. Kreuter, in "Colloidal Drug Delivery Systems" (J. Kreuter, Ed.), pp. 219–342. Dekker, New York, 1994.
3. J. Kreuter and P. Speiser, *J. Pharm. Sci.* 65, 1624 (1976).
4. P. Couvreur, L. Grislain, V. Lenaert, F. Brasseur, P. Guiot, and A. Biernacki, in "Polymeric Nanoparticles and Microparticles" (P. Guiot and P. Couvreur, Eds.), pp. 27–93. CRC Press, Boca Raton, 1986.
5. V. Labhasetwar, C. Song, and R. J. Levy, *Adv. Drug Del. Rev.* 24, 63 (1997).
6. R. Arshady, *Polym. Eng. Sci.* 33, 865 (1993).
7. J. R. Ford, A. A. Morfesis, and R. L. Rowell, *J. Colloid Interface Sci.* 105, 516 (1985).
8. G. Ohshima and T. Kondo, *J. Colloid Interface Sci.* 130, 281 (1989).
9. R. Buscall, in "Science and Technology of Polymer Colloid" (G. W. Poechlein, R. H. Ottewill, and J. W. Goodwin, Eds.), Vol. 2, pp. 279–313. Nijhoff, The Hague, 1983.
10. F. Hoshino, T. Fujimoto, H. Kawaguchi, and Y. Ohtsuka, *Polym. J.* 19, 241 (1987).
11. B. Vincent, in "Science and Technology of Polymer Colloid" (G. W. Poechlein, R. H. Ottewill, and J. W. Goodwin, Eds.), Vol. 2, pp. 235–352. Nijhoff, The Hague, 1983.
12. T. Nashima and K. Furusawa, *Colloids Surf.* 55, 149 (1991).
13. B. Vincent, J. Edwards, S. Emmett, and A. Jones, *Colloids Surf.* 18, 261 (1986).
14. Y. Almong, S. Reich, and M. Levy, *Br. Polym. J.* 14, 131 (1987).
15. C. K. Ober and K. P. Lok, *Marcomolecules* 20, 268 (1987).
16. P. A. Kramer, *J. Pharm. Sci.* 63, 1646 (1974).
17. M. J. Rajaonarivory, C. Vauthier, G. Couarraze, F. Puisiex, and P. Couvreur, *J. Pharm. Sci.* 82, 912 (1993).
18. P. Couvreur, M. Roland, and P. Speiser, U.S. Patent 4, 329, 332, 1982.
19. P. Couvreur, M. Roland, and P. Speiser, U.S. Patent 4, 489, 055, 1984.
20. P. Couvreur, in "Microencapsulation: Methods and Industrial Applications" (S. Benita, Ed.), Vol. 73, pp. 183–211. Dekker, New York, 1996.
21. I. C. Eromosele, D. C. Pepper, and B. Ryan, *Makromol. Chem.* 190, 1613 (1989).
22. E. F. Donnelly, *Polymer Lett. Ed.* 15, 399 (1977).
23. D. C. Pepper, *J. Polymer Sci.: Polymer Symp.* 62, 65 (1978).
24. J. M. Rollot, P. Couvreur, L. Roblot-Treupel, and F. Puisieux, *J. Pharm. Sci.* 75, 361 (1986).
25. "Emulsions and Nanosuspensions for the Formulations of Poorly Soluble Drugs" (R. H. Müller, S. Benita, and B, Böhm, Eds.). Medpharm Scientific, Stuttgart, 1998.
26. P. Bernoulli, in "Pharmazeutische Technolgie" (S. Sucker, H. Fuchs, and P. Speiser, Eds.), pp. 101–102. Gerog Thieme, Stuttgart, 1978.
27. B. Böhm, Ph.D. Thesis, Freie Universität Berlin, Germany, 1999.
28. R. H. Müller, *Arch. Pharm.* 322, 700 (1989).
29. R. H. Müller and B. Böhm, in "Emulsions and Nanosuspensions for the Formulations of Poorly Soluble Drugs" (R. H. Müller, S. Benita, and B. Böhm, Eds.). Medpharm Scientific, Stuttgart, 1998.
30. K. P. Krause, O. Kayser, K. Mäder, R. Gust, and R. H. Müller, *Int. J. Pharm.* 196, 169 (2000).
31. R. H. Müller, K. Peters, R. Becker, and B. Kruss, *Int. Symp. Control. Rel. Bioact. Mater.* 22, 574 (1995).
32. R. H. Müller, R. Becker, B. Kruss, and K. Peters, U.S. Patent 5, 858, 410, 1998.
33. R. H. Müller, C. Jacobs, and O. Kayser, *Adv. Drug. Del. Rev.* 47, 3 (2001).
34. R. Gref, Y. Minamitake, M. T. Peracchia, V. Trubetskoy, V. Torchlin, and R. Langer, *Science* 263, 1600 (1994).
35. B. Seijo, E. Fattal, L. Roblot-Treupel, and P. Couvreur, *Int. J. Pharm.* 62, 1 (1990).
36. L. Illum, M. A. Khan, E. Mak, and S. S. Davis, *Int. J. Pharm.* 30, 17 (1986).
37. V. Guise et al., *Pharm. Res.* 7, 736 (1990).
38. J. L. Grangier, M. Puygrenier, J. C. Gauthier, and P. Couvreur, *J. Control. Release* 15, 3 (1991).
39. C. Chavany et al., *Pharm. Res.* 9, 441 (1992).
40. R. H. Muller, C. Lherm, J. Herbort, and P. Couvreur, *Biomaterials* 11, 590 (1990).
41. V. Lenaerts et al., *Biomaterials* 5, 65 (1978).
42. P. Couvreur, B. Kante, M. Roland, and P. Speiser, *J. Pharm. Sci.* 68, 1521 (1979).
43. M. N. V. Ravi Kumar, Neeraj Kumar, A. J. Domb, and M. Arora, *Adv. Polym. Sci.* 160, 45 (2002).
44. A. Vila, A. Sanchez, M. Tobio, P. Calvo, and M. J. Alonso, *J. Control. Release* 78, 15 (2002).
45. J. Davda and V. Labhasetwar, *Int. J. Pharm.* 233, 51 (2002).
46. M. N. V. Ravi Kumar, *J. Pharm. Pharm. Sci.* 3, 234 (2000).
47. D. M. Schachter and J. Kohn, *J. Control. Release* 78, 143 (2002).
48. J. M. Harris, in "Topics in Applied Poly(ethylene glycol) Chemistry: Biochemical and Biomedical Applications," pp. 1–14. Plenum Press, New York, 1992.
49. T. Minko, P. Kopeckova, V. Pozharov, and J. Kopecek, *J. Control. Release* 54, 223 (1998).
50. R. Duncan, *Anti-cancer Drugs* 3, 175 (1992).
51. M. Yokoyama, G. S. Kwon, T. Okano, Y. Sakurai, T. Seto, and K. Kataoka, *Bioconj. Chem.* 3, 295 (1992).
52. S. Stolink, L. Illum, and S. S. Davis, *Adv. Drug. Deliv. Rev.* 16, 195 (1995).
53. Y. Tabata, Y. Murakami, and Y. Ikada, *J. Control. Release* 50, 123 (1998).
54. J. Kreuter, *J. Control. Release* 16, 169 (1991).
55. H.-S. Yoo, K.-H. Lee, J.-E. Oh, and T.-G. Park, *J. Control. Release* 68, 419 (2000).
56. S. Mitra, U. Gaur, P. C. Gosh, and A. N. Maitra, *J. Control. Release* 74, 317 (2001).
57. K. A. Janes, M. P. Fresneau, A. Marazuela, A. Fabra, and M. J. Alonso, *J. Control. Release* 73, 255 (2001).
58. M. S. Newman, G. T. Colbem, P. K. Working, C. Engbers, and A. A. Amantea, *Cancer Chemother. Pharmacol.* 43, 1 (1999).
59. P. Steerenberg et al., *Cancer Chemother. Pharmacol.* 21, 299 (1988).
60. R. Prez-Soler et al., *Clin. Cancer Res.* 3, 373 (1997).
61. K. N. J. Burger et al., *Nature Med.* 8, 81 (2002).
62. K. Cifcti, H. Kas Suheyla, A. Hincal Atilla, T. Meral Ercan, and S. G. O'Ruacan, *Int. J. Pharm.* 131, 73 (1996).
63. H. J. Jeon, Y. Jeong, M. K. Jang, Y. H. Park, and Y. W. Nah, *Int. J. Pharm.* 207, 99 (2000).
64. H.-Y. Kwon, J.-Y. Lee, S.-W. Choi, Y. Jang, and J.-H. Kim, *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 182, 123 (2001).
65. L. Mu and S. S. Feng, *J. Controlled Release* 80, 129 (2002).
66. N. S. Santos-Magalhaes, A. Pontes, V. M. W. Pereira, and M. N. P. Caetano, *Int. J. Pharm.* 208, 71 (2000).
67. J. Jagur-Grodzinski, *React. Funct. Polym.* 39, 99 (1999).
68. G. S. Kwon and T. Okano, *Adv. Drug Del. Rev.* 21, 107 (1996).
69. C. Scholz, M. Iijima, Y. Nagasaki, and K. Kataoka, *Macromolecules* 26, 7295 (1995).
70. S. B. La, T. Okano, and K. Kataoka, *J. Pham. Sci.* 85, 85 (1995).
71. X. Zhang and H. M. Burt, *Pharm. Res.* 12, S265 (1995).
72. S. E. Dunn, A. G. A. Coombes, M. C. Garnett, S. S. Davis, M. C. Davis, and L. J. Illum, *J. Control. Release* 44, 65 (1997).
73. S. M. Moghimi, I. S. Muir, L. Illum, S. S. Davis, and V. KolBacnofen, *Biochim. Biophys. Acta* 1179, 157 (1993).
74. H. R. Allock, S. R. Ducher, and A. G. Scopelianos, *Biomaterials* 15, 563 (1994).

75. H. R. Allock, T. J. Fuller, D. P. Mack, K. Matsumura, and K. M. Smeltz, *Macromolecules* 10, 824 (1997).
76. J. Vandorpe and E. Schacht, *Biomaterials* 18, 1147 (1997).
77. R. Gref, A. J. Domb, P. Quellec, T. Blunk, R. H. Müller, J. M. Verbavatz, and R. Langer, *Adv. Drug Del. Rev.* 16, 215 (1995).
78. J. S. Hrkach, M. T. Peracchia, A. J. Domb, N. Lotan, and R. Langer, *Biomaterials* 18, 27 (1997).
79. A. G. A. Coombes, S. Tasker, M. Lindblad, K. Holmgreen, and V. Hoste, *Biomaterials* 18, 1153 (1997).
80. C. Damage, C. Michel, M. Aprahamian, P. Couvreur, and J. P. Devissaguet, *J. Control. Release* 13, 233 (1990).
81. C. Damage, D. Hillaire-Buys, R. Puech, A. Hoeltzel, C. Michel, and G. Ribes, *Diab. Nutr. Metab.* 8, 3 (1995).
82. N. Alkhouri-Fallouh, L. Roblot Treupel, and H. Fessi, *Int. J. Pharm.* 28, 125 (1986).
83. F. Leonard, R. K. Kulkarni, G. Brandes, J. Nelson, and J. Cameron, *J. Appl. Polym. Sci.* 10, 259 (1996).
84. C. Damge, H. Vranckx, P. Balschmidt, and P. Couvreur, *J. Pharm. Sci.* 86, 1403 (1997).
85. M. Aboubakar, F. Puisieux, P. Couvreur, and C. Vauthier, *Int. J. Pharm.* 183, 63 (1999).
86. G. Lambert, E. Fattal, H. Pinto-Alphandary, A. Gulik, and P. Couvreur, *Pharm. Res.* 17, 707 (2000).
87. R. Gref, Y. Minamitake, M. T. Peracchia, and R. Langer, in "Microparticulate Systems for the Delivery of Proteins and Vaccines" (S. Cohen and H. Bernstein, Eds.), pp. 279–306. Dekker, New York.
88. M. T. Peracchia, D. Desmaële, P. Couvreur, and J. d'Angelo, *Macromolecules* 30, 846 (1997).
89. M. T. Peracchia, E. Fatal, D. Desmaële, M. Besnard, J. P. Noël, J. M. Gomis, M. Appel, J. d'Angelo, and P. Couvreur, *J. Control. Release* 60, 121 (1999).
90. B. Stella, Silvia, A., M. T. Peracchia, D. Desmaële, J. Hoebeke, M. Renoir, J. d'Angelo, L. Cattel, and P. Couvreur, *J. Pharm. Sci.* 89, 1452 (2000).
91. C. Rouzes, R. Gref, M. Leonard, D. De Sousa, and E. Dellacherie, *J. Biomed. Mater. Res.* 50, 557 (2000).
92. A. Moore, E. Marecos, A. Bogdanov, Jr., and R. Weissleder, *Radiology* 214, 568 (2000).
93. C. Passirani, L. Ferrarini, G. Barratt, J. P. Devissaguet, D. Labarre, *J. Biomater. Sci. Polym. Ed.* 10, 47 (1999).
94. I.-S. Kim, Y.-I. Jeong, and S.-H. Kim, *Int. J. Pharm.* 205, 109 (2000).
95. G. S. Kwon, S. Suwa, M. Yokoyana, T. Okano, Y. Sakuri, and K. Kataoka, *J. Control. Release* 29, 17 (1994).
96. I. Oh, K. Lee, H.-Y. Kwon, Y.-B. Lee, S.-C. Shin, C.-S. Cho, and C.-K. Kim, *Int. J. Pharm.* 181, 107 (1999).
97. C.-S. Cho, S.-W. Kim, and T. Komoto, *Macromol. Rapid. Commun.* 191, 981 (1997).
98. Y. I. Jeong, J. B. Cheon, S.-H. Kim, J. W. Nah, and Y.-M. Lee, *J.Control. Release* 51, 169 (1998).
99. J. Kreuter, *Biomed. Sci. Tech.* 31 (1998).
100. M. Amiji and K. Park, in "Polymers of Biological Significance" (S. W. Shalaby, Y. Ikada, R. Langer, and J. Williams, Eds.), ACS Symposium Series 540. Am. Chem. Soc., Washington, DC, 1994.
101. C. M. Lehr, J. A. Boustra, E. H. Schacht, and J. E. Junginer, *Int. J. Pharm.* 78, 43 (1992).
102. P. Calvo, C. Remunan-Lopez, J. L. Vila-Jato, and M. J. Alonso, *J. Appl. Polym. Sci.* 63, 125 (1997).
103. J. Chang and S. Y. Zhongguo, *Chinese J. Biomed. Eng.* 15, 102 (1996).
104. A. P. Kaplun, L. B. Son, Y. M. Krasnopolsky, and V. I. Shvets, *Vopr. Med. Khim.* 45, 3 (1999).
105. C. Schwarz and W. Mehnert, *J. Microencapsulation* 16, 205 (1999) and references therein.
106. Y. S. Kim and K. S. Kim, *Yakehe Hakhoechi* 28, 249 (1998) [in Korean].
107. W. Mehnert and K. Mäder, *Adv. Drug Deliv. Rev.* 47, 165 (2001).
108. R. H. Müller, K. Mäder, and S. Gohla, *Eur. J. Pharm. Biopharm.* 50, 161 (2000).
109. O. Olbrich, U. Bakowsky, C.-M. Lehr, R. H. Müller, and C. Kneuer, *J. Control. Release* 77, 345 (2001).
110. J. H. Lee, J. Kopecek, and J. D. Andrade, *J. Biomed. Mater. Res.* 23, 351 (1989).
111. I.-S. Kim and S.-H. Kim, *Int. J. Pharm.* 226, 23 (2001).
112. I.-S. Kim, Y.-I. Jeong, C.-S. Cho, and S.-H. Kim, *Int. J. Pharm.* 211, 1 (2000).
113. I.-S. Kim, Y.-I. Jeong, and S.-H. Kim, *Int. J. Pharm.* 205, 109 (2000).
114. I.-S. Kim, S.-H. Kim, and C.-S. Cho, *Macromol. Rapid Commun.* 21, 1272 (2000).
115. H. F. M. Cremers, J. P. Lens, L. Seymour, and J. Feijen, *J. Control. Release* 36, 167 (1995).
116. E. Wanger, M. Zenke, M. Cotton, H. Beug, and M. L. Birnstiel, *Proc. Natl. Acad. Sci. USA* 87, 3410 (1990).
117. S. Katayose and K. Kataoka, *Bioconjugate Chem.* 8, 702 (1997).
118. J. Choi, E. Lee, Y.-H Choi, Y.-J. Jeong, and J.-S. Park, *Bioconjugate Chem.* 10, 62 (1999).
119. V. Toncheva, M. A. Wolfert, P. R. Dash, D. Oupicky, K. Ulbrich, L. W. Seymour, and E. H. Schacht, *Biochim. Biophys. Acta* 1380, 354 (1998).
120. S. Asayama, M. Nogawa, Y. Takei, T. Akaike, and A. Maruyama, *Bioconjugate Chem.* 9, 476 (1998).
121. X. H. Zhou, A. L. Kilbanav, and L. Huang, *Biochim. Biophys. Acta* 1065, 8 (1991).
122. W. Wang, L. Tetley, and I. F. Uchegbu, *Langmuir* 16, 7859 (2000).
123. C. Kneuer, M. Sameti, E. G. Haltner, T. Schiestel, H. Schirra, H. Schmidt, and C. M. Lehr, *Int. J. Pharm.* 196, 257 (2000).
124. C. Kneuer, M. Sameti, U. Bakowsky, T. Schielstel, H. Schirra, H. Schmidt, and C. M. Lehr, in BIONA report 14 (A. Wisser and W. Nachtigall, Eds.), Akad. Lit., Mainz; GTBB, Saarbrücken, pp. 64–65, 2000.
125. E. W. Alton and D. M Geddies, *Gene Therapy* 2, 88 (1995).
126. A. Smith, *Nature* 49, 807 (1995).
127. M. Cotton and E. Wagner, *Curr. Opin. Biotechnol*. 4, 705 (1993).
128. A. Singhal and L. Huang, in "Gene Therapeutics: Methods and Applications of Direct Gene Transfer" (J. A. Wolff, Ed.). Birkhauser, Boston, 1994.
129. V. L. Troung-Le, S. M. Walsh, E. Schweibert, H. Q. Mao, W. B. Guggino, J. T. August, and K. W. Leong, *Arch. Biochem. Biophys.* 361, 47 (1999).
130. H. Cohen, R. J. Levy, J. Gao, I. Fishbein, V. Kousaev, S. Sosnowski, S. Slomkowski, and G. Golomg, *Gene Therapy* 7, 1896 (2000).
131. Y. Capan, B. H. Woo, S. Gebrekidan, S. Ahmed, and P. P. DeLuca, *J. Control. Release* 60, 279 (1999).
132. S. Ando, D. Putnam, D. W. Pack, and R. Langer, *J. Pharm. Sci.* 88, 126 (1999).
133. P. L. Felgner, Y. Barenholz, J. P. Behr, S. H. Cheng, P. Cullis, L. Huang, J. A. Jessee, L. W. Seymour, F. Szoka, A. R. Thierry, E. Wagner, and G. Wu, *Human Gene Therapy* 8, 11 (1997).
134. E. Wagner, M. Zenke, M. Cotten, H. Beug, and M. L. Birnstiel, *Proc. Natl. Acad. Sci. USA* 87, 3410 (1990).
135. O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J. P. Behr, *Proc. Natl. Acad. Sci. USA* 92, 7297 (1995).
136. P. van de Wetering, J. Y. Cherng, H. Talsma, D. J. Crommelin, and W. E. Hennink, *J. Control. Release* 53, 145 (1998).
137. J. Haensler and F. C. Szoka, Jr., *Bioconjug. Chem.* 4, 372 (1993).
138. P. Erbacher, S. Zou, T. Bettinger, A. M. Steffan, and J. S. Remy, *Pharm. Res.* 15, 1332 (1998).
139. C. Kneuer, M. Sameti, U. Bakowsky, T. Schiestel, H. Schirra, H. Schmidt, and C. M. Lehr, *Bioconj. Chem.* 11, 926 (2000).

140. A. Maruyama, T. Ishihara, J. S. Kim, S. W. Kim, and T. Akaike, *Bioconj. Chem.* 8, 735 (1997).
141. C. Olbrich, U. Bakowsky, C. M. Lehr, R. H. Müller, and C. Kneuer, *J. Control. Release* 77, 345 (2001).
142. M. Ogris, S. Brunner, S. Schüller, R. Kircheis, and E. Wagner, *Gene Therapy* 6, 595 (1999).
143. P. D. Scholes, A. G. A. Coombes, M. C. Davis, L. Illum, and S. S. Davis, in "Controlled Drug Delivery: Challenges and Strategies" (K. Park, Ed.), pp. 73–106. American Chemical Society, Washington, DC, 1997.
144. J. Kreuter, Nanoparticles, in "Colloidal Drug Delivery Systems" (J. Kreuter, Ed.), pp. 219–342. Dekker, New York, 1994.
145. E. J. Van Hoogdalem, A. G. De Boer, and D. D. Breimer, *Pharm. Therapy* 44, 407 (1989).
146. V. H. L. Lee and A. Yamamoto, *Adv. Drug Deliv. Rev.* 4 171 (1990).
147. A. Yamamoto and S. Muranishi, *Adv. Drug Deliv. Rev.* 28, 275 (1997).
148. T. Fujita, T. Fujita, K. Morikawa, H. Tanaka, O. Iemura, A. Yamamoto, and S. Muranishi, *Int. J. Pharm.* 134, 47 (1996).
149. A. Fix, *J. Control. Release* 6, 151 (1987).
150. V. H. L. Lee, *J. Control. Release* 13, 213 (1990).
151. S. Muranishi, *Ther. Drug Carrier Syst.* 7, 1 (1990).
152. J. Kreuter, *Adv. Drug Deliv. Rev.* 7, 71 (1996).
153. P. Couvreur and F. Puisieux, *Adv. Drug Deliv. Rev.* 10, 141 (1993).
154. P. Couvreur, C. Dubernat, and F. Puisieux, *Eur. J. Pharm. Biopharm.* 41, 2 (1995).
155. I. Morishita, M. Morishita, K. Takayama, Y. Machida, and T. Nagai, *Int. J. Pharm.* 78, 9 (1992).
156. M. Fukunaga, M. M. Miller, and L. J. Deftos, *Horm. Metab. Res.* 23, 166 (1991).
157. A. Matsuzaka, M. Morishita, K. Takayama, and T. Nagai, *Biol. Pharm. Bull.* 18, 1718 (1995).
158. C. Damgé, C. Michel, M. Aprahamian, and P. Couvreur, *Diabetes* 37, 246 (1988).
159. C. Damgé, C. Michel, M. Aprahamian, P. Couvreur, and J. P. Devissaguet, *J. Control. Release* 13, 233 (1990).
160. C. Michel, M. Aprahamian, L. Defontaine, P. Couvreur, and C. Damgé, *J. Pharm. Pharmacol.* 43, 1 (1991).
161. M. Roques, C. Damgé, C. Michel, C. Staedel, G. Cremel, and P. Hubert, *Diabetes* 41, 451 (1992).
162. Y. Kawashima, H. Yamamoto, H. Takeuchi, and Y. Kuno., *Pharm. Dev. Technol.* 5, 77 (2000).
163. H. Takeuchi, H. Yamamoto, T. Niwa, T. Hino, and Y. Kawashima, *Pharm. Res.* 13, 896 (1996).
164. S. Sakuma, N. Suzuki, H. Kikuchi, K. Hiwatari, K. Arikawa, A. Kishida, and M. Akashi, *Int. J. Pharm.* 149, 93 (1997).
165. S. Sakuma, N. Suzuki, H. Kikuchi, K. Hiwatari, K. Arikawa, A. Kishida, and M. Akashi, *Int. J. Pharm.* 158, 69 (1997).
166. S. Sakuma, Y. Ishida, R. Sudo, N. Suzuki, H. Kikuchi, K. Hiwatari, A. Kishida, M. Akashi, and M. Hayashi, *Int. J. Pharm.* 159, 181 (1997).
167. S. Sakuma, R. Sudo, N. Suzuki, H. Kikuchi, M. Akashi, and M. Hayashi, *Int. J. Pharm.* 177, 161 (1999).
168. T. Niwa, H. Takeuchi, T. Hino, M. Nohara, and Y. Kawashima, *Int. J. Pharm.* 121, 45 (1995).
169. J. Kreuter, *Adv. Drug Deliv. Rev.* 47, 65 (2001).
170. P. Couvreur, B. Kante, M. Roland, and P. Speiser, *J. Pharm. Sci.* 12, 1521 (1979).
171. M. Brightman, in Physiology and pharmacology of the blood–brain barrier, in "Handbook of Experimental Pharmacology" (M. W. B. Bradbury, Ed.), Vol. 103, pp. 1–22. Springer, Berlin, 1992.
172. L. Grislain, P. Couvreur, V. Lenaerts, M. Roland, D. Deprez-Decampeneere, and P. Speiser, *Int. J. Pharm.* 15, 335 (1983).
173. P. Couvreur, L. Grislain, V. Lenaerts, F. Brasseur, P. Guiot, and A. Biernacki, in "Polymeric Nanoparticles and Microspheres" (P. Guoit and P. Couvreur, Eds.), pp. 27–93. CRC Press, Boca Raton, FL, 1986.
174. R. Alyautdin, D. Gothier, V. Petrov, D. Kharkevich, and J. Kreuter, *Eur. J. Pharm. Biopharm.* 41, 44 (1995).
175. U. Schroeder and B. A. Sabel, *Nanoparticles Brain Res.* 710, 121 (1996).
176. P. Ramge, J. Kreuter, and B. Lemmer, *Chronobiol. Int.* 17, 767 (1999).
177. A. E. Gulyaev, S. E. Gelperina, I. N. Skidan, A. S. Antropov, G. Y. Kivman, and J. Kreuter, *Pharm. Res.* 16, 1564 (1999).
178. U. Schroeder, P. Sommerfeld, and B. A. Sabel, *Peptides* 19, 777 (1998).
179. U. Schroeder, P. Sommerfeld, S. Ulrich, and B. A. Sabel, *J. Pharm. Sci.* 87, 1305 (1998).
180. G. Borchard, K. L. Audus, F. Shi, and J. Kreuter, *Int. J. Pharm.* 110, 29 (1994).
181. R. N. Alyautdin, A. Reichel, R. Löbenberg, P. Ramge, J. Kreuter, and D. J. Begley, *J. Drug Target.* 9, 209 (2001).
182. P. Ramge, "Untersuchungen zur Überwindung der Blut-Hirn-Schranke mit Hilfe von Nanopartikeln," pp. 93–98. Shaker Verlag, Aachen, 1998.
183. L. Fenart, A. Casanova, B. Dehouck, C. Duhem, S. Slupek, R. Cecchelli, and D. Betbeder, *J. Pharmacol. Exp. Ther.* 291, 1017 (1999).
184. Y. Tabata, Y. Inoue, and Y. Ikada, *Vaccine* 14, 1677 (1996).
185. H. Nakase et al., *J. Pharmacol. Exp. Ther.* 301, 59 (2002).
186. H. Nakase et al., *J. Pharmacol. Exp. Ther.* 292, 15 (2000).
187. A. Lamprecht et al., *J. Pharmacol. Exp. Ther.* 299, 775 (2001).
188. A. Lamprecht, U. Schafer, and C. M. Lehr, *Pharm. Res.* 18, 788 (2001).
189. A. C. Stewart and S. Burnett, *J. Virology* 68, 4669 (1994).

Porphyrin-Based Chemical Sensors

Roberto Paolesse

Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma "Tor Vergata," 00133 Roma, Italy; and Instituto per la Microelettronica e Microsistemi, IMM-CNR, 00133 Rome, Italy

Federica Mandoj, Alessia Marini

Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma "Tor Vergata," 00133 Rome, Italy

Corrado Di Natale

Dipartimento di Ingegneria Elettronica, Università di Roma "Tor Vergata," 00133 Rome, Italy; and Instituto per la Microelettronica e Microsistemi, IMM-CNR, 00133 Rome, Italy

CONTENTS

1. Introduction
2. Gas Phase Optical Sensors
3. Liquid Phase Optical Sensors
4. Electrochemical Sensors
5. Nanogravimetric Sensors
6. Miscellaneous
7. Sensor Arrays
Glossary
References

1. INTRODUCTION

In the last few years there has been a huge demand to monitor different chemical environments, such as urban indoor and outdoor atmospheres, food aromas, etc. Chemical sensors are among the most promising devices to be exploited for these applications, because they have the great advantage of allowing an online measure suitable for remote control [1].

The structure of a generic chemical sensor can be ideally divided in two subunits: the sensing material and the transducer. The sensing material interacts with the chemical species present in the environment by changing some of its physico-chemical properties, while the transducer transforms these variations in a readable signal, generally an electric signal.

The sensor performances in term of sensitivity, reproducibility, and selectivity strictly depend on the properties of the sensing materials. For this reason great effort has been made to develop sensing materials with improved properties [2].

From this point of view, the exploitation of organic compounds as sensing materials is particularly advantageous. The progress made in designing synthetic receptors [3] allows the orientation of the sensor selectivity toward different classes of compounds via modulation of weak interactions occurring between the sensing material and the analytes.

Among the different classes of receptors developed, porphyrins and metalloporphyrins represent one of the most promising, because of the richness of their properties, their stability, and the development of the chemistry of these macrocycles that allow the possibility to modulate their properties by synthetic modifications or by changing the coordinated metal. In these applications porphyrins mimic their functions in biological systems, where they are able, for example, to reversibly bind oxygen. Because generally target analytes are also good ligands for metal ions, porphyrins represent a perfect match for the properties required to sense materials.

For this reason research in this area has experienced significant growth in the last decade and porphyrin based chemical sensors are now going to be as famous as those based on related phthalocyanines. In this chapter we want to highlight their exploitation in sensors based on different transduction mechanisms and devoted to detecting analytes in both liquid and gaseous phase.

2. GAS PHASE OPTICAL SENSORS

Optical sensors are based on the variation of one optical property, such as absorption or luminescence, of the sensing material as a consequence of the interaction with the target

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

analyte. A general setup of an optical sensor is reported in Figure 1. Porphyrins have been called the “pigment of life” and for these peculiar optical properties it is quite obvious that they have been considered as sensing materials for optical sensors. The first examples reported in the literature were related to gas detection and in particular to the development of oxygen sensors.

2.1. Oxygen Sensors

These devices are based on the quenching of the phosphorescence of some metalloporphyrins due to their interaction with molecular oxygen. The dependence of phosphorescence quenching to the oxygen concentration is described by the Stern–Volmer relationship

$$I_o/I = 1 + K_{sv}(PO_2)$$

where I_o represents the native phosphorescence of the lumophore in the absence of oxygen. K_{sv} is the Stern–Volmer constant and represents a useful parameter to compare the sensor performances. K_{sv} is obtained experimentally from the Stern–Volmer equation taking the partial pressure of oxygen at which the initial sensor phosphorescence is decreased by 50% ($I_o/I = 2$). In Table 1 we report a list of the K_{sv} values of the some porphyrin based oxygen sensors.

The most widely used metalloporphyrins for these applications are Pt and Pd derivatives, due to their high phosphorescence quantum yields and short lifetimes. The detection limit of these porphyrin based sensors can be lower than 0.1 Torr of oxygen partial pressure and for this reason they can be exploited for vacuum control, for determination of surface pressure in wind-tunnel tests [4], or for determination of the metabolic activity of cell cultures [5].

In chemical sensors the sensing material should be deposited as thin films onto the transducer surface. In these devices porphyrins are generally deposited dispersed into a supporting matrix. Although is not directly involved in the sensing mechanism, the matrix is of critical importance for sensor performance, because its oxygen permeability determines the detection limit of the resulting sensor.

Another fundamental feature to be considered in optical devices is the photostability of the sensing material.

To obtain stable and reliable devices it is necessary to improve the photostability of the lumophore and to obtain a perfect match between matrix and lumophore properties and for this reason a large part of the research in this area has been devoted to improving sensor performances by increasing porphyrin stabilities and by tuning the matrix characteristics.

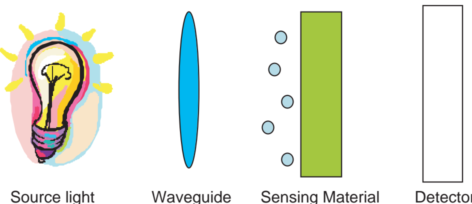


Figure 1. General setup of an optical sensor.

Table 1. Sensing properties of porphyrin-based luminescent oxygen sensors.

Porphyrin	PO_2^a (Torr)	Support	Ref.
PtOEP	53.9	PS	[7]
	47.6	PS	[28]
	0.48	PS	[6]
	73	SR	[10]
	0.18	TMSP	[14]
	4.3	PS-PFS	[15]
	2.5	PS-TFEM	[17]
	2.6	IBM-TFEM	[17]
	166.7	IBM	[17]
	10	PS	[16]
	11	PS	[17]
	12.5	glass	[22]
	1.8	PEGMA	[30]
	33.3	PVC	[16]
	3.85	PS-PFS	[15]
	13.3	PDS	[16]
	1.2	IBM-TFPM	[17]
	28.6	PS	[18]
	12.9	CAB	[13]
	0.943	CAB	[13]
	204	PMMA	[13]
	2.82	PMMA	[13]
PdOEP	0.56	PS-PFS	[15]
	0.76	PS	[16]
	0.18	PEGMA	[30]
	0.55	Sty-PFS	[15]
	1.51	CAB	[13]
	0.253	CAB	[13]
	8.0	PMMA	[13]
	0.238	PMMA	[13]
PtOEPK	49.2	PS	[7]
	56.9	PS	[8]
	32	PS	[13]
PdOEPK	5.6	PS	[13]
PdCP	3.57	SR	[13]
	7.2	PS	[13]
	27.1	PMMA	[13]
PdTPP	2.58	LB (arachidic acid)	[13]
PtTDCPP	8.0	SR	[10]
PtTFMPP	3.7	SR	[10]
$PtBr_8TPP$	6.4	SR	[10]
PtTFPP	11.1	PS	[16]
	55.6	PS	[28]
PtTCPP	0.61	alumina	[24]
PdTCPP	0.22	alumina	[24]
PdalkTCPP	0.33	alumina	[25]
Poly-PtTPP	1.54	ITO	[26]
Poly-PdTPP	0.83	ITO	[26]
ZnTFPP	1.64	PS	[29]

[a] $PO_2 = 1/K_{sv}$.

The first metalloporphyrins exploited in these sensors were PtOEP (see Table 2 for abbreviations) and PdOEP (Fig. 2) [6, 7], which showed significant photobleaching under the operative conditions [6]. From this point of view one of the approaches has been to modify the porphyrin by introducing different substituents, in order to increase its stability under operative conditions.

Table 2. List of abbrevations.

OEP	2,3,7,8,12,13,17,18-octaethylporphyrin
OEPK	2,3,7,8,12,13,17,17-octaethyl-18-oxo-porphyrin ketone
OEPDK	3,3,7,8,13,13,17,17-octaethyl-2,12-dioxo-porphyrin diketone
TPP	5,10,15,20-tetraphenylporphyrin
Br_8TMP	5,10,15,20-tetramesityl- β-octabromo-porphyrin
PVC	poly(vinyl chloride)
PS	poly(styrene)
CAB	cellulose acetate butyrate
PMMA	poly(methyl methacrylate)
TMSP	trimethylsilyl-1-propine
EC	ethyl cellulose
TCPP	5,10,15,20-tetrakis(4-carboxyphenyl)porphyrin
TFPP	5,10,15,20-tetrakis(pentafluorophenyl)porphyrin
CP	coproporphyrin
TEHOPP	5,10,15,20-tetrakis(3,4-bis[2-ethylhexyloxy] phenyl)porphyrin
TDSAPP	5,10,15,20-tetrakis(4-*N*-dodecyl-sulfonylamidophenyl) porphyrin
TSPP	5,10,15,20-tetrakis(4-stereamidophenyl)porphyrin
PBMA	poly(buthyl methacrylate)
PVP	poly(vinyl pyrrolidone)
TMPyP	5,10,15,20-tetrakis(4-*N*-methylpyridyl)porphyrin
DTMABP	2,3,7,8,12,13,17,18-octamethyl-5,15-bis-(4-trimethylammonium-yl-phenyl)porphyrin
TSPP	5,10,15,20-tetrakis(4-sulfonatophenyl)porphyrin
T(4-APP)	5,10,15,20-tetrakis(4-aminophenyl)porphyrin
PU	poly(urethane)
PP-IX	protoporphyrin IX
T(2-APP)	5,10,15,20-tetrakis(2-aminophenyl)porphyrin
TETMAPP	5,15-bis(2-aminophenyl)-2,8,12,18-tetraethyl-3,7,13,17-tetramethylporphyrin
TMHPP	5,10,15,20-tetrakis(3-methoxy-4-hydroxyphenyl)porphyrin
TMPP	5,10,15,20-tetrakis(4-methoxylphenyl)porphyrin
SR	silicon rubber
PFS	poly(styrene-co-pentafluorostyrene)
PDS	poly(dimethylsyloxane)
TFEM	trifluoroethylmethacrylate
IBM	isobutylmethacrylate
PEGMA	poly(ethylene glycol methacrylate)
ITO	indium tin oxide
o-NPOE	2-nitrophenyl octyl ether
TDMAC	tridodecylmethylammoniumchloride
KTFPB	potassium tetrakis[bis(3,5-trifluoromethyl)phenyl]borate
DBS	dibutyl sebacate
AAO	ascorbic acid oxidase
OA	oleic acid
AP	acetophenone
DOP	dioctyl phthalate
NaTPB	sodium tetraphenylborate

Oxochlorin, generally called porphyrin ketone (OEPK) (Fig. 3), has been proposed by Papkovsky and co-workers [8]; the dye showed higher photochemical stability to respect OEP, conserving the optical properties useful for phosphorescence sensing, with only shortened lifetime values. Following this approach the same author proposed also dioxobacteriochlorin, or octaethylporphyrin-diketone (OEPDK) (Fig. 4), as sensing material, with the hope that the introduction of a second oxo group at the peripheral positions resulted in a further increase of the photostability of the resulting porphyrin [9]. This result was effectively achieved, but, on the other hand, Pd and Pt complexes of these porphyrins showed weak phosphorescence and reduced quantum yields, making them unusable for oxygen sensing.

Following a similar approach, β-halogenated TPPs have been proposed as sensing materials [10]. These porphyrins have been widely used as oxidation catalysts, where they were shown to be more resistant to degradation than TPP. The same effect was obtained for photostability and PtBr_8TMP was indeed more stable than PtTPP under operative conditions.

On the other hand, significant effort has been devoted to improving the characteristics of the supporting matrix, to improve both stability and sensitivity of the resulting sensor. The matrix can be either organic or inorganic, but the most common materials used to encapsulate porphyrins are organic polymers. A general requirement for these polymers is to be oxygen permeable with a low diffusion barrier for the gas. The permeability to oxygen of course strongly influences the characteristic of the resulting sensor.

A careful choice of polymer and plasticizer is necessary to obtain reliable sensors. Of course polymers or additives should not interfere with the sensing mechanism of the porphyrin; for example the widely used plasticizer 2-nitrophenyl octyl ether cannot be exploited in luminescence based sensors, because it acts as a quencher of the porphyrin luminescence, interfering in the sensing mechanism [11].

Plasticizers are necessary to increase the oxygen permeability of the organic polymers and for this reason they strongly influence the working range of the resulting sensor. Detailed studies of the plasticizer influence on the sensor characteristics have been carried out for Pd and Pt complexes of OEP and OEPK [12]. These studies showed that addition of plasticizers to PVC, PS, CAB and PMMA polymers allows the modulation of the O_2 pressure working range of the sensor, by effectively increasing the oxygen permeability of the polymer matrix. On the other hand, addition

Figure 2. Molecular structure of OEP.

Figure 3. Molecular structure of OEPK.

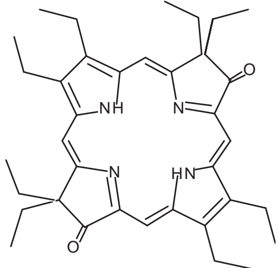


Figure 4. Molecular structure of OEPDK.

of plasticizers can induce a deviation from the linearity of the Stern–Volmer relationship at higher oxygen concentration. This feature has been attributed to the presence of two oxygen accessible sites in the polymeric thin films; one is easily accessible and the other one is an oxygen site difficult to access. The different oxygen sites where the porphyrin can be present result in different phosphorescence quenching constants and consequently in the nonlinear Stern–Volmer plot [13]. This effect is not favorable, because makes calibration of the sensors more difficult.

An increase of the oxygen permeability can resolve this problem and for this reason novel polymeric matrices have been tested in recent years to obtain sensors with improved performances [14–18]. Poly(TMSP) demonstrated high oxygen permeability, and the detection limit for PtOEP dispersed in poly(TMSP) film was less than 0.3% [14]. Furthermore the Stern–Volmer plot was linear in the oxygen low concentration range and PtOEP-poly(TMSP) based sensors were promising for trace oxygen detection. PtOEP-poly(TMSP) films showed a 3% decrease after 12 h of continuous irradiation, indicating a promising photostability, although this aspect was not completely characterized [14].

Another promising approach was related to the exploitation of fluoropolymers, because the presence of fluorine increases the oxygen permeability of the resulting matrix. Furthermore these polymers are also stable toward photooxidation, due the electron-withdrawing character of the fluorine. Fluoropolymers are obtained by co-polymerization of a fluorinated monomer with a standard precursor, such as styrene [15] or acrylate [16–19], in different compositions. The high oxygen permeability of these fluoropolymers resulted in a good linearity of the Stern–Volmer plots for both PtOEP and PdOEP based sensors. For PdOEP sensors the phosporescence of the sensing material was completely quenched for 20% oxygen concentration. Furthermore the photostability of these sensing films was enhanced and a maximum of 3% decrease of the starting intensity was noted after 24 h of continuous irradiation [16–18].

Nafion® has also been used as supporting matrix for Pt, Pd, and Rh complexes of water-soluble porphyrins [19]. The resulting sensing films showed linear Stern–Volmer plots, and promising sensitivity and photostability for practical applications, considering also the simple procedure for the immobilization of the lumophores into the supporting matrix.

Recently Douglas and Eaton [20] studied the response characteristics of Pd and Pt complexes of OEP encapsulated in different polymeric matrices, showing that it is possible to develop luminescent oxygen sensors with a very broad concentration working range, by combination of polymers with different oxygen permeability and Pd and Pt OEP, having different lifetimes. Deviations from the linearity of the Stern–Volmer plots have been modeled by a Freundlich-like isotherm, which can be used to obtain simple and precise calibration data. The heterogeneity inducing nonlinear Stern–Volmer behavior was attributed to a distribution of oxygen permeability present in the polymers and not to distribution of different lumophore lifetime. The same authors showed that ambient humidity strongly influenced the sensing behavior of PtOEP based luminescent sensors when a hydrophilic polymer, such as EC or CAB, was used as supporting matrix, while no significant effects were observed for hydrophobic matrices, such as silicone rubber, PVC, or PS [21]. Luminescence decay kinetics studies in the presence of oxygen showed that the influence of humidity is in the change of gas permeability, rather than modification of the porphyrin distribution sites in the polymer. Addition of plasticizers strongly reduced the humidity effects on the sensing behavior.

Although polymers are the most widely used matrices, it is also possible to disperse lumophore into inorganic supports, such as glasses. The advantages of glasses can be related to their transparency, chemical stability, and photostability. The sol–gel technique is a simple way to disperse an organic dye into a glass matrix. Lee and Okura prepared PtOEP doped silica gel glasses by basic hydrolysis of tetraethylethyl orthosilicate [22]. The addition of a surfactant, Triton X-100, was necessary to give homogeneity to the resulting monoliths. Optical properties of PtOEP in silica glass were similar to those observed in acetone solution, indicating that the silica matrix does not interfere with the optical sensing of PtOEP. Sol–gel films were dried at room temperature, at 150 and 180 °C. The best sensitivity was obtained at 150 °C, although good results were obtained at room temperature, showing that the sintering step is not necessary. Higher drying temperature resulted in a marked decrease of the sensor performance, attributed by the authors to some variation of the sol–gel microenvironment, although decomposition of the dye cannot be excluded. Stern–Volmer plots showed deviations from the linearity, evidencing the presence of different oxygen accessible sites in the matrix. High response time and reversibility were observed for such sensors, and these superior characteristics were attributed to the microporous texture of the silica matrix. The stability of the PtOEP doped glasses was optimal in the dark, while under irradiation the results obtained were similar to those observed for polymeric matrices.

Amao et al. explored also the possibility to avoid the need for a supporting matrix, depositing a monolayer of the sensing dye by a chemisorption technique [23, 24]. Pt and Pd complexes of TCPP (Fig. 5) were deposited onto alumina plates using the carboxy functional groups to directly attach the lumophore to the inorganic surface. Excitation spectra were similar to those observed in ethanol, an indication of the absence of aggregation between macrocycles in the chemisorbed film. Phosphorescence of the films was effectively quenched by oxygen; Stern–Volmer plots were linear at low oxygen concentration, while significant nonlinearity

Figure 5. Molecular structure of TCPP.

was observed for higher concentrations. A good photostability was also observed, indicating that these films are suitable for oxygen sensing devices. To improve the sensitivity of such a device, the same authors deposited monolayers of a modified Pd porphyrin bearing a peripheral alkyl chain substituent (Fig. 6) [25].

PtTPP and PdOEP thin films were also deposited by a electropolymerization technique [26]. Indium tin oxide glass slides were used as substrates and PtTPP polymeric films were deposited by a CV technique. In the case of PdTPP it was necessary to add 2,6-di*tert*-butylpyridine as catalyst for the electropolymerization process. The sensitivities of the developed devices were comparable to similar oxygen sensors, but the long-term photostability was greatly improved.

The electropolymerization deposition technique is particularly promising for the fabrication of miniaturized devices, because the sensing film can be directly deposited at the end of the optical fiber.

The measurement of luminescence intensity is the most exploited technique, because it is a simple technique. However, it can be affected by different parameters, the most important being photobleaching or leaching of the dye and the intrinsic sample fluorescence. These problems can be overcome by measuring the luminescence decay time of the sensing dye, because this measure is independent of the emitted light intensity and it is compatible with fiber optics exploitation. In this case the Stern–Volmer equation is related to the lumophore lifetime:

$$\tau_o/\tau = 1 + K_{sv} pO_2$$

Following this approach, different examples of oxygen sensor based on phosphorescence lifetime decay have been reported [27–29]. For example Trettnak and co-workers reported for example an oxygen sensor comprised of simple and low-cost optoelectronic components, such as light emitting diodes (LEDs) and photodetectors [27], working in both liquid and gaseous phase. The exploited lumophore was the PtOEPK supported in a polystyrene matrix.

Amao et al. reported an oxygen sensor based on the triplet–triplet reflectance quenching using laser flash photolysis [28, 29]. PtOEP and PtTFPP immobilized on PS matrices were used as sensing dyes and the authors obtained the same results by using diffuse reflectance laser flash photolysis and the classical luminescence lifetime measurements by laser excitation [28]. Diffuse reflectance laser flash photolysis can also allow the use of nonphosphorescent compounds as oxygen sensing dye, and this has been successfully demonstrated by the exploitation of different Zn porphyrins immobilized in various polymeric matrices [29].

More recently Douglas end Eaton pointed out that careful attention should be paid for the Stern–Volmer analysis of data using conventional phosphorimeters or instruments incorporating pulsed xenon lamps [30], because of systematic errors in quantitative emission intensity measurements due to the fact that the total emission of the lumophore is not completely within the instrument gate time [31].

2.2. Nitrogen Oxides Sensors

Oxygen sensors based on phosphorescence quenching generally do not suffer interference due to the presence of other gases, with the only exception being nitrogen oxides, which are powerful luminescence quenchers. For this reason, Papkovski and co-workers explored the potential exploitation of PtOEP immobilized on PS matrices and PdCP-III Langmuir–Blodgett films as NO_x optical sensors. For both films NO_x acted as a strong luminescence quencher, but in this case a fast chemical interaction of the dye with NO_x and consequent film degradation ruled out their exploitation in the sensor field [32].

More recently the exploitation of Langmuir–Blodgett (LB) films of substituted porphyrins as optical sensors for NO_2 detection has been deeply studied [33–37]. The sensing mechanism is based on the variations of the optical absorption spectra of LB films of TEHOPP, bearing the 2-ethylhexyloxyphenyl substituents at the peripheral positions (Fig. 7). The deposition of LB films was performed at a higher rate than the conventional method; this resulted in an inhomogeneous morphology of the resulting film, constituted by grain assemblies of porphyrin, and incomplete coverage of the sensor substrate. This structure, however, is not detrimental for sensor applications, because it led to porous films, which allows easy diffusion of the gas in the bulk of the sensing material.

The change in the optical spectrum consisted of a redshift of the Soret band with the appearance of a new band

Figure 6. Molecular structure of PdATTP.

Figure 7. Molecular structure of TEHOPP.

around 700 nm, which is consistent with a protonation of the macrocycle core. The optical is highly sensitive to low concentration of NO_2, reaching the ppm level. The 60% of relative absorbance change (at 434 nm) was obtained for 4.4 ppm of NO_2 [33, 34], with the original spectrum fully recovered after exposure to NO_2, by fluxing with N_2, indicating a reversible interaction. The optical response is also sensitive to the temperature and decrease with the increase of the temperature. This effect was used to increase the recovery rate after the measure, by heating the substrate at 350 °C for a few seconds [33, 34]. The adsorption isotherm followed the Langmuir model.

The sensing behavior of two asymmetrically substituted porphyrins, obtained by substitution of two peripheral 2-ethylhexyloxyphenyl substituents by 4-nitrophenyl or 4-aminophenyl groups, toward NO_2 detection was also studied [35]. While both porphyrins showed in the presence of NO_2 the same change in the optical spectra recorded in solution, similar to those observed for the symmetrical porphyrin, the LB films of these porphyrins showed a different behavior, and only the 4-aminophenyl substituted porphyrin was sensitive to NO_2. This different behavior was ascribed to a different arrangement of the porphyrin rings in the LB films, with the 4-nitrophenyl substituted porphyrins having a molecular orientation that did not allow a diffusion of the gas and consequently the interaction with the macrocycles [35].

The dependence of the sensing performances to the LB film thickness was investigated for multilayer films of TEHOPP ranging from 1 to 20 excursions. Comparison of the response time taken at 50% and 90% of the overall response allowed the interpretation of the sensing mechanism in terms of surface and bulk interaction of the LB film with the gas. The response time reached a minimum value for films obtained by 5–10 excursions; after that the response magnitude of the sensing layer can increase, but the effective sensitivity of the film can decrease [36].

2.3. Other Gas Sensors

Porphyrin based optical sensors are not only limited to oxygen or nitrogen oxide detection. Similar devices have been tested, for example, for the detection of chlorine [38–40]. The first approach was based on the fluorescence changes of TPP encapsulated on silicone rubber, induced by exposure to chlorine [38]. The change observed can be attributed also in this case to the formation of the porphyrin dication; due to the insolubility of the porphyrin dication into the silicone matrix, the measurement should be carried in a kinetic mode and a faster reversibility can be obtained by temperature control, by heating the film after each measurement. This optical sensor cannot discriminate between chlorine, hydrogen chloride, or nitrogen dioxide, because all of these gases induce the protonation of the sensing film, which is the sensing mechanism of this device.

A similar protonation process was observed after exposure to gaseous chlorine by recording optical absorption spectra of monolayers of TDSAPP deposited onto glass plates [39]. In this case the influence of porphyrin orientation on the sensor behavior was observed, with a faster response for monolayer deposited at lower surface pressure, where porphyrins are assumed to be oriented flat on the surface and consequently more accessible to interaction with gaseous chlorine. Also in this case heating the substrate after the measurement reduced the recovery time of the device. Similar behavior was observed for monolayers of TSPP [40].

Changes in the optical absorption spectra were exploited in a similar fashion for HCl detection [41–43]. TPP and substituted analogs were encapsulated in different polymeric matrices and changes in the reflectance spectra induced by protonation of the macrocyclic cores were used for the detection of gaseous HCl. An increase of the sensor responses was observed when electron-releasing substitutents, such as OH groups, were introduced at the peripheral phenyl groups of the porphyrin rings, reaching a detection limit in the sub-ppm range [42]. However, this synthetic modification also induced longer response and recovery times and to overcome this problem, different alkoxy substituted TPPs encapsulated in various polymeric matrices were tested as sensing elements [43]. The results obtained showed that saturation of the sensors was reached for 4 ppm of HCl concentration and that the sensor behavior was mostly controlled by the polymeric matrices and not by the length of the peripheral alkoxy chain. Best results were obtained by using PBMA as supporting matrix [43].

Leray and co-workers used water-soluble porphyrins (Fig. 8) encapsulated in PVP or PMMA matrices for benzene detection [44]. These porphyrins are know to interact with benzene in aqueous solution; the changes induced in the optical spectra by interaction with benzene were similar in solution and in the polymeric matrices and consisted in a hypochromic effect. Best results were achieved with DTMABP and PVP as matrix, although the sensitivity and response time were not sufficient for exploitation in urban air monitoring, although the sensitivity was comparable to other devices reported in the literature [44].

Figure 8. Molecular structure of TMPyP and DTMABP.

The same author proposed sol–gel grafted Co porphyrins for amine detection [45]. Grafting of the porphyrin to the sol–gel matrix was achieved by introducing triethoxysilyl groups at the peripheral positions of the porphyrin ring (Fig. 9). Grafting the porphyrin resulted in the oxidation of the coordinated cobalt atom to the +3 oxidation state. Diffusion of amines was possible in solution and in the gaseous phase and resulted in optical changes characteristics of axial coordination to the metal center. The same interaction was not possible in the case of Co(II)TMPyP, which was insensitive to the presence of amines; this result was attributed to a strong interaction of the cationic porphyrin to the anionic centers of the inorganic matrix. Response time of the developed sensor was very long, due to a slower diffusion rate than those observed in polymeric matrices [45].

Figure 9. Molecular structure of silyl modified TPP.

Very recently Qin and co-workers proposed an optical sensor for the detection of amines in the gaseous phase, based on In(OEP)OH as sensing material [46]. The sensing mechanism operating in such a device is similar to that observed in potentiometric sensors later described (see Section 4) and it is based on a monomer–dimer equilibrium of In(OEP)OH in polymeric films. In the presence of lipophilic anion, in fact, this equilibrium can be expressed as

$$2\,\mathrm{In(OEP)^+} + \mathrm{H_2O} + \mathrm{R^-} \rightleftharpoons [\mathrm{In(OEP)_2OH}]\mathrm{R} + \mathrm{H^+}$$

Amines can break the dimer by coordination to the metal center:

$$[\mathrm{In(OEP)_2OH}]\mathrm{R} \rightleftharpoons \mathrm{In(OEP)OH} + [\mathrm{In(OEP)RNH_2}]\mathrm{R}$$

This reaction can be easily monitored by optical absorption spectra, because the blueshifted Soret band of the dimer ($\lambda = 390$ nm) disappeared with the concomitant formation of the absorption of the $\mathrm{In(OEP)^+}$ ($\lambda = 406$ nm). The presence of water is necessary for the formation of the dimer and all the measurements were carried at 80% of relative humidity. In these conditions the sensor response to the different amines depends on the relative partition coefficient of the amine into the polymeric films and to their ability as coordinating ligand for the InOEP. In the best situation the detection limit can reach 0.1 ppm in the case of butylamine, while for the less lipophilic ammonia it is in the range of 10 ppm [46].

3. LIQUID PHASE OPTICAL SENSORS

3.1. Heavy Metal Ions

Porphyrin based optical sensors have been developed to detect Hg and other heavy metals in solution [47–51]. These sensors are based on the well-known ability of porphyrins to coordinate a wide range of different metals in the inner core. Coordination of the metal induces changes in both the absorption and emission spectra that can be used for the sensing mechanism.

Plaschke and co-workers immobilized water soluble TSPP into silica glasses by the sol–gel technique [47]. Interaction with Hg(II) ion was monitored by following the emission of Hg porphyrin complex at 612 nm. The acid or basic catalysis for the preparation of the sol–gel gave films with different characteristics. The acidic prepared films were more stable in terms of dye leaching but were insensitive to the presence of Hg ions, while the basic catalyzed films gave the opposite behavior. This result was attributed to the different morphologies of the films, with the acidic prepared films more packed and less permeable to the diffusion of the Hg ions. For this reason a dextran covalently linked porphyrin was prepared and used in the basic sol–gel process; in this case a better stability of the resulting film was observed. An ideal detection limit of 2 μg/L of Hg(II) was obtained for the basic catalyzed films, although the instability of the film

toward dye leaching prevented the exploitation for sensing applications [47].

To overcome the problem of the sensing element leaching, the same group developed a polymeric matrix with a covalently bound porphyrin [48]. The polymeric matrix was a hydroxyethyl substituted PMMA, deposited onto a PMMA disk. The change of absorption spectra after metal coordination was exploited to develop the optical sensor. Cd(II) and Hg(II) ions were the target analytes and a detection limits of 63 and 20 μg/L were obtained respectively for these ions, with a response time of 10 min. The different ions can be easily discriminated because they have different absorption bands and for this reason a simultaneous determination is possible. Due to the covalent binding the sensor was stable over 6 weeks of working period, while a simple dispersed porphyrin was completely leached from the polymeric membrane within some days.

TMpyP was adsorbed in sol–gel matrix or grafted onto a glass surface to develop optical sensors for the detection of Hg(II), Pb(II), and Cd(II) ions, by following changes both in the absorption or in the emission spectra [49]. In sol–gel the detection limit observed was 10^{-3} mol/L, but the porphyrin–metal ion interaction was strongly limited by the ion entrapping action of the silica matrix. When the porphyrin was grafted onto glass plates by covalent bonding (Fig. 10), shorter response time and higher sensitivity were obtained to respect the bulk sol–gel material.

An optical sensor for the detection of Hg(II) ions has been developed by encapsulating TPP into a PVC membrane [50]. Detection of Hg was carried out by following the fluorescence quenching of the porphyrin operated by coordination of the metal ion. More recently an increase in the sensor performances has been obtained by using a porphyrin dimer as sensing element [51]. The presence of two porphyrin rings in the sensing material allowed a higher sensitivity and selectivity for the Hg(II) ion in the pH range of 2.4–8. A detection limit of 5.2×10^{-7} mol/L with a response time of 9 min was obtained. The presence of a wide range of different ions, including Cd and Pb, did not affect the Hg detection, with a relative error less than 5%, acceptable for practical applications.

Figure 10. Molecular structure of TPyP grafted on silica surface.

3.2. Anions

Metalloporphyrins have axial sites available for interactions with different analytes. This approach has been exploited for anion detection, by following optical changes due to the interaction of metalloporphyrins with the target analytes.

Solid-state phosphorescent probes for the detection of sulphite have been developed by immobilization of a Pt(II)CP-I covalent conjugate with bovine serum albumin on a preactivated biodyne ABC membrane [52]. A flow-injection system for the determination of sulphite was developed by placing the membrane in a flow cell and using a fiber optic phosphorescent detector. Quenching of the phosphorescence was observed at acidic pH and was attributed to the formation of SO_2. The detection limit reported for this system was 10^{-5} mol/L.

Changes of the optical absorption spectra of metal complexes of T(4-APP) or TMPyP immobilized in thin films of Nafion® were exploited for the detection of different anions, such as iodide and thiocyanate [53]. The selectivity and sensitivity of such a films depended on both the coordinated metal ion and the porphyrin used. Co(II)TMPyP gave fast and reversible responses, with thiocyanate, showing also a good long-term stability.

Polymeric membranes doped with In(III) porphyrin complexes and a lipophilic dichlorofluorescein (DCF) derivative were studied for the development of anion optical sensors [54]. With InOEPCl DCF acted as axial ligand, and interaction of the chromophores induced modification of the absorption spectrum of the diad. Anions competed with DCF as axial ligand and this resulted in an increase of the Soret band absorption, which can be used for optical sensing. In the case of a In(III) complex of a picket-fence porphyrin, DCF was not coordinated, as evidenced the absence of modification of the absorption spectrum. This complex was exploited for the development of acetate anion or acetic acid optical sensors.

More recently, Zhang and co-workers have developed an optical chloride sensor based on the previously described monomer–dimer equilibrium of In(OEP)OH in plasticized PVC films [55]. In the absence of chloride anion, the blueshifted Soret band of the In porphyrin dimer was observed. Chloride acts as axial ligand for In porphyrin and increasing the amount of chloride induced the breaking of the dimer with the formation of the monomeric complex. This resulted in an increase of the redshifted Soret band of the monomeric species. A good selectivity toward a wide range of different anions was observed, although the interference of salicylate was observed.

3.3. Other Analytes

PVC membranes doped with TPP and MnTPPCl were used to develop an optical fiber optic sensor for the determination of berberine, an anti-inflammatory drug for the heart [56]. The sensing mechanism was based on the fluorescence quenching of TPP-MnTPPCl operated by berberine. The sensor operated in the 7.5×10^{-7} to 5.6×10^{-4} mol/L working range of berberine concentration, with fast response time and fair selectivity toward common interfering species. The exploitation of the TPP-MnTPPCl mixture gave better results than the use of the simple TPP.

4. ELECTROCHEMICAL SENSORS

4.1. Potentiometric Sensors

Among the different examples of porphyrin based chemical sensors, their exploitation as ionophores in the development of ion selective electrodes (ISE) is probably one of the most popular. These electrodes (Fig. 11) are generally comprised of a polymeric membrane, where the sensing materials are dispersed, an internal filling solution, and an electrode for the electrical connection to the external circuit. Such sensors respond to the activity of target ions by an intrinsic ion-exchange mechanism between the organic polymeric phase and the aqueous phases. In Table 3 we report the detection limits of some porphyrin based ISEs reported in the literature. The basic theory describing the working mechanism of such a sensor has been developed [58], but the contribution of the different constituents (e.g. ionophore, additives, etc.) is still a matter of further researches. The membrane, usually comprised of a PVC disk, has ion-exchange properties that depend on the ionic sites present in the polymeric structure. The sensor responds to the activities of the ions present in the sample and the selectivity depends on the relative solubility of the ions into the polymeric membrane solvent (plasticizer). It is the ionophore that drives the selectivity of the sensor toward the target analyte. Under certain conditions [58], the response of the ISE to the target analyte concentration follows the Nernst equation:

$$E = E^o + RT/zF \ \ln(a_i)$$

The developed coordination chemistry of metalloporphyrins has attracted interest for their exploitation in the development of anion-selective sensors. In the absence of metalloporphyrins, there are no specific interactions and the selectivity of ISE depends on the anion lipophilicities and results in the so-called Hofmeister series. The presence of metalloporphyrin in the membrane induces deviations on this series, which depend in large part on the coordination of the anion to the metalloporphyrin. After the first report by Chaniotakis co-workers [59], the importance of coordination interactions has been demonstrated, in terms of both the nature of the central metal and the substituents present in the porphyrin ring [60, 61]. Metalloporphyrins can act as neutral or charged ion carriers, depending on the oxidation state of the coordinated metals. While coordinated metals in a +2 oxidation state lead to neutral carriers, and metals in a +4 oxidation state lead to charged carriers, with metals in a +3 oxidation state can give either neutral or charged carriers, depending on both the coordination number of the metal and the nature of the incoming axial ligand (neutral or anionic) [62]. The performances of metalloporphyrin based ISEs depend also on the addition on lipophilic ionic additives, such as triphenylborate or trialkylammonium salts, which improve the potentiometric behavior of the resulting sensors, in terms of selectivity and long-term stability. The nature of the positive effect of these additives has been studied in detail and both the nature and the amount of the ionic additive to be added for the membrane compositions can be rationally predicted [63, 64].

Furthermore, the origin of the so-called “super-Nernstian” behavior reported for several metalloporphyrin-based ISEs has recently been clarified [65]. These non-Nernstian responses are characteristic of anion-selective electrodes based on Mn(III), Sn(IV), In(III), and Ga(III) porphyrins and are derived from a monomer–dimer equilibrium occurring in the PVC membrane. The occurrence of this equilibrium and the degree of dimerization have been demonstrated by ultravoilet-visible spectroscopy, following the blueshifted Soret band characteristic of the hydroxy-bridged dimer. Furthermore In and Ga complexes of picket-fence porphyrins, which do not dimerize, showed slightly sub-Nernstian slopes. A theoretical model based on the dimerization equilibrium was shown to be able to predict the super-Nernstian behavior of the resulting ISEs.

Different metalloporphyrins have been proposed in recent years for the development of ISEs devoted to the detection of various analytes [62, 66–69]. A free base TPP modified with appended urea group (Fig. 12) has been proposed as ionophore to develop an ISE for acetate and it was demonstrated to be able to measure the concentration of acetic acid in vinegar samples [70].

Among the shortcomings observed with these metalloporphyrin-PVC based ISEs, one of the most important is their reduced long-term stability because of the ionophore leaching from the membrane. To avoid this problem, a covalent binding of the porphyrin to the polymeric matrix has been proposed [71, 72]. This immobilization effectively overcame the problem of ionophore leaching but also contributed to reducing the interference of lipophilic anion, such as salicylate, because of size exclusion from the polymeric matrix. This steric effect also prevented the super-Nernstian behavior due to metalloporphyrin dimerization and the resulting Mn(III) porphyrin showed a Nernstian behavior to iodide.

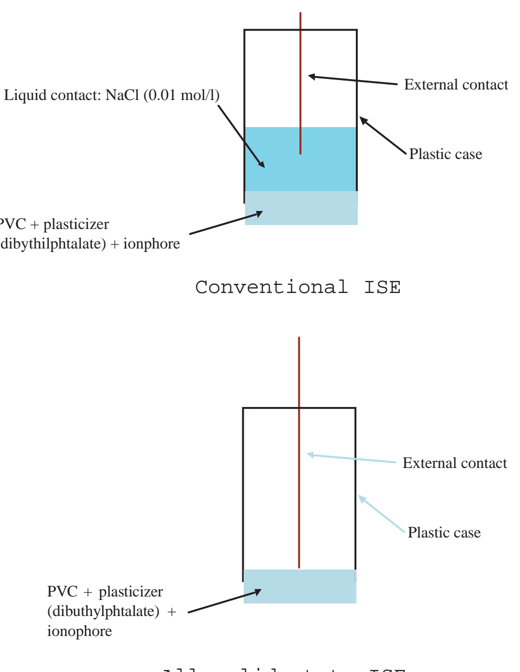


Figure 11. General structure of ISEs.

Table 3. Sensing properties of porphyrin-based ISEs.

Porphyrin	Analyte	Membrane	Ionic additive	Detection limit (mol/L)	Ref.
CoTPPCl	NO_2^-	PVC/*o*-NPOE	—	7.9×10^{-6}	[62]
		PU	—	10^{-5}	[73]
		PU	TDMAC	10^{-5}	
		PVC	TDMAC	10^{-5}	
InOEPCl	Cl^-	PU	—	10^{-4}	
		PU	TDMAC	10^{-2}	
		PU	KTFPB	3×10^{-5}	
		PVC	KTFPB	3×10^{-5}	
		SR/*o*-NPOE		2.5×10^{-4}	[74]
		SR	DBS	10^{-4}	
GaOEPCl	F^-	PVC	KTFPB	3.2×10^{-4}	[73]
		PU	—	10^{-3}	
		PU	KTFPB	6.3×10^{-4}	
MnOEPCl	Cl^-	SR/*o*-NPOE	*o*-NPOE	7.9×10^{-4}	[74]
		SR	DBS	1.3×10^{-3}	
MnTPPCl	Cl^-	SR/*o*-NPOE	*o*-NPOE	3.2×10^{-4}	[74]
		SR	DBS	1.6×10^{-3}	
	SCN^-	PVC/DOP	NaTPB	5×10^{-8}	[68]
$ZrOEPCl_2$	F^-	PVC/*o*-NPOE	KTFPB	3.2×10^{-5}	[70]
$ZrTPPCl_2$	F^-	PVC/*o*-NPOE	KTFPB	10^{-5}	[70]
$SnTPPCl_2$	salycilate	PVC/*o*-NPOE	NaTFPB	10^{-4}	[67]
	salycilate	PVC/*o*-NPOE	—	1.6×10^{-4}	
OMoTPP(OEt)	salycilate	PVC/*o*-NPOE	NaTFPB	10^{-5}	[67]
UPFP	acetate	PVC/*o*-NPOE	TDDMACl	3×10^{-5}	[70]
MnPPIXCl-PS	I^-	PVC/*o*-NPOE		10^{-5}	[71]
TSPP	I^-	polypyrrole	—	10^{-6}	[82]
TAPP	Zn^{2+}	PVC/AP	OA	3×10^{-5}	[81]
PPIXDME	Zn^{2+}	PVC/DOP	NaTPB	1.5×10^{-5}	[80]

Although PVC is the most widely used membrane, some examples of different matrices have been proposed for particular applications. Malinowska and co-workers have studied the potentiometric behavior of ISEs based on In(III), Ga(III), and Co(III) porphyrin complexes encapsulated in PU membranes, in comparison with that observed with conventional PVC ISEs [73]. The same sensing mechanism is observed in both membranes, although the presence of endogeneous cationic sites in PU, instead of anionic sites as in PVC, strongly influences the potentiometric response if no lipophilic salt additives are added.

Silicon rubber has also been proposed for the development of all-solid-state ISEs [74]. In these devices the membrane is directly attached to the electrode surface, without the presence of the internal solution (Fig. 11). Mn(III) and In(III) porphyrins encapsulated in SR matrices were used to develop chloride selective electrodes, suitable for the measure of chloride in clinical samples. Silicon rubber is necessary for the development of all-solid-state sensors because of its adhesion properties to the electrode surfaces. It was necessary to add plasticizers in order to obtain potentiometric responses. In(III)OEPCl gave the best results as ionophore, because of reduced interference from salicylate, and an InOEPCl based ISE was demonstrated to be suitable for serum chloride measurement.

Other than inorganic anions, metalloporphyrin based ISEs have been exploited for the measurement of particular target analytes, such as aminoacids [75, 76], and drugs [77, 78].

Free base porphyrins have also been used as ionophores for cation selective electrodes [79, 80] although cation sensing is far less developed than anion detection. TPP and TTP encapsulated in PVC membranes have been used to develop Ni(II)-selective sensors [79], although only moderate selectivity was effectively achieved. More recently two examples of Zn(II)-selective electrodes have been reported in the literature [80, 81] In the first example the PP-IX dimethyl ester was encapsulated into a PVC membrane, along with

Figure 12. Molecular structure of urea-functionalized TPP.

sodium tetraphenylborate and dioctyl phtalate as plasticizer [80]. The resulting Zn-ISE showed fast response time, good long-term stability, pH range 2.1–4.0, and a working concentration range of 10^{-5} to 10^{-1} mol/L. The second sensor was developed using T(2-APP) as ionophore in PVC membranes [81]. This ISE showed good selectivity toward other metal ions in a pH range 3.0–6.0 and it was successfully applied for the determination of Zn in pharmaceutical samples.

The development of ISEs that do not include a polymeric support has also been investigated. Alternate deposition of water soluble anionic TSPP and cationic polypyrrole was used to obtain the deposition of a multilayer film onto a 2-aminothiol modified silver electrode [82]. This chemical modified electrode (CME) was used as iodide ISE, showing a Nernstian behavior and detection limit of 10^{-6} mol/L, with fast response time and low resistance.

An interesting approach for the development of all-solid-state ISEs has been the modification of the electrode surfaces by an electropolymerization technique [83–85] T(2-APP) and the related TETMAPP were used as precursors to deposit polymeric films onto the surfaces of different electrodes. Free base and Fe(III), Ni(II), Co(II), and Cu(II) complexes of T(2-APP) were deposited by an electropolymerization technique onto the surfaces of different metal electrodes [83]. The films obtained were not conductive, probably because of the three-dimensional, spongelike structure of the polymeric film. They were tested as ISEs for the detection of different anions and, in the case of the free base, both anions and cations. It is interesting to note that useful potentiometric data were obtained when the electrode was positively polarized at positive potential before use, so inducing positive charge on the film. These ISEs generally showed sub-Nernstian slopes for iodide ions, with detection limits ranging from 10^{-4} to 10^{-3} mol/L. However, different parameters influenced the responses of these ISEs and for this reason it was sometimes difficult to give a rationale for the results obtained. In the case of the electropolymerized electrodes developed by Volf and co-workers [84], the anion selectivity pattern shown was attributed to coordination interactions in the case of polymeric films containing metalloporphyrins, and to ion-exchange polyporphyrin free base. When applied to the determination of amino acids [85], a remarkable selectivity for cysteine was observed.

4.2. Amperometric and Voltammetric Sensors

The working mechanisms of these devices are based on the well known ability of metalloporphyrins to catalyze oxidation or reduction of different substrates. Thin films of metalloporphyrins have been deposited onto electrode surfaces to develop chemical modified electrodes (CME) and the electrocatalytic action of the developed CMEs has been used for the detection of different analytes. Particular attention has been devoted to the detection of NO, starting from the seminal paper of Malinski and Taba [86]. NiTMHPP (Fig. 13) was deposited by the electropolymerization technique to give conductive films onto the surface of carbon fiber electrodes. An additional layer of Nafion® was necessary to avoid the interference of anion present in real matrices, such as nitrite. A major advantage of this approach was the extremely small dimension in which the electrodes can be

Figure 13. Molecular structure of TMHPP.

fabricated. The sensor is miniaturized by thermal sharpening of the electrode carbon fiber. The diameter of the carbon fiber is reduced to 500 nm and the fiber is mounted into a truncated needle (Fig. 14) [87]. The tip of the fiber (500 nm length) is coated with the polyNiTMHPP by electropolymerization and then an additional layer of Nafion® is added by a dipping technique. The resulting CMEs were demonstrated to be useful for the detection of NO in biological samples, because the developed nanosensors induced minimal damage for biological tissues, so allowing both *in vitro* and *in vivo* measurements [86–89]. In Table 4 we report examples of detection limits of some sensors reported in the literature. Furthermore the small size of these sensors was necessary to allow the monitoring of transient concentration of NO produced in physiological processes, because the half-life of this reactive molecule is very short in biological samples. While the miniaturization and the exploitation of these nanosensors has been greatly improved, the exact mechanism of electrochemical NO detection by the poly-NiTMHPP sensors is still unclear. In the first reports by Malinski and Taba and coworkers the electrocatalytic role of the coordinated Ni(II) ion has been involved for the oxidative reaction of NO, and coordination of NO and formation of the Fe-nitrosyl complex [Fe(III)P]NO was confirmed by cyclic voltammetry and spectroscopic data by Bedioui and co-workers [90]. On the other hand the same authors showed that in the case of Ni

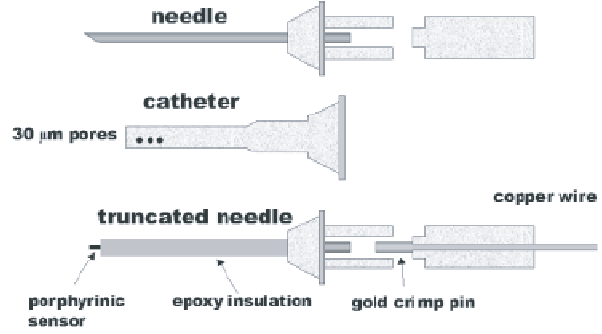


Figure 14. Sketch of the NiTMHPP-based nanoelectrode used for NO measurement. Reprinted with permission from [89], T. Malinski et al., *Eur. Surg.* 34, 79 (2002). © 2002, Blackwell.

Table 4. Porphyrin-based NO sensors.

Porphyrin	Detection limit (Nmol/L)	Additional layer	Ref.
NiTHMPP	10	Nafion®	[86]
	1.5		[88]
	27	Nafion® + *o*-PD	[95]
	85	eugenol	[94]
	1		[89]
	0.5	Nafion®, AAO, polylisine	[96]
CoTPPCl	1.2	Nafion®	[88]
MnTCPPCl	100	Poly-pyrrole	[88]

complexes the coordination of NO to the metalloporphyrin is probably not involved in the electrocatalytic process [91]. The hypothesis that the role of the porphyrin layer could not be related to the axial coordination chemistry of the Ni ion was later confirmed, showing that the nonmetalated poly-TMHPP coated CME presented the same electroctalytic activity toward NO detection [92]. The same authors showed that the thickness of the porphyrin layer influenced the efficiency of the NO detection, and a tentative oxidation mechanism, where the porphyrin polymer acted as a three-dimensional structure, was proposed [93].

The external Nafion® has been added to increase the selectivity toward interfering anions present in biological matrices, such as nitrite, which cannot diffuse through the Nafion layer. On the other hand Nafion can increase the interference of cation species of biological relevance, such as catecholamines. For this reason, Ciszewski and Milczarek proposed the exploitation of electropolymerized eugenol as external layer, in order to preserve the selectivity toward anions and to reduce the interference of cationic species [94]. Pontié and co-workers developed a NO multilayer microelectrode composed of a poly-NiTMHPP layer, Nafion®, and *o*-phenylenediamine [95]. The developed sensor showed a detection limit as low as 40 nmol/L, a NO concentration expected in some biological system.

Mitchell and Michaelis reported a composite multilayer electrode having poly-NiTHMPP as sensing material and layers of Nafion®, acorbic acid oxidase, and either polylisine or polypyridinium to preclude interferences by physiological concentrations of both anionic and cationic electroactive species [96]. The resulting microdevice was used for *in-situ* NO detection, showing a NO detection limit of 8 nmol/L for *in vitro* and 173 nmol/L for *in vivo* tissue sample analyses.

Several metalloporphyrins were later exploited to develop CMEs devoted to the determination of different analytes [97–99]. Reducing anions such as nitrite and sulfite has been determined by dip coating modified glassy carbon electrodes with a tetraruthenated cobalt porphyrin (Fig. 15) [97]. These electrodes showed excellent performance for the determination of nitrite and sulphite, even at ppb level, both in conventional and in flow injection conditions. The same authors developed modified glassy carbon electrodes by deposition of an electrostatically assembled film of the tetraruthenated cobalt porphyrin and Zn(II)TSPP [98]. These electrodes were successfully exploited for nitrite detection in flow injection analysis with a detection limit of 10^{-5} mol/L. The same electrode can be used for nitrate ion detection

Figure 15. Molecular structure of the cobalt complex of tetraruthenated-TPP.

after reduction to nitrite in a reductor column containing copperized cadmium. Graphite electrodes functionalized by cobalt(II)TPP permitted the detection of a wide range of different organohalides, including compounds that are industrial pollutants [99]. A further advantage of these electrodes was the absence of interference from oxygen.

5. NANOGRAVIMETRIC SENSORS

Nanogravimetric sensors are generally based on the piezoelectric properties of crystalline quartz. Quartz crystal is cut along a certain symmetrical axis, usually AT cut, to obtain material able to sustain bulk electroacoustical oscillation at frequencies from 5 to 30 MHz. The quartz property making it interesting as a sensor is that the resonance frequency is, in a limited linear range, inversely proportional to the mass gravitating onto the surfaces of the quartz, as described by the Sauerbrey law:

$$\Delta f = -K_q \Delta m$$

This behavior is exploited to turn quartz into a chemical sensor when some chemically interactive material, able to capture molecules from the environment, is used as coating.

The resolution of these transducers is generally in the order of nanograms and for this reason such devices have been called quartz microbalances (QMBs).

In recent years metalloporphyrins have been proposed as sensing elements in mass transducers for the detection of volatile organic compounds (VOCs) [100–103].

Different from optical and electrochemical transducers, a QMB does not influence the selectivity of the sensing layer because, since they respond to mass variations, any kind of material–analyte interaction is recorded.

In the case of metalloporphyrins, both specific interaction, such as coordination to the metal center or π–π interaction with the aromatic system, and nonspecific interaction contribute to the sensing mechanism, so allowing the detection of virtually any kind of VOC. This feature has been confirmed by studying the adsorption isotherm of metalloporphyrin coated QMBs [100, 101]. In the case of ligands that cannot give specific interaction, such as alkanes, the

adsorption isotherm is linear and depends on the solubility of the analyte into the porphyrin matrix, according to the Henry law. When specific interaction can occur, the isotherm showed a Langmuir-like behavior at low concentrations, while at higher concentrations, after the saturation of the specific sites only nonspecific interaction occur and the isotherm became linear [100–103].

Because of the described sensing mechanism, the selectivity of the metalloporphyrins as sensing materials can be finely tuned by modifications of both coordinated metal and peripheral substituents [100–103]. A huge number of the elements present in the periodic table have been coordinated to the porphyrins and furthermore the organic chemistry of these compounds is well developed, and a wide range of different substituents can be introduced at their peripheral positions. All together these characteristics increase the versatility of these molecules as sensing elements for QMB sensors, all showing outstanding properties of these materials in terms of stability, sensitivity, and reproducibility.

Among the different factors influencing the selectivity, the coordinated metal was demonstrated to be one of the most important and this influence was interpreted according to Pearson's principle [104]. In Figure 16 the sensitivity properties of some metalloporphyrins toward model VOCs are reported. The versatility of porphyrins as a chelating system allows the development of a large set of different sensors by changing the coordinated metal. Furthermore the selectivity is influenced by tuning the peripheral substitution patterns of the macrocycle, thus influencing the electronic distribution on the π-aromatic system of the porphyrin and modifying the interaction of the metalloporphyrin with the axial ligands [104].

Another important aspect is represented by the technique chosen for the deposition of metalloporphyrin thin film onto the quartz surface, because the morphology of the film can also influence sensing of the material–analyte interaction.

Because of the stability and versatility of porphyrins, different techniques are available for this purpose and in the case of QMBs, the following have been successfully exploited: (a) spray casting, (b) Langmuir–Blodgett, (c) self-assembled monolayers, and (d) electropolymerization.

Spray casting is the simplest technique and consists of the dissolution of the metalloporphyrin in an organic solvent and then this solution is sprayed onto the QMB surface.

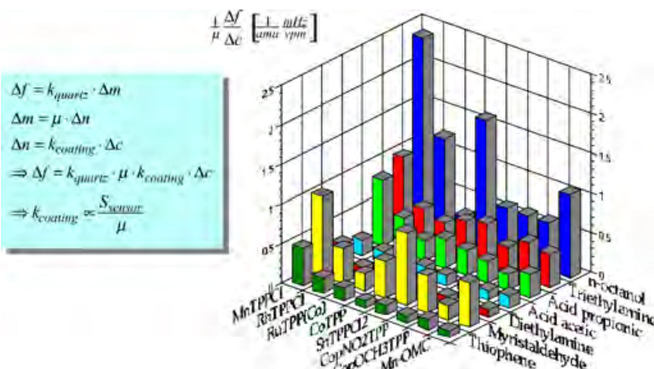


Figure 16. Sensitivities of metalloporphyrin-coated QMB toward different VOCs.

On the other hand, this technique does not control the morphology of the growing film and the reproducibility of the film characteristics can be reduced [104].

Langmuir–Blodgett is a well-known method and allows the deposition of ordered multilayers of porphyrins. It was advantageous to study in detail the mechanism of the metalloporphyrin–analyte interaction [105]. On the other hand the porphyrin should be functionalized in order to have the amphiphilic character necessary for the LB deposition. Alternatively, porphyrin can be deposited in a mixture of fatty acids, but this can be a problem in the case of QMBs, because the adsorption properties of the fatty acid matrix cannot be ignored.

The self-assembled monolayer technique consists of the preparation of functionalized surfaces by chemisorption of functionalized metalloporphyrins onto inorganic substrates. In the case of QMBs, porphyrins bearing sulphide groups at the peripheral positions (Fig. 17) have been chemisorbed onto the gold pad of a QMB [106]. The performances of these sensors have been measured with respect to model volatile compounds of interest for practical applications. While the sensitivities showed by these sensors were higher compared to the corresponding casting coated sensors, the presence of a single layer of the sensing material induced a saturation of the specific sites at very low concentration of the analytes. In Figure 18 the sensitivities of these sensors toward different VOCs have been reported.

Very recently, a chemisorbed chiral porphyrin diad (Fig. 19) has been used for enantiodiscrimination [107]. The enantiodiscrimination properties of such a sensors toward the enantiomeric pairs of chiral analytes have been studied. While in the case of analytes bearing donor ligand atoms no significant enantioselectivity was observed, a significant degree of chiral discrimination has been observed in the case of limonene. In Figure 20 we report the responses of the diad toward saturated vapors of different compounds. The results obtained fitted nicely with the changes observed on the solution visible and CD spectra of the diad. This result

Figure 17. Molecular structure of thiol-functionalized TPP.

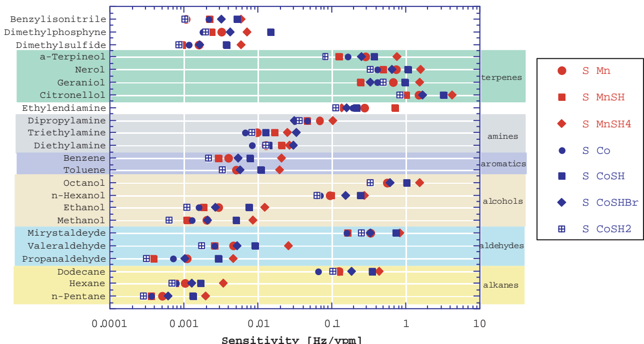


Figure 18. Sensitivities of self-assembled monolayers of metalloporphyrin-coated QMBs toward different VOCs.

was particularly encouraging for the potential development of enantioselective chemical sensors.

The suitability of electropolymerization as a deposition technique has been demonstrated for the development of CMEs. In the case of QMBs, the quartz has been used as a working electrode in an electrochemical cell to deposit thin films of metal complexes of PP-IX [108]. The resulting sensors showed promising performances for the detection of VOCs.

QMB sensors have also been used to investigate the molecular recognition properties of cast films of Zn(II) complexes of porphyrin-resorcinol derivatives (Fig. 21) [109]. These compounds are known to form hydrogen bonded inclusion complexes with appropriate guests and the selectivity showed by these QMB sensors depended on both the formation of hydrogen bond with peripheral OH groups and coordination to the central Zn ion. Furthermore a size dependence in guest selectivity was observed.

One of the drawbacks of QMBs is generally their low resolution for some particular applications. To overcome this problem it has been attempted, maintaining the mass transduction principle, to use surface acoustic wave (SAW) devices as transducers [110]. These devices exploit the propagation of SAWs along layered structures consisting of a piezoelectric substrate coated by a thin film of the sensing material. Changes produced by the adsorption event on the sensing material properties can affect both the phase velocity and the propagation loss of the acoustic wave, although the first is the most generally used. This effect

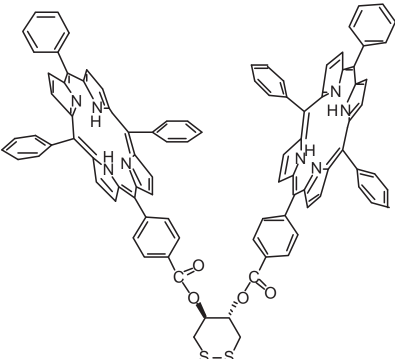


Figure 19. Molecular structure of the porphyrin diad.

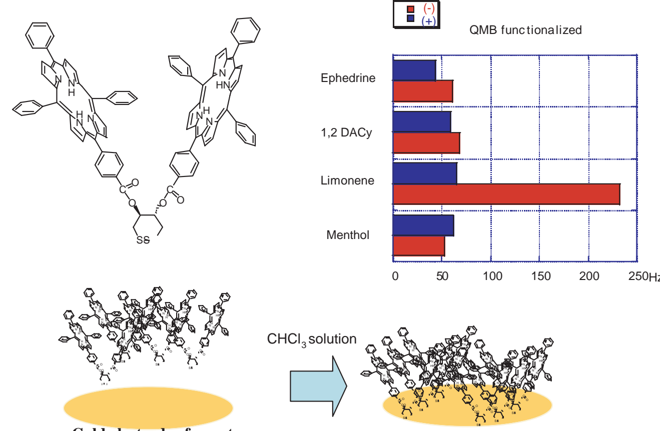


Figure 20. The responses of diad-coated QMB toward saturated vapors of enantiomeric pairs of chiral compounds.

can be easily converted in a frequency shift, provided that a device is configured as a SAW oscillator. Metalloporphyrin-coated SAWs showed positive results for VOC detection, representing an interesting solution for applications in which a high-resolution transduction is necessary, although this approach is more difficult for practical applications, due to the need for more complicated electronic circuits and read-out instruments [110].

Cobalt(II) complexes of picket-fence or picnic-basket porphyrins have been exploited as SAW sensing coatings for the detection of O_2 [111]. The resulting devices were stable toward oxidation and were shown to bind molecular oxygen following a Langmuir isotherm. Although the responses obtained were lower than those observed in solution, the developed SAWs were demonstrated to be useful for the measurement of a wide range of oxygen partial pressures.

6. MISCELLANEOUS

The richness of properties of porphyrins has allowed the exploitation of other kind of transduction mechanisms, although they are not common as these previously described.

The deposition of thin films of metalloporphyrins on the gate of metal oxide semiconductor field effect transistors

Figure 21. Molecular structure of the Zn complex of resorcinol porphyrin.

(MOSFETs) led to the development of the so-called chemFET [112–114]. LB films of different metal complexes of a functionalized T(4-APP) (Fig. 22) used as sensing layers of chemFET were tested for the detection of ambient gases such as NO_2, NH_3, CO, and H_2S. Also in this case the coordinated metal induced the selectivity of the sensor and only the Co derivative was sensitive to NO_2, with 0.5 ppm as the detection limit [112, 113]. More recently metalloporphyrins were deposited on the MOSFET gates by a spray-coating technique or by self-assembly, and the resulting sensors were exploited for the detection of VOC [114]. The performances of the sensors strongly depended on the gate thickness and, among the deposition techniques, the deposition of monolayers showed better sensitivity and stability.

The Kelvin probe has also been used as a transducer to test sensing properties of metalloporphyrin thin films [115]. The Kelvin probe is a well-established method for the measurement of the work function and it has been extensively studied for the characterization of metals and inorganic semiconductor surfaces. It consists of a noncontact, nondestructive vibrating capacitor inducing an alternate current between the sample and a vibrating tip, as a consequence of the contact potential difference resulting from the different work functions of the sample and the reference tip. From the chemical sensor point of view, the Kelvin probe is a simple and relatively fast technique to monitor the variation of surface potential during the adsorption and desorption of gases from and to the vapor phase. Interaction of metalloporphyrin monolayers with different VOCs gave significant Kelvin probe signals, demonstrating the suitability of this technique as a transducer [115]. Dimers of Ni(II)OEP units linked by a butadiyne bridg (Fig. 23) were deposited in mixture with arachidic acid as LB-thin films by Arnold and co-workers [116]. These films were used as sensing layers in chemoresistive sensors and tested for the detection of ammonia. This sensor worked at low temperature (50 °C) and showed a detection limit of 30 ppm.

A novel oxygen sensor has been developed by chemisorption of a disulphide bearing two imidazole groups onto the surface of CdSe; this ligand was used to hinge Fe(III)PP-IX units [117, 118] The sensing event, based on the metalloporphyrin interaction with molecular oxygen, is communicated to the semiconductor surface by the ligand and the transduction mechanism was based on change of the photoluminescence of the semiconductor. The detection limit obtained was in the 0.1 atm oxygen pressure range [118].

Figure 22. Molecular structure of modified T(4-APP).

Figure 23. Molecular structure of the Ni complex of OEP-butadiyne dimer.

A similar approach was used for the immobilization of FeTPPCl on the GaAs surface and the resulting device was able to detect NO in solution at very low concentration (3×10^{-5} mol/L) [119]. This device, in fact, changes its electrical resistivity after adsorption of molecules on the semiconductor surface. Higher resolutions were obtained when hemin was directly chemisorbed, through the peripheral carboxylic chain, to the GaAs surface [120]. This device was able to detect NO at a concentration of 10^{-6} mol/L in physiological solutions and for this reason it represents a fast and simple method for the biosensing of NO.

7. SENSOR ARRAYS

Since its beginning chemical sensor science has been devoted to the development of chemical sensors with improved properties in terms of sensitivity, selectivity, reproducibility, and stability. However, while the sensitivity, reproducibility, and stability are properties necessary to be satisfied, the selectivity needs to be further discussed.

An absolute selectivity of the sensing material (i.e., the ability to detect only a single analyte in any complex mixture) leads to a specific sensor. This kind of sensor is required when we have to qualitatively and quantitatively detect a target compound, such as a toxin, but it is not helpful when we have to assess the quality of an indoor environment or judge the edibility of a food sample. Furthermore it should be considered that to obtain absolute selectivity is a very difficult task and biosensors, where the almost unique catalytic properties of enzymes were exploited, fulfilled in some cases these requirements, whereas the development of supramolecular chemistry concepts has been used to improve the selectivity characteristics of chemical sensors.

For this reason a different concept of selective sensors was introduced at the end of the 1980s [121]. In this approach the sensors should not be specific, but they have to interact with almost all the chemical compounds present in the environment, but with different sensitivities. These sensors should be used in an array configuration and in analogy to the mammalian olfaction these arrays offered the chance to develop artificial olfaction systems.

Although an absolute selectivity is strictly no longer essential, it is, however, of great importance that the sensors

have different sensitivities toward the analytes characterizing the samples under measurement; in this way the sensor responses will not be completely correlated and the maximum of chemical information can be extracted by the sensor array, which can provide recognition and classification of the different measured matrices.

Porphyrins are particularly suitable for this kind of application, because they showed interesting properties as sensing materials and their properties can be tuned by synthetic modifications. For example their selectivity can be modified by changing the coordinated metal and by tuning the peripheral substitution patterns of the macrocycle, thus influencing the electronic distribution on the π-aromatic system of the porphyrin and modifying the interaction of the metalloporphyrin with the axial ligands. Furthermore different transduction mechanisms can be exploited with porphyrins and it is possible to develop sensors operating both in the liquid and in the gaseous phase.

7.1. Gas Phase Sensor Arrays: Electronic Noses

Since 1995 a sensor array based on metalloporphyrin-coated QMBs has been developed. The performance of the developed electronic nose, called LibraNose, has been tested in different fields of applications, ranging from food to clinical analysis [122–132].

For food analyses the objectives of the experiments were different, depending on the particular matrices analyzed, ranging from the recognition of storage time to assessment of food quality.

In the case of the analysis of cod fish and veal meat [122, 123], the aim of the analysis was the classification and recognition of days of storage. In both cases the electronic nose provided a good classification of the data, which was in perfect agreement with the food aging. More recently codfish fillet aging has been studied over a period of 17 days and it was demonstrated that the electronic nose followed the variation of the headspace chemical composition [129]. Furthermore it was showed that misclassification errors were due to the nonlinear evolution of the headspace chemical composition during the spoilage process and not to lack of resolution of the sensor array.

For red wine the electronic nose was exploited to monitor the aroma modification when the wine was exposed to the air; a significant modification was observed three hours after the opening of the bottle, with a qualitative agreement with the expected variation of the taste: it is well known that red wine needs contact with air to enhance its taste [123].

A significant result was obtained in the analysis of tomato paste; in this case the aim of the experiment was the classification of different tomato pastes produced by two different cultivations: biological and traditional. The data obtained with the electronic nose were compared with those obtained by a human panel of trained tasters. Data analysis indicated a good agreement between the classification obtained by the electronic nose and the human panel, but electronic nose showed a lower dispersion of the data and consequently a better definition of classes [124, 125]. A similar comparison was carried out for the analysis of UHT milk samples [123]. In this test, 13 different commercially available brands of UHT milk were taken into consideration. Each of them was analyzed with the electronic nose, while the panel searched the presence and the intensity of a specific sensation, which is a drawback for this particular product: the aroma of burned milk. There was perfect agreement between sensorial analysis and electronic nose results and this was surprising because sensorial analysis was aimed at evaluating only one of the aroma features, while the electronic nose data were related to the whole aroma composition. The same approach was recently followed for the analysis of fruit samples, such as peaches and nectarines, with the aim to define a sensorial profile for these fruits [132].

A different experiment was aimed to measure the aroma decay of oranges over a period of time of about one month, while in the case of apples the study was devoted to the measure of mealiness and skin damage. Results obtained showed that electronic nose was able to correctly predict the amount of defects (apple) and storage days (for oranges) [128].

Electronic nose and spectroscopic techniques were recently integrated to give a virtual instrument devoted to the measurement of some analytical parameters of peach samples [127]. The fusion of visible spectra and electronic nose data gave an interesting method to probe fruit samples in a nondestructive way.

Different olive oils have also been classified by electronic nose analysis [130]. Two different electronic noses were exploited in this study and their data were compared and integrated.

Although food analysis has been the most explored field of application for electronic noses, more recently the medical field was also taken into consideration and the application of the electronic nose to detect diseases was proposed [133–136]. The hypothesis of these studies was that body odor is correlated with human health. Human odor is in fact correlated with the combined action of both gland and bacterial population and any change of the complex equilibrium regulating the organism can induce a change in the nature and amount of volatile organic compounds excreted by the organism. On the other hand medicine recognizes that some pathologies produce unpleasant characteristic odors (for example diabetes and hepatic diseases). As a consequence the exploitation of electronic nose, which is a noninvasive probe, can represent an eligible tool for a preliminary diagnosis of diseases. In particular, the sensor array was able to detect 5a-androst-16-en-3-one, a steroid present in human sweat and supposed to be a male pheromone, with a resolution comparable to that of human olfaction [133]. In a different paper correlation between electronic nose measurements and the ovulation cycle phases in fertile women has been claimed [135].

The metalloporphyrin-QMB sensor array was also exploited for the analysis of the headspace of urines belonging to patients affected by kidney diseases [134]. The electronic nose showed good performance in distinguishing samples containing blood, one of the pathologies induced by the kidney disease, also giving some indication of the blood amount.

More recently, LibraNose has been exploited for the diagnosis of lung cancer [136]. Volatile organic compounds present in expired breath may give a potential source of

information about general metabolic conditions and, in particular, on lung physiology; and breath analysis carried out by combined gas chromatography–mass spectroscopy evidenced that different volatile patterns occur in diseased individuals. The exploitation of the electronic nose to analyze breath from individuals affected by lung cancer resulted in a good classification among the groups of individuals according to their known pathologies, discriminating them from a sample of healthy people measured as reference. A number of patients were measured after the surgical removal of the cancer mass. In this case the nose considered these patients to be healthy. These preliminary results demonstrated a very promising field of application for electronic noses, with an enormous impact for the early diagnosis of diseases.

Other than QMBs, the same authors proposed an electronic nose based on metalloporphyrin based optical sensors [137]. The optoelectronic nose was developed by deposition of thin films of two metalloporphyrins, CoTPP and RhTPPCl, a Mn(III) corrole and a sapphyrin as sensing materials onto a transparent substrate. Commercial, low-price blue-LEDs and photodetectors were used to develop the sensor array. In order to reduce as much as possible the influence of LED intensity fluctuations, a differential measurement strategy was chosen and one position was left free of porphyrin and used as blank reference. In Figure 24 a sketch of the array is reported. The array was exposed to different concentrations of VOCs and was able to discriminate among them. If compared with QMB sensors, optical transduction was demonstrated to give higher resolutions for the same sensing material (Table 5).

Very recently an optical sensor array was constituted of LB films of FeTPPCl, MnTPPCl, CoTMPP, and CoOEP co-deposited with arachidic acid [138]. The system used four LEDs operating at different wavelengths (645, 615, 601, and 546 nm) as light sources. Fiber optics were used to illuminate LB films and to receive the reflected lights. The responses of such a sensors were fast and reversible and the array was able to discriminate the aroma of three different capsicum samples.

Following a different approach, Suslick and co-workers have recently reported a colorimetric electronic nose based on metalloporphyrins [139–142]. The array is formed by spots of different metalloporphyrins deposited onto a silica

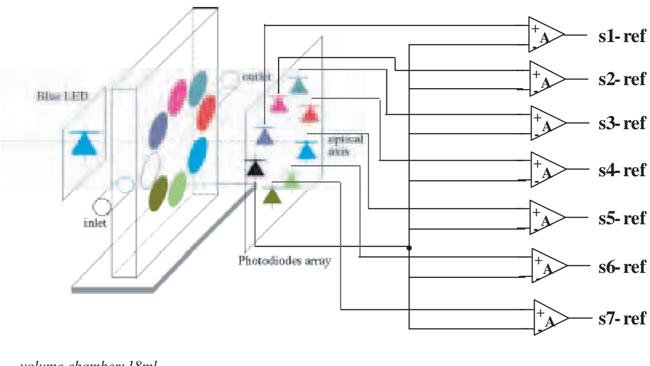


Figure 24. Sketch of the porphyrin-based optical sensor array. Reprinted with permission from [137], C. Di Natale et al., *Sens. Actuators B* 65, 220 (2000). © 2000, Elsevier Science.

support. The image of the array was digitalized by a scanner before and after the interaction with VOC and color change profiles represent the response of the array and are characteristic of the analyzed VOC. In Figure 25 the colorimetric changes observed have been reported. This electronic nose approach has been called SmellSeeing™ [140, 141]. Resolutions down to 2 ppm were obtained for such an array and in some case responses in the range of 100 ppb were recorded. The inherent simplicity, the absence of complicated data analysis, the high sensitivity, and the fast and reversible sensor responses make this approach particularly attractive. Furthermore this approach did not suffer from humidity variation, a parameter of significant importance for practical applications.

Exploitation of this approach to the detection of bacterial growth has been recently reported [142]. The SmellSeeing™ array was able to detect a wide range of VOCs, which can be related to bacterial growth, in the ppb range. This results in the detection of culture of *E. coli* bacteria in short time; promising results were also obtained with the growth of skin and mouth bacteria, making this array particularly promising for the development of real-time, portable, and easy-to-use sensors for bacteria identification.

7.2. Liquid Phase Sensor Arrays: Electronic Tongues

While electronic noses are an established research and development topic, surprisingly, only a few examples of the relative concept applied to liquid samples have been developed, although multivariate analysis with arrays of ion selective electrodes were proposed in the late 1980s [143] and the first electronic tongue, based on ion selective electrodes functionalized with lipid membranes, was proposed at the beginning of the 1990s by Hayashi et al. [144].

The excellent performance of metalloporphyrins in vapor phase chemical sensors led to the exploration of the development of liquid phase sensors based on metalloporphyrins as sensing material, where the metalloporphyrins were immobilized into PVC membranes. The developed electrodes were integrated in an array to develop an electronic tongue based on potentiometric sensors [145, 146]. The developed electronic tongue was integrated with a metalloporphyrin-QMB based electronic nose in the fields of clinical and food analysis. In this way it was possible to extract information from both the liquid and the headspace of the analyzed matrix, resulting in a net increase of performances. This is the first approach to develop a complete artificial sensorial system. More recently CMEs were fabricated by the electropolymerization of different metal complexes of TMHPP on glassy carbon electrodes, using the cyclic voltammetric method [147]. The resulting electrodes were used as potentiometric sensors and demonstrated higher resolution and stability than the corresponding PVC based electrodes. These CMEs were assembled in an array configuration and utilized to analyze several samples. In particular, this array was able to discriminate compounds, each one used as a model of the five classical tastes. These preliminary results were a basic step for the development of porphyrin based taste sensor systems for a variety of applications.

Table 5. Resolution of optical and QMB based porphyrin sensors.

	VOC range (ppm)	$CopNO_2TPP$		RhTPPCl		MnOMC	
		Δc optic.	Δc QMB	Δc optic.	Δc QMB	Δc optic.	Δc QMB
Acetic acid	70–200	0.28 ppm	2 ppm	0.2 ppm	4.1 ppm	0.1 ppm	1 ppm
Hexane	1000–4000	50 ppm	50 ppm	50 ppm	50 ppm	16.6 ppm	50 ppm
Ethanol	500–1500	5.2 ppm	16 ppm	1.7 ppm	16 ppm	1.3 ppm	8 ppm
Methanol	1000–3000	10 ppm	50 ppm	4.1 ppm	50 ppm	2.2 ppm	50 ppm
Triethylamine	500–2000	2.8 ppm	5.5 ppm	1.4 ppm	5 ppm	1.1 ppm	3 ppm

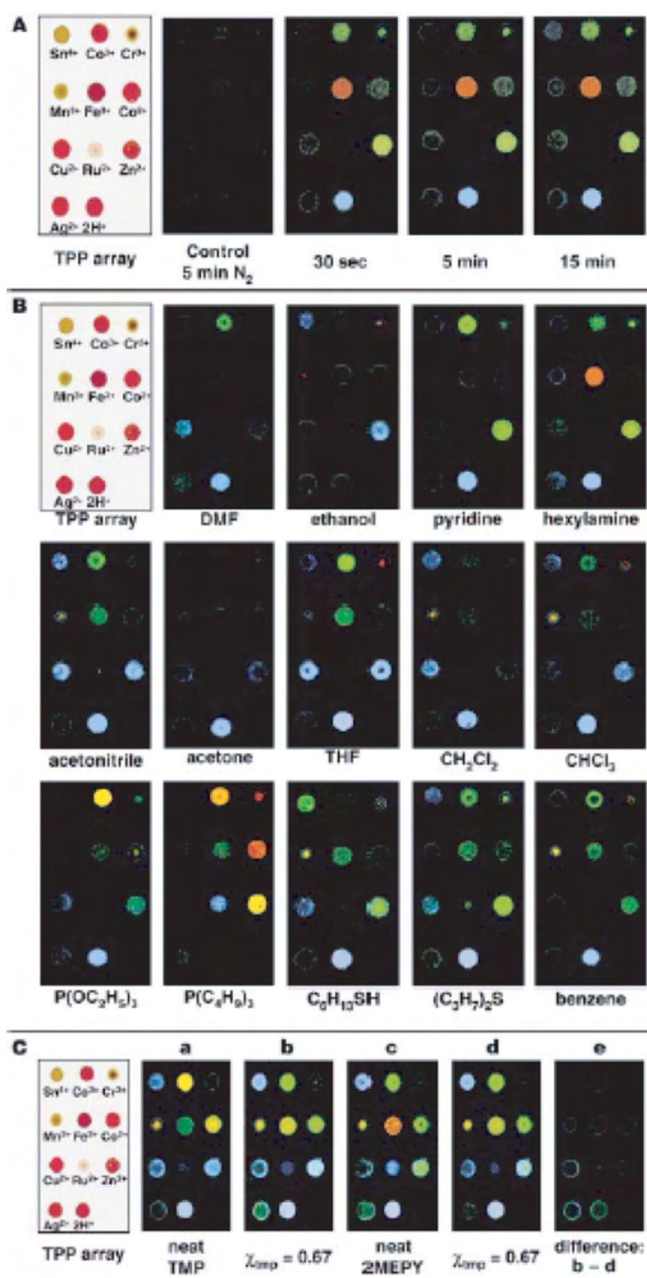


Figure 25. Color changes of metalloporphyrins array after VOC exposure. Reprinted with permission from [139], N. A. Rakow and K. S. Suslick, *Nature* 406, 710 (2000). © 2000, Macmillan.

7.3. Data Analysis for Sensor Arrays

Artificial sensorial systems are attempts to reproduce the functions of the natural chemical senses (olfaction and taste). Currently, a sort of standard electronic nose model common to the available implementations of artificial olfaction may be identified. In this model three major subsystems may be identified: sampling, sensor array, and data analysis. While sensor arrays are oriented to exploit the analogies with natural receptors (e.g., cross-selectivity and redundancy) the data analysis does not take advantage of natural data treatment paradigms. The principles of data processing of natural olfaction receptor signals are still rather unknown and the differences in terms of sensors signals make them hardly applicable to process sensor data, although the topic is evolving and first attempts to derive analysis methodologies from natural paradigms have already been proposed [148].

Electronic nose data analysis is rather based on techniques borrowed from other disciplines and chosen among that vast class of algorithms known as pattern recognition.

Pattern recognition in chemistry appeared at the end of the 1960s. Among the first applications the use of the Learning Machine on behalf of the analytical chemists at the Washington State University has to be mentioned. They classified molecules according to molecular structural categories used to form patterns [149].

In the gas sensing area, multivariate methods of pattern recognition are commonly required when sensor arrays, composed of real, nonselective, cross-sensitive sensors, are utilized. Pattern recognition, exploiting the cross-correlation, extracts information contained in the sensor outputs ensemble. Only a few of the manifold of pattern recognition techniques introduced in other disciplines have been utilized ([150–152] offer reviews of the most practiced ones).

Formally, pattern recognition may be defined as "the mapping of a pattern from a given pattern space into a class-membership function" [153].

A pattern can be defined as an ordered set of real numbers. In analytical chemistry, the response of a multichannel instrument to one sample (e.g., gas chromatographs and spectrophotometers) is considered to form patterns. For a sensor array, the responses of the sensors to a sample take the name sensor pattern.

Class-membership space represents either a quantitative or qualitative set of quantities. When quantification is considered the class-membership function turns out to be a vectorial function defined in a metric space. In the case of quality recognition, the class membership is an ensemble of

abstract sets to which each measured sample is assumed to belong. In the first case we prefer to talk of multicomponent analysis while the term pattern recognition is usually referred to for the second case.

Mathematically, a pattern is a vector belonging to its own vector space. The vector space where the sensor responses are represented is the sensor space. The simplest of these representations considers one scalar response for each sensor. This response is called the sensor feature and feature extraction is the operation that extracts, from a sensor response, a number of synthetic descriptors to form patterns.

When a sensor (independently from its working principle) is exposed to a gas or a mixture of gases, it gives a response depending on the nature of the gas and its concentration. Almost the totality of the examples reported in literature are considered, as the sensor response is either the absolute or the relative change in sensor signal measured in two steady-state conditions: in the absence and in the presence of the gas and when all the transients vanished.

Pattern recognition applied to several disciplines and practical problems produced a huge number of algorithms and techniques that are, in principle, applicable to classify electronic nose data. Several of these techniques have been actually utilized and frequently appear in literature. On the other hand, there is not yet an analysis approach based on natural olfaction paradigms. This makes the field almost totally dependent on the developments achieved in other fields. A comparison among the various techniques has also been attempted by some authors, but since the variables determining electronic nose performances are very numerous (choice of sensors, samples, sampling systems, ...) the results achieved are scarcely indicative enough for a sure direction that is valid in general.

Nonetheless some of these techniques became classics of this field (such as scaling and normalization, PCA, and discriminant analysis) and nonetheless, they are often used without a deep understanding of the hypotheses on which they are based and the potentiality of these methods is not yet fully exploited. As for the other components of electronic noses (e.g., sampling system) a rethinking of the assumptions, the implementations, and the interpretation of the methods and solution adopted is highly advised for meaningful improvement in the field.

GLOSSARY

Chemical sensor A device that transforms chemical information, ranging from the concentration of a specific sample component to total composition analysis, into an analytically useful signal (IUPAC). It is composed of two components: a sensing material and a transducer system.

Electronic nose An analytical instrument which comprises an array of electronic chemical sensors with partial specificity and an appropriate pattern-recognition system, capable of recognizing simple or complex odors [by Gardner and Bartlett, *Sensors Actuators* 18–19, 21, (1994)]. An important component of the electronic nose is also a sampling system, which introduces the analytes into the sensor chamber.

Electronic tongue A multisensor instrument for liquid analysis, based on broad selective liquid-phase chemical sensor or arrays and pattern-recognition system.

Metalloporphyrin Metal complex of a porphyrin. The four nitrogen atoms of the inner core of the macrocycle represent a square planar chelating system where the metal ion is coordinated. The coordination sphere of the metal can be completed by additional ligands coordinated at the axial positions.

Porphyrin A tetrapyrrolic macrocycle in which the four pyrrole rings are joined through their α-carbon atoms by four methine bridges. They possess an 18-electron aromatic system, with a conjugation path analogous to that of the annulene series. They are the active nucleus of hemoglobin and chlorophylls.

Sensor array A collection of sensors in which the sensors are all exposed to the same sample, and the responses of the sensors should be interpreted together.

Transducer A device that transforms energy from one kind of system (for example, chemical) to another (for example, physical). In the case of chemical sensors, the interaction of the analyte with the sensing material produces a change in physical/chemical properties, and the transducer generates a proportionally related electrical or optical signal, suitable for further processing.

REFERENCES

1. W. Gopel and K. D. Schierbaum, in “Sensors” (W. Gopel, T. A. Jones, M. Kleitz, J. Lundstrom, and T. Seiyama, Eds.), Vol. 2. Wiley–VCH, Weinheim, 1991.
2. J. W. Grate, *Chem. Rev.* 100, 2627 (2000).
3. J.-M. Lehn, “Supramolecular Chemistry.” Wiley–VCH, Weinheim, 1995; J. L. Atwood and J.-M. Lehn, “Comprehensive Supramolecular Chemistry.” Pergamon Press, Oxford, 1996.
4. J. Kavandi, J. Callis, M. Gouterman, G. Khalil, D. Wright, E. Green, D. Burns, and B. McLachlan, *Rev. Sci. Instrum.* 61, 3340 (1990).
5. T. C. O’Riordan, D. Buckley, V. Ogurtsov, R. O’Connor, and D. B. Papkovsky, *Anal. Biochem.* 278, 221 (2000).
6. D. B. Papkovski, *Sens. Actuators B* 11, 293 (1993).
7. D. B. Papkovski, *Sens. Actuators B* 29, 213 (1995).
8. D. B. Papkovski, G. V. Ponomarev, W. Trettnak, and P. O’Leary, *Anal. Chem.* 67, 4112 (1995).
9. D. B. Papkovski and G. V. Ponomarev, *Spectrochim. Acta* 57, 1897 (2001).
10. W. W.-S. Lee, K.-Y. Li, Y.-B. Leung, C.-S. Chan, and K. S. Chan, *J. Mater. Chem.* 3, 1031 (1993).
11. D. B. Papkovski, G. J. Mohr, and O. S. Wolfbeis, *Anal. Chim. Acta* 337, 201 (1997).
12. P. Hartmann and W. Trettnak, *Anal. Chem.* 68, 2615 (1996).
13. A. Mills and A. Lepre, *Anal. Chem.* 69, 4653 (1997).
14. Y. Amao, K. Asai, I. Okura, H. Shinohara, and H. Nishide, *Analyst* 125, 1911 (2000).
15. Y. Amao, T. Miyashita, and I. Okura, *Analyst* 125, 871 (2000).
16. Y. Amao, T. Miyashita, and I. Okura, *Anal. Chim. Acta* 421, 167 (2000).
17. Y. Amao, K. Asai, T. Miyashita, and I. Okura, *Polym. Adv. Technol.* 11, 705 (2000).
18. Y. Amao, T. Miyashita, and I. Okura, *Reac. Funct. Pol.* 47, 49 (2001).
19. V. V. Vasil’ev and S. M. Borisov, *Sens. Actuators B* 82, 272 (2002).
20. P. Douglas and K. Eaton, *Sens. Actuators B* 82, 200 (2002).
21. P. Douglas and K. Eaton, *Sens. Actuators B* 82, 94 (2002).
22. S.-K. Lee and I. Okura, *Analyst* 122, 81 (1997).
23. Y. Amao and I. Okura, *Analyst* 125, 1601 (2000).

24. Y. Amao, K. Asai, and I. Okura, *J. Porphyrins Phthalocyanines* 4, 179 (2000).
25. Y. Amao, K. Asai, K. Miyakawa, and I. Okura, *J. Porphyrins Phthalocyanines* 4, 19 (2000).
26. A. S. Homes-Smith, A. Hamill, M. Campbell, and M. Uttamlal, *Analyst* 124, 1463 (1999).
27. W. Trettnak, C. Kolle, F. Reininger, C. Dolezal, P. O'Leary, and R. A. Binot, *Adv. Space Res.* 10, 1465 (1998).
28. Y. Amao, K. Asai, and I. Okura, *J. Porphyrins Phthalocyanines* 4, 292 (2000).
29. T. Furuto, S.-K. Lee, Y. Amao, K. Asai, and I. Okura, *J. Photochem. Photobiol. A* 132, 81 (2000).
30. G. Di Marco and M. Lanza, *Sens. Actuators B* 63, 42 (2000).
31. P. Douglas and K. Eaton, *Sens. Actuators B* 82, 48 (2002).
32. D. B. Papkovski, I. V. Desyaterik, G. V. Ponomarev, I. N. Kurochin, and T. Korpela, *Anal. Chim. Acta* 310, 233 (1995).
33. T. H. Richardson, C. M. Dooling, O. Worsfold, L. T. Jones, K. Kato, K. Shinbo, F. Kaneko, R. Treggoning, M. O. Vysotsky, and C. A. Hunter, *Thin Solid Films* 393, 259 (2001).
34. T. H. Richardson, C. M. Dooling, O. Worsfold, L. T. Jones, K. Kato, K. Shinbo, F. Kaneko, R. Treggoning, M. O. Vysotsky, and C. A. Hunter, *Colloid Surfaces A* 198–200, 843 (2002).
35. J. M. Pedrosa, C. M. Dooling, T. H. Richardson, R. K. Hyde, C. A. Hunter, M. T. Martin, and L. Camacho, *Langmuir* 18, 7594 (2002).
36. C. M. Dooling and T. H. Richardson, *Mater. Sci. Eng. C* 22, 269 (2002).
37. K. Kato, C. M. Dooling, K. Shinbo, T. H. Richardson, F. Kaneko, R. Treggoning, M. O. Vysotsky, and C. A. Hunter, *Colloid Surfaces A* 198–200, 811 (2002).
38. M. G. Baron, R. Narayanaswamy, and S. C. Thorpe, *Sens. Actuators B* 29, 358 (1995).
39. T. Richardson, V. C. Smith, R. A. W. Johnstone, A. J. F. N. Sobral, and A. M. d'A.Rocha-Gonsalves, *Thin Solid Films* 327–329, 315 (1998).
40. C. D. George, T. Richardson, M. E. Hofton, C. M. Vale, M. G. M. Neves, and J. A. S. Cavaleiro, *Mater. Sci. Eng. C* 8–9, 559 (1999).
41. Y. Sadaoka, N. V. Rodriguez, and P. Tagliatesta, *Chem. Lett.* 7, 509 (1995).
42. Y. Sadaoka, N. V. Rodriguez, and P. Tagliatesta, *Chem. Lett.* 10, 907 (1995).
43. K. Nakagawa, Y. Sadaoka, H. Supriyatno, A. Kubo, C. Tsutsumi, and K. Tabuchi, *Sens. Actuators B* 76, 42 (2001).
44. I. Leray, M.-C. Vernières, and C. Bied-Charreton, *Sens. Actuators B* 54, 243 (1999).
45. D. Delmarre and C. Bied-Charreton, *Sens. Actuators B* 62, 136 (2000).
46. W. Qin, P. Parzuchowski, W. Zhang, and M. E. Meyerhoff, *Anal. Chem.* in press.
47. M. Plaschke, R. Czolk, and H. J. Ache, *Anal. Chim. Acta* 304, 107 (1995).
48. D. Radloff, C. Matern, M. Plaschke, D. Simon, J. Reichert, and H. J. Ache, *Sens. Actuators B* 35–36, 207 (1996).
49. D. Delmarre, R. Méallet, C. Bied-Charreton, and R. B. Pansu, *J. Photochem. Photobiol. A* 124, 23 (1999).
50. W. H. Chan, R. H. Yang, and K. M. Wang, *Anal. Chim. Acta* 444, 261 (2001).
51. X.-B. Zhang, C.-C. Guo, Z.-Z. Li, G.-L. Shen, and R.-Q. Yu, *Anal. Chem.* 74, 821 (2002).
52. D. Papkovsky, M. A. Uskova, G. V. Ponomarev, T. Korpela, S. Kulmala, and G. G. Guilbault, *Anal. Chim. Acta* 374, 1 (1998).
53. R. Czolk, *Sens. Actuators B* 30, 61 (1996).
54. E. Wang, C. Romero, D. Santiago, and V. Syntilas, *Anal. Chim. Acta* 433, 89 (2001).
55. W. Zhang, E. Malinowska, P. Parzuchowski, and M. E. Meyerhoff, *Anal. Chem.* 74, 4548 (2002).
56. X.-B. Zhang, Z.-Z. Li, C.-C. Guo, S.-H. Chen, G.-L. Shen, and R.-Q. Yu, *Anal. Chim. Acta* 439, 65 (2001).
57. W. E. Morf, "The Principle of Ion Selective Electrodes and of Membrane Transport." Elsevier, New York, 1981.
58. E. Bakker, P. Bühlmann, and E. Pretsch, *Chem. Rev.* 97, 3083 (1997).
59. N. A. Chaniotakis, A. M. Chasser, M. E. Meyerhoff, and J. T. Groves, *Anal. Chem.* 60, 185 (1988).
60. D. Gao, J. Gu, R.-Q. Yu, and G.-D. Zheng, *Anal. Chim. Acta* 302, 263 (1995).
61. E. Bakker, E. Malinowska, R. D. Schiller, and M. E. Meyerhoff, *Talanta* 41, 881 (1994).
62. E. Malinowska and M. E. Meyerhoff, *Anal. Chim. Acta* 300, 33 (1995).
63. U. Schaller, E. Bakker, U. E. Spichiger, and E. Pretsch, *Anal. Chem.* 66, 391 (1994).
64. S. Amemiya, P. Bühlmann, E. Pretsch, B. Rusterholz, and Y. Umezawa, *Anal. Chem.* 72, 1618 (2000).
65. E. D. Steinle, S. Amemiya, P. Bühlmann, and M. E. Meyerhoff, *Anal. Chem.* 72, 5766 (2000).
66. P. Bühlmann, E. Pretsch, and E. Bakker, *Chem. Rev.* 98, 1593 (1998).
67. E. Malinowska, J. Niedzìólka, E. Rozniecka, and M. E. Meyerhoff, *J. Electroanal. Chem.* 514, 109 (2001).
68. J. H. Khorasani, M. K. Amini, H. Motaghi, S. Tangestaninejad, and M. Moghadam, *Sens. Actuators B* 87, 448 (2002).
69. E. Malinowska, L. Górski, and M. E. Meyerhoff, *Anal. Chim. Acta* 468, 133 (2002).
70. S. Amemyia, P. Bühlmann, Y. Umezawa, R. C. Jagessar, and D. H. Burns, *Anal. Chem.* 71, 1049 (1999).
71. K.-C. Oh, K.-A Kim, I. R. Paeng, D. Baek, and K.-J. Paeng, *J. Electroanal. Chem.* 468, 98 (1999).
72. K.-C. Oh, S. M. Lim, I. R. Paeng, D. Baek, and K.-J. Paeng, *J. Electroanal. Chem.* 506, 42 (2001).
73. E. Malinowska, J. Niedziolka, and M. E. Meyerhoff, *Anal. Chim. Acta* 432, 67 (2001).
74. I. J. Yoon, J. H. Shin, I. R. Paeng, H. Nam, G. S. Cha, and K.-J. Paeng, *Anal. Chim. Acta* 367, 175 (1998).
75. M. K. Amini, S. Shahrokhian, and S. Tangestaninejad, *Anal. Chem.* 71, 2502 (1999).
76. M. K. Amini, S. Shahrokhian, and S. Tangestaninejad, *Analyst* 124, 1319 (1999).
77. X.-B. Zhang, C.-C. Guo, J.-B. Xu, G.-L. Shen, and R. Q. Yu, *Analyst* 125, 867 (2000).
78. I. H. A. Badr, M. E. Meyerhoff, and S. S. M. Hassan, *Anal. Chim. Acta* 321, 11 (1996).
79. V. K. Gupta, A. K. Jain, L. P. Singh, and U. Khurana, *Anal. Chim. Acta* 355, 33 (1997).
80. V. K. Gupta, A. Kumar, and R. Mangla, *Sens. Actuators B* 76, 617 (2001).
81. A. R. Fakhari, M. Shamsipur, and K. Ghanbari, *Anal. Chim. Acta* 460, 177 (2002).
82. C. Sun, J. Zhao, H. Xu, Y. Sun, X. Zhang, and J. Shen, *Talanta* 46, 15 (1998).
83. E. M. Bruti, M. Giannetto, G. Mori, and R. Seeber, *Electroanalysis* 11, 565 (1999).
84. R. Volf, T. V. Shishkanova, P. Matejka, M. Hamplová, and V. Král, *Anal. Chim. Acta* 381, 197 (1999).
85. R. Volf, T. V. Shishkanova, and V. Král, *J. Incl. Phen. Macroc. Chem.* 35, 111 (1999).
86. T. Malinski and Z. Taba, *Nature* 358, 676 (1992).
87. T. Malinski, in "The Porphyrin Handbook" (K. M. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 6, p. 231. Academic Press, New York, 2000.
88. F. Bedoui and N. Villeneuve, *Electroanalysis* 15, 5 (2003), and references therein.
89. T. Malinski, J. Nanobashvili, and I. Huk, *Eur. Surg.* 34, 79 (2002).
90. F. Bedioui, S. Trevin, V. Albin, M. G. G. Villegas, and J. Devynck, *Anal. Chim. Acta* 341, 177 (1997).

91. S. Trevin, F. Bedioui, and J. Devynck, *Talanta* 43, 303 (1996).
92. A. Ciszewski, E. Kubaszewski, and M. Lozinski, *Electroanalysis* 8, 293 (1996).
93. A. Ciszewski, G. Milczarek, E. Kubaszewski, and M. Lozinski, *Electroanalysis* 10, 628 (1998).
94. A. Ciszewski and G. Milczarek, *Electroanalysis* 10, 791 (1998).
95. M. Pontié, F. Bedioui, and J. Devinck, *Electroanalysis* 11, 845 (1999).
96. K. M. Mitchell and E. K. Michaelis, *Electroanalysis* 10, 81 (1998).
97. K. Araki, L. Angnes. C. M. N. Azevedo, and H. E. Toma, *J. Electroanal. Chem.* 397, 205 (1995).
98. J. R. Caetano da Rocha, L. Angnes, M. Bertotti, K. Araki, and H. E. Toma, *Anal. Chim. Acta* 452, 23 (2002).
99. D. J. Dobson and S. Saini, *Anal. Chem.* 69, 3532 (1997).
100. J. A. J. Brunink, C. Di Natale, F. Bungaro, F. A. M. Davide, A. D'Amico, R. Paolesse, T. Boschi, M. Faccio, and G. Ferri, *Anal. Chim. Acta* 325, 53 (1996).
101. C. Di Natale, A. Macagnano, G. Repole, A. D'Amico, R. Paolesse, and T. Boschi, *Mater. Sci. Eng. C* 5, 209 (1998).
102. C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, C. Goletti, and A. D'Amico, *Sens. Actuators B* 52, 162 (1998).
103. A. D'Amico, C. Di Natale, R. Paolesse, A. Macagnano, and A. Mantini, *Sens. Actuators B* 65, 209 (2000).
104. C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, P. Mari, and A. D'Amico, *Sens. Actuators B* 68, 319 (2000).
105. C. Di Natale, R. Paolesse, A. Macagnano, V. I. Troitsky, T. S. Berzina, and A. D'Amico, *Anal. Chim. Acta* 384, 249 (1999).
106. R. Paolesse, C. Di Natale, A. Macagnano, F. Davide, and T. Boschi, A. D'Amico, *Sens. Actuators B* 47, 70 (1998).
107. R. Paolesse, D. Monti, L. La Monica, M. Venanzi, A. Froiio, S. Nardis, C. Di Natale, E. Martinelli, and A. D'Amico, *Chem. Eur. J.* 8, 2476 (2002).
108. R. Paolesse, C. Di Natale, V. Campo Dall'Orto, A. Macagnano, A. Angelaccio, N. Motta, A. Sgarlata, J. Hurst, I. Rezzano, M. Mascini, and A. D'Amico, *Thin Solid Films* 354, 245 (1999).
109. K. Ariga, K. Endo, Y. Aoyama, and Y. Okahata, *Colloids Surfaces A* 169, 177 (2000).
110. C. Caliendo, P. Verardi, E. Verona, A. D'Amico, C. Di Natale, G. Saggio, M. Serafini, R. Paolesse, and S. E. Huq, *Smart Mater. Struct.* 6 (1997).
111. D. M. Oglesby, B. T. Upchurch, B. D. Leighty, J. P. Colmann, X. Zhang, and P. C. Hermann, *Anal. Chem.* 66, 2745 (1994).
112. L. Sun, C. Gu, K. Wen, X. Chao, T. Li, G. Hu, and J. Sun, *Thin Solid Films* 210–211, 486 (1992).
113. C. Gu, L. Sun, T. Zhang, and T. Li, *Thin Solid Films* 284–285, 863 (1996).
114. M. Andersson, M. Holmberg, I. Lundström, A. Lloyd-Spetz, P. Mårtensson, R. Paolesse, C. Falconi, E. Proietti, C. Di Natale, and A. D'Amico, *Sens. Actuators B* 77, 567 (2001).
115. A. D'Amico, C. Di Natale, R. Paolesse, A. Mantini, C. Goletti, F. Davide, and G. Filosofi, *Sens. Actuators B* 70, 254 (2000).
116. D. P. Arnold, D. Manno, G. Micocci, A. Serra, A. Tepore, and L. Valli, *Langmuir* 13, 5951 (1997).
117. G. Ashkenasy, A. Ivanisevic, R. Cohen, C. E. Felder, D. Cahen, A. B. Ellis, and A. Shanzer, *J. Am. Chem. Soc.* 122, 1116 (2000).
118. A. Ivanisevic, A. B. Ellis, G. Ashkenasy, A. Shanzer, and Y. Rosenwaks, *Langmuir* 16, 7852 (2000).
119. D. G. Wu, G. Ashkenasy, D. Shvarts, R. V. Ussyshkin, R. Naaman, A. Shanzer, and D. Cahen, *Angew. Chem. Int. Ed.* 39, 4496 (2000).
120. D. G. Wu, D. Cahen, P. Graf, R. Naaman, A. Nitzan, and D. Shvarts, *Chem. Eur. J.* 7, 1743 (2001).
121. K. C. Persaud and G. Dodd, *Nature* 299, 352 (1982).
122. C. Di Natale, J. A. J. Brunink, F. Bungaro, F. Davide, A. D'Amico, R. Paolesse, T. Boschi, M. Faccio, and G. Ferri, *Meas. Sci. Techn.* 7, 1 (1996).
123. C. Di Natale, A. Macagnano, F. Davide, A. D'Amico, R. Paolesse, T. Boschi, M. Faccio, and G. Ferri, *Sens. Actuators B* 44, 521 (1997).
124. C. Di Natale, A. Macagnano, A. Mantini, E. Tarizzo, A. D'Amico, R. Paolesse, T. Boschi, F. Sinesio, F. M. Bucarelli, E. Moneta, and G. B. Quaglia, *Sens. Actuators B* 50, 246 (1998).
125. F. Sinesio, C. Di Natale, G. Quaglia, F. Bucarelli, E. Moneta, A. Macagnano, R. Paolesse, and A. D'Amico, *J. Sci. Food Agric.* 80, 63 (2000).
126. A. D'Amico, C. Di Natale, A. Macagnano, F. Davide, A. Mantini, E. Tarizzo, R. Paolesse, and T. Boschi, *Biosens. Bioel.* 13, 711 (1998).
127. C. Di Natale, M. Zude-Sasse, A. Macagnano, R. Paolesse, B. Herold, and A. D'Amico, *Anal. Chim. Acta* 459, 107 (2002).
128. C. Di Natale, A. Macagnano, R. Paolesse, and A. D'Amico, *Biotechnol. Agron. Soc. Environ.* 5, 159 (2001).
129. C. Di Natale, G. Olafsdottir, S. Einarsson, E. Martinelli, R. Paolesse, and A. D'Amico, *Sens. Actuators B* 77, 572 (2001).
130. C. Di Natale, A. Macagnano, S. Nardis, R. Paolesse, C. Falconi, E. Proietti, P. Siciliano, R. Rella, A. Taurino, and A. D'Amico, *Sens. Actuators B* 78, 303 (2001).
131. C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, E. Proietti, and A. D'Amico, *Sens. Actuators B* 78, 26 (2001).
132. C. Di Natale, A. Macagnano, E. Martinelli, E. Proietti, R. Paolesse, L. Castellati, S. Campani, and A. D'Amico, *Sens. Actuators B* 77, 561 (2001).
133. C. Di Natale, A. Macagnano, R. Paolesse, A. Mantini, E. Tarizzo, and A. D'Amico, *Sens. Actuators B* 65, 216 (2000).
134. C. Di Natale, A. Mantini, A. Macagnano, D. Antuzzi, R. Paolesse, and A. D'Amico, *Physiol. Measur.* 20, 1 (1999).
135. A. Mantini, C. Di Natale, A. Macagnano, R. Paolesse, A. Finazzi-Agro', and A. D'Amico, *Critical Rev. Biomed. Eng.* 28, 481 (2000).
136. C. Di Natale, E. Martinelli, R. Paolesse, R. Baciocchi, G. D'Arcangelo, C. Roscioni, A. Finazzi-Agro', and A. D'Amico, *Biosens. Bioelec.*, in press.
137. C. Di Natale, D. Salimbeni, R. Paolesse, A. Macagnano, and A. D'Amico, *Sens. Actuators B* 65, 220 (2000).
138. M. M. Salleh, Akrajas, and M. Yahaya, *Thin Solid Films* 417, 162 (2002).
139. N. A. Rakow and K. S. Suslick, *Nature* 406, 710 (2000).
140. K. S. Suslick and N. A. Rakow, U.S. Patent 6, 368, 558, 2002.
141. K. S. Suslick and N. A. Rakow, in "Artificial Chemical Sensing" (J. R. Stetter and W. R. Penrose, Eds.), p. 8. Electrochem. Soc. Pennington, NJ, 2001.
142. M. E. Kosal, W. B. McNamara III, A. Sen, and K. S. Suslick, in "IX ISOEN," Rome, 2002.
143. M. Otto and J. D. R. Thomas, *Anal. Chem.* 57, 2647 (1985).
144. H. Hayashi, M. Yamanaka, K. Toko, and K. Yamafuji, *Sens. Actuators B* 2, 205 (1990).
145. C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, L. Lvova, A. Rudnitskaya, and Y. Vlasov, *Sens. Actuators B* 64, 15 (2000).
146. C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, M. Ubigli, A. Legin, L. Lvova, A. Rudnitskaya, and Y. Vlasov, *Sens. Actuators B* 69, 342 (2000).
147. C. Di Natale, R. Paolesse, M. Burgio, E. Mazzone, and A. D'Amico, *Sens. Actuators B* in press.
148. T. Pearce, *Biosystems* 41, 43 (1997).

149. B. R. Kowalski and S. Wold, in "Handbook of Statistics" (P. R. Krishnaiah and L. N. Kanal, Eds.), Vol. 2, p. 673, North-Holland, Amsterdam, 1982.
150. A. Hierlemann, M. Schweizer, U. Weimar, and W. Göpel, in "Sensors Update" (W. Göpel, J. Hesse, and H. Baltes, Eds.), Vol. 2, VCH, Weinheim, 1995.
151. E. L. Hines, E. Llobet, and J. W. Gardner, *IEE Proc. Circuits Devices Syst.* 146, 297 (1999).
152. P. C. Jurs, G. A. Bakken, and H. E. McClelland, *Chem. Rev.* 100, 2649 (2000).
153. C. Di Natale, F. Davide, and A. D'Amico, *Sens. Actuators B* 23, 111 (1995).

www.aspbs.com/enn

Preparation of Vesicles (Liposomes)

Peter Walde

ETH Zürich, Zürich, Switzerland

CONTENTS

1. What are Vesicles (Liposomes)?
2. Vesicles and the Liquid Crystalline States of Surfactants
3. Methods for Preparing Normal Vesicles
4. Preparation of Reversed Vesicles
5. Characterizations and Applications of Vesicles
6. Concluding Remarks
Glossary
References

1. WHAT ARE VESICLES (LIPOSOMES)?

1.1. Introduction

Vesicles—more precisely "*normal* vesicles"—are a particular type of polymolecular aggregate (polymolecular assembly) of certain amphipathic molecules, formed in aqueous solution. A vesicle is composed of one or more closed shells (which are usually 4–5 nm thick) that entrap a small volume of the aqueous solution in which the vesicle is formed. Vesicles are often spherical (under osmotically balanced conditions) and can have diameters between about 20 nm and more than 0.1 mm. If an analogous type of aggregate is formed in a water-immiscible, apolar solvent, the aggregate is called a *reversed vesicle* (see Section 4). In the following, the term "vesicle" or "lipid vesicle" always stands for a normal type of vesicle and *not* for a reversed vesicle.

For the sake of simplicity, a so-called *uni*lamellar vesicle is first considered. It is a closed lamella with an aqueous interior. The lamella is composed of amphipathic molecules, compounds that comprise at least two opposing parts, a hydrophilic part (which is soluble in water) and a hydrophobic part (which is not soluble in water but is soluble in an organic solvent that is not miscible with water, in this context also called "oil"). Amphipathic molecules have a "sympathy" as well as an "antipathy" for water. Because of the mixed affinities within the same chemical structure, amphipathic molecules are also called *amphiphiles* (meaning "both loving," water as well as oil). They are *surfactants*, which stands for "surface active agents" and means that they accumulate at the surface of liquids or solids. The accumulation of surfactant molecules on the surface of water (at the water-air interface) leads to a reduction in the surface tension of water, as a result of an alteration of the hydrogen bonds between the interfacial water molecules.

The aqueous solution in which vesicles form is present outside of the vesicles as well as inside. Figure 1 is a schematic representation of a unilamellar spherical vesicle formed by the amphiphile POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) in water at room temperature. The vesicle drawn in Figure 1 is assumed to have an outer diameter of 100 nm. The bilayer has a thickness of only about 4 nm, which corresponds in a first approximation to about two times the length of an extended POPC molecule.

Lipid vesicles in the size range of 0.1 μm (=100 nm) can be visualized by electron microscopy, for example, by the freeze-fracture technique [1, 2], or by cryofixation [2, 3] (see Fig. 2A and B). For vesicles in the micrometer range, light microscopy can be applied [4] (see Fig. 2C).

Based on simple geometric considerations, one can calculate the approximate number of lipid molecules present in a particular defined vesicle. In the case of the unilamellar vesicle shown in Figure 1 (outer diameter 100 nm), about 8.1×10^4 POPC molecules form the shell of one vesicle, and all of these molecules are held together by noncovalent bonds. The single lamella of the giant vesicle shown in Figure 2C (diameter ~ 60 μm) contains about 2.6×10^{10} POPC molecules. The shell is a molecular bilayer with an arrangement of the POPC molecules in such a way that the *hydrophobic* ("water-hating") acyl chains are in the interior of the bilayer and the *hydrophilic* ("water-loving") polar head groups are on the two outer sites of the bilayer, in direct contact with either the trapped water inside the vesicle or with the bulk water in which the vesicle is dispersed. Since the bilayer shell in a sphere is necessarily curved, the number of amphiphiles constituting the inner layer is expected to be smaller than the number of amphiphiles present in the outer layer. For the 100-nm vesicle of Figure 1, the calculated number of POPC molecules is 3.74×10^4 in the inner layer and 4.36×10^4 in the outer layer, assuming a mean head group area of one POPC molecule of 0.72 nm^2 [5]

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

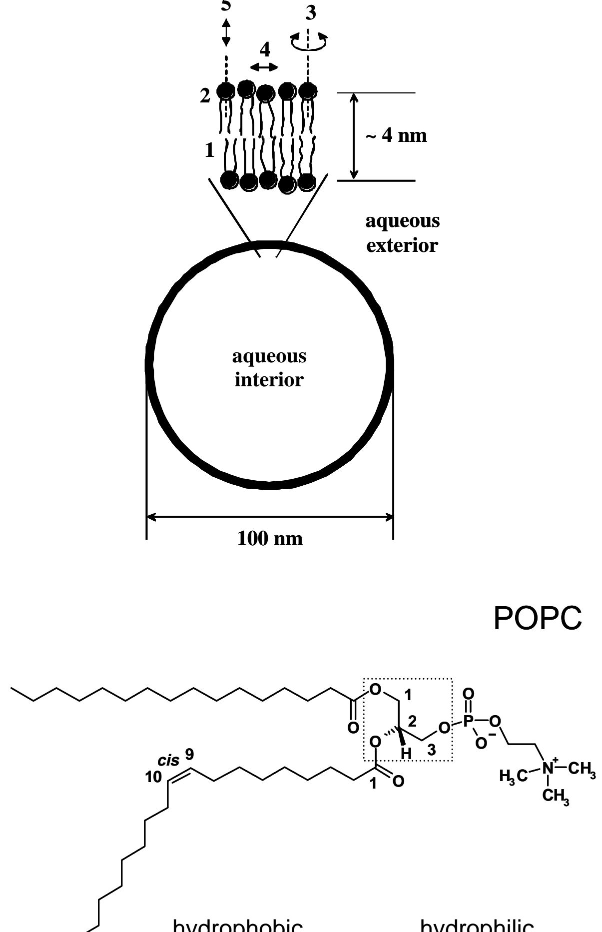


Figure 1. Schematic representation of the cross section through a particular unilamellar, spherical vesicle that has an assumed outer diameter of 100 nm and is formed in water from the surfactant POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine). POPC constituting the single, closed lamellar shell of the vesicle is represented as a filled circle to which two tails are connected. The tails stand for the two hydrophobic (water-insoluble) chains of POPC, and the filled circle symbolizes the hydrophilic (water-soluble) phosphocholine head group. POPC in the vesicle bilayer is present in the fluid-disordered state, above T_m (see Section 2.3). The chemical structure of POPC, which is a naturally occurring glycerophospholipid, is also shown. The glycerol moiety with its stereospecifically numbered carbon atoms is localized inside the dotted rectangle. The important molecular motions above T_m are indicated: (1) conformational transitions in the hydrophobic tails; (2) conformational transitions in the head group; (3) rotational diffusion about the axis perpendicular to the surface of the bilayer; (4) lateral diffusion within the bilayer plane; (5) vertical vibrations, out of the bilayer plane; and (not shown) collective undulations of the membrane. See text for nomenclature details and [392] for a detailed description of the vesicle membrane dynamics.
S

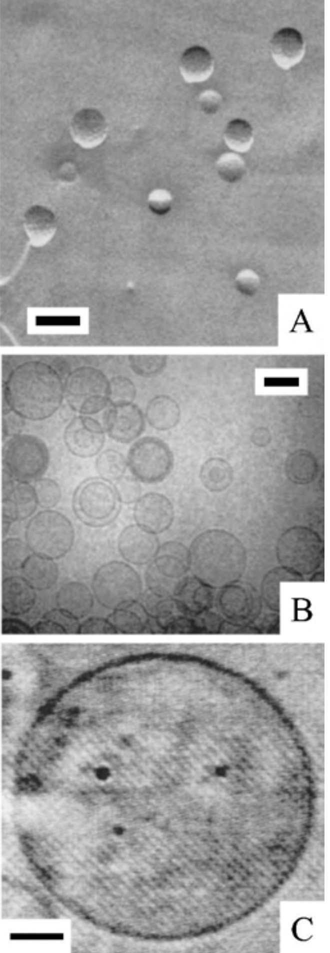


Figure 2. Electron and light microscopic visualization of vesicles prepared from POPC in water at ~ 25 °C. (A and B) Transmission electron micrographs of a suspension containing LUVs (FAT-VET$_{100}$), prepared by the extrusion technique (see Section 3.8 and Fig. 3). The samples were analyzed by the freeze-fracture technique (A) and by cryo-fixation (B). The length of the bars corresponds to 100 nm. (C) Light micrograph (differential interference contrast mode) of a single GUV prepared by the electroformation method (see Section 3.3). Length of the bar: 10 μm. The electron micrographs were taken by M. Müller, E. Wehrli, and N. Berclaz, Service Laboratory for Electron Microscopy, in the Department of Biology at the ETH Zürich. The light micrograph was taken by R. Wick at the Department of Materials Science at the ETH Zürich.

and a bilayer thickness of 3.7 nm [6]. This is certainly a rough estimation, and the real situation in a curved bilayer is always asymmetric with different packing conditions (and mean surfactant head group areas) in the inner and outer layers [7, 8]. Whether vesicles are thermodynamically stable or not (see Sections 2 and 6) depends critically on whether the chains are flexible enough to accommodate these asymmetric constraints [7].

The mean head group area of a surfactant, abbreviated a_0, corresponds to the area that is occupied on average by the polar head group packed within the vesicle's bilayer as a consequence of the actual molecular size and the interaction between the neighboring surfactant molecules. A simple geometric model that, in addition to a_0, takes into account the overall geometric shape of the surfactant as a critical packing parameter (or "shape factor"), p $(=v/(a_0 l_c))$, has been developed [7–10] and successfully applied in a useful simple theory toward an understanding of molecular self-assembling systems at large [11]; v is the volume of the hydrophobic portion of the molecule and l_c is the critical

length of the hydrophobic tails, effectively the maximum extent to which the chains can be stretched out. According to this simple theory, lipid bilayers and vesicles can be prepared if p for a particular amphiphile has a value between 0.5 and 1.0 (conditions under which the surfactants will pack into flexible, curved, or planar bilayers).

1.2. Terminology

There are a large number of amphiphiles that form vesicles. The most intensively investigated are certain lipids present in biological membranes, glycerophospholipids, lipids that contain a glycerol-3-phosphate unit. Actually, the geometric structure of vesicles as *spherulites*—which is the term originally used [12]—containing one (or more) concentric bilayer shell(s) was elaborated for the first time with vesicle preparations made from (mixtures of) naturally occurring glycerophospholipids [12, 13]. This is why it has been proposed to call these aggregates *lipid vesicles* [14] or *liposomes* (actually meaning "fat bodies") [15, 16]. Since the type of aggregate shown in Figures 1 and 2 has not so much to do with a "fat body," it is in principle more appropriate to use the term *lipid vesicle* or just *vesicle* instead of *liposome* [17]. In any case, all of the terms *liposome, lipid vesicle,* and *vesicle* are used here for the same type of polymolecular aggregate. Sometimes, however, one also finds the term *synthetic vesicles* [18, 19], referring to vesicles formed by synthetic, often charged, nonnatural surfactants. Others use the term *vesicle* exclusively for a closed *unilamellar* (not multilamellar) aggregate of amphiphiles [20, 21], such as the one shown schematically in Figure 1. Other names appearing in the literature are *niosomes* (vesicles prepared from nonionic surfactants) [22, 23], *polymer vesicles* or *polymersomes* (vesicles prepared from polymeric surfactants) [24–26], and so on. Table 1 summarizes different terms to describe a particular type of vesicle. Whenever one uses one of these terms, one should specify how it is used and how it is actually defined, to avoid any possible confusion or misunderstanding.

1.3. Nomenclature and Chemical Structures of Vesicle-Forming Glycerophospholipids

Although the basic principles of vesicle formation are for all types of vesicle-forming amphiphiles in the end the same—independently of whether they are charged, neutral, or polymeric—the descriptions in Sections 2 and 3 focus on just one particular group of phospholipids, the so-called phosphatidylcholines (PCs). POPC (see Fig. 1) is a particular PC, namely 1-palmitoyl-2-oleoyl-phosphatidylcholine, more precisely, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (also called 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphorylcholine). Since the nomenclature of glycerophospholipids (such as POPC) is often not so well known to those not working with this particular type of biological molecule, a short overview of the main nomenclature rule is given next. The nomenclature of lipids outlined and used here has been proposed by the International Union of Pure and Applied Chemistry–International Union of Biochemistry (IUPAC-IUB) [27, 28].

POPC is a chiral phospholipid with one chiral center at that carbon atom that is localized in the middle of the glycerol moiety; POPC belongs to the *glycerophospholipids*, the quantitatively most important structural group within the class of phospholipids. Glycerophospholipids have a *glycerol* backbone to which a *phosphate* group is bound through a phosphoric acid ester linkage to one of the glycerol hydroxyl groups. To designate the configuration of this glycerol derivative, the carbon atoms of the glycerol moiety are *numbered stereospecifically* (indicated in the chemical name as prefix -*sn*-). If the glycerol backbone is written in a Fischer projection (see, for example, [29]) in such a way that the three carbon atoms are arranged vertically and the hydroxyl group connected to the central carbon atom is pointing to the left, then the carbon atom on top is C-1, the carbon atom in the middle is C-2 (this is the actual chiral center common to all glycerophospholipids), and the carbon atom at the bottom is C-3. With this convention, the chemical structure of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine is defined. It is the naturally occurring form of the two possible enantiomers. The mirror image of POPC is 2-oleoyl-3-palmitoyl-*sn*-glycero-1-phosphocholine. In the Cahn-Ingold-Prelog nomenclature system (also called the R/S convention) [29], the configuration of C-2 in POPC is R. (Another shorthand description of POPC sometimes used is PamOleGro*P*Cho or PamOlePtdCho [27, 28]; Pam stands for palmitoyl, Ole for oleoyl, Gro for glycerol, P for phosphate, Cho for choline, and Ptd for phosphatidyl).

All naturally occurring diacylglycerophospholipids that have a zwitterionic phosphocholine head group are also called 3-*sn*-phosphatidylcholines, or just *phosphatidylcholines*. In the biochemical and biophysical literature, phosphatidylcholine is also called *lecithin* (or L-α-lecithin because of its stereochemical relationship to the naturally occurring L-α-glycerol-phosphate). Therefore, POPC is a lecithin. The name *lecithin* implies that egg yolk (*lekithos* in Greek) contains large amounts of phosphatidylcholines. According to the IUPAC-IUB, the use of the term *lecithin* is permitted but not recommended [27]. It is indeed better to avoid using this term since it may have another meaning in the food technology literature. The International Lecithin and Phospholipid Society (ILPS) of the American Oil Chemists Society (AOCS) defines *lecithin* as a mixture of lipids obtained from animal, vegetable, or microbial sources, which includes PCs but also contains a variety of other substances, such as sphingosylphospholipids, triglycerides, fatty acids, and glycolipids [30].

Egg yolk is one of the cheapest commercial sources for the isolation of phosphatidylcholines; the other one is soybeans. Phosphatidylcholines from egg yolk contain a number of chemically different phosphatidylcholines. All of these PCs have, however, the same glycerol backbone and the same polar head group (phosphocholine). They only differ in the acyl chains esterified with the glycerol hydroxyl groups at C-1 and C-2 (see Table 2) [31].

Many studies on the preparation and characterization of lipid vesicles have been carried out with egg yolk PC. On average, in position C-1 in egg yolk PC is often palmitic acid, and in position C-2, oleic acid [31]. POPC is therefore a representative PC molecule for egg yolk PCs. In contrast to egg yolk PCs, however, POPC has a well-defined chemical

Table 1. List of some of the terms used to describe a particular type of vesicle.

Term	Meaning and use of the term in the literature
Algosome	Vesicle prepared on the basis of 1-*O*-alkylglycerol [485].
Archaesome	Vesicles prepared from archaebacterial, bolaamphiphilic lipids [359, 486].
Bilosome	Vesicle prepared from a particular mixture of non-ionic surfactants (1-monopalmitoyl-glycerol), cholesterol, dihexadecylphosphate (5:4:1 molar ratio), and bile salt (particularly deoxycholate) [487].
Catanionic vesicle	Vesicle prepared from a mixture of a cationic and an anionic surfactant [59, 488].
Cerasome	Vesicle with a silicate framework on its surface [366, 489].
Ethosome	Vesicle that contains in the final preparation a considerable amount of ethanol (prepared by a particular method described in Section 3.20) [229–231].
Fluorosome	SUV containing a fluorescent dye embedded in its bilayer to monitor the entry of molecules into the bilayer [490–492].
Hemosome	Hemoglobin-containing vesicle [493].
Immunoliposome	Vesicle as a drug delivery system that contains on the external surface antibodies or antibody fragments as specific recognition sites for the antigen present on the target cells [69, 373, 494].
Lipid vesicle	Vesicle prepared from amphiphilic lipids [31, 69].
Liposome	Vesicle prepared from amphiphilic lipids [31, 69].
Magnetoliposome	Vesicle containing magnetic nanoparticles (e.g., magnetite Fe_3O_4) [495–498].
Marinosome	Vesicle based on a natural marine lipid extract composed of phospholipids (PCs and phosphatidylethanolamines) containing a high amount (~65%) polyunsaturated aycl chains [499].
Niosome	Vesicle prepared from non-ionic surfactants [23, 500]. In some cases, at room temperature polyhedral niosomes exist, which transform into spherical niosomes upon heating, cholesterol addition, or sonication [501–503].
Novasome	Oligo- or multilamellar vesicle prepared by a particular technology that involves the addition of vesicle-forming surfactants in the liquid state (at high temperature) to an aqueous solution (Section 3.10) [185].
Phospholipid vesicle	Vesicle prepared from (amphiphilic) phospholipids [71].
PLARosome	Phospholipid-alkylresocinol liposome: phospholipid vesicle containing resorcinolic lipids or their derivatives [504].
Polymer vesicle	Vesicle prepared from polymeric amphiphiles, such as di- or triblock copolymers [24, 26].
Polymerized vesicle	Vesicle prepared from polymerizable amphiphiles that were (partially) polymerized after vesicle formation [33, 352, 353, 505].
Polymersome	Vesicle prepared from polymeric amphiphiles, such as di- or triblock copolymers [24, 26].
Proliposomes	A preparation that upon mixing with an aqueous solution results in the formation of vesicles. The preparation contains vesicle-forming amphiphiles and an alcohol (ethanol, glycerol, or propyleneglycol) (see Section 3.19). Dry (ethanol-free) granular preparations of vesicle-forming amphiphiles, which upon hydration lead to vesicle formation, are also called proliposomes (Section 3.19) [506].
Proniosomes	A dry, granular product containing mainly (but not exclusively) non-ionic surfactants which, upon the addition of water, disperses to form MLVs [507].
Reversed vesicle	Inverted vesicle formed in a water-immiscible, apolar solvent in the presence of a small amount of water (Section 4) [508].
Spherulite	Onion-like vesicle prepared with spherulite technology, which involves the use of shear forces (Section 3.11).
Sphingosome	Vesicle prepared on the basis of sphingolipids present in human skin [69, 509].
Stealth liposome	Sterically stabilized vesicle, achieved through the use of co-amphiphiles that have PEG (poly(ethyleneglycol))-containing hydrophilic head groups [510–512]. Stealth liposomes are not so easily detected and removed by the body's immune system (they are long-circulating in the blood). The name stems from an analogy to the American "Stealth bomber" aircraft, which is not easily detected by radar. Alternatively to PEG, polysaccharides have also been used [373].
Synthetic vesicle	Vesicle prepared from synthetic surfactants (surfactant mixtures) that are not present in biological membranes. The surfactants usually have a single hydrophobic tail [17, 488].
Toposome	Vesicle that has a surface that is site-selectively (toposelectively) modified in a stable manner at specific and deliberate locations (e.g., through chemical modifications or chemical functionalizations) [513].
Transfersome	Ultradeformable ethanol-containing mixed lipid/detergent vesicle claimed to transfer water-soluble molecules across human skin (Section 3.28) [275, 276].
Ufasome	Vesicle prepared from unsaturated fatty acid/soap mixtures [330].
Vesicle	General term to describe any type of hollow, surfactant-based aggregate composed of one or more shells. In the biological literature, the term *vesicle* is used for a particular small, membrane-bounded, spherical organelle in the cytoplasm of an eukaryotic cell [97].
Virosome	Vesicle containing viral proteins and viral membranes, reconstituted from viral envelopes, the shells that surround the virus [69, 514–516].

Note: In this chapter, all of the terms listed in the table are called vesicles (or lipid vesicles), independent on the chemical structure of the amphiphiles (surfactants) constituting the vesicle shell(s).

structure. For more basic studies, POPC may be more suited than the egg yolk PC mixture. For applications, however, the cheaper egg yolk PCs may be advantageous.

Although lipid vesicles prepared from egg yolk PCs are similar in many respects to vesicles prepared from POPC, the properties of POPC vesicles at a particular fixed temperature may be very different from those of the chemically related DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine) vesicles, for example. The reason for this is outlined in Section 2.

Table 2. Main approximate fatty acid content in egg yolk and soybean PCs (see [31]).

		Relative abundance in phospholipids					
		Egg yolk			Soybeans		
Fatty acid (trivial name)	Abbreviation	Total (%)	at *sn*-1 (%)	at *sn*-2 (%)	Total (%)	at *sn*-1 (%)	at *sn*-2 (%)
Hexadecanoic acid (palmitic acid)	C 16:0	~35	~69	~2	~17	~34	
Octadecanoic acid (stearic acid)	C 18:0	~14	~26	~1		~8	
Cis-9-octadecenoic acid (oleic acid)	C 18:1*c*9	~27	~5	~49	~23	~30	~16
Cis, cis-9,12-octadecadienoic acid (linoleic acid)	18:2*c*9*c*12				~48	~24	~71
Cis, cis-6,9-octadecadienoic acid	18:2*c*6*c*9	~6		~11			
All *cis*-9,12,15-octadecatrienoic acid	18:3*c*9*c*12*c*15				~9	~4	~13
All *cis*-5,8,11,14,17-eicosapentaenoic acid	20:5*c*5*c*8*c*11*c*14*c*17	~4		~7			
All *cis*-4,7,10,13,16,19-docosahexaenoic acid	22:6*c*4*c*7*c*10*c*13*c*16*c*19	~13		~25			

Note: The abbreviation 16:0, for example, indicates that the linear fatty acid has 16 carbon atoms without any double bonds; 18:1*c*9 indicates that the linear fatty acid is composed of 18 carbon atoms with one *cis* double bond in position 9,10 (starting at position 9), where the carboxy C atom is carbon number 1.

1.4. There Are Not Only Unilamellar Vesicles

Vesicles are not only classified by the chemical structure of the molecules constituting the vesicle shell(s) as reported in Table 1, but also according to their size, lamellarity and morphology, and method of preparation (see Table 3). Small unilamellar vesicles (SUVs) have one lamella and diameters of less than about 50 nm. So-called large unilamellar vesicles (LUVs) have one lamella and diameters between about 50 nm and about 500 nm (see Fig. 1 and Fig. 2A and B). Giant vesicles (GVs) can be observed by light microscopy and have diameters of more than about 0.5–1 μm (Fig. 2C). Oligolamellar vesicles (OLVs) have a few and multilamellar vesicles (MLVs) have many concentrically arranged lamellae. Multivesicular vesicles (MVVs) contain nonconcentrically arranged vesicles within a larger vesicle.

As described in detail in Section 3, the preparation of vesicle suspensions generally involves the use of a particular technique, a particular preparation method. Depending on the technique applied, the vesicle suspensions are characterized by a certain degree of homogeneity, a certain mean size and mean lamellarity, and a certain trapped volume. The trapped volume is the aqueous volume that is encapsulated by the lipid vesicles, expressed as microliters of aqueous solution per micromole of lipid (=liters/mol). A trapped volume of 1 μl/μmol means that in a vesicle suspension containing 1 μmol lipid, for example, only 1 μl of the aqueous solution is trapped by the vesicles. The spherical unilamellar POPC vesicle shown in Figure 1 has a trapped volume of about 3 μl/μmol, as calculated based on simple geometric considerations. This means that in a vesicle suspension prepared from 10 mM POPC (=7.6 g/liter), only 30 μl out of 1 ml is trapped by the vesicles (3 volume %). The total lipid-water interfacial area in this vesicle suspension is 4.3×10^3 m^2!

Since, in many cases, the vesicle suspensions prepared by one particular technique are not further characterized with respect to mean size, size distribution, and lamellarity, the vesicles are just named according to the method used. Examples include REVs (reversed-phase evaporation vesicles, vesicles prepared by the so-called reversed-phase evaporation method), VETs (vesicles prepared by the so-called extrusion technique), etc.; see some entries in Tables 1 and 3.

2. VESICLES AND THE LIQUID CRYSTALLINE STATES OF SURFACTANTS

2.1. Introduction

Since in most cases, lipid vesicles can be considered as dispersions of a liquid crystalline state of a vesicle-forming surfactant, it is useful to give a short introduction to some of the liquid crystalline phases of amphiphilic molecules, particularly focusing on the so-called L_α- and L_β- or L'_β-phases, which are considered to be the relevant thermodynamic equilibrium states of most glycerophospholipids under the conditions in which vesicle formation is observed. For a recent excellent general review on surfactant liquid crystals, see [32].

Table 3. List of some of the abbreviations often used for a particular type of ve sicle.

Abbreviation	Meaning of the abbreviation	Characteristics
DRV	Dehydrated-rehydrated vesicle (or dried-reconstituted vesicle)	Vesicle prepared by the dehydration-rehydration method (Section 3.7) [31]
EIV	Ethanol-injected vesicle	Vesicle prepared by the ethanol injection method (Section 3.18)
FATMLV	Vesicle prepared by repeatedly freezing and thawing a MLV suspension	Equilibration and homogenization procedure (Section 3.6) [133]
v FPV	Vesicle prepared with a French press	Unilamellar vesicle or OLV prepared with a French press for vesicle size homogenization (Section 3.8) [72]
GUV	Giant unilamellar vesicle	Unilamellar vesicle with a diameter larger than about 500 nm
GV	Giant vesicle	Vesicle with a diameter larger than about 500 nm
IFV	Interdigitation-fusion vesicle	Vesicle prepared by the interdigitation-fusion method (Section 3.21) [232, 517]
LUV	Large unilamellar vesicle	Unilamellar vesicle with a diameter between about 50 nm and 500 nm
LUVET	Large unilamellar Vesicle prepared by the extrusion technique	Vesicles prepared by the extrusion technique are usually large and mainly unilamellar (Section 3.8)
MLV	Multilamellar vesicle	The vesicle contains several concentrically arranged lamellae osmotically stressed after formation because of an exclusion of solute molecules during their formation [76]
MVL	Multivesicular liposome	A large vesicle that contains internal, nonconcentrically arranged vesicular compartments; also called MVV [211, 212]
MVV	Multivesicular vesicle	A large vesicle that contains internal, nonconcentrically arranged vesicular compartments; also called MVL [210]
OLV	Oligolamellar vesicle	The vesicle contains a few concentrically arranged lamellae
REV	Reversed-phase evaporation vesicle	Vesicle prepared by the reversed-phase evaporation technique (Section 3.14)
RSE vesicle	Rapid solvent exchange vesicle	Vesicle prepared by the rapid solvent exchange method [238]
SPLV	Stable plurilamellar vesicle	Similar to MLV but not osmotically stressed after its formation [76]
SUV	Small (or sonicated) unilamellar vesicle	Unilamellar vesicle with a diameter of less than about 50 nm, as typically obtained by sonicating MLVs (Section 3.5)
ULV	Unilamellar vesicle	Vesicle with only one lamella
UV	Unilamellar vesicle	Vesicle with only one lamella
VET	Vesicle prepared by the extrusion technique	Vesicles prepared by the extrusion technique are usually large and mainly unilamellar (Section 3.8)

Note: The abbreviations are based on the size and morphology, the lamellarity, or method of preparation.

2.2. Lamellar Phase and "Gel Phase"

A liquid crystalline state (also called the "mesophase") of a substance is a state between a pure crystal (characterized by a high order of rigid molecules) and a pure liquid (characterized by rapid molecular motions of disordered molecules) [32]. There are dozens of different liquid crystalline states, all characterized by a different degree of molecular mobility and order. Liquid crystals can be produced either by heating a particular crystalline solid—called a "thermotropic liquid crystal"—or by dissolution of particular substances in a solvent—called a "lyotropic liquid crystal." Many surfactants in water form lyotropic liquid crystalline phases, such as the lamellar phase (L_{α}), the so-called gel phase (L_{β} or L'_{β}), the normal or reversed hexagonal phase (H_I or H_{II}), or one or more of the known cubic phases (Ia3d, Pn3m, Im3m). The type of phase formed depends on the chemical structure of the surfactant used and on the experimental conditions (such as concentration and temperature, or the presence of other compounds).

In the L_{α}-phase (also called the liquid-analogue [33] or liquid-disordered state [34, 35]), the surfactant molecules are arranged in bilayers, frequently extending over large distances (1 μm or more) [32]. The hydrophobic chains are rather disordered, with a lot of *gauche* conformations in the saturated hydrocarbon parts of the hydrophobic chains, making the bilayers fluid, characterized by fast lateral and rotational diffusions of the surfactant molecules, similar to a liquid. Comparable molecular motions are also present in the liquid-disordered state of vesicles. In the case of SUVs prepared from POPC, for example, the lateral diffusion coefficient seems to be on the order of $3–4 \times 10^{-8}$ cm^2/s, as determined between 5 °C and 35 °C [36]. The rotational correlation time is on the order of 10^{-9} to 2×10^{-8} s [37].

The L_{β} (or L'_{β}-)-phase of surfactant molecules (also called solid-analogue [33] or solid-ordered state [34, 35]) closely resembles the L_{α}-phase in the sense that the surfactant molecules are also arranged in bilayers. The viscosity is very high, however. This is a consequence of the rigidity of the individual surfactant molecules which are mostly present with all-*trans* conformations in the saturated hydrocarbon parts of the hydrophobic chains. The motion of the molecules is rather restricted, similar to the molecules in a crystal. To specify the relative arrangement of the lipid molecules, the gel phase may be abbreviated as L'_{β} or L_{β},

depending on whether the alkyl chains are tilted (P'_{β}) or not titled (L_{β}) with respect to the normal of the lipid bilayer. If it is not known whether the chains are tilted, L_{β} is often used as a general abbreviation.

The phase behavior of a number of phosphatidylcholines [38, 39] and a number of other lipids and lipid mixtures [38, 40] has been determined and reviewed.

2.3. Main Phase Transition Temperature T_m of Glycerophospholipids

In the case of conventional glycerophospholipids, either DPPC or POPC, the P'_{β}-phase is formed at thermodynamic equilibrium at temperatures at least 5–10° below a lipid specific temperature called the main phase transition temperature (or lamellar chain melting temperature) (T_m). T_m is also called the lamellar gel-to-liquid crystalline phase transition temperature and can be determined, for example, as the endothermic peak maximum in heating scans of differential scanning calorimetry (DSC) measurements [41–43]. Above T_m the lipids are in the L_{α}-phase.

Between the L_{β}- (or P'_{β}-) phase and the L_{α}-phase, an intermediate gel phase, called the ripple phase (abbreviated as P'_{β}) is often observed at high water content in the case of PCs [44]. This particular lipid phase takes its name from the fact that in freeze-fracture electron micrographs, a "ripple" structure can be seen if the lipid dispersion is rapidly frozen from the particular temperature interval in which the ripple phase is formed [44–47]. The transition from the "gel phase" to the "ripple phase" is called *pretransition*.

With respect to certain practical aspects in the methods for lipid vesicle preparations described in Section 3, the T_m value of the lipid (or the lipid mixtures) used is important to know. In the case of dilute POPC-water systems, for example, T_m is around −3 °C [48, 49]. If the water content is decreased below ~10 wt%, T_m increases above 0 °C, until it reaches a value of 68 °C in the anhydrous system [49].

A list of different T_m values for a number of dilute aqueous phosphatidylcholine systems (MLVs) is given in Table 4. For a more detailed list of T_m values, including other glycerophospholipids, see [38, 40, 50]. Please note that in the case of phospholipids with charged head groups, the T_m values depend on the degree of protonation and may depend considerably on the chemical nature of the counter-ions present [51]. Furthermore, measurements carried out with SUVs give values about 4–5° lower than the T_m values obtained from MLVs [43, 52, 53].

Table 4. Main $P_{\beta'}$-L_{α} phase transition temperature (T_m) values of dilute aqueous dispersions of certain common bilayer-forming phosphatidylcholines, data taken from [39] and [518] (for soybean PCs).

Phosphatidylcholine	T_m (°C)
DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine), 14:0/14:0	23.6 ± 1.5
DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine), 16:0/16:0	41.3 ± 1.8
DSPC (1,2-distearoyl-*sn*-glycero-3-phosphocholine), 16:0/16:0	54.5 ± 1.5
POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine), 16:0/18:1*c*9	−2.5 ± 2.4
SOPC (1-stearoyl-2-oleoyl-*sn*-glycero-3-phosphocholine), 18:0/18:1*c*9	6.9 ± 2.9
DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine), 18:1*c*9/18:1*c*9	−18.3 ± 3.6
Egg yolk PCs (see Table 2)	−5.8 ± 6.5
Soybean PCs (see Table 2)	−15 ± 5
Hydrogenated soybean PCs	51–52

Note: 16:0/18:1*c*9, for example, indicates that the linear acyl chain at *sn*-1 has 16 carbon atoms without any double bonds; the linear acyl chain at *sn*-2 has 18 carbon atoms with one *cis* double bond in position 9,10 (see also Table 2).

3. METHODS FOR PREPARING NORMAL VESICLES

3.1. Introduction

The thermodynamic equilibrium state of glycerophospholipids (and many other bilayer-forming amphiphiles) in water (or in a particular aqueous solution) is—probably under most experimental conditions—a stacked bilayer arrangement of the surfactant molecules, either as L_{α}-phase (above T_m) or as L_{β}-, P'_{β}- (or $P_{\beta'}$)-phase (below T_m) in equilibrium with excess aqueous phase (see Section 2).

Upon dispersing in an aqueous solution, vesicles generally form from an amphiphile (or a mixture of amphiphiles) that forms a lamellar L_{α}-phase at thermodynamic equilibrium. Depending on how the dispersion is actually prepared (in other words, which method or technique is applied), the vesicles formed by the dispersed amphiphiles are either very heterogeneous or rather homogeneous and are mainly small (below about 50 nm), mainly large (between about 50 nm and 500 nm), or mainly very large (above about 500 nm). It all depends on the lipid (or lipid mixture) used, on the aqueous solution, and particularly on the preparation method.

In the following, the principles of some of the best known and widely used methods for the preparation of lipid vesicle suspensions—mainly on a laboratory scale of a few milliliters up to about 100 ml—will be described. For each method, a different vesicle preparation with different typical general characteristics is obtained.

It is important to point out once more that in most of the cases the resulting vesicle suspension is not at thermodynamic equilibrium, but represents only a metastable, kinetically trapped state. The equilibrium phases are L_{α}, $P_{\beta'}$, L_{β} or L'_{β}, as discussed in Section 2. A particular vesicle dispersion prepared is therefore physically (with respect to vesicle size and lamellarity) not indefinitely stable; it may slowly transform into the thermodynamically most stable state (stacked bilayers), as a result of a so-called aging process [54, 55]. This aging may occur either through the fusion of vesicles or because of an exchange of amphiphiles that are not aggregated (free, monomeric surfactant) [55]. This latter process—called Ostwald ripening (in analogy to the corresponding process occurring in emulsion systems)—is expected to be particularly relevant in the case of synthetic short-chain amphiphiles with a high monomer solubility (10^{-8} M). In the case of DPPC (and probably also POPC), the monomer solubility is on the order of 10^{-10} M [56], which means that aging through an Ostwald ripening process is less likely in these cases.

Although often only metastable, vesicle suspensions may be stable for a prolonged period of time, for example, for days, weeks, or months, provided that the vesicle-forming amphiphiles are chemically stable during this period of time [57, 58].

In the presence of other amphiphilic molecules (cosurfactants), the situation may change, particularly if the cosurfactant tends to form micellar aggregates, characterized by a packing parameter $p \sim 1/3$ (relatively large head group cross-sectional area, a_0). In this case mixed micelles may exist at thermodynamic equilibrium (in equilibrium with cosurfactant monomers), if the micelle-forming surfactant is present to a large extent. Such mixed micellar systems are used as a starting solution in the case of the so-called detergent depletion method described in Section 3.27. Furthermore, there are also known cases where there appear to be thermodynamically stable vesicles (particularly composed of mixtures of surfactants) [59, 60] (see Section 6).

The presence of cholesterol (or other sterols)—molecules that are water insoluble and do not form vesicles alone—may also influence the properties of lipid vesicles, depending on the amount of cholesterol present [61, 62]. In the case of DPPC, for example, up to 33 mol% cholesterol, the T_m value of hydrated bilayers changes only slightly [63]. With increasing cholesterol concentration, however, the phase transition temperature is completely eliminated at 50 mol% (1:1 molar ratio of DPPC to cholesterol). The fluidity of the bilayer membrane is thereby changed, resulting in an increase in the fluidity below T_m of the PC and a decrease in fluidity above T_m, a state of the membranes that is intermediate between solid-ordered and liquid-disordered (see Section 2.3). This state is called liquid-ordered [34].

For all of the methods outlined in the following, more detailed descriptions can be found in the original literature cited. Furthermore, most of the generally known methods have already been summarized before—more or less extensively—in review articles or books about lipid vesicles (liposomes) [14, 31, 64–75].

Most of the methods can be roughly divided into two groups:

(i) Methods that are based on the simple swelling of initially dried, preorganized lipids and the mechanical dispersion and mechanical manipulation of the dispersed bilayers (Sections 3.2–3.12).

(ii) Methods that involve the use of (a) a cosolvent in which the lipids are soluble (Sections 3.14–3.26 and 3.30), (b) an additional non-bilayer-forming "helper amphiphile," a coamphiphile (Sections 3.27 and 3.30), or (c) certain ions that influence the initial aggregational state of the lipids (Sections 3.13 and 3.29). All three type of molecules control the assembly process of the bilayer-forming amphiphile in a particular way during the vesicle preparation process, and all three types of molecules may at the end be difficult to remove completely from the final vesicle suspensions.

3.2. MLVs, GUVs, or Myelin Figures Formed by Thin Lipid Film Hydration

One of the easiest ways to prepare a vesicle suspension is to disperse a dried lipid film in an aqueous solution [13, 76, 77]. The vesicle-forming and swellable [78] amphiphile is first dissolved in an organic solvent in which the amphiphile is soluble (usually chloroform). This solution is then placed inside a round-bottomed flask, and the solvent is completely removed by rotatory evaporation under reduced pressure, followed by high-vacuum drying overnight. The remaining amphiphiles form a dry thin film that is oriented in such a way as to separate hydrophilic and hydrophobic regions from each other [31]. If an aqueous solution is added to this film at a temperature above the main phase transition temperature T_m (see Section 2.3), the head groups of the dry lipids become hydrated and hydrated bilayers form. The hydration and swelling process is usually speeded up by gentle or vigorous shaking (using a vortex mixer), thereby dispersing the bilayers in the aqueous solution, resulting in the formation of mainly MLVs, which are very heterogenous with respect to size and lamellarity. Lipid film thickness and extent of shaking have an influence on the properties of the resulting vesicle suspension. The interlamellar repeat distance in equilibrated, completely hydrated PC bilayers above T_m is around 6.5 nm [79, 80]. This means that in a MLV formed from POPC as example, the aqueous space between two neighboring lipid lamellae is about 2.5 nm thick (taking into account a bilayer thickness of 4 nm). On average, a MLV may be composed of up to 10 bilayers [81].

The formation of closed bilayers (vesicles), in contrast to open bilayers, can be easily understood on the simple basis that interactions between the hydrophobic chains of the amphiphiles and the water molecules—as would be the case in open bilayers—are energetically unfavorable and therefore rather unlikely.

The preparation of MLVs by the dispersal of a dried lipid film (also called hand-shaken MLVs [31]) is often a first step in the preparation of more defined vesicle suspensions (see, for example, Sections 3.5 and 3.8). With respect to the equilibration of water-soluble molecules between the bulk aqueous medium and the inner aqueous compartments of MLVs, ionic species may be unevenly distributed [76]. A more even distribution can be achieved by applying freeze/thaw cycles (see Section 3.6).

The experimental conditions under which a dried lipid film is hydrated affect the resulting lipid aggregates obtained. In the case of phospholipid mixtures containing 90 mol% PC and 10 mol% of a negatively charged phospholipid (egg yolk phosphatidylglycerol, bovine brain phosphatidylserine, or bovine heart cardiolipin), the dried lipid film prepared inside a test tube can first be prehydrated at 45 °C with water-saturated nitrogen gas for 15–25 min. Afterward, an aqueous solution containing water-soluble molecules to be entrapped (e.g., 100 mM KCl and 1 mM $CaCl_2$) is gently added, and the tube is sealed under argon and incubated at 37 °C for 10–15 h. During this incubation, the lipid film is completely stripped from the glass surface and forms vesicular aggregates as a kind of white, floating precipitate in the aqueous solution. The analysis of this

precipitate indicated the presence of many mainly unilamellar giant vesicles (not MLVs) with diameters on the order of tens of micrometers up to more than 300 μm [82]. In addition, much smaller vesicles, large multilamellar vesicles as well as myelin figures (see below) and undispersed lipid material, could also be observed [82]. In the case of egg PC alone (no negatively charged phospholipids present), no giant unilamellar vesicles (GUVs) formed under the experimental procedure used [83]. It therefore seems that electrostatic repulsions between the charged head groups facilitate the formation of unilamellar membranes by opposing the intrinsic adhesive force between the membranes [83]. If divalent cations (1–30 mM Ca^{2+} or Mg^{2+}) are present, giant vesicle formation is also observed with zwitterionic phospholipids alone (POPC), with the use of a procedure almost identical to the one just described [84]. Divalent cations seem to promote giant unilamellar vesicle formation in the case of POPC due to a binding of the ions to the free phosphate oxygen of the lipid head group, which is known to alter the mean head group conformation [85] and the fluidity of the lipid bilayer [86], and which makes a zwitterionic PC molecule positively charged overall [87].

In an independent study and with a different experimental procedure, the formation of giant vesicles from PCs (DOPC or soybean PCs) in the presence of Mg^{2+} ($<$10 mM) has also been observed [87].

The general role of ions (including buffer ions)—as well as dissolved gas molecules—in vesicle formation and in physical chemistry at large is an open question [88] (see Section 6).

The preparation of MLVs by thin lipid film hydration generally involves the use of round-bottomed flasks and gentle shaking. If flat-bottomed flasks are used instead and if additionally the lipid swelling process occurs undisturbed (no shaking) above T_m of the lipids [89] over a period of several hours, the vesicles formed are no longer mainly multilamellar but rather mainly unilamellar, with diameters between 0.5 and 10 μm [90] or even at 300 μm [91]. These are GUVs.

The preparation of GUVs by simple hydration of certain amphiphiles deposited from an organic solution (e.g., a chloroform-methanol mixture) on a flat surface can be considerably improved to yield a higher fraction of unilamellar vesicles by using a roughened flat disk of Teflon (polytetrafluoroethylene) [92]. The vesicles formed (e.g., from DMPC after a slow swelling at a temperature of 30–35 °C) can be harvested by gentle pipette aspiration for further individual investigation and micromanipulation [92, 93]. The sizes of the vesicles thus formed are in the range of 20–40 μm [93, 94].

The slow swelling of dried phosphatidylcholines with the addition of water had already been observed and investigated by light microscopy in the middle and at the end of the nineteenth century by Rudolf Virchow [95] and Otto Lehmann [96]. The elongated, tubular structures that are observed to grow from a deposit of PC molecules (e.g., egg yolk PCs) with the addition of water under undisturbed conditions (no shaking) are called myelin figures, as named by Virchow [95] while making these observations with a (PC-containing) lipid extract from myelin, the isolating lipid sheath surrounding the elongated portion of nerve cells [97]. Myelin figures are structures dozens of micrometers long and a few micrometers thick. In the case of egg yolk PCs and 25 °C, water addition leads to the formation of myelin figures that have a diameter of about 20 μm and a length of several hundred micrometers [98]. Myelin figures are cylindrical, rodlike structures composed of many lamellae formed by the amphiphiles, stacked coaxially around the rod axis, with water between all of the bilayers [98, 99].

The conventional solid surface on which the lipid film is initially dried—glass, as described above—has been replaced by a support made from microcrystals (such as zeolite X with a crystallite size of 400 nm) [100]. Using very dilute mixtures of egg yolk PC and the positively charged amphiphile hexadecyltrimethylammonium bromide, the swelling in water (or 5 mM NaCl) of the very thin dried film deposited on zeolite X led neither to the formation of MLVs nor to GUVs or myelin figures, but to uniform SUVs with diameters around 22 nm [100]. In general, the size distribution can be controlled to some extent by the topography of the surface upon which the phospholipid film is deposited [101].

3.3. GUVs Prepared by the Electroformation Method

If the swelling of PCs (or other phospholipids) in water (or in an aqueous solution with low ionic strength) is carried out in a controlled way by applying an alternating electric field, instead of myelin figures, GUVs form, with diameters typically between 5 and 200 μm, depending on the chemical structure of the lipids used, on the lipid composition, on the swelling medium, and on the external electric field parameters [102–104]. The vesicles are often formed under a light microscope in a specially designed cell on platinum wires that are positioned at a certain fixed distance [103, 104]. This so-called electroformation method has been proved to be rather powerful for the investigation (including microinjection [105, 106]) of individual GUVs of defined size by light microscopy [103–110].

The mechanism of GUV electroformation in an alternating electric field is not fully understood. It is likely, however, that the electroosmotic motion of the water molecules is responsible for a controlled swelling and separation of the bilayers deposited on the platinum wires, leading to the formation of unilamellar vesicles on or close to the metal wires [103, 109]. The individual giant vesicles formed seem to be connected to the platinum wires through thin lipid bilayer tubes (so-called tethers [111, 112]) and possibly myelin-like protrusions [113].

In comparison with other methods for the preparation of giant unilamellar PC vesicles, the electroformation method offers several advantages with respect to reproducibility and later vesicle manipulations [114].

3.4. MLVs Prepared by Hydration of Spray-Dried Lipids

Instead of the preparation of a thin lipid film first, followed by hydration (Sections 3.2 and 3.3), the lipids can first be spray-dried and then hydrated [115]. With respect to the reproducibility and mass production of MLVs, the spray-drying method has several advantageous over the thin-film dispersion method, although it seems that lipids with low T_m

(e.g., egg yolk PCs) cannot be used, because of adhesion to the containers in which the vesicles are prepared [115].

In the spray-drying method, the vesicle-forming amphiphiles are first dissolved in chloroform or methylene chloride (in which, advantageously, mannitol is dispersed) and then spray-dried with a commercial spray-dryer. Vesicle formation is observed after hydration of the spray-dried lipids and vortexing above T_m of the amphiphile used. The presence of the mannitol prevents adhesion of the (saturated) lipids to the reaction containers and leads to a better hydration behavior. (The role of sugars in surfactant assembly at large has hardly been touched on.)

The vesicles prepared are rather heterogeneous with respect to size and lamellarity. One of the homogenization procedures described below may follow the spray-drying and hydration step.

3.5. SUVs (and Possibly LUVs) Prepared by Sonication (and Storage)

The treatment of a MLV suspension with ultrasound at a temperature at least about 5 °C above T_m leads to a homogenization of the vesicles by reducing the size of most of the vesicles to probably the smallest possible diameter (about 20 nm in the case of egg yolk PC mixtures [10, 116] or POPC), due to simple molecular packing considerations. The vesicles thus prepared are unilamellar and called SUVs, an abbreviation that stands for sonicated unilamellar vesicles or small unilamellar vesicles. If the sonication is performed below T_m, structural defects within the bilayers are observed that result in an increased bilayer permeability [117, 118].

So-called probe sonication (the insertion into the vesicle suspension of a metal rod that releases ultrasound waves from the tip of the rod) [119–121] is more efficient than bath sonication [31, 122]. In both cases, however, the sizes of the vesicles (20–50 nm usually) in the sonicated suspension depend on the sonication time [122], on the lipid composition, and on whether other compounds (such as cholesterol) are present [14, 119, 123]. Furthermore, the vesicle suspension may not be free from larger MLVs, and a separation step, for example, size exclusion chromatography (using, for example, sepharose 4B) or high-speed centrifugation, is usually necessary [119, 121, 124]. A centrifugation step may also be used in the case of probe sonication to remove the possibly released metal particles from the tip of the metal rod [125].

The ultrasound used in the treatment of MLVs consists of pressure waves with frequencies around 20 kHz, and the ultrasound propagation gives rise to periodic changes in local pressure and temperature [126]. Therefore, the heat generated during sonication has to be controlled and compensated for by cooling, otherwise the vesicle-forming lipids may undergo a partial chemical degradation [125, 127]. Furthermore, it has been reported that nonvesicular aggregates may be generated as a result of the high-energy input into the system, and the ultrasound treatment may lead to a drastic decrease in the surface tension of sonicated MLV suspensions due to the accelerated movement of lipid molecules form the vesicles to the water-air interface [128]. This process is usually much slower in the case of vesicles that have been prepared under milder conditions, such as by extrusion (Section 3.8) [128].

Based on this latter observation, more uniform and reproducible SUV preparations may be obtained by annealing at elevated temperatures (50 °C in the case of DMPC:1,2-dimyristoyl-*sn*-glycero-3-phosphoserine vesicles), followed by a removal of possibly present large aggregates by centrifugation at $100{,}000 \times g$ (30 min at 37 °C in the particular case) [129].

If carefully prepared SUVs from the saturated DPPC at a concentration in the range of 20–120 mM (in a 20 mM piperazine-N',N''-bis(2-ethanesulfonic acid) buffer, pH 7.4, containing 10 mM NaCl) are kept at a temperature considerably lower than the T_m of DPPC (which is around 41 °C; see Table 4), then the vesicles fuse into uniform LUVs with a diameter of about 70 nm (after 7 days at 4 °C) or 95 nm (after 3–4 weeks at 4 °C) [130]. At higher temperatures (21 °C) but still below T_m, the fusion process is slower [131], whereas at 50 °C (above T_m), no appreciable fusion occurs over a period of at least 5 days [130]. A similar fusion of SUVs below T_m is observed for DSPC, resulting in vesicles of about 60- and 100-nm diameter [132], with an apparently higher fusion rate [130].

The implications of these experimental observations are twofold: (a) LUVs of about 70 or 100 nm can be prepared from saturated PCs by simple storage of SUVs below T_m. (b) SUVs should not be stored below T_m, if one likes to keep the vesicle sizes small; otherwise the vesicles may fuse to form larger vesicles.

In general, SUVs are often prepared in a first step to fuse them in a second step to LUVs by using another methodology, for example, in the case of the cochleate-cylinder method (see Section 3.13) or in the case of the interdigitation-fusion method (see Section 3.21).

3.6. MLVs and MVVs (and Possibly LUVs) Prepared by Repetitive Freezing and Thawing

A MLV suspension—as prepared by the thin-film dispersion method—is put, for example, inside a thick-walled test tube and repetitively (3–6 or, better, 10 times) completely frozen in liquid nitrogen (at −195 °C) and then thawed in a water bath kept at a temperature above T_m. In this way, the vesicle suspension undergoes a sort of equilibration procedure caused by the water crystals and transient rigidification of the lipid molecules. It is a kind of homogenization process, and it has been reported that the population of MVVs may increase [133] and that the amount of very small vesicles tends to decrease [134, 135], depending on the lipid used and on the salt content [136]. It may also be that freeze-thaw cycles lead to a fragmentation of large MLVs into large or small unilamellar vesicles [136, 137]; it all very much depends on the experimental conditions, such as type and concentration of lipid (or lipid mixture) and salt content [136, 138].

A further effect of freezing and thawing of vesicle suspensions is the removal of dissolved gas [88], which has not yet been explored. It is known that removal of dissolved gas strongly affects hydrophobic interactions and colloidal stability, as well as the structure of water [88, 139].

Freeze-thaw cycles have been applied to a series of phospholipid vesicular suspensions, particularly focusing on phospholipids with unsaturated acyl chains (DOPC) and DOPC/DOPA mixtures (DOPA is the abbreviation for the negatively charged 1,2-dioleoly-*sn*-glycero-3-phosphate) [136]. It has been shown that freeze-thaw cycles can be applied successfully to the preparation of mainly LUVs starting from MLVs, depending on the experimental conditions [136]. With the use of a 0.1 M Hepes (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer of pH 8.0 containing 0.1 M KCl and 5 mM EDTA, 10 freeze-thaw cycles of aqueous suspensions of DOPC (or POPC) MLVs, for example, led to suspensions that contained about 98% unilamellar vesicles with diameters below 200 nm. The remaining 2% were multilamellar. The vesicles obtained by repetitively freezing and thawing MLVs are often abbreviated as FATMLVs [140]. In the case of DOPC/DOPA (80:20 mol%), almost no MLVs could be detected after the freeze-thaw procedure; the vesicles again had diameters below 200 nm after 10 freeze-thaw cycles [136]. After 50 cycles, the vesicle diameters decreased to below 100 nm [136]. In the case of saturated PCs (DMPC or DPPC), MLV suspensions remained essentially multilamellar, even after 10 freeze-thaw cycles under the conditions used [132].

Fusion of SUVs into larger vesicles upon repetitive freezing and thawing [31, 136, 140, 141] seems to occur efficiently only in pure water (without added salt) [136]. Furthermore, freeze-thaw cycles may not only affect the mean vesicle size and lamellarity, but also the entrapment efficiency for water-soluble molecules [142–145]. In some cases, the trapping efficiency may be dramatically increased [142]. Furthermore, the composition of the aqueous solution may generally influence the freeze-thaw behavior of the vesicles; all of the freeze-thaw effects depend on the experimental conditions [136, 138].

Depending on the phospholipid(s) used, freezing and thawing of SUVs may lead to the formation of giant MLVs with diameters ranging from 10 to 60 μm in the presence of salt (30–500 mM KCl, in the case of phospholipid mixtures containing egg yolk phosphatidylethanolamine and bovine brain phosphatidylserine, 7:3, mol/mol, at pH 7.0, for example) [146].

Although the process is not completely understood, freeze-thaw cycles often precede a further downsizing of the vesicles, for example, by extrusion (Section 3.8).

3.7. MLVs Prepared by Dehydration and Rehydration

The dehydration of a vesicle suspension followed by a controlled rehydration at a temperature above T_m leads to the fusion of small vesicles present in the suspension, resulting finally in the formation mainly of MLVs [147–149]. This procedure is based on the instability of dried SUVs and a lipid lamellae formation (SUV fusion) during the water removal. The resulting vesicle suspensions are referred to as dehydration-rehydration vesicles or dried-reconstituted vesicles (DRVs) [31]. DRVs are often prepared to obtain high entrapment efficiencies (high encapsulation yields) for water-soluble enzymes and other water-soluble molecules. The entrapment efficiency is defined as the percentage amount of the encapsulated molecule in relation to the total amount of molecules present during the vesicle formation and entrapment process.

In a typical, originally described preparation of DRVs [148], SUVs of the appropriate lipid composition are first prepared in distilled water by probe sonication. After centrifugation to remove large vesicles and titanium particles released from the sonicator probe (see Section 3.5), the SUV suspension is mixed with an equal volume of an aqueous solution of the compounds to be trapped. After rapid freezing of the mixture, it is lyophilized with the use of a commercial freeze-dryer at reduced pressure. After freeze-drying, the preparation is rehydrated with a volume of distilled water equivalent to one-tenth of the total volume of the initial SUV suspension. The use of a small volume necessarily has a positive effect on the entrapment yield. The rehydration is usually aided by light vortexing and equilibration for 30 min. Nonentrapped molecules are removed by centrifugation or dialysis after an appropriate dilution of the vesicle preparation [148].

Since DRVs are generally not uniform with respect to size and lamellarity, a second downsizing process often follows a DRV preparation, such as extrusion (Section 3.8) or microfluidization (Section 3.9).

The essential process in the preparation of DRVs is the *disintegration* of the initially prepared SUVs during the dehydration step, which finally results in high entrapment yields of the DRVs. If the experimental conditions during dehydration (during freeze-drying) are such that the vesicles remain largely intact, the rehydration will not lead to exceptionally high entrapment yields of DRVs. One particular such case is the use of lyoprotectants, chemicals that protect the lipid vesicles against drying stress (also called cryoprotectants [150]), such as trehalose [151], glucose [152], sucrose [153], or maltose [154], used typically at concentrations of 10% (wt/wt) [153]. For a particular lyoprotectant, the lyoprotection effect is very much dependent on the type of vesicle-forming lipid used, on the bilayer composition, on the size of the vesicles, on the temperature at which the vesicles are kept before rehydration, and on the freezing rate [153, 155, 156]. Lyoprotectants are thought to prevent the increase in the T_m of the lipid during dehydration (see Section 2.3), to bind water molecules and to interact with the polar head groups of the lipids. In the latter case, water molecules around the head groups would be replaced by the lyoprotectants, thereby protecting the vesicles from aggregation and fusion during the freeze-drying process [157, 158]. In the particular case of POPC, the presence of sucrose or sorbitol results in a value of T_m in an almost dry state of less than 6 °C, in comparison with about 60 °C in the absence of lyoprotectant [159] (see Section 2.3).

Lyoprotectants are only effective in protecting the lipid vesicles if the proper concentrations of the added lyoprotectants are used. If, for example, the potentially effective lyoprotectant sucrose is used at a concentration below the one known to preserve the stability of vesicles during dehydration [157], DRVs can be prepared from SUVs, which are characterized by high entrapment yields and a size no greater than that of the initially prepared SUVs, but also not as large as the size obtained in the absence of sucrose: 90–200 nm in the presence of low sucrose concentrations

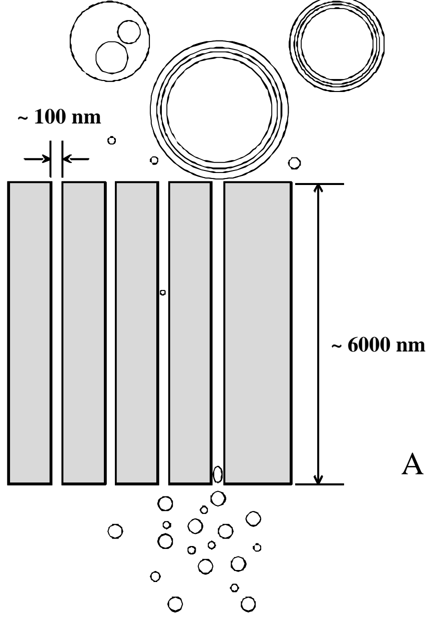
(<150 mM, with an optimal entrapment efficiency at a molar ratio of sucrose to lipid of 1), in comparison with about 1–6 μm in the absence of sucrose [160].

3.8. LUVs Prepared by the Extrusion Technique

This method is one of the most popular for the reproducible preparation of rather homogeneous vesicle suspensions containing LUVs, often with a mean diameter of about 70 nm or 100 nm [161–163]. The principle of the method is the following (see Fig. 3): a MLV suspension is passed under moderate pressure repetitively (usually 10 times) at a temperature at least 5 °C above the T_m [164, 165] through the pores of track-etched polycarbonate membranes. The pores are almost cylindrical, and vesicles (unilamellar or multilamellar) that are larger than the mean pore diameter are reduced in size and lamellarity during the passage through the pores, resulting in a final mean vesicle size that corresponds in a first approximation to the mean size of the pores. Vesicles smaller than the pore diameters pass through the pores without significant size change. Usually a MLV suspension is first passed 10 times through polycarbonate membranes with a relatively large mean pore diameter of 400 nm, followed by a passage 10 times through membranes with 200-nm pores and, finally, 100-nm pores. If desired, membranes with 50-nm pores or even 30-nm pores can be used for final extrusions. The corresponding vesicle suspensions obtained are abbreviated as VET_{400}, VET_{200}, VET_{100}, VET_{50}, or VET_{30}. VET stands for "vesicles prepared by the extrusion technique." The subscript indicates the pore size of the membranes used for final filtrations. The mean vesicle diameter of VET_{200} is usually less than 200 nm because of the presence of vesicles considerably smaller than 200 nm in the original vesicle suspension [166, 167], unless the very small vesicles are first eliminated by freeze-thaw cycles [162, 163] (see Section 3.6). If freezing-thawing is used as a homogenization method before extrusion, the resulting vesicles are often abbreviated as FAT-VET (e.g., FAT-VET_{100}: repetitively frozen and thawed vesicles that have been repetitively extruded through polycarbonate membranes with mean pore diameters of 100 nm for final extrusions).

In the early stages of the development of the extrusion technique, freeze-thaw cycles were not performed [166, 168, 169], and, depending on the application, it may even be better to avoid it, for example, in the case of the preparation of vesicles containing entrapped enzymes that may be inactivated during freeze/thaw cycles [75].

VET_{100}- or FAT-VET_{100} suspensions are generally rather monodisperse, containing vesicles with a mean diameter close to 100 nm. The mean vesicle diameter of VET_{50} or VET_{30} is generally larger than 50 nm or 30 nm, respectively. This observation can be explained on the basis of the possible mechanism by which the vesicles transform during the passage through the pores [170–172]: most likely, spherical vesicles change their shape to cylindrical structures within the pores. Because of a velocity profile inside the pores—with a low velocity close to the pore wall and a high velocity toward the center of the pores—the cylinders pearl off ellipsoidal vesicles, which then relax in size and transform into spheres upon leaving the pores. Figure 3 shows a schematic



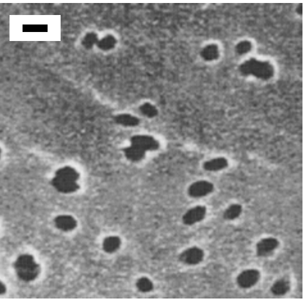


Figure 3. Illustration of the extrusion technique, which is the most popular technique for the reproducible preparation of sub-micrometer-sized unilamellar lipid vesicles from a heterogeneous multilamellar vesicle suspension (see Section 3.8). (A) Schematic representation showing (more or less) cylindrical pores with a mean diameter of 100 nm (as an example), spanning a polycarbonate membrane of about 6 μm thickness; MLVs; a MVV; and LUVs before and LUVs after repetitive extrusions (passages through the pores). (B) Transmission electron micrograph of a commercial polycarbonate membrane (Nucleopore) with pores that have a nominal mean diameter of 100 nm. Length of the bar: 100 nm. The electron micrograph was taken by E. Wehrli, Service Laboratory for Electron Microscopy, at the Department of Biology at the ETH Zürich.

drawing of the extrusion method (Fig. 3A) together with an electron micrograph of a top view of a section of a 100-nm polycarbonate membrane (Fig. 3B).

There are several commercially available devices that have been developed for the preparation of extruded vesicle suspensions [173] with volumes between 0.2 ml and 1.0 ml (e.g., LiposoFast [174]), between 1.0 ml and 10 ml (called the Extruder [163]), between 100 ml and 800 ml (the Extruder), or between 200 ml and 1 l (e.g., the Maximator [175]). A comparison between lipid vesicles prepared with the different devices (and the same set of lipids) has been performed [173], and it has been found—with the exception of FAT-VET_{50}—that the vesicles prepared with the LiposoFast were smaller than the vesicles prepared with the Maximator,

most likely because of higher flow rates and pressure drops. In the case of FAT-VET$_{50}$, the vesicles had the same mean diameter (around 50–60 nm) [173].

Since the vesicles prepared by the extrusion techniques are large (LUVs as defined in Table 3) and mainly unilamellar, extruded vesicle suspensions are often abbreviated as LUVET (e.g., LUVET$_{100}$, meaning large unilamellar vesicles prepared by the extrusion technique, using for final extrusions polycarbonate membranes with 100-nm pore diameters) [161]. FAT-VET$_{200}$ and particularly FAT-VET$_{400}$ suspensions usually contain a considerable amount of oligolamellar vesicles [162].

The extrusion technique can be applied as final down sizing and homogenization in principle for any type of vesicle suspension, MLVs, DRVs (Section 3.7), or REVs (Section 3.14), resulting in MLV-VET, DRV-VET, or REV-VET suspensions, respectively.

With a so-called French press cell—a device that does not contain any polycarbonate membranes at all but is often applied for the disintegration of biological cells—a vesicle suspension can be "extruded" through a small orifice at high pressure, resulting in the formation of unilamellar or oligolamellar vesicles with diameters between about 30 and 80 nm, depending on the pressure [31, 66, 71, 176, 177]. These vesicles are referred to as vesicles prepared by the French press (FPVs) [72]. The mechanism of the change in vesicle size in a French press is very different from the actual extrusion technique using polycarbonate membranes [66]. In the French press, shear forces seem to play a particular role [66] (see also Section 3.11).

3.9. LUVs and OLVs Prepared by the Treatment of a MLV Suspension with a Microfluidizer

Vesicle suspensions containing a mixture of mainly large MLVs can be downsized at a temperature above T_m with a high-pressure homogenizer (a microfluidizer) [178–181]. The resulting vesicles usually have a relatively narrow size distribution centered around a value between about 50 and 300 nm [179–181]. The mean size of the vesicles obtained (LUVs and OLVs) depends on the experimental conditions, such as homogenization pressure [182], number of cycles [181–183], ionic strength [181], lipid concentration [183], and lipid composition [182]. Relatively high lipid concentrations can be used, and a large-scale production is possible [181, 184]. To achieve high entrapment yields, a microfluidization step can be performed on vesicles first prepared by the dehydration-rehydration method, which itself results in high entrapment yields [180] (see Section 3.7).

3.10. Preparation of Oligolamellar and Multilamellar Novasomes

Novasomes, also called novasome vesicles [185], can be manufactured from a variety of amphiphiles (including PCs and many nonphospholipid surfactants) by a unique and cheap process on a laboratory scale (a few milliliters) or for industrial bulk applications.

The bilayer-forming amphiphile (including any bilayer-modulating molecules to be added, such as cholesterol) is first heated above the melting temperature (above about 70 °C in the case of POPC; see Section 2.3) to give a liquid. This liquid is then injected at high velocity (10–50 m/s) through small channels with 1–3-mm diameters (or through a needle) with turbulent mixing into excess aqueous phase and immediately cooled to room temperature. The rapid injection leads to the formation of tiny droplets of the amphiphiles that are quickly converted into small aggregates. The subsequent cooling under conditions of continued turbulence then leads to the formation of vesicles (the Novasomes) within milliseconds [185]. Novasomes are OLVs or MLVs with diameters typically between 200 nm and 1 μm, depending on the experimental conditions. The size of the vesicles is controlled by the chemical structure of the amphiphile used and by the hydrodynamic shearing during the fusion of the small aggregates during the cooling process. Furthermore, the diameter of the exit channel plays a role [185].

3.11. Preparation of Multilamellar Spherulites

Spherulites (also called onions or multilamellar spherulites) are relatively uniform, densely packed MLVs that are prepared by shearing of a lyotropic lamellar phase under controlled conditions [186, 187]. The MLVs formed show very little polydispersity and can have mean diameters somewhere within 100 nm and 20 μm [188] or within 200 nm and 50 μm [187], depending on the formulation (i.e., the chemical structure of the amphiphile and possible additives) and the shear rate.

The starting lamellar phase is concentrated and composed of a stack of bilayers of the amphiphiles, separated by water layers. This lamellar phase can be composed of a large number of different amphiphiles or mixtures of amphiphiles [189–192], including soybean PCs [193]. During the shearing, which is performed in specially prepared Couette cells [186], the lamellar phase reorganizes into spherical or polyhedral microdomains [186], which can be dispersed in an excess amount of aqueous phase. The process is called spherulite technology [186, 187, 189]. Spherulites can be considered droplets of dispersed lamellar phase [188, 193], and the technology allows the entrapment of a variety of different water-soluble compounds at high yields, including enzymes [194], nucleic acids [195], and metal ions for nanoparticle synthesis [187, 190].

The formation of spherulites is a direct consequence of global packing constraints [196].

3.12. MLVs Prepared by the Bubble Method

This method is based on the bubbling of an inert gas (nitrogen) over several hours through a coarse dispersion of (initially) nonhydrated lipid particles (containing phospholipids or non-ionic surfactants) and results in vesicle preparations with a clearly non-uniform size distribution [197]. The actual bubbling unit consists of a round-bottomed flask with three necks, one used for a water-cooled reflux, one for a thermometer, and one for the gas supply. A continuous stream of gas bubbles is generated at the bottom of the flask. A coarse vesicle dispersion is first formed usually with a

homogenizer, and the gas bubbling is carried out at a temperature above the T_m of the lipids (actually at 80 °C in the case of a mixture of hydrogenated soybean PCs and dicetylphosphate (10:1, mol:mol). The T_m of hydrogenated soybean PCs is about 51–52 °C (see Table 4). Depending on the experimental conditions, the resulting vesicles have a mean diameter between 200 and 500 nm [197].

3.13. LUVs and GUVs Prepared by the Cochleate Cylinder Method

A so-called cochleate cylinder is a precipitate of a cylindrical, cigar-like phospholipid aggregate that looks like a snail with a spiral shell (its Greek name is *cochleate* [198]).

In the standard procedure, cochleate cylinders form upon stepwise addition of Ca^{2+} to SUVs prepared by bath sonication from phospholipid mixtures containing negatively charged amphiphiles, such as phosphatidylserine (or phosphatidylglycerol) [31, 199].

Cochleate cylinders are rolled bilayers that do not contain an interlamellar aqueous space. The divalent cation brings the negatively charged lipids into close contact and excludes water. A majority of the lipid present must therefore be negatively charged [198], and an excess of Ca^{2+} with respect to the negative charges present has to be added.

Instead of adding Ca^{2+} to the SUVs prepared, the vesicle suspension can also be dialyzed against an aqueous solution that contains the required amount of Ca^{2+} ions [198]. Furthermore, instead of SUVs, unilamellar vesicles prepared by the detergent-depletion method (see Section 3.27), with D-glucopyranoside as a detergent, can be applied [198].

Once cochleate cylinders are prepared, LUVs are formed under the appropriate conditions upon removal of the Ca^{2+} ions by the complexing agent EDTA (ethylenediaminetetraacetate) at a controlled pH of around 7.4. Experimentally, the cochleate cylinders are first separated from the bulk Ca^{2+}-containing solution by centrifugation to form a tight pellet, which is further used and made free from lipid-bound Ca^{2+} by one of three different procedures, with the use of EDTA (direct EDTA addition, rotatory dialysis, or agarose plug diffusion [198]). EDTA binds Ca^{2+}, which leads to an unrolling of the bilayers and then to a formation of (not very uniform) unilamellar vesicles with diameters in the range of below 1 μm to about 10 μm, depending on the experimental conditions used, such as the Ca^{2+} complexation procedure (see above) [198]: while the direct EDTA addition results in vesicle sizes below 1 μm, rotatory dialysis yields vesicles with an average size of 0.5–1 μm, and the agarose plug diffusion method gives vesicles below 1 μm or 5–10 μm, depending on the procedure.

3.14. LUVs, OLVs, MLVs, and SPLVs Prepared by the Reversed-Phase Evaporation Technique

For vesicles prepared by the so-called reversed-phase evaporation technique, cosolvents are used [31, 200]. The bilayer-forming amphiphile (e.g., POPC) is first dissolved in a water-immiscible organic solvent of low boiling point (e.g., diethylether, isopropylether, or mixtures of these ethers with chloroform or methanol). This solution is then mixed with an aqueous solution in which the final vesicles are to be formed. After vortexing and brief sonication in a bath-type sonifier, a reverse emulsion (a so-called water-in-oil emulsion, abbreviated w/o emulsion) is formed in which aqueous droplets are stabilized for a certain period of time by the amphiphiles at the interface between the droplets and the bulk organic solution. The organic solvent is then removed under reduced pressure at a temperature above T_m. During the solvent removal, the reversed emulsion collapses and is transferred into bilayered vesicles. The resulting vesicles (abbreviated REVs) are often unilamellar or oligolamellar vesicles [200], and the size is usually not very uniform, ranging from about 200 nm to 1 μm [200] or even more. Therefore, a homogenization step, such as extrusion (Section 3.8), often follows REV preparation [169, 201].

Vesicles prepared by the reversed-phase evaporation technique are useful for obtaining high encapsulation yields for water-soluble molecules (similar to the DRV preparation described in Section 3.7).

In a modification of the originally developed REV method, the experimental conditions are altered in such a way that the phospholipid/water ratio in the water-in-oil emulsion is changed; and the vesicles formed are no longer mainly uni- or oligolamellar, but mainly MLVs with sizes above 600 nm, thereby allowing higher encapsulation yields for water-soluble compounds [202–204].

With the aim of applying the principles of the REV method to large scale productions, the original method has been modified with the use of the nontoxic and cheap supercritical fluid CO_2 [205]. CO_2 has a critical temperature of 31 °C and a critical pressure of 73.8 bar [205], above which CO_2 exists as liquid. For the preparation of DPPC vesicles, the temperature was set at 60 °C (well above T_m; see Table 4) and the pressure was kept at 200 bar [205]. The resulting suspension contains mainly unilamellar vesicles with diameters between ~0.1 and 1.2 μm and high trapping efficiencies [205].

Another procedure that is rather similar to the originally developed preparation of REVs has been described; it resulted in the formation of stable MLVs (SPLVs, stable plurilamellar vesicles [76], to distinguish them from the MLVs obtained by the thin-film method, as described in Section 3.2). In the case of SPLVs, a dry phospholipid film is first dissolved in diethyl ether. After the addition of a buffer solution, the two-phase mixture formed is emulsified with a sonication bath, during which a gentle stream of nitrogen gas is passed over the mixture, until the ether is almost completely evaporated. The resulting mass ("cake") is resuspended in buffer solution, followed by a pelleting by centrifugation and washing with buffer [76].

3.15. LUVs Prepared From W/O- and W/O/W-Emulsions

This multistep procedure for the formation of vesicles [206, 207] is very similar to the reversed-phase evaporation method (Section 3.14). It starts with the preparation of a w/o emulsion containing the water-soluble molecules to be entrapped in the final vesicle preparation. The emulsion is formed from soybean PCs, cholesterol, and benzene

(or methylene chloride). This w/o emulsion is then emulsified with another aqueous phase to form a water-in-oil-in-water emulsion (w/o/w emulsion), paralleled by evaporation of the solvent molecules, which is speeded up by mechanical agitation and a stream of nitrogen gas [207]. The mainly unilamellar vesicles thus obtained have a reported mean diameter of about 400 nm [207]. The size of the vesicles (apparently always between 50 and 500 nm) depends on the experimental conditions, such as the intensity of the second emulsification step (to form the w/o/w emulsion) [206] and the chemical nature of the organic solvent used [207]. With increasing boiling point of the solvent, the mean size of the vesicles tends to decrease [207].

3.16. MLVs, GUVs, and MVVs Prepared by the Solvent-Spherule (W/O/W Emulsion) Method or DepoFoam Technology

This particular method is called the solvent-spherule method by its inventors [208], since solvent spherules (surfactant-stabilized o/w emulsion droplets) are the starting system from which the oil (the solvent) is removed in a particular way, resulting in the formation of micrometer-sized MLVs [208]. The method is conceptually similar to the methods described in Sections 3.14 and 3.15.

The vesicle-forming amphiphiles (necessarily containing a small amount of negatively charged surfactants) are first dissolved in a 1:1 (v/v) mixture of chloroform and diethyl ether. This lipid solution is placed under the surface of an aqueous solution (5% glucose) with a glass capillary pipette. After agitation for about 1 min, surfactant-stabilized solvent spherules (droplets) form in the aqueous phase (w/o/w emulsion). The volatile solvent mixture is then removed in a particular way by careful dropwise addition to a flask to which a stream of nitrogen gas is added. The flask is kept at 37 °C and gently swirled. The average size of the MLVs formed is affected by the lipid concentration and the size of the lipid spherules formed by mechanical agitation.

In a modification of the method, which involves the additional use of triolein and a certain complex way of mixing, evaporation and centrifugation steps, GUVs in the 5–10-μm size range can be prepared, depending on the experimental conditions (strength and duration of the initial vortexing to form the o/w emulsion droplets) [31, 209].

In a further variation of the method, multivesicular vesicles (MVVs), with sizes between about 5 and 30 μm, can be prepared by a similar stepwise procedure (which involves pelleting by centrifugation), using a particular lipid mixture and, as solvents chloroform, diethyl ether and triolein [31, 210]. The technology for the preparation of this type of MVV is known as DepoFoam technology [211].

DepoFoam technology is a double emulsification process that has been developed based on initial observations [210] of the formation of MVVs. The vesicles formed are multivesicular vesicles (also called multivesicular liposomes), micrometer-sized vesicles that contain internal, nonconcentrically arranged compartments. The internal packing is comparable to the way gas bubbles are packed in a gas-liquid foam [211, 212]. The contacts between the compartments exhibit a tetrahedral coordination.

The first step in the formation of DepoFoam MVVs is the preparation of a w/o emulsion by dissolving a mixture of vesicle-forming amphiphiles (e.g., phospholipids) containing at least one neutral lipid (e.g., triolein) in one or more volatile, water-immiscible organic solvents (e.g., chloroform) and the addition of an aqueous solution containing water-soluble molecules to be entrapped in the final vesicles formed. In a second step, the w/o emulsion is mixed with a second aqueous solution, followed by mechanical mixing to yield solvent spherules suspended in the second aqueous phase (a w/o/w emulsion [212]). The organic solvent is then removed from the spherules by evaporation at reduced pressure or by passing a stream of nitrogen gas over or through the suspension [211]. The properties of the MVVs formed (such as captured volume) depend on the experimental conditions, such as molar fraction of the neutral lipid [210].

The presence of a neutral lipid like triolein is important since it allows a particular type of compartmentation [211]. Triolein acts as a hydrophobic space filler at bilayer intersection points and stabilizes these junctions. Furthermore, triolein is also present as oil droplets dispersed in the encapsulated aqueous space [211].

3.17. GVs Prepared from an Organic/Aqueous Two-Phase System

Giant vesicles—claimed to be unilamellar [213]—can be prepared rapidly by first dissolving the vesicle-forming amphiphiles in a chloroform-methanol solution in a round-bottomed flask, and then adding carefully along the flask walls an aqueous solution (water or buffer) that may also contain water-soluble molecules to be entrapped. After removal of the organic solvent in a rotatory evaporator under reduced pressure and at elevated temperature (40 °C), a suspension is obtained that contains many giant vesicles with diameters up to 50 μm, which can be removed for further investigation [213, 214].

3.18. SUVs and LUVs Prepared by the Ethanol Injection Method

This method is a rather simple one that uses ethanol as a cosolvent and does not require homogenization devices [31, 215, 216]. The bilayer-forming amphiphile (e.g., POPC or another PC) is first dissolved at a certain concentration in ethanol (or methanol [217]). A transparent solution is obtained. If a small amount of this ethanolic (or methanolic) lipid solution is now rapidly added at a temperature above the T_m to an aqueous solution, the formation of vesicles is observed. The reason for vesicle formation is the miscibility of ethanol (or methanol) with water and the migration of alcohol molecules—originally surrounding the lipid molecules—away from the lipids into the bulk solution. Depending on the experimental conditions (e.g., lipid concentration in the alcohol, speed of adding the alcoholic lipid solution, final concentration in the aqueous suspension, and stirring rate), the vesicles formed are more or less homogeneous with respect to size and lamellarity. The most important factor seems to be the concentration of the lipid in the alcohol injected into the buffer solution.

Vesicles prepared from DMPC (at 35 °C) or DPPC (at 55 °C), for example, were mainly unilamellar vesicles with diameters between about 30 nm and about 120 nm if the PC concentration in the ethanolic solution was varied between about 3 mM and about 40 mM (with an ethanol concentration in the buffer after the addition of 2.5–7.5%, v/v) [210]. Similar results were obtained with a mixture of soybean phospholipids [218].

In the case of POPC injected as a methanolic solution, POPC concentrations in methanol up to 25 mM result in vesicles with diameters between 40 and 70 nm [217].

The ethanol (or methanol) present in the final vesicle suspension may be removed almost completely by dialysis if required [216].

One of the limitations of the method is the limited solubility of the phospholipids in the alcohol (e.g., 40 mM soybean PC in ethanol [218]). This necessarily results in relatively dilute vesicle suspensions (a few millimolar). Furthermore, whereas entrapment yields for ethanol-soluble substances are high [218], the encapsulation efficiency for water-soluble compounds is low [218], unless a sophisticated cross-flow injection technique is used [219], which also allows an upscaling to at least several 100 ml [220].

The ethanol injection method has also been combined with high-speed homogenization [221, 222], thereby allowing the preparation of uniform vesicles with a diameter of 170–200 nm on an industrial scale [222].

In a further modification of the method, an ethanolic lipid solution is not injected into an aqueous solution, but water is poured into a concentrated lipid-ethanol solution, followed by the removal of the ethanol in an evaporator and the addition of water [223]. The particular lipid mixture contained a defined amount of soybean PC, cholesterol, β-sitosterol β-D-glucoside, and oleic acid, and the resulting polydisperse vesicles had mean diameters between about 150 nm and 1.3 μm, depending on the experimental conditions [223].

3.19. ULVs and OLVs Prepared by the Proliposome Method

This method is related to the ethanol injection method described in Section 3.18 in the sense that ethanol is also used as a cosolvent. An initial mixture (called proliposome mixture [224]) containing vesicle-forming amphiphiles (egg PCs [225], soybean PCs, or hydrogenated soybean PCs [224]), ethanol (or glycerol or propyleneglycol), and water is converted into vesicles by a dilution step [224–226]. It is a method that seems to be particularly applicable to the bulk production of lipid vesicles.

The vesicles (liposomes) only form after water addition since the proliposome mixture does not contain enough water to trigger vesicle formation. The proliposome mixture is probably built up of extended hydrated lipid bilayers that are separated by an ethanol-rich hydrophilic medium [224].

The encapsulation efficiencies for water-soluble or bilayer soluble compounds are rather high [224, 226]. The vesicles formed by the proliposome method may be predominantly unilamellar or oligolamellar vesicles with a broad size distribution between 20 nm and about 400 nm [218]. The mean size and lamellarity of the vesicles obtained seem, however, to depend on the actual experimental conditions. In the case of vesicles formed from a "proliposome mixture" containing egg yolk PC:ethanol:water at a ratio of 100:80:20 (w/w/w), the mean size varied between 100 nm and 1.2 μm, and most vesicles were oligo- or multilamellar [225].

In a modification of the actual proliposome method, a large-scale production of lipid vesicles could be achieved by diluting about 10–20 times—with an aqueous phase using a dynamic mixing device above T_m of the lipids—a water-miscible solvent mixture composed of N-methylpyrrolidone and *tert*-butyl alcohol (1:4, v/v) containing POPC:1,2-dioleoyl-*sn*-glycero-3-phosphoserine (7:3, mol/mol) and a drug to be entrapped [227]. The resulting mean size of the mainly unilamellar vesicles formed after the dilution varied with the composition of the aqueous phase between about 50 and 150 nm [227].

Please note that the term "proliposome" has also been used for particular dry granular phospholipid preparations which, upon dispersion in water, result in the formation of MLVs [31, 228]. These preparations do not contain ethanol at all.

3.20. Preparation of Multilamellar Ethosomes

With a third method in which ethanol plays an important role, so-called ethosomes can be prepared. Ethosomes are lipid vesicles that contain in the final preparation a considerable amount of ethanol. Ethosomes are prepared by first dissolving a phospholipid (such as soybean PC mixtures) in ethanol. Water is then slowly added in a fine stream with constant mixing to a specially prepared container, followed by an equilibration of the system at 30 °C. The final vesicle preparation contains 2% (w/w) soybean PCs and 30% (w/w) ethanol and seems to be particularly useful in pharmaceutical applications for drug transport across the skin (transdermal drug delivery) [229–231]. The vesicles formed are mainly MLVs and are apparently relatively monodisperse, with a mean reported diameter of about 150 nm [229]. The size of the vesicles seems to increase with decreasing ethanol concentration [229]. At 20% (w/w) ethanol, the mean ethosome diameter is around 190 nm; at 45% (w/w) it is around 100 nm [229]. The mean vesicle size is also dependent on the lipid concentration. At 30% (w/w) ethanol, the vesicles' mean diameter varies from about 120 nm to about 250 nm on going from 0.5% (w/w) to 4% (w/w) soybean PCs [229].

In transdermal drug delivery applications, the ethanol present in an ethosome preparation may act as a skin permeation enhancer because of the interaction with the lipid layers of the skin's horny layer (stratum corneum), thereby allowing the passage of drugs across the skin.

3.21. GVs Prepared by the Interdigitation-Fusion Method

A fourth method that uses ethanol is based on the fact that under certain conditions certain glycerophospholipids are known to form bilayers that have interpenetrated (interdigitated) hydrophobic chains. This means that the methyl groups localized at the end of the hydrophobic chains of a monolayer in a bilayer are in contact with the methylene groups of the hydrophobic chains of the other monolayer

and vice versa. Such interdigitated structures (interpenetrated lamellar sheets) are formed upon the addition of ethanol to SUVs prepared from specific saturated PCs (e.g., DPPC), at a temperature below T_m. When the temperature is increased above T_m, the interdigitated lamellar sheets fuse and transform into vesicles that are mainly unilamellar and have diameters above 1 μm (called IFVs, vesicles prepared by the interdigitated fusion method [232]).

The trapped volume of the IFVs depends on the chemical structure of the lipid, the concentration of ethanol used to induce interdigitation fusion, and the size of the precursor SUVs [232].

3.22. ULVs and MLVs Prepared by the Coacervation Technique

The starting system in this method of vesicle preparation is a mixture of naturally occurring egg yolk phospholipids (including about 81% PCs), an alcohol in which the phospholipids are soluble (methanol, ethanol, *n*-propanol, or 2-propanol), and water [233]. Under appropriate conditions, phase separation is observed in this three-component system, corresponding to a region in the phase diagram that is related to a so-called coacervation. *Coacervate* is an old term used in colloid chemistry [234]. It refers to a system in which an amphiphile-rich aqueous phase is in equilibrium with an amphiphile-poor aqueous phase. It seems that coacervates actually correspond to the so-called sponge phase (the L_3-phase), a disordered version of the bilayered bicontinuous cubic phase [235].

After the initial coacervation system is dialyzed against water, vesicles form that seem to be either relatively homogeneous and unilamellar vesicles (in the case of methanol or propanol) or mainly MLVs (in the case of ethanol), ranging in size from about 100 nm to 1 μm, depending on the experimental conditions used [233].

3.23. Vesicles Prepared by the Supercritical Liposome Method

With the use of a specially designed, technically rather complex apparatus, vesicles with an average size of about 200 nm can be prepared by mixing at low pressure an aqueous solution with supercritical CO_2 (kept at high pressure (25 MPa) and 60 °C) containing the vesicle-forming amphiphile (POPC:cholesterol, 7:3, mol/mol). The mean vesicle sizes vary with the experimental conditions, such as geometric dimensions of the important parts of the apparatus [236].

3.24. Vesicles Prepared by the Ether Injection Method

Vesicles can be prepared by slowly injecting (at 0.2 ml/ml) a diethyl ether/phospholipid solution into an aqueous phase that has been warmed to a temperature (55 °C) above the boiling point of diethyl ether [31, 237]. The diethyl ether vaporizes upon contact with the aqueous phase, and the dispersed lipids form preferentially (but not entirely) unilamellar vesicles. These vesicles can then be sized down by extrusion (see Section 3.8) or simple Millipore filtration. In the latter case, the reported vesicle diameters are in the range of 100 to 300 nm [237].

3.25. OLVs Prepared by the Rapid Solvent Exchange Method

The rapid solvent exchange (RSE) method has been specifically designed for the preparation of vesicles from phospholipid-cholesterol mixtures containing high amounts of cholesterol [238]. The method is based on a rapid transfer of the vesicle-forming lipids from an organic solvent to an aqueous buffer solution in which the vesicles are meant to be formed. This rapid solvent exchange avoids the transient formation of solid lipid mixtures, which often demix (phase separate) and result in inhomogeneous vesicle preparations.

The lipids and the membrane soluble additives (i.e., cholesterol) are first dissolved in a solvent that is not miscible with water (e.g., chloroform or methylene chloride). This lipid solution is then added to an aqueous solution above the T_m of the lipids in a particular manner at reduced pressure with a specially designed apparatus in such a way that the solvent is rapidly (within 1 min) and almost completely evaporated, as a result of pressure changes during the injection process.

Vesicles prepared by the RSE strategy from POPC, for example, are oligolamellar with an expected lamellar repeat distance of 6.5 nm (see Section 3.2) and an entrapped volume of about 4.5 liters/mol [238].

3.26. Vesicles Prepared from an Initial O/W Emulsion

In this simple method [239], an o/w emulsion is first formed by bath sonication from a vesicle-forming amphiphile (egg yolk PC and cholesterol), an aqueous solution (containing the water-soluble molecules to be entrapped), and *n*-decane. This o/w emulsion is then transferred to a second aqueous solution, which gives two separated phases, an upper organic phase and a lower aqueous phase. The two-phase system is centrifuged at 3500 × *g* for 10 min, resulting in the movement of the amphiphiles from the organic phase into the aqueous phase and, as a consequence, the formation of vesicles in the lower aqueous phase, which is separated from the upper phase. The resulting vesicles, which may contain small amounts of *n*-decane in the bilayer, have a diameter in the range of 50–200 nm and are characterized by relatively high encapsulation yields [239].

3.27. ULVs Prepared by the Detergent-Depletion Method

If a bilayer-forming lipid is mixed in an aqueous solution with a micelle-forming surfactant (often called detergent, from the Latin word *detergere*, meaning to wipe off or to clean) under such conditions that the detergent molecules "dominate," mixed detergent/lipid micelles are formed [240–243]. These aggregates are composed of bilayer-forming amphiphiles as well as micelle-forming amphiphiles and are disc-like, sheet-like, or cylindrical structures.

In the detergent-depletion method (also called the detergent dialysis or detergent removal method), the starting system from which the vesicles are formed is mixed detergent/lipid micelles. The micelle-forming detergent molecules (with their large a_0; see Section 1.1) are expected to be distributed in the mixed-micelle aggregate in such a way that they particularly occupy the highly curved edges of the aggregates [31]. The micelle-forming surfactant is also present in a relatively high amount in the bulk phase as nonaggregated, monomeric detergent, at a concentration corresponding in a first approximation to a value a bit lower than the detergent's cmc, the critical concentration for micelle formation determined separately under comparable conditions.

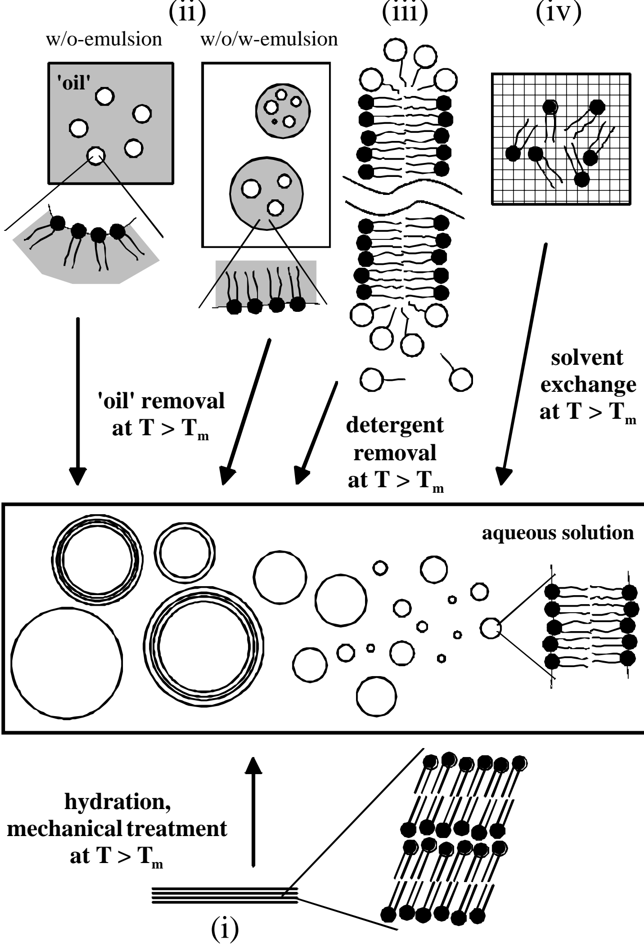


Figure 4. Simplified schematic representation of the principal pathways for the preparation of (normal) lipid vesicles in the case of "conventional amphiphiles" (surfactants that do not form a true vesicle phase at thermodynamic equilibrium). The pathways involve, as starting state of the amphiphiles, (i) preorganized dry lipids, which are hydrated and (possibly) mechanically manipulated above the T_m of the lipids; (ii) preorganized lipids in w/o emulsions (or w/o mircoemulsions or reversed micelles) or w/o/w emulsions prepared in a volatile solvent that is removed during the vesicle preparation procedure above T_m of the lipids; (iii) preorganized lipids in the presence of micelle-forming detergents (mixed detergent/lipid micelles) existing in dynamic equilibrium with free detergent monomers that are removed during the vesicle preparation procedure above the T_m of the lipids; or (iv) lipids dissolved in a solvent that is miscible with water and is exchanged with water during the vesicle preparation procedure above the T_m of the lipids. Once a vesicle suspension is formed, the mean vesicle size and size distribution can always be altered by mechanical treatments above T_m.

The amount of monomeric detergent in the mixed micellar system is important, as it is this nonaggregated amphiphile that is removed from the solution during the detergent removal process, which finally leads to the formation of vesicles [31, 66, 69, 71, 244–247]. The principle of the detergent-depletion method is the following: mixed detergent/lipid vesicles, present in rapid equilibrium with detergent monomers, are put into a dialysis bag or another dialysis device [31] at a temperature above T_m of the lipid used [248]. The dialysis membrane is characterized by a permeability for the monomers, whereas the much larger mixed micelles cannot pass the membrane. Then, at a temperature above T_m [248], the dialysis device is put in contact with a buffer solution in which the mixed micelles were formed. Since the monomers can pass the dialysis membrane, the amount of monomers in the solution inside the dialysis device continuously and slowly decreases, and detergent monomers move from the mixed micellar aggregate into the bulk solution. This process continues until the amount of detergents in the micellar aggregates is so low that mixed micelles can no longer exist, and extended mixed bilayer fragments (sheets) and finally mixed lipid/detergent vesicles form. Extensive dialysis leads to the formation of vesicles that are almost (but not necessarily completely [249]) free from detergent molecules. These vesicle suspensions are often to a large extent unilamellar and have a narrow size distribution. The mean size depends on the experimental conditions, such as type of detergent used, initial lipid and detergent concentrations in the mixed micellar solution, and speed of detergent removal [247, 250–252].

Table 5 lists detergent molecules that are often used for the preparation of vesicles by the detergent-depletion method, together with characteristic size ranges of the mainly unilamellar vesicles formed.

In the case of a system containing egg yolk PC and the bile salt sodium taurochenodeoxycholate (which aggregates itself stepwise into a particular type of unconventional, small micelles [253–255]), the mixed micelle-mixed vesicle transformation process—initiated by a rapid dilution process—has been investigated by time-resolved static and dynamic light-scattering measurements [256]. The scattering data analysis indicates that the key kinetic steps during vesicle formation are the rapid appearance of disc-like intermediate micelles, followed by growth of these micelles and closure of the large discs formed into vesicles [256].

In addition to detergent removal through dialysis, gel permeation chromatography [256, 257] (which is based on the partitioning of detergent monomers into the pores of a swollen gel matrix) or so-called Bio Beads [246, 247, 258, 259] (which bind detergent monomers) can also be applied.

In the case of saturated PCs like DMPC and DPPC and octyl-β-D-glucopyranoside, the originally developed detergent-dialysis method has been modified slightly because of the relatively high T_m value of these lipids (see Table 4) [248]. The important modification is a slow dilution step before the actual dialysis procedure [248], resulting in mainly unilamellar vesicles with a mean diameter of 98 nm

Table 5. Some of the detergents most often used for the preparation of vesicles by the detergent depletion method (Section 3.27) and approximate mean sizes of the vesicles formed in the case of egg yolk PCs.

Detergent	cmc at 25 °C (mM)	Method for removing the detergent	Reported approximate vesicle diameters (nm)	Refs.
Sodium cholate	~11	Gel permeation chromatography	30	[257]
		Dialysis	60	[244]
			70	[252]
			60–80	[245]
			80–100	[251]
			50–150	[519]
Sodium glycocholate	~10	Dilution and dialysis	30–100	[143]
Sodium deoxycholate	~4	Dialysis	150	[252]
Sodium chenodeoxycholate	~5	Dialysis	160	[252]
n-Octyl-*β*-D-glucopyranoside	~23	Dialysis	180	[245, 524]
			230	[525]
		Bio Beads	300–500	[246]
			250	[526]
$C_{12}EO_8$ (*n*-dodecyl octaethylenglycol monoether)	~0.08–0.09	Bio Beads	60–90	[247]
			25–80	[527]
			120	[526]
CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propane sulfonate)	~5–10	Dialysis	380	[519]

Note: The size of the vesicles may very much depend on the experimental conditions (see text). In particular, the resulting sizes may also depend on the presence of bilayer soluble substances (e.g., cholesterol [245, 246]) or in particular cases (cholate) on the presence of divalent cations (e.g., Ca^{2+}) [519]. The approximative cmc values given in the table are taken for the bile salts from [520]; for $C_{12}E_8$ from [521, 522]; for CHAPS from [522, 523]; and for *n*-octyl-*β*-D-glucopyranoside from [522].

(DMPC) and 94 nm (DPPC) under the corresponding conditions used [248].

It is quite generally possible to first simply dilute the mixed micellar system, followed by dialysis to completely remove the detergent [260]. If the detergent is not completely removed, the vesicle preparation by simple dilution will always contain detergent molecules, even if the dilution is 100- or 200-fold [250].

In a particular case [261], SUVs with a mean diameter of 23 nm were first prepared by sonication from egg yolk PC at a concentration of 20 mM. The detergent sodium deoxycholate was then added to give an aqueous mixture at a ratio of deoxycholate to PC of 1:2 (mol/mol). This mixture contained vesicles that were considerably larger than the SUVs used, because of the uptake of the detergent molecules. Deoxycholate was then removed to about 96–98% first by gel filtration and then almost completely by a second gel filtration. The final preparation—containing less than one deoxycholate molecule per PC molecule—were dispersed unilamellar vesicles with a mean diameter of 100 nm [261]. The same mean vesicle sizes were also obtained if instead of SUVs a dry egg yolk PC film was treated with deoxycholate, followed by bulk sonication and detergent removal [261].

Large-scale production of vesicles by detergent dialysis is possible (e.g., [262]), and commercial devices are available under the trade names Liposomat and Mini Lipoprep.

The detergent depletion method is the method of choice for the reconstitution of water-insoluble membrane-associated proteins, which in a first step are extracted from the biological membrane by a mild detergent that does not lead to an irreversible protein denaturation [258, 263]. The membrane protein-containing mixed detergent/lipid micelles are then converted to membrane protein-containing vesicles by one of the detergent removal techniques described above.

3.28. (Mixed) Vesicles Prepared by Mixing Bilayer-Forming and Micelle-Forming Amphiphiles

As mentioned in Section 3.27, mixed lipid/detergent vesicles form transiently during detergent removal from detergent/lipid micelles. Such mixed vesicles can also be prepared by adding to preformed lipid vesicles (prepared by any method) above T_m an appropriate amount of a particular detergent [264, 265] or by simply diluting a mixed detergent/lipid micellar solution [266]. Examples of detergents that have been used include sodium cholate [267], *n*-octyl-*β*-D-glucopyranoside [268], or 3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate (CHAPS) [269]. The size of the resulting mixed lipid/detergent vesicles depends on the experimental conditions, such as the size (and lamellarity) of the initial vesicles, the chemical structure of the lipid, the lipid concentration, the chemical structure of the detergent, and the detergent concentration.

Phosphatidylcholines like POPC, SOPC, DMPC, DPPC, or the mixtures extracted from egg yolk or soybeans are the amphiphiles whose mixed lipid/detergent vesicle (bilayer) formation has been studied most extensively. These PCs all have two long hydrophobic acyl chains containing 14 to 18 or 22 carbon atoms (see Tables 2 and 4).

PCs with two short acyl chains, containing fewer than 10 carbon atoms, do not form bilayers (vesicles) in dilute aqueous solution, but rather micelles (as long as the concentration is kept above the cmc) [270, 271]. While a very short-chain PC (1,2-dibutanoyl-*sn*-glycero-3-phosphocholine) forms more or less spherical micelles, an increase in the chain length up to eight carbon atoms leads to the formation of extended, elongated micellar structures. All of these short-chain PCs are detergents in the sense of the term used here (Section 3.27) [272], and unilamellar vesicles composed of long-chain PCs and short-chain PCs can easily be prepared, for example, by the addition of a micellar solution of 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (final concentration 5 mM) to a DPPC dispersion (final concentration 20 mM) in an aqueous solution [273]. The resulting vesicles are rather stable and have mean diameters below as well as above 1 μm [274], depending on the experimental conditions, such as the ratio of the short-chain PC to the long-chain PC and cholesterol content [274].

One particular type of mixed lipid/detergent vesicular system (also containing ethanol)—not prepared by detergent dialysis, but by simple detergent and lipid mixing—are the so-called Transfersomes, vesicles that are claimed to be able to transfer water-soluble drugs through the skin to the blood circulation (transdermal drug delivery), particularly through the outermost physical barrier of the skin, the horny layer, called the stratum corneum. Transfersomes seem to be ultradeformable and may squeeze through the stratum corneum pores, which have diameters only one-tenth of the size of the Transfersomes. In this way, encapsulated water-soluble drugs may be transported across the skin [231]. Transfersomes are prepared by mixing an ethanolic solution of soybean PCs with (for example) sodium cholate to yield a suspension typically containing 8.7% (w/w) soybean PCs and 1.3% (w/w) cholate as well as approximately 8.5% (v/v) ethanol [275, 276]. For a reduction in vesicle size to about 500 nm, the suspension was treated by sonication, freeze-thaw cycles, and processing with a homogenizer.

In the case of the particular non-ionic industrial amphiphile Brij 30 (*n*-dodecyl tetraethylene glycol ether, abbreviated as $C_{12}E_4$ or $C_{12}EO_4$, also called Laureth-4), vesicles have been prepared by simple dilution from so-called hydrotrope solutions, which contained in addition to $C_{12}EO_4$ the micelle-forming surfactant sodium xylene sulfonate [277]. Through a detailed elaboration and analysis of the three-component phase diagram containing $C_{12}EO_4$, sodium xylene sulfonate, and water, the conditions under which vesicles form have first been established. Appropriate starting conditions were then chosen, and a corresponding dilution resulted finally in the desired vesicular system, which contained not only the vesicle-forming amphiphile, but also the hydrotrope (as constituent of the aqueous phase) [277].

3.29. Vesicles Prepared from Lipids in Chaotropic Ion Solutions

This method is related to the detergent-depletion method described in Section 3.27. Chaotropic substances (e.g., urea, guanidinium hydrochloride, potassium thiocyanate, trichloroacetate, or trichlorobromate) are known to disturb (break) the structure of water, and lipids are soluble in chaotropic solutions [278]. Solutions of trichloroacetate, for example, dissolve PCs as micellar solution [279], through a binding of the ions to the head groups, which leads to an increase in a_0 (see Section 6). If a SUV suspension prepared from phospholipids in buffer solution by probe sonication is mixed with an aqueous solution containing sodium trichloroacetate, a micellar solution is first obtained (at about 1–3 M sodium trichloroacetate in the case of 0.1–0.5 mM phospholipids). This solution is then extensively dialyzed against a buffer solution to remove all of the chaotropic ions, resulting in the formation of uni- or oligolamellar vesicles with diameters between 10 and 20 μm, although many small vesicles are also present [278]. If freeze-thaw cycles are used, the amount of chaotropic ions needed to generate giant vesicles is lowered considerably [278].

3.30. GVs Prepared from a W/O Emulsion with the Help of a Detergent

This method involves the use of nonphospholipid amphiphiles in a four-step procedure [280]. First, a w/o emulsion is formed by homogenization of an aqueous solution, *n*-hexane and a mixture of soybean phospholipids and the non-ionic surfactant Span 80 (sorbitan monooleate). Second, *n*-hexane is removed by rotatory evaporation from the emulsion under reduced pressure, resulting in the formation of a water-in-lipid mixture. Third, this mixture is mixed with an aqueous solution containing a water-soluble detergent (such as sodium dodecyl sulfate, hexadecytrimethylammonium bromide, or Tween 80 (POE (20) sorbitan monooleate, which is a modified sorbitan monooleate that contains on average a total of 20 hydrophilic oxyethylene units). Fourth, the highly water-soluble detergent is removed by dialysis, leading to the formation of micrometer-sized vesicles.

4. PREPARATION OF REVERSED VESICLES

There seems to be a kind of symmetry among most self-organized surfactant aggregates with regard to the distribution of the hydrophilic and the hydrophobic parts of the surfactants in the aggregate [281–283]. There are (normal) micelles (L_1-phase) and reversed micelles (L_2-phase), a (normal) hexagonal phase (H_I) and a reversed hexagonal phase (H_{II}), and (normal) emulsions (oil-in-water, o/w) and reversed emulsions (water-in-oil, w/o) [235]. In the case of vesicular aggregates, the existence of reversed vesicles has been demonstrated [283], for example, for a system containing the non-ionic surfactant tetraethyleneglycol dodecyl ether (*n*-dodecyl tetraethyleneglycol monoether, abbreviated $R_{12}EO_4$, $C_{12}EO_4$, or $C_{12}E_4$), the solvent *n*-dodecane, and water [281, 282]. These reversed vesicles form first upon the preparation of a mixture composed of 1 wt% water and 99 wt% of a *n*-dodecane solution which contains 2.5 wt% $C_{12}EO_4$ [281]. After equilibration at 25 °C, a two-phase system forms that is composed of a lamellar liquid crystalline

phase that is in equilibrium with an excess *n*-dodecane phase. When these two phases are mixed by hand-shaking, a heterogeneous dispersion of mainly multilamellar reversed vesicles forms with diameters of the reversed vesicles from less than 1 μm to 10–20 μm [281]. Strong vortexing leads to a reduction in size below 200 nm [282]. Without water, no reversed vesicles form.

A reversed vesicle formed in an oil (in a water-immiscible, apolar solvent) is composed of an oily core and one or several closed "reversed" surfactant bilayer shells. The reversed bilayers are organized in such a way that the hydrated hydrophilic head groups of the amphiphiles are facing toward the center of the reversed bilayer, while the hydrocarbon chains are in contact with the oil [281].

The particular type of reversed vesicles formed from $R_{12}EO_4$, water, and *n*-dodecane seem to be rather unstable and convert back into the thermodynamically stable two-phase system within hours or days [281]. Considerably more stable reversed vesicles can be obtained from a mixture of the non-ionic surfactants sucrose monoalkanoate (which is, to at least 95%, a monoester containing 10 wt% tetradecanoyl, 40 wt% hexadecanoyl, and 50 wt% octadecanoyl chains [283]), hexaethyleneglycol hexadecyl ether (abbreviated $R_{16}EO_6$, $C_{16}EO_6$, or $C_{16}E_6$), *n*-decane, and water [284]. The two surfactants are mixed at a weight ratio of sucrose monoalkanoate to $C_{16}EO_6$ of 85:15, and vesicle formation is observed at typically 3 wt% surfactant in *n*-decane with a typical weight ratio of surfactant to water of 6 [283]. Treatment of the vesicle suspension with a probe sonicator at 30 °C leads to a reduction in the reversed vesicle size to about 150–300 nm, depending on the experimental conditions, such as the water-to-surfactant ratio and the surfactant concentration [284].

The formation by vortexing of micrometer-sized reversed vesicles (and reversed myelin figures) has also been observed in *m*-xylene at a surfactant-to-water weight ratio around 0.8 for a number of polyethyleneglycol oleoyl ethers (abbreviated $R_{18:1}EO_x$ or $C_{18:1}EO_x$, with mean values of x varying between 10 and 55) [285]. Furthermore, it seems that some phospholipids can also form reversed vesicles under particular conditions, as reported in the case of a system containing soybean phosphatidylethanolamine and triolein, saturated with water through vapor equilibration [286]. The particles formed—which are most likely reversed vesicles—have diameters in the range of 500 nm to 1.7 μm [286].

5. CHARACTERIZATIONS AND APPLICATIONS OF VESICLES

5.1. Characterizations of Vesicles

There are a number of typical properties that characterize a vesicle suspension, particularly the mean vesicle size, the vesicle size distribution, the mean number of lamellae per vesicle, and the chemical and physical stability [287]. All of these properties depend on (i) the chemical structure of the amphiphiles (or mixtures of amphiphiles) from which the vesicles are formed; (ii) the solution in which the vesicles are formed (e.g., salt and buffer content in the case of (normal) vesicles and aqueous solutions); and (iii) the method of vesicle preparation (see Section 3). A minimal characterization of a vesicle suspension prepared is always needed. In many cases, a routine size measurement is appropriate, particularly by photon correlation spectroscopy (PCS), also called dynamic light scattering or quasi-elastic light scattering. PCS provides information about the homogeneity of a vesicle suspension, and it is relatively straightforward to find out whether the vesicles in a preparation are relatively monodisperse or whether they are polydisperse. For monodisperse and spherical vesicles, the mean hydrodynamic diameter of the vesicles can be determined relatively easily. In the case of polydisperse vesicle preparations and/or nonspherical vesicles, the interpretation of the scattering curves is more delicate, and it is recommended to apply additional methods as well, such as electron microscopy. A most probable vesicle size can be obtained by cryotransmission electron microscopy (cryo-TEM), although this method is limited to sizes below about 250 nm, and the sizes and morphologies obtained have to be taken with caution [288–290]. The use of different independent methods is always recommended.

Table 6 lists a number of methods that are used for characterizing vesicle suspensions. In many cases, fluorescently labeled amphiphiles are used in the vesicle membranes at a concentration of 1–2 mol%. These amphiphiles contain fluorescent groups, and the behavior of these molecules is then taken as a measure for the behavior of the bulk vesicle-forming amphiphiles, for example, for the determination of the lateral diffusion coefficient of the amphiphiles. Since the fluorescent groups are sometimes rather bulky, the kinetic constants obtained with these labeled molecules may be different from the actual membrane-forming lipids [291, 292]. The same is true for nitroxide-labeled amphiphiles used for ESR (electron spin resonance) measurements [292] (see Table 6).

For the detection of the internal aqueous volume in (normal) vesicles—usually expressed as microliter of trapped aqueous solution per micromole of amphiphile (μl/μmol)—dye molecules are often used, which can be easily quantified spectroscopically. Meaningful results, however, can only be obtained if the dye molecules do not interact with the vesicle membrane (no incorporation inside the membrane, no adsorption to the membrane).

In the case of permeability measurements, the use of dye molecules is convenient, and the release of vesicle-trapped molecules into the external medium can easily be quantified. In this case, however, one should always be aware that the permeability measured is the permeability of the particular dye molecules used under the particular experimental conditions. In this respect, NMR methods that are based on the use of paramagnetic shift reagents are better. The external addition of the shift reagents allows a distinction between the particular solute molecules present inside the vesicles and the solute molecules present outside of the vesicles, presuming that the shift reagent does not permeate the membranes. This has first to be proved. The permeability may very much depend on the experimental conditions, such as the presence of certain buffer species [293].

The list in Table 6 is certainly not complete. However, it gives a few hints about the general principles and problems of the vesicle characterization and lists some of the methods and techniques used. Depending on the vesicle preparation, a particular characterization may be more useful than

Table 6. Some of the principles and analytical methods used for characterizing the properties of vesicles (see also [66, 287]).

Property	Methods used and comments	Ref.
Mean vesicle size and size distribution	*PCS* Analysis of the time dependence of intensity fluctuations in scattered laser light due to the Brownian motion of the vesicles, which is related to the mean hydrodynamic radius (R_h) of the vesicles. If the vesicle suspension is very polydisperse and/or contains non-spherical vesicles (e.g., caused by an osmotic imbalance between the inside and the outside of the vesicles), the size analysis is rather complex and difficult	[2, 31, 528–531]
	SLS Average scattering intensities are measured as a function of the scattering angle and the vesicle concentration, allowing the determination of the mean radius of gyration (R_g) of the vesicles	[532–534]
	FFEM Replicas of fractured vesicles are analyzed. The vesicles are fractured at low temperature (-100 °C) under conditions where the water is in an uncharacterized (amorphous), "glassy" solid state. Nonequatorial fracturing leads to replicas that do not represent the "true" vesicle size	[1, 2, 136, 162, 300, 535]
	Cryo-TEM The vesicles are directly observed at low temperature (-170 °C) after rapid freezing ($\sim 10^6$ °C/s)—under conditions where the water is in an amorphous solid state—to represent the state of the vesicles at the temperature from which the sample is cooled. The observed diameters of the vesicles correspond to the "true" vesicle diameters. Only applicable for vesicles with sizes below ~250–500 nm. Not free of "artefacts" ("artifacts"), i.e., formation of microstructures due to specimen preparation, electron optics, or radiolytic effects [288–290]	[2, 3, 536]
	LM Only for GVs	[4, 66, 89, 91, 536]
	Size exclusion chromatography Conventional and high-performance liquid chromatography (HPLC) separation based on the principle that the partitioning behavior of vesicles in the pores of a solid column matrix depends on the size of the vesicles	[66, 386, 537]
	NMR Size determinations possible from ^{31}P-NMR spectra of phospholipid vesicles that have sizes below 1 μm with the help of a comparison with simulated spectra. For spectra simulation, the dynamics of the vesicles and the phospholipids in the vesicle membranes are considered	[136]
	AFM The vesicles have to be deposited on a solid surface, which may lead to vesicle deformations and vesicle aggregations	[538, 539]
	SANS Measurements in D_2O that may influence the vesicle size *via* altered head group interactions. Relative complex analysis involving fitting of experimental data	[540, 541]
Bilayer thickness	*X-ray* Based on the interaction of X-rays with the electrons of the amphiphiles in the vesicles	[80, 300, 542]
	Cryo-TEM	[543]
	SANS Measurements in D_2O. Bilayer thickness determinations possible in the gel, in the ripple, and in the lamellar phase of the vesicle-forming surfactants (see Section 2)	[540, 541, 544–546]
Vesicle shape and morphology	*LM* Only for GVs. Detection of MVVs	[547–552]
	FFEM Detection of MVVs	
	Negative staining EM Heavy metal ions have to be added to the vesicle dispersion, which always alters the vesicles. Furthermore, the vesicles are observed in a dry state	[12]

continued

Table 6. Continued

Property	Methods used and comments	Ref.
	Cryo-TEM Only for vesicles with sizes below ~250–500 nm	[553]
Lamellarity	*Cryo-TEM* Only for vesicles with sizes below ~250–500 nm	[3, 293]
	Negative Staining EM Addition of heavy metals (osmium tetroxide) needed. The number of lamellae detected may not represent the "true" number of lamellae	[12, 31]
	NMR ^{31}P-NMR in the case of phospholipid vesicles. Externally added membrane-impermeable Mn^{2+} influences the ^{31}P-NMR signal of the phospholipids in the outermost monolayers by broadening the resonance beyond detection. The resonances of the inner phospholipids are unaffected. Paramagnetic shift reagents may also be used (e.g., Pr^{3+} and Eu^{3+})	[161, 162, 293, 554]
	Fluorescence quenching An appropriate amphiphilic fluorescent probe molecule is used in the vesicle preparation, and the fluorescence of the amphiphiles in the outermost layers is quenched by adding to the vesicles a membrane-impermeable reagent	[555]
	SAXS Based on the interaction of X-rays with the electrons of the amphiphiles in the vesicles	[556]
	Chemical modification Functional groups present in the hydrophilic head groups of the amphiphiles in the external layer(s) of the vesicle and exposed to the bulk aqueous solution are chemically modified by externally added reagents that do not permeate the membranes and therefore do not react with the amphiphiles that are present in the inner layer(s) of the vesicles with their head groups exposed to the trapped aqueous space. The chemically modified amphiphiles are then quantified	[31]
Lipid domains in the membranes (rafts)	*Two-photon fluorescence microscopy* Use of membrane soluble fluorescent probes. Only for GVs	[557]
	NMR ^{31}P-, ^{13}C-, or ^{2}H-NMR	[558, 559]
	Confocal fluorescence microscopy Distinction between solid-ordered and fluid-disordered domains in vesicle membranes made with two fluorescent probes that have different affinities for the two domains. Only for GVs	[560]
Phase transition temperature (T_m)	*DSC* A vesicle sample and an inert reference are heated independently to maintain an identical temperature in each. In endothermic solid-ordered/liquid-ordered phase transitions, heat is required in excess in the vesicle sample over the heat required to maintain the same temperature in the reference	[43, 300, 540]
	Fluorescence Fluorescence polarization measurements of a membrane soluble fluorescent probe (e.g., 1,6-diphenylhexatriene, DPH)	[31, 561, 562]
	CD Induced circular dichroism below the T_m of an achiral probe (1,6-diphenylhexatriene, DPH) embedded inside the vesicle membrane constituted by chiral amphiphiles. No CD signal above T_m	[563]
Mobility of the amphiphiles in the vesicle membranes	*NMR* ^{1}H-, ^{2}H-, ^{13}C-, ^{31}P-NMR	[37, 564–567]
	ESR Use of spin-labeled amphiphiles	[568]
	Quasi-elastic neutron scattering	[569, 570]

continued

Table 6. Continued

Property	Methods used and comments	Ref.
	Fluorescence Use of fluorescent amphiphiles	[571]
	Fluorescence correlation spectroscopy Analysis of the translational diffusion of a fluorescent probe. Only for GVs	[560]
	Fluorescence recovery after photobleaching Lateral diffusion of fluorescently labeled amphiphiles determined by first bleaching the lipids with an intense laser pulse, and then analyzing the fluorescence recovery kinetics in the bleached area, due to the diffusion of unbleached molecules into the previously bleached area	[37]
Molecular conformation of the amphiphiles in the vesicle membranes	*NMR* ^{13}C- or ^{2}H-NMR using labeled phospholipids	[300, 559, 564, 565, 572]
	FT-IR and *Raman* Determination of the equilibrium conformational characteristics of the amphiphiles	[37, 300, 573]
	CD For vesicles composed of chiral amphiphiles	[563]
Membrane fluidity and order	*Fluorescence* Use of fluorescent membrane probes. Measurements of static or time-resolved fluorescence anisotropy	[37]
	NMR Use of partially and specifically deuterated amphiphiles	[300, 564]
	ESR Use of amphiphiles that have a nitroxide radical. The ESR spectrum of these amphiphiles is sensitive to the motions of the molecules	[37]
Transmembrane lipid diffusion (flip-flop)	*ESR* Use of spin-labeled amphiphiles (nitroxide radical)	[292, 574]
	Fluorescence Use of fluorescently labeled amphiphiles	[292]
Surface charge	*Zeta potential measurements, microelectrophoresis* Problematic, since the method is derived from classical theories of the double layer that do not include specific ion effects (see Section 6). Limited to particular background salt conditions (usually NaCl)	
Internal volume and entrapment (encapsulation) yield	*Use of dye molecules* The dye molecules (often fluorescent) are water-soluble and should not interact with the vesicle membrane. The dye molecules are entrapped during vesicle preparation, and the amount of entrapped dye molecules is determined quantitatively either after separation of the nonentrapped molecules from the vesicles or by addition to the vesicles of a membrane-impermeable reagent, which leads to a complete quenching of the fluorescence of the externally present dyes (e.g., calcein with quencher Co^{2+}) [575]. Simple dilution of the externally present dye may also be possible [577]	[31, 66, 575–577]
	NMR ^{17}O-NMR of the water oxygen. Addition of the membrane-impermeable paramagnetic shift reagent $DyCl_3$ to the vesicles below T_m leads to a shift in the ^{17}O resonance corresponding to the external water. The internal water peak remains the same. Above T_m, only one peak is observed because of rapid water equilibration	[554]
Membrane permeability	*Use of dyes or radioactively labeled molecules* Determination of the release of dye molecules entrapped inside the aqueous interior of the vesicles as a function of time under particular storage conditions	[31, 578]
	NMR Use of ^{1}H- or ^{17}O-NMR and an externally added membrane-impermeable paramagnetic shift reagent (Mn^{2+}, Pr^{3+}) to distinguish between internal and external permeants	[579–582]

continued

Table 6. Continued

Property	Methods used and comments	Ref.
	Use of ion-selective electrodes The ions present outside of the vesicles can be detected	[583]
Chemical stability	*Thin-layer chromatography* Analysis of possible degradation product due to hydrolysis and oxidation reactions	[72]
	High-performance liquid chromatography Analysis of possible degradation product due to hydrolysis and oxidation reactions	[72]
Physical stability	*Turbidity, PCS, FFEM, Cryo-TEM*, and others Analysis of possible size changes during storage due to vesicle aggregation and fusion	[66, 69, 584]

Abbreviations: AFM, atomic force microscopy; CD, circular dichroism; cryo-TEM, cryo-transmission electron microscopy (also called cryo-fixation); DSC, differential scanning calorimetry; EM, electron microscopy; ESR, electron spin resonance (also called electron paramagnetic resonance, EPR); FFEM, freeze-fracture electron microscopy; FT-IR, Fourier transformed infrared; LM, light microscopy; NMR, nuclear magnetic resonance; PCS, photon correlation spectroscopy (also called dynamic light scattering (DLS) or quasi-elastic light scattering (QELS)); SANS, small-angle neutron scattering; SAXS, small-angle X-ray scattering; SLS, static (or classical) light scattering.

another, and certain methods may not be applied at all (e.g., ^{31}P-NMR obviously cannot be used for the determination of the lamellarity of vesicles that are not composed of phosphorous-containing amphiphiles).

5.2. Applications of Vesicles

Lipid vesicles are used successfully in many different fields as interesting and versatile submicrometer- or micrometer-sized compartment systems [69, 294–297]. This wide applicability of vesicles and the broad interest in vesicles can be understood at least on the basis of the following four reasons:

(i) Lipid vesicles can be considered membrane or biomimetic systems [298, 299], since the molecular arrangement of conventional vesicle-forming amphiphiles in a vesicle is a (more or less curved) bilayer, like the lipid matrix in biological cell membranes [97, 300] or in the outer coat of certain viruses [97].
(ii) Vesicles prepared from amphilphiles present in biological systems allow applications as biocompatible and biodegradable systems.
(iii) Water-soluble as well as certain water-insoluble molecules can be entrapped inside the aqueous or hydrophobic domains of the vesicles, allowing the use of vesicles as carrier systems and nanometer- or micrometer-sized reaction compartments.
(iv) Vesicles can be prepared not only from the conventional PC type of bilayer-forming amphiphiles, but also from a large number of different nonphospholipid surfactants (or mixtures of surfactants), allowing the preparation of application-tailored and specifically designed and functionalized systems.

Among the more than 30,000 known surface active compounds [301], a large number of surfactants and surfactant mixtures (including many nonphospholipid amphiphiles) have been reported to form vesicles. The basic principles that lead to the formation of vesicles are the same for all, that is, the requirements of (i) an effective packing parameter $p\ (=v/a_0 l)_{\text{eff}} \approx 1$ (see Section 1.1), (ii) chain flexibility ($T > T_m$), and (iii) sufficiently low amphiphile concentration (global packing constraints) [7]. Examples of vesicle-forming amphiphiles include

- cationic di-*n*-alkyldimethylammonium ions [302, 303] (such as di-*n*-dodecyldimethylammonium bromide (DDAB) [304] and other di-*n*-dodecyldimethylammonium halides [305] and di-*n*-octadecyldimethylammonium chloride [306–308] or bromide [308, 309] or other counter-ions [310]);
- the cationic oleyldimethylaminoxide [311];
- anionic phospholipids (such as egg yolk phosphatidic acid mixtures and ox brain phosphatidylserine mixtures [312, 313]);
- anionic di-*n*-alkylphosphates [314–316] (e.g., di-*n*-dodecylphosphate [317], di-*n*-hexadecylphosphate [318–322], and various di-polyprenylphosphates [323, 324]);
- anionic linear and branched monoalkylphosphates (such as different polyprenylphosphates [325], *n*-dodecylphosphate [326], 6-propylnonylphosphate [327], 4-butyloctylphosphate [327], and 2-pentylheptylphosphate [327]);
- anionic tridecyl-6-benzene sulfonate in the presence of salt (sodium chloride) [328];
- anionic fatty acid/soap mixtures (e.g., *n*-octanoic acid/*n*-octanoate [329], *n*-decanoic acid/*n*-decanoate [329–331], oleic acid/oleate [332–334]);
- anionic surfactant/alcohol mixtures (such as sodium *n*-dodecylsulfate/*n*-dodecanol [329, 335] and sodium *n*-decanoate/*n*-decanol [336], or sodium oleate/*n*-octanol [337], which can form a highly viscous phase of densely packed vesicles);
- mixtures of cationic and anionic surfactants, so-called catanionic mixtures (such as *n*-hexadecyltrimethylammonium tosylate/sodium *n*-dodecylbenzenesulfonate [20, 59], *n*-hexadecyltrimethylammonium bromide/sodium *n*-octylsulfate [338–342], and sodium *n*-dodecylsulfate and di-*n*-dodecyldimethylammonium bromide [343, 344]);
- ganglioside GM3 [345, 346];
- phosphatidylnucleosides (such as 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocytidine [347, 348]);
- diblock copolymers [24] (such as the ethylene oxide (EO)/butylene oxide (BO) diblock copolymers EO_6BO_{11}, EO_7BO_{12}, $EO_{11}BO_{11}$, $EO_{14}BO_{10}$, and $EO_{19}BO_{11}$ [349]);

- triblock copolymers (such as the ethylene oxide (EO)/propylene oxide(PO)/ethylene oxide (EO) triblock copolymer $EO_5PO_{68}EO_5$ (called Pluronic L121) [350] and the polymerizable poly(2-methyloxazoline)/ poly(dimethylsiloxane)/ poly(2-ethyloxazoline) triblock copolymers [351]);
- polymerizable amphiphiles [33, 352, 353];
- perfluorated surfactants [354] (such as short-chain perfluorophosphocholines [355] and perfluoroalkyl PCs [356]);
- bolaamphiphiles, which are membrane-spanning amphiphiles with two hydrophilic head groups at the two ends of the molecule and actually form vesicles containing not bilayered but monolayered shells [357–360];
- gemini surfactants that contain two hydrophobic tails and two hydrophilic head groups linked together with a short linker [361, 362];
- industrial, not very well defined surfactant/cosurfactant mixtures (such as *N*-methyl-*N*-alkanoyl-glucamine/ octanol or oleic acid mixtures [363]);
- calixarene-containing [364] or cryptand-containing amphiphiles [365];
- fullerene-containing amphiphiles with two hydrophobic chains and two charged head groups [366];
- double-chain amphiphiles with a polar alkoxysilyl head group, allowing the preparation of a kind of organic-inorganic hybrid vesicle, called cerasomes [367];
- triple-chain amphiphiles containing three hydrophobic chains and two charged polar head groups [365, 368];

and many more (e.g., [33, 358, 369]).

Finally, complex vesicle-based surfactant aggregates can be prepared (e.g., large vesicles composed of one or more types of amphiphiles may contain vesicles of another type of amphiphile), based on principles that include the specific molecular recognition between different types of vesicles initially prepared by one of the established methods described in Section 3 [370–372].

It is rather obvious that one consequence of the fact that vesicle formation is observed from so many different classes of amphiphiles (or mixtures of amphiphiles) is the very broad range of applications in very different fields. Vesicles are applied—or investigated for potential applications—at least in the following areas:

- in pharmacology and medicine [69, 73, 296, 297, 373–378] as parenteral or topical drug delivery systems [297, 379–381], in the treatment of infectious diseases, in anticancer therapy, as gene delivery systems, as immunoadjuvants, and as diagnostics;
- in immunoassays [382–385];
- in chromatographic separations using immobilized vesicles [386, 387];
- in cosmetics as formulations for water and nutrient delivery to the skin [388–391];
- in a variety of biophysical investigation of biological membrane components, including the reconstitution and use of membrane-soluble proteins [4, 69, 258, 263, 392];
- in research on membrane-soluble ion channels [393–399];
- in research related to the question of the origin and evolution of life, as models for the precursor structures of the first cells [400];
- in research aimed at constructing artificial (or minimal) cells [401, 402], for example, for potential biotechnological applications [402];
- in food technology and nutrition as carrier systems for food additives and ingredients and for the control of certain food processes (e.g., cheese ripening) [403–405];
- in agrochemistry [406];
- as nanometer- or micrometer-sized bioreactors containing catalytically active enzymes [75, 407–411];
- in nanoparticle technology [190, 299, 412–414], for example, for the preparation of semiconductor particles [412];
- in catalytic processes as simple models for enzymes [415–420] or simple models of other protein functionalities (e.g., as catalysts for the unfolding [421] or folding [387, 422] of proteins);
- in biosensor developments [423–425], particularly for the controlled preparation of bilayers adsorbed to solid surfaces [426–429];
- in the extraction of heavy metal ions with the help of functionalized, metal-sorbing vesicles [430];
- as supramolecular, nanostructured polymeric materials (as polymerized vesicles) [431–433];
- in biomineralization [434–436];
- as templates for the synthesis of inorganic mesoporous materials [437, 438] or biomaterials [436] and in the preparation of hollow polymer capsules [356, 439–441];
- as templates for modifying the distribution of reaction products, for example, in reactions that lead to products that are only sparingly soluble in the absence, but soluble in the presence, of vesicles [442, 443];
- as a medium for the preparation of size-defined polymer particles from monomers that are soluble in the vesicle shell [444];
- as supramolecular, self-assembly-based devices [33, 298, 299, 358, 445], for example, for the conversion of light energy into chemical energy (artificial photosynthesis) [446, 447], for signaling and switching [448, 449], for the construction of molecular wires [450], and for a number of different redox processes [418];

and many more (e.g., [295, 299]).

One illustration of vesicles loaded with water-soluble molecules is shown in Figure 5. A cryo-transmission electron micrograph of POPC vesicles containing the protein ferritin is shown. The vesicles were prepared in the presence of ferritin, and the nonentrapped protein molecules were separated from the vesicles by size exclusion chromatography after vesicle preparation [451, 452]. Since this particular protein has a dense iron core with a size of about 8 nm, it is visible by electron microscopy, and the actual number of protein molecules per individual vesicle can be directly counted. This type of loaded vesicles can be used, for example, in basic studies on vesicle transformation processes [451, 452]. If the vesicles contain catalytically active proteins (enzymes), they may be used in drug delivery or as small bioreactors [75, 410, 411].

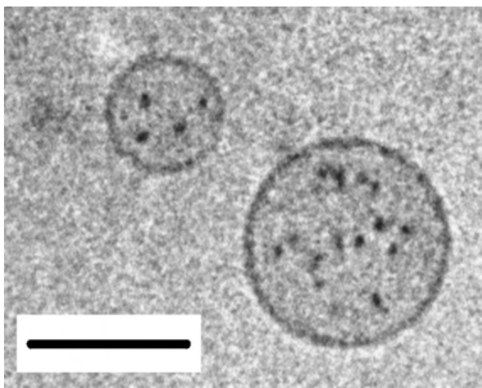


Figure 5. Cryo-transmission electron micrograph of two unilamellar vesicles that have been prepared from POPC by the reversed-evaporation technique (see Section 3.14) loaded with the iron storage protein ferritin, followed by extrusion ($REV\text{-}VET_{100}$). Nonentrapped ferritin molecules were removed by size-exclusion chromatography. Each black spot inside the vesicles represents one iron core (the core of an individual protein). The length of the bar corresponds to 100 nm. The electron micrograph was taken by M. Müller and N. Berclaz, Service Laboratory for Electron Microscopy, at the Department of Biology at the ETH Zürich. See [451, 452] for experimental details and for an application of ferritin-containing vesicles in the investigation of vesicle transformation processes.

Conceptually, an interesting principle of vesicle applications is vesicle transformation (at the site of application) to another type of surfactant assembly as a result of temperature changes that lead to changes in the surfactant packing parameter p (see Section 1.1). One particular example is the transformation of a 1-monoolein MLV suspension (L_α-phase) into a 1-monoolein bicontinuous cubic phase [9, 453–456].

There is no doubt that the number of applications will increase with increasing molecular complexity of the vesicular systems. Vesicles—and other surfactant assemblies (e.g., hexagonal and cubic phases)—will be applied more and more in all fields related to nanoscience and nanotechnology. The title of a recent publication in the field of vesicle drug delivery is an example of one of the directions in which vesicle research and application may go: "Biotinylated Stealth Magnetoliposomes" [457], a particular vesicle preparation that combines steric stabilization (Stealth) with molecular recognition (biotin) and magnetic nanoparticle properties. In other words, it is likely that the complexity of the vesicles prepared and investigated will increase to make them functional, possibly by combining the principles of supramolecular chemistry and surfactant self-assembly, to prepare nanometer- or micrometer-sized synthetic systems that may carry out some of the functions biological systems do rather efficiently [458, 459].

6. CONCLUDING REMARKS

There are many different methodologies that have been described in the literature for the preparation of lipid vesicles. Some (but not all), and certainly the most prominent and well known, are mentioned in this review, focusing on normal lipid vesicles (Section 3), although reversed vesicles are briefly mentioned as well (Section 4).

As outlined in Sections 1 and 2, the use of a number of different terms in the field of vesicles (and surfactant assemblies at large) is often rather confusing, and confusion even exists over the general use of the term "surfactant." It is worth pointing out that a reduction in the surface tension of water by surfactants can be achieved not only with classical micelle-forming, single-chain amphiphiles (such as sodium *n*-dodecyl sulfate, SDS) [460], but also with typical vesicle-forming double-chain phospholipids (such as SOPC or DPPC). Certainly, the kinetics and extent of adsorption of a particular surfactant at the water-air interface very much depend on the precise experimental conditions [461]. In the case of DPPC, for example, the equilibrium surface tension and rate of monolayer formation at the water-air interface depend on the temperature [462]. Below T_m, in the solid-ordered state of the molecules (see Section 2), the decrease in the surface tension of water by DPPC is considerably lower than that above T_m [462]. However, there is no doubt that glycerophospholipids like SOPC, POPC, or DPPC are surfactants in the sense of the definition of this term given in Section 1.1.

Since in dilute aqueous solution many of the bilayer-forming surfactants known (at least the most studied phospholipids) at thermodynamic equilibrium do not form a true vesicle phase (with a defined vesicle size, size distribution, and lamellarity), but under the appropriate conditions rather form a lamellar phase (L_α, L_β, L'_β or P'_β) with extended stacked bilayers, the preparation of vesicle dispersions necessarily involves the use of a particular method, a particular technology.

Depending on the experimental conditions, as mentioned in Section 3, depending on the chemical structure of the amphiphile or mixtures of amphiphiles used, the amphiphile concentration, the substances that may be encapsulated in the vesicle's aqueous interior or in the vesicle membrane, etc., there may be one particular method that is more advantageous in comparison with others.

Some of the methods described in Section 3 involve the use of organic solvents. This may be a problem for certain applications. Some methods can be scaled up to a bulk preparation [287], some methods involve mechanical treatments that may cause an inactivation of sensitive molecules one may like to entrap [75]; some methods are cheaper than others; some methods lead to very small vesicles (SUVs); other methods lead mainly to MLVs or LUVs or GUVs, or even MVVs; some methods are fast; some methods are relatively time consuming, etc.

There is no "best method." The choice of a certain method that may be useful very much depends on the particular problem one is trying to solve [67]. There are, at least in the case of the conventional type of phospholipids (or in the case of lipid mixtures containing phospholipids, particularly PC), a few general findings that are often valid; but one should be aware of the existence of possible exceptions:

- The hydration of a dried thin lipid film above the T_m of the lipids usually leads to the formation of MLVs if the film is dispersed vigorously (vortexing or hand-shaking) (see Section 3.2).
- The hydration of a dried thin lipid film above the T_m of the lipids usually leads to the formation of GUVs if

the film is undisturbed while being hydrated slowly (see Section 3.2).

- The electroformation method can be used for the preparation of GUVs with diameters between about 10 and 50 μm (or more) in the case of certain lipids (and lipid mixtures) and buffers with low ionic strength (Section 3.3).
- Relatively homogeneous preparations of SUVs with diameters around 50 nm or below can often be obtained by probe sonication (Section 3.5), although degradation of the surfactants and consequences of the possibly present metal particles (which may be a source of nucleation for vesicle transformation processes) have to be taken into account.
- Rather homogeneous preparations of LUVs with diameters around 50 or 100 nm can be obtained by the extrusion technique as FAT-VET$_{50}$ or FAT-VET$_{100}$ (Section 3.8, Fig. 3).
- The dehydration-rehydration method usually results in high encapsulation yields (Section 3.7).
- Freeze-thaw cycles may lead to vesicle size homogenization and solute equilibration between the external bulk solution and the trapped aqueous solution (Section 3.6).
- The detergent-depletion method often results in rather uniform vesicles with sizes below 100 nm, although the possibility of an incomplete removal of the detergent should be considered (Section 3.27, Table 5).
- The use of volatile cosolvents (oils) during vesicle preparation is often based on the principle that either w/o or w/o/w emulsions are formed from which the solvent is removed, limiting the solvents to all those that have a boiling point considerably lower than the boiling point of water (Sections 3.14–3.17, 3.24, and 3.25).

The physicochemical properties (such as chemical and physical stability, membrane permeability) of the vesicles formed very much depend on the surfactant (or surfactant mixtures) used. Depending on the surfactant, the characteristic properties may be very different from those of the conventional PC type of vesicles. Furthermore, the whole equilibrium phase behavior may be different, and cases are known where even true vesicle phases seem to form "spontaneously" [7, 463–465] and seem to exist at thermodynamic equilibrium [7, 59, 60, 463]. In these cases, vesicles of a certain size and lamellarity just form by mixing, independent of the method of preparation. The stability of the vesicles is understood in the case of the so-called catanionic vesicles (which are composed of a mixture of positively and negatively charged amphiphiles) on the basis of a most likely uneven distribution of the amphiphiles in the two bilayers. This allows the required differences in curvature and molecular packing in the two halves of a bilayer [60, 466], similar to the case of lipid vesicles prepared from mixtures of long-chain and short-chain PCs (Section 3.28) [273, 274]. With respect to the spontaneity in the formation of this type of vesicles, it has been argued, however, that shear forces present during the preparation (mixing of solutions) play an important role in vesicle formation [467].

An interesting case of unilamellar vesicles as thermodynamic equilibrium state has been described in the case of certain ionized phospholipids (e.g., 1,2-dimyristoyl-*sn*-glycero-3-phosphoglycerol, DMPG) [468–472]. Under the experimental conditions used, unilamellar DMPG vesicles apparently only form at the critical temperature (T^*) of 31.6 °C [471], which is different from T_m. Above T^* MLVs are formed; below T^* the phospholipids arrange into a sponge phase [471]. Further investigations are needed to fully understand this critical phenomenon and to clarify whether this unilamellar vesicle formation is a particular case or whether it can be more generally observed.

With the exception of a few cases—such as the detergent depletion method (Section 3.27) (e.g., [256])—the general mechanism of vesicle formation is not yet completely understood in its details, although general principles have been elaborated [7, 69, 473–477].

From a more practical point of view, and by looking at all of the methods described in Section 3, it is evident that the vesicles often form from a preorganized state of the lipids. This preorganization may be

(i) Lamellar sheets present in a dry film deposited on a solid surface (Sections 3.2 and 3.3) or in the ethanolic pro-liposome state (Section 3.19).
(ii) W/o or w/o/w emulsion droplets (Sections 3.14, 3.15, and 3.30).
(iii) Mixed micelles in the case of the detergent-depletion method (Section 3.27) or micelles in the case of chaotropic ion solutions (Section 3.29).

In a few cases, the vesicle formation is triggered as a result of a direct contact of a nonorganized state of the lipids with an aqueous environment, as in the case of the Novasome technology (Section 3.10), the ethanol injection method (Section 3.18), or the ether injection method (Section 3.24). A simple summarizing schematic representation of some of the different pathways for the formation of (normal) lipid vesicles is given in Figure 4. It is expected that more methods for the preparation of vesicles will be developed, although the general pathways may remain the same. The range of pre-organized starting systems from which vesicles can be formed may be expanded, and more will probably soon be understood for the vesicle formation at large.

(Normal) vesicles are just a particular state of aggregation of surfactants (or surfactant mixtures) in an aqueous solution that are topologically closed with an internal aqueous space. Vesicles can be unilamellar or multilamellar, and the general principles of surfactant assembly have been outlined [7–11]. As mentioned above, vesicles sometimes appear to be thermodynamically stable, sometimes not. Like all self-assembled amphiphile aggregates—micelles, microemulsions, cubic phases, even biological membranes, etc.—the formation of vesicles depends on only a few things: (i) local curvature, (ii) global packing constraints (including interaggregate interactions), and (iii) flexibility of the hydrophobic chain(s). All of the methods described essentially depend on satisfying these criteria.

Physicochemical conditions of inside and outside of vesicles often differ greatly. This is a consequence of the closed topology. This fact, together with adverse packing conditions, can often result in a stable state of so-called supra-self-assembly (e.g., surfactant micelles existing inside surfactant *vesicles*) [478–480].

The basic principles that underlie self-assembly of vesicles are quite general and are firmly based in thermodynamics and statistical mechanics [7, 9, 10, 481]. The requirements for “designing” a vesicle are conceptually simple: a packing parameter close to unity (which means effectively a double-chain surfactant or mixed single-chain surfactants), flexible hydrophobic chains (a temperature above T_m), and control of inter- and intraaggregate interactions. Local and global packings are the key principles to consider.

Vesicles, however, are a very small part of a much larger class of self-assembled surfactant aggregates that include cubic phases, which are usually bicontinuous structures of zero average curvature. “Bicontinuous” means that both the aqueous and “oily” parts of the structure are continuously connected over the whole system. Cubic phases or bicontinuous structures in general are ubiquitous in biology for directing biochemical traffic [11]. Likewise, hexagonal phases and microtubules are close to lamellar (and vesicular) phases in a phase diagram [11].

Although these things are known and are even beginning to be understood, quantitative predictions remain a problem. This can be traced to the fact that the underlying theory of molecular forces that underpins physical chemistry and colloid and surface science is flawed [482, 483]. Previous theories cannot deal with specific ion effects (so-called Hofmeister series), dissolved gas, and other solutes that change the water structure. There is rapid development in this area at the moment, which is likely to provide predictability in vesicle design [482–484], which is certainly what one is aiming for.

GLOSSARY

Amphiphile A molecule that comprises at least two opposing parts, a *solvophilic* (for example “hydrophilic”, meaning water-loving) and a *solvophobic* (for example “hydrophobic”, meaning water-hating). Amphiphiles are surfactants.

Detergent A surfactant that in dilute aqueous solution forms micelles, spherical or non-spherical aggregates that contain in the interior of the aggregate the hydrophobic part of the surfactant and on the surface the hydrophilic part of the surfactant.

Glycerophospholipid A particular phospholipid that contains a glycerol backbone to which a phosphate group is bound.

Lamellar phase, L_α A particular liquid crystalline equilibrium state of surfactant molecules, also called liquid-disordered state. In an aqueous environment, the surfactant molecules are arranged in layers in which the hydrophobic part of a surfactant is in the interior of the layer and the hydrophilic part on the two surfaces of the layer, exposed to the aqueous environment.

Liposome Vesicle prepared from amphiphilic lipids.

Phase transition temperature, T_m Characteristic temperature (also called lamellar chain melting temperature) of surfactants that form a lamellar phase. Above the phase transition temperature, the surfactant molecules are in a liquid-disordered state, below in a solid-ordered state.

Phosphatidylcholine, PC A particular glycerophospholipid that contains a choline group in the hydrophilic part, bound to the phosphate.

Phospholipid A surfactant that is present in some of the biological membranes and contains at least one phosphate group.

Reversed vesicle Inverted vesicle formed in a water-immiscible, apolar solvent in the presence of a small amount of water.

Surfactant A molecule that is surface active, meaning that it accumulates at the surface of liquids or solids. Surfactants are amphiphiles.

Vesicle General term to describe any type of hollow, surfactant-based aggregate composed of one or more shells. In the biological literature, the term *vesicle* is used for a particular small, membrane-bounded, spherical organelle in the cytoplasm of an eukaryotic cell.

ACKNOWLEDGMENTS

The author thanks Barry W. Ninham (Department of Applied Mathematics, Australian National University, Canberra, Australia) for extensive discussions on the principles of surfactant self-assembly and for critically reading the manuscript. Manuscript reading and literature advice by Pasquale Stano (Department of Materials Science, ETH, Zürich, Switzerland) are also acknowledged. Furthermore, discussions with Saša Svetina (Institute of Biophysics, University of Ljubljana, Slovenia), Martien A. Cohen Stuart, Frans A. M. Leermakers, and Mireille M. A. E. Claessens (all from the Laboratory of Physical Chemistry and Colloid Science, Wageningen University, Wageningen, The Netherlands) are appreciated.

REFERENCES

1. H. W. Meyer and W. Richter, *Micron* 32, 615 (2001).
2. S. U. Egelhaaf, E. Wehrli, M. Müller, M. Adrian, and P. Schurtenberger, *J. Microsc.* 184, 214 (1996).
3. M. Almgren, K. Edwards, and G. Karlsson, *Colloids Surf. A: Physicochem. Eng. Asp.* 174, 3 (2000).
4. P. L. Luisi and P. Walde, Eds., Giant Vesicles, Perspectives in Supramolecular Chemistry, Vol. 6. Wiley, Chichester, 2000.
5. B. A. Cornell, J. Middlehurst, and F. Separovic, *Biochim. Biophys. Acta* 598, 405 (1980).
6. Y. Tahara and Y. Fujiyoshi, *Micron* 25, 141 (1994).
7. D. J. Mitchell and B. W. Ninham, *J. Chem. Soc., Faraday Trans. 2* 77, 601 (1981).
8. B. W. Ninham and D. F. Evans, *Faraday Discuss. Chem. Soc.* 81, 1 (1986).
9. J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham, *Biochim. Biophys. Acta* 470, 185 (1977).
10. J. N. Israelachvili, “Intermolecular and Surface Forces,” 2nd ed. Academic Press, London, 1992.
11. S. Hyde, S. Andersson, K. Larsson, Z. Blum, T. Landh, S. Lidin, and B. W. Ninham, “The Language of Shape.” Elsevier, Amsterdam, 1997.
12. A. D. Bangham and R. W. Horn, *J. Mol. Biol.* 8, 660 (1964).
13. A. D. Bangham, M. M. Standish, and J. C. Watkins, *J. Mol. Biol.* 13, 238 (1965).
14. F. Szoka, Jr. and D. Papahadjopoulos, *Annu. Rev. Biophys. Bioeng.* 9, 467 (1980).

15. G. Sessa and G. Weissmann, *J. Lipid Res.* 9, 310 (1968).
16. S. C. Kinsky, J. Haxby, C. B. Kinsky, R. A. Demel, and L. L. M. van Deenen, *Biochim. Biophys. Acta* 152, 174 (1968).
17. D. D. Lasic, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. I, p. 1. CRC Press, Boca Raton, FL, 1996.
18. R. T. Hamilton and E. W. Kaler, *J. Colloid Interface Sci.* 116, 248 (1987).
19. E. M. A. Pareira, D. F. S. Petri, and A. M. Carmona-Ribeiro, *J. Phys. Chem. B* 106, 8762 (2002).
20. E. W. Kaler, K. L. Herrington, A. K. Murthy, and J. A. N. Zasadzinski, *J. Phys. Chem.* 96, 6698 (1992).
21. R. G. Laughlin, *Colloids Surf. A: Physicochem. Eng. Asp.* 128, 27 (1997).
22. A. T. Florence, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, p. 157. CRC Press, Boca Raton, FL, 1984.
23. I. F. Uchegbu and A. T. Florence, *Adv. Colloid Interface Sci.* 58, 1 (1995).
24. B. M. Discher, Y.-Y. Won, D. S. Ege, J. C.-M. Lee, F. S. Bates, D. E. Discher, and D. A. Hammer, *Science* 284, 1143 (1999).
25. L. Luo and A. Eisenberg, *J. Am. Chem. Soc.* 123, 1012 (2001).
26. D. E. Discher and A. Eisenberg, *Science* 297, 967 (2002).
27. IUPAC-IUB Commission on Biochemical Nomenclature, *Eur. J. Biochem.* 2, 127 (1967); *Eur. J. Biochem.* 79, 11 (1977); *Biochem. J.* 171, 21 (1978).
28. J. R. Silvius, in "Phospholipids Handbook" (G. Cevc, Ed.), p. 1. Dekker, New York, 1993.
29. A. Streitwieser, C. H. Heathcook, and E. M. Kosower, "Introduction to Organic Chemistry," 4th ed. Macmillian, New York, 1992.
30. A. Wendel, in "Kirk-Othmer Concise Encyclopedia of Chemical Technology," Vol. 15, 4th ed., p. 192. Wiley, New York, 1995.
31. R. R. C. New, Ed., "Liposomes: A Practical Approach." Oxford University Press, New York, 1990.
32. S. Hassan, W. Rowe, and G. J. T. Tiddy, in "Handbook of Applied Surface and Colloid Chemistry" (K. Holmberg, Ed.), Vol. I, p. 466. Wiley, Chichester, 2002.
33. H. Ringsdorf, B. Schlarb, and J. Venzmer, *Angew. Chem. Int. Ed.* 27, 113 (1988).
34. C. Trandum, P. Westh, K. Jørgensen, and O. G. Mouritsen, *Biophys. J.* 78, 2486 (2000).
35. R. M. R. S. Mesquita, E. Melo, T. E. Thompson, and W. L. C. Vaz, *Biophys. J.* 78, 3019 (2000).
36. B.-S. Lee, S. A. Mabry, A. Jonas, and J. Jonas, *Chem. Phys. Lipids* 78, 103 (1995).
37. A. Blume, in "Phospholipids Handbook" (G. Cevc, Ed.), p. 455. Dekker, New York, 1993.
38. The Lipid Data Bank at http://www.ldb.chemistry.ohio-state.edu/ (accessed October 2002).
39. R. Koynova and M. Caffrey, *Biochim. Biophys. Acta* 1376, 91 (1998).
40. R. Koynova and M. Caffrey, *Chem. Phys. Lipids* 115, 107 (2002).
41. S. Marbey-Gaud, in "Liposomes: From Physical Structure to Therapeutic Applications" (C. G. Knight, Ed.), p. 105. Elsevier/North-Holland Biomedical Press, Amsterdam, 1981.
42. R. N. A. H. Lewis and R. N. McElhaney, in "The Structure of Biological Membranes" (P. Yeagle, Ed.), p. 73. CRC Press, Boca Raton, FL, 1992.
43. R. L. Biltonen and D. Lichtenberg, *Chem. Phys. Lipids* 64, 129 (1993).
44. J. Katsaras, S. Tristam-Nagle, Y. Liu, R. L. Headrick, E. Fontes, P. C. Mason, and J. F. Nagle, *Phys. Rev. E.* 61, 5668 (2000).
45. A. J. Verkleij and J. de Gier, in "Liposomes: From Physical Structure to Therapeutic Applications" (C. G. Knight, Ed.), p. 83. Elsevier/North-Holland Biomedical Press, Amsterdam, 1981.
46. W.-J. Sun, S. Tristram-Nagle, R. M. Suter, and J. F. Nagle, *Proc. Natl. Acad. Sci. USA* 93, 7008 (1996).
47. K. Sengupta, V. A. Raghunathan, and J. Katsaras, *Europhys. Lett.* 49, 722 (2000).
48. W. Curatolo, B. Sears, and L. J. Neuringer, *Biochim. Biophys. Acta* 817, 261 (1985).
49. D. V. Lynch and P. L. Steponkus, *Biochim. Biophys. Acta* 984, 267 (1989).
50. G. Cevc, in "Phospholipids Handbook" (G. Cevc, Ed.), p. 939. Dekker, New York, 1993.
51. G. Cevc and D. Marsh, "Phospholipid Bilayers: Physical Principles and Models." Wiley, New York, 1987.
52. P. Machy and L. Leserman, "Liposomes in Cell Biology and Pharmacology." John Libbey Eurotext Ltd./INSERM, London, 1987.
53. D. Lichtenberg, E. Freire, C. F. Schmidt, Y. Barenholz, P. L. Felgner, and T. E. Thompson, *Biochemistry* 20, 3462 (1981).
54. H. Madani and E. W. Kaler, *Langmuir* 6, 125 (1990).
55. U. Olsson and H. Wennerström, *J. Phys. Chem B* 106, 5135 (2002).
56. R. Smith and C. Tanford, *J. Mol. Biol.* 67, 75 (1972).
57. M. Grit and D. J. A. Crommelin, *Chem. Phys. Lipids* 62, 113 (1992).
58. N. J. Zuidam, H. K. M. E. Gouw, Y. Barenholz, and D. J. A. Crommelin, *Biochim. Biophys. Acta* 1240, 101 (1995).
59. E. W. Kaler, A. K. Murthy, B. E. Rodriguez, and J. A. Zasadzinski, *Science* 245, 1371 (1989).
60. S. A. Safran, P. Pincus, and D. Andelman, *Science* 248, 354 (1990).
61. Y. Barenholz, *Prog. Lipid Res.* 41, 1 (2002).
62. H. Ohvo-Rekilä, B. Ramstedt, P. Leppimäki, and J. P. Slotte, *Prog. Lipid Res.* 41, 66 (2002).
63. P. R. Cullis, D. B. Fenske, and M. J. Hope, in "Biochemistry of Lipids, Lipoproteins and Membranes" (D. E. Vance and J. E. Vance, Eds.), p. 1. Elsevier, Amsterdam, 1996.
64. A. D. Bangham, M. W. Hill, and N. G. A. Miller, *Methods Membr. Biol.* 1, 1 (1974).
65. M. J. Hope, M. B. Bally, L. D. Mayer, A. S. Janoff, and P. R. Cullis, *Chem. Phys. Lipids* 40, 89 (1986).
66. D. Lichtenberg and Y. Barenholz, *Methods Biochem. Anal.* 33, 337 (1988).
67. M. C. Woodle and D. Papahdjopoulos, *Methods Enzymol.* 171, 193 (1989).
68. U. K. Nässander, G. Storm, P. A. M. Peeters, and D. J. A. Crommelin, in "Biodegradable Polymers as Drug Delivery Systems" (M. Chasin and R. Langer, Eds.), p. 261. Dekker, New York, 1990.
69. D. D. Lasic, "Liposomes: from Physics to Applications." Elsevier, Amsterdam, 1993.
70. G. V. Betagri, S. A. Jenkins, and D. L. Parsons, "Liposome Drug Delivery Systems." Technomic Publishing, Lancaster, PA, 1993.
71. H. Hauser, in "Phospholipids Handbook" (G. Cevc, Ed.), p. 603. Dekker, New York, 1993.
72. D. J. A. Crommelin and H. Schreier, in "Colloidal Drug Delivery Systems" (J. Kreuter, Ed.), p. 73. Dekker, New York, 1994.
73. Y. Barenholz and D. J. A. Crommelin, in "Encyclopedia of Pharmaceutical Technology" (J. Swarbrick and J. C. Boylan, Eds.), Vol. 9, p. 1. Dekker, New York, 1994.
74. S. Vemuri and C. T. Rhodes, *Pharm. Acta Helv.* 70, 95 (1995).
75. P. Walde and S. Ichikawa, *Biomol. Eng.* 18, 143 (2001).
76. S. M. Gruner, R. P. Lenk, A. S. Janoff, and M. J. Ostro, *Biochemistry* 24, 2833 (1985).
77. W. R. Perkins, S. R. Minchey, M. J. Ostro, T. F. Taraschi, and A. S. Janoff, *Biochim. Biophys. Acta* 943, 103 (1988).
78. R. W. Corkery and S. T. Hyde, *Langmuir* 12, 5528 (1996).
79. R. P. Rand and V. A. Parsegian, *Biochim. Biophys. Acta* 988, 351 (1989).
80. J. F. Nagle and S. Tristram-Nagle, *Biochim. Biophys. Acta* 1469, 159 (2000).
81. D. Lichtenberg and T. Markello, *J. Pharm. Sci.* 73, 122 (1984).
82. K. Akashi, H. Miyata, H. Itoh, and K. Kinosita, Jr., *Biophys. J.* 71, 3242 (1996).
83. E. Evans and D. Needham, *J. Phys. Chem.* 91, 4219 (1987).

84. K. Akashi, H. Miyata, H. Itoh, and K. Kinosita, Jr., *Biophys. J.* 74, 2973 (1998).
85. M. F. Brown and J. Seelig, *Nature* 269, 721 (1977).
86. R. Kataoka, S. Aruga, S. Mitaku, K. Kinosita, Jr., and A. Ikegami, *Biophys. Chem.* 21, 277 (1985).
87. N. Magome, T. Takemura, and K. Yoshikawa, *Chem. Lett.* 205 (1997).
88. M. Alfridsson, B. Ninham, and S. Wall, *Langmuir* 16, 10087 (2000).
89. R.-M. Servuss, *Z. Naturforsch.* 43c, 938 (1988).
90. J. P. Reeves and R. M. Dowben, *J. Cell Physiol.* 73, 49 (1969).
91. P. Mueller, T. F. Chien, and B. Rudy, *Biophys. J.* 44, 375 (1983).
92. D. Needham and E. Evans, *Biochemistry* 27, 8261 (1988).
93. D. Needham and D. Zhelev in "Giant Vesicles," Perspectives in Supramolecular Chemistry (P. L. Luisi and P. Walde, Eds.), Vol. 6, p. 103. Wiley, Chichester, 2000.
94. D. Needham and D. V. Zhelev, in "Vesicles," Surfactant Science Series (M. Rosoff, Ed.), Vol. 62, p. 373. Dekker, New York, 1996.
95. R. Virchow, *Arch. Pathol. Anat. Physiol. Klin. Med.* 6, 562 (1854).
96. O. Lehmann, "Die neue Welt der flüssigen Kristalle." Akad. Verlagsgesellschaft, Leipzig, 1911.
97. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Molecular Biology of the Cell," 4th ed. Garland Science, New York, 2002.
98. I. Sakurai and Y. Kawamura, *Biochim. Biophys. Aca* 777, 347 (1984).
99. I. Sakurai, T. Suzuki, and S. Sakurai, *Biochim. Biophys. Acta* 985, 101 (1989).
100. D. D. Lasic, J. Kidric, and S. Zagorc, *Biochim. Biophys. Acta* 896, 117 (1987).
101. D. D. Lasic, *J. Colloid Interface Sci.* 124, 428 (1988).
102. M. I. Angelova and D. D. Dimitrov, *Faraday Discuss. Chem. Soc.* 81, 303 (1986).
103. M. I. Angelova, in "Giant Vesicles," Perspectives in Supramolecular Chemistry (P. L. Luisi and P. Walde, Eds.), Vol. 6, p. 27. Wiley, Chichester, 2000.
104. P. Bucher, A. Fischer, P. L. Luisi, T. Oberholzer, and P. Walde, *Langmuir* 14, 2712 (1998).
105. R. Wick, M. I. Angelova, P. Walde, and P. L. Luisi, *Chem. Biol.* 3, 105 (1996).
106. J. M. Holopainen, M. I. Angelova, and P. K. J. Kinnunen, *Biophys. J.* 78, 830 (2000).
107. M. I. Angelova, S. Soléau, P. Méléard, J.-F. Faucon, and P. Bothorel, *Prog. Colloid Polym. Sci.* 89, 127 (1992).
108. C. Dietrich, M. I. Angelova, and B. Pouligny, *J. Phys. II France* 7, 1651 (1997).
109. F. M. Menger and M. I. Angelova, *Acc. Chem. Res.* 31, 789 (1998).
110. P. Méléard, C. Gerbeaud, T. Pott, and M. D. Mitov in "Giant Vesicles," Perspectives in Supramolecular Chemistry (P. L. Luisi and P. Walde, Eds.), Vol. 6, p. 185. Wiley, Chichester, 2000.
111. B. Bozic, S. Svetina, B. Zekš, and R. E. Waugh, *Biophys. J.* 61, 963 (1992).
112. L. Mathivet, S. Cribier, and P. F. Devaux, *Biophys. J.* 70, 1112 (1996).
113. V. Kralj-Iglic, G. Gomišcek, J. Majhenc, V. Arrigler, and S. Svetina, *Colloids Surf. A: Physicochem. Eng. Aspects* 181, 315 (2001).
114. L. A. Bagatolli, T. Parasassi, and E. Gratton, *Chem. Phys. Lipids* 105, 135 (2000).
115. H. Kikuchi, H. Yamauchi, and S. Hirota, *Chem. Pharm. Bull.* 39, 1522 (1991).
116. C. Huang and J. T. Mason, *Proc. Natl. Acad. Sci. USA* 75, 308 (1978).
117. R. Lawaczek, M. Kianosho, J. L. Giraudet, and S. I. Chan, *Nature* 256, 584 (1975).
118. R. Lawaczek, M. Kianosho, and S. I. Chan, *Biochim. Biophys. Acta* 443, 313 (1976).
119. C. Huang, *Biochemistry* 8, 344 (1969).
120. O. Bakouche, D. Gerlier, J.-M. Letoffe, and P. Claudy, *Biophys. J.* 50, 1 (1986).
121. M. P. Sheetz and S. I. Chan, *Biochemistry* 11, 4573 (1972).
122. D. Paphadjopoulos and J. C. Watkins, *Biochim. Biophys. Acta* 135, 639 (1967).
123. S. M. Johnson, *Biochim. Biophys. Acta* 307, 27 (1973).
124. Y. Barenholz, D. Gibbes, B. J. Litman, J. Goll, T. E. Thompson, and F. D. Carlson, *Biochemistry* 16, 2806 (1977).
125. H. Hauser, *Biochem. Biophys. Res. Commun.* 45, 1049 (1971).
126. S. Mitaku and T. Date, *Biochim. Biophys. Acta* 688, 411 (1982).
127. R. A. Klein, *Biochim. Biophys. Acta* 210, 486 (1970).
128. R. Qiu and R. C. MacDonald, *Biochim. Biophys. Acta* 1191, 343 (1994).
129. W. H. Pitcher III and W. H. Huestis, *Biochem. Biophys. Res. Commun.* 296, 1352 (2002).
130. M. Wong, F. H. Anthony, T. W. Tillack, and T. E. Thompson, *Biochemistry* 21, 4126 (1982).
131. S. E. Schullery, C. F. Schmidt, P. Felgner, T. W. Tillack, and T. E. Thompson, *Biochemistry* 19, 3919 (1980).
132. A. L. Larrabee, *Biochemistry* 18, 3321 (1979).
133. M. J. Hope, R. Nayar, L. D. Mayer, and P. R. Cullis, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 123. CRC Press, Boca Raton, FL, 1993.
134. U. Pick, *Arch. Biochem. Biophys.* 212, 186 (1981).
135. K. Anzai, M. Yoshida, and Y. Kirino, *Biochim. Biophys. Acta* 1021, 21 (1990).
136. M. Traïkia, D. E. Warschawski, M. Recouvreur, J. Cartaud, and P. F. Devaux, *Eur. Biophys. J.* 29, 184 (2000).
137. R. C. MacDonald, F. D. Jones, and R. Qui, *Biochim. Biophys Acta* 1191, 362 (1994).
138. R. C. MacDonald and R. I. MacDonald, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 209. CRC Press, Boca Raton, FL, 1993.
139. N. F. Bunkin, O. A. Kiseleva, A. V. Lobeyev, T. G. Movchan, B. W. Ninham, and O. I. Vinogradova, *Langmuir* 13, 3024 (1997).
140. N. Oku and R. C. MacDonald, *Biochemistry* 22, 855 (1983).
141. M. Kasahara and P. C. Hinckle, *J. Biol. Chem.* 252, 7384 (1977).
142. L. D. Mayer, M. J. Hope, P. R. Cullis, and A. S. Janoff, *Biochim. Biophys. Acta* 817, 193 (1985).
143. C. J. Chapman, W. L. Erdahl, R. W. Talor, and D. R. Pfeiffer, *Chem. Phys. Lipids* 55, 73 (1990).
144. C. J. Chapman, W. E. Erdahl, R. W. Taylor, and D. R. Pfeiffer, *Chem. Phys. Lipids* 60, 201 (1991).
145. P. A. Monnard, N. Berclaz, K. Conde-Frieboes, and T. Oberholzer, *Langmuir* 15, 7504 (1999).
146. K. Higashi, S. Suzuki, H. Fujii, and Y. Kirino, *J. Biochem.* 101, 433 (1987).
147. D. W. Deamer and G. L. Barchfeld, *J. Mol. Evol.* 18, 203 (1982).
148. C. Kirby and G. Gregoriadis, *BioTechnology* 2, 979 (1984).
149. R. L. Shew and D. W. Deamer, *Biochim. Biophys. Acta* 816, 1 (1985).
150. T. D. Madden and N. Boman, in "Liposomes: Rational Design" (A. S. Janoff, Ed.), p. 261. Dekker, New York, 1999.
151. P. R. Harrigan, T. D. Madden, and P. R. Cullis, *Chem. Phys. Lipids* 52, 139 (1990).
152. T. Suzuki, H. Komatsu, and K. Miyajima, *Biochim. Biophys. Acta* 1278, 176 (1996).
153. E. C. A. van Winden, W. Zhang, and D. J. A. Crommelin, *Pharm. Res.* 14, 1151 (1997).
154. K. Miyajima, *Adv. Drug Delivery Rev.* 24, 151 (1997).
155. W. Zhang, A. C. A. van Winden, J. A. Bouwstra, and D. J. A. Crommelin, *Cryobiology* 35, 277 (1997).
156. E. C. A. van Winden and D. J. A. Crommelin, *J. Controlled Release* 58, 69 (1999).
157. J. H. Cowe and L. M. Cowe, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 229. CRC Press, Boca Raton, FL, 1993.

158. J. D. Castile and K. M. G. Taylor, *Int. J. Pharm.* 188, 87 (1999).
159. K. L. Koster, M. S. Webb, G. Bryant, and D. V. Lynch, *Biochim. Biophys. Acta* 1193, 143 (1994).
160. B. Zadi and G. Gregoriadis, *J. Liposome Res.* 10, 73 (2000).
161. M. J. Hope, M. B. Bally, G. Webb, and P. R. Cullis, *Biochim. Biophys. Acta* 812, 55 (1985).
162. L. D. Mayer, M. J. Hope, and P. R. Cullis, *Biochim. Biophys. Acta* 858, 161 (1986).
163. M. J. Hope, R. Nayar, L. D. Mayer, and P. R. Cullis, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 123. CRC Press, Boca Raton, FL, 1993.
164. A. M. Brendzel and I. F. Miller, *Biochim. Biophys. Acta* 596, 129 (1980).
165. R. Nayar, M. J. Hope, and P. R. Cullis, *Biochim. Biophys. Acta* 986, 200 (1989).
166. H, Jousma, H. Talsma, F. Spies, J. G. H. Joosten, H. E. Junginger, and D. J. A. Crommelin, *Int. J. Pharm.* 35, 263 (1987).
167. V. Dorovska-Taran, R. Wick, and P. Walde, *Anal. Biochem.* 240, 37 (1996).
168. F. Olson, C. A. Hunt, F. C. Szoka, W. J. Vail, and D. Papahadjopoulos, *Biochim. Biophys. Acta* 557, 9 (1979).
169. F. Szoka, F. Olson, T. Heath, W. Vail, E. Mayhew, and D. Papahadjopoulos, *Biochim. Biophys. Acta* 601, 559 (1980).
170. S. G. Clerc and T. E. Thompson, *Biophys. J.* 67, 475 (1994).
171. D. G. Hunter and B. J. Frisken, *Biophys. J.* 74, 2996 (1998).
172. B. J. Frisken, C. Asman, and P. J. Patty, *Langmuir* 16, 928 (2000).
173. N. Berger, A. Sachse, J. Bender, R. Schubert, and M. Brandl, *Int. J. Pharm.* 223, 55 (2001).
174. R. C. MacDonald, R. I. MacDonald, B. P. M. Menco, K. Takeshita, N. K. Subbarao, and L.-R. Hu, *Biochim. Biophys. Acta* 1061, 297 (1991).
175. T. Schneider, A. Sachse, G. Rössling, and M. Brandl, *Drug Dev. Ind. Pharm.* 20, 2787 (1994).
176. Y. Barenholz, S. Amselem, and D. Lichtenberg, *FEBS Lett.* 99, 210 (1979).
177. R. L. Hamilton, J. Goerke, L. S. S. Guo, M. C. Williams, and R. J. Havel. *J. Lipid Res.* 21, 981 (1980).
178. E. Mayhew, R. Lazo, W. J. Vail, and A. M. Green, *Biochim. Biophys. Acta* 775, 169 (1984).
179. H. Talsma, A. Y. Özer, L. van Bloois, and D. J. A. Crommelin, *Drug Dev. Ind. Pharm.* 15, 197 (1989).
180. G. Gregoriadis, H. da Silva, and A. T. Florence, *Int. J. Pharm.* 65, 235 (1990).
181. J.-C. Vuillemard, *J. Microencapsulation* 8, 547 (1991).
182. M. M. Brandl, D. Bachmann, M. Drechsler, and K. H. Bauer, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 49. CRC Press, Boca Raton, FL, 1993.
183. R. Barnadas-Rodríguez and M. Sabés, *Int. J. Pharm.* 213, 175 (2001).
184. F. L. Sorgi and L. Huang, *Int. J. Pharm.* 144, 131 (1996).
185. D. F. H. Wallach and J. R. Philippot, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 141. CRC Press, Boca Raton, FL, 1993.
186. O. Diat and D. Roux, *J. Phys. II France* 3, 9 (1993).
187. F. Gauffre and D. Roux, *Langmuir* 15, 3070 (1999).
188. T. Gulik-Krzywicki, J. C. Dedieu, D. Roux, C. Degert, and R. Laversanne, *Langmuir* 12, 4668 (1996).
189. O. Diat, D. Roux, and F. Nallet, *J. Phys. II France* 3, 1427 (1993).
190. D.-W. Kim, S.-G. Oh, S.-C. Yi, S.-Y. Bae, and S.-K. Moon, *Chem. Mater.* 12, 996 (2000).
191. J. I. Escalante, M. Gradzielski, H. Hoffmann, and K. Mortensen, *Langmuir* 16, 8653 (2000).
192. T. D. Le, U. Olson, and K. Mortensen, *Phys. Chem. Chem. Phys.* 3, 1310 (2001).
193. F. Gauffre and D. Roux, *Langmuir* 15, 3070 (1999).
194. A. Bernheim-Grosswasser, S. Ugazio, F. Gauffre, O. Viratelle, P. Mahy, and D. Roux, *J. Chem. Phys.* 112, 3424 (2000).
195. N. Mignet, A. Brun, C. Degert, B. Delord, D. Roux, C. Hélène, R. Laversanne, and J.-C. François, *Nucleic Acids Res.* 28, 3134 (2000).
196. P. André, B. W. Ninham, and M. P. Pileni, *New J. Chem.* 25, 563 (2001).
197. H. Talsma, M. J. van Steenbergen, J. C. H. Borchert, and D. J. A. Crommelin, *J. Pharm. Sci.* 83, 276 (1994).
198. S. Gould-Fogerite and R. J. Mannino, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 67. CRC Press, Boca Raton, FL, 1993.
199. D. Papahadjopoulos, W. J. Vail, K. Jacobson, and G. Poste, *Biochim. Biophys. Acta* 394, 483 (1975).
200. F. Szoka, Jr. and D. Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* 75, 4194 (1978).
201. N. Düzgünes, J. Wilschut, K. Hong, R. Fraley, C. Perry, D. S. Friend, T. L. James, and D. Papahadjopoulos, *Biochim. Biophys. Acta* 732, 289 (1983).
202. C. Pidgeon, A. H. Hunt, and K. Dittrich, *Pharm. Res.* 3, 23 (1986).
203. C. Pidgeon, S. McNeely, T. Schmidt, and J. E. Johnson, *Biochemistry* 26, 17 (1987).
204. C. Pidgeon, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 99. CRC Press, Boca Raton, FL, 1993.
205. K. Otake, T. Imura, H. Sakai, and M. Abe, *Langmuir* 17, 3898 (2001).
206. F. Ishii, A. Takamura, and H. Ogata, *J. Dispersion Sci. Technol.* 9, 1 (1988).
207. F. Ishii, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 111. CRC Press, Boca Raton, FL, 1993.
208. S. Kim, R. E. Jacobs, and S. H. White, *Biochim. Biophys. Acta* 812, 793 (1985).
209. S. Kim and G. M. Martin, *Biochim. Biophys. Acta* 646, 1 (1981).
210. S. Kim, M. S. Turker, E. Y. Chi, S. Sela, and G. M. Martin, *Biochim. Biophys. Acta* 782, 339 (1983).
211. S. Mantripragada, *Prog. Lipid Res.* 41, 392 (2002).
212. M. S. Spector, A. Zasadzinksi, and M. B. Sankaram, *Langmuir* 12, 4704 (1996).
213. A. Moscho, O. Orwar, D. T. Chiu, B. P. Modi, and R. N. Zare, *Proc. Natl. Acad. Sci. USA* 93, 11443 (1996).
214. D. T. Chiu, C. F. Wilson, F. Ryttsén, A. Strömberg, C. Farre, A. Karlsson, S. Nordholm, A. Gaggar, B. P. Modi, A. Moscho, R. A. Garza-López, O. Orwar, and R. N. Zare, *Science* 283, 1892 (1999).
215. S. Batzri and E. D. Korn, *Biochim. Biophys. Acta* 298, 1015 (1973).
216. J. M. H. Kremer, M. W. J. v. d. Esker, C. Pathmamanoharan, and P. H. Wiersema, *Biochemistry* 16, 3932 (1977).
217. A. S. Domazou and P. L. Luisi, *J. Liposome Res.* 12, 205 (2002).
218. M. Pons, M. Foradada, and J. Estelrich, *Int. J. Pharm.* 95, 51 (1993).
219. A. Wagner, K. Vorauer-Uhl, G. Kreismayr, and H. Katinger, *J. Liposome Res.* 12, 271 (2002).
220. A. Wagner, K. Vorauer-Uhl, G. Kreismayr, and H. Katinger, *J. Liposome Res.* 12, 259 (2002).
221. R. W. Kriftner, in "Liposome Dermatics" (O. Braun-Falco, H. C. Korting, and H. I. Maibach, Eds.), p. 203. Springer-Verlag, Berlin, 1992.
222. R. Naeff, *Adv. Drug Delivery Rev.* 18, 343 (1996).
223. Y. Maitani, H. Soeda, W. Junping, and K. Takayama, *J. Liposome Res.* 11, 115 (2001).
224. W. P. Williams, S. Perrett, M. Golding, J.-P. Arnaud, and M. Michez, in "Phospholipids: Characterization, Metabolism, and Novel Biological Application" (G. Cevc and F. Paltauf, Eds.), p. 181. AOCS Press, Champaign, IL, 1995.
225. S. Perrett, M. Golding, and W. P. Williams, *J. Pharm. Pharmacol.* 43, 154 (1991).
226. P. Williams, *SÖFW J.* 118, 377 (1992).
227. U. Isele, P. van Hoogevest, R. Hilfiker, H.-G. Capraro, K. Schieweck, and H. Leuenberger, *J. Pharm. Sci.* 83, 1608 (1994).

228. N. I. Payne, R. Timmins, V. Ambrose, M. D. Ward, and F. Ridgway, *J. Pharm. Sci.* 75, 325 (1986).
229. E. Touitou, N. Dayan, L. Bergelson, B. Godin, and M. Eliaz, *J. Controlled Rel.* 65, 403 (2000).
230. N. Dayan and E. Touitou, *Biomaterials* 21, 1879 (2000).
231. B. W. Barry, *Eur. J. Pharm. Sci.* 14, 101 (2001).
232. P. L. Ahl, L. Chen, W. R. Perkins, S. R. Minchey, L. T. Boni, T. F. Taraschi, and A. S. Janoff, *Biochim. Biophys. Acta* 1195, 237 (1994).
233. F. Ishii, A. Takamura, and Y. Ishigami, *Langmuir* 11, 483 (1995).
234. F. M. Menger and B. M. Sykes, *Langmuir* 14, 4131 (1998).
235. D. F. Evans and H. Wennerstöm, "The Colloidal Domain: Where Physics, Chemistry, Biology and Technology Meet," 2nd ed. Wiley-VCH, New York, 1999.
236. L. Frederiksen, K. Anton, P. van Hoogevest, H. R. Keller, and H. Leuenberger, *J. Pharm. Sci.* 86, 921 (1997).
237. D. W. Deamer, *Ann. N.Y. Acad. Sci.* 308, 250 (1978).
238. J. T. Buboltz and G. W. Feigenson, *Biochim. Biophys. Acta* 1417, 232 (1999).
239. L. Zhang, J. Hu, and Z. Lu, *J. Colloid Interface Sci.* 190, 76 (1997).
240. A. Helenius and K. Simons, *Biochim. Biophys. Acta* 415, 29 (1975).
241. D. Lichtenberg, R. J. Robson, and E. A. Dennis, *Biochim. Biophys. Acta* 737, 285 (1983).
242. D. Lichtenberg, *Biochim. Biophys. Acta* 821, 470 (1985).
243. J. Lasch, *Biochim. Biophys. Acta* 1241, 269 (1995).
244. M. H. W. Milsmann, R. A. Schwendener, and H.-G. Weder, *Biochim. Biophys. Acta* 512, 147 (1978).
245. O. Zumbühl and H.-G. Weder, *Biochim. Biophys. Acta* 640, 252 (1981).
246. J. Philippot, S. Mustafschiev, and J. P. Liautard, *Biochim. Biophys. Acta* 734, 137 (1983).
247. M. Ueno, C. Tanford, and J. A. Reynolds, *Biochemistry* 23, 3070 (1984).
248. R. A. Parente and B. R. Lentz, *Biochemistry* 23, 2353 (1984).
249. T. M. Allen, A. Y. Romans, H. Kercret, and J. P. Segrest, *Biochim. Biophys. Acta* 601, 328 (1980).
250. M. Rotenberg and D. Lichtenberg, *J. Colloid Interface Sci.* 144, 591 (1991).
251. J. S. Tauskela, M. Akler, and M. Thompson, *Anal. Biochem.* 201, 282 (1992).
252. M. Wacker and R. Schubert, *Int. J. Pharm.* 162, 171 (1998).
253. N. Funasaki, S. Hada, and S. Neya, *J. Phys. Chem. B* 103, 169 (1999).
254. N. Funasaki, M. Nomura, S. Ishikawa, and S. Neya, *J. Phys. Chem. B* 104, 7745 (2000).
255. Y. Li, J. F. Holzwarth, and C. Bohne, *Langmuir* 16, 2038 (2000).
256. J. Leng, S. U. Egelhaaf, and M. E. Cates, *Europhys. Lett.* 59, 311 (2002).
257. J. Brunner, P. Skrabal, and H. Hauser, *Biochim. Biophys. Acta* 455, 322 (1976).
258. J.-L. Rigaud, D. Levy, G. Mosser, and O. Lambert, *Eur. Biophys. J.* 27, 305 (1998).
259. J. R. Philippot and J. P. Liautard, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 1, 2nd ed., p. 81. CRC Press, Boca Raton, 1993.
260. P. Schurtenberger, N. Mazer, S. Waldvogel, and W. Känzig, *Biochim. Biophys. Acta* 775, 111 (1984).
261. H. G. Enoch and P. Strittmatter, *Proc. Natl. Acad. Sci. USA* 76, 145 (1979).
262. R. Peschka, T. Purmann, and R. Schubert, *Int. J. Pharm.* 162, 177 (1998).
263. J.-L. Rigaud, B. Pitard, and D. Levy, *Biochim. Biophys. Acta* 1231, 223 (1995).
264. D. Lichtenberg, E. Opatowski, and M. M. Kozlov, *Biochim. Biophys. Acta* 1508, 1 (2000).
265. M. Almgren, *Biochim. Biophys. Acta* 1508, 146 (2000).
266. M. Ollivon, S. Lesieur, C. Grabielle-Madelmont, and M. Paternostre, *Biochim. Biophys. Acta* 1508, 34 (2000).
267. S. Almog, T. Kushnir, S. Nir, and D. Lichtenberg, *Biochemistry* 25, 2597 (1986).
268. M. Paternostre, O. Meyer, C. Grabielle-Madelmont, S. Lesieur, M. Ghanam, and M. Ollivon, *Biophys. J.* 69, 2476 (1995).
269. C. Sun and M. Ueno, *Colloid Polym. Sci.* 278, 855 (2000).
270. M. N. Jones and D. Chapman, "Micelles, Monolayers, and Biomembranes." Wiley-Liss, New York, 1995.
271. P. Walde, E. Blöchliger, and K. Morigaki, *Langmuir* 15, 2346 (1999).
272. H. Hauser, *Biochim. Biophys. Acta* 1508, 164 (2000).
273. N. E. Gabriel and M. F. Roberts, *Biochemistry* 23, 4011 (1984).
274. N. E. Gabriel and M. F. Roberts, *Biochemistry* 25, 2812 (1986).
275. G. Cevc, A. Schätzlein, and G. Blume, *J. Controlled Release* 36, 3 (1995).
276. G. Cevc, D. Gebauer, J. Stieber, A. Schätzlein, and G. Blume, *Biochim. Biophys. Acta* 1368, 201 (1998).
277. S. E. Friberg, H. Yang, L. Fei, S. Sadasivan, D. H. Rasmussen, and P. A. Aikens, *J. Dispersion Sci. Technol.* 19, 19 (1998).
278. N. Oku and R. C. MacDonald, *Biochim. Biophys. Acta* 734, 54 (1983).
279. N. Oku and R. C. MacDonald, *J. Biol. Chem.* 258, 8733 (1983).
280. S. Matsumoto, M. Kohda, and S.-I. Murata, *J. Colloid Interface Sci.* 62, 149 (1977).
281. H. Kunieda, K. Nakamura, and D. F. Evans, *J. Am. Chem. Soc.* 113, 1051 (1991).
282. H. Kunieda, K. Nakamura, H. T. Davis, and D. F. Evans, *Langmuir* 7, 1915 (1991).
283. H. Kunieda, N. Kanai, A. Uemoto, and I. Tobita, *Langmuir* 10, 4006 (1994).
284. U. Olsson, K. Nakamura, H. Kunieda, and R. Strey, *Langmuir* 12, 3045 (1996).
285. H. Kunieda, K. Shigeta, and M. Suzuki, *Langmuir* 15, 3118 (1999).
286. H. Mays, M. Almgren, A. Dedinaite, and P. M. Claesson, *Langmuir* 15, 8072 (1999).
287. Y. Barenholz, in "Medical Applications of Liposomes" (D. D. Lasic and D. Papahadjopoulos, Eds.), p. 545. Elsevier, Amsterdam, 1998.
288. J. R. Bellare, H. T. Davis, L. E. Scriven, and Y. Talmon, *J. Electron Microsc. Tech.* 10, 87 (1988).
289. B. Klösgen and W. Helfrich, *Eur. Biophys. J.* 22, 329 (1993).
290. Y. Talmon, *Ber. Bunsenges. Phys. Chem.* 100, 364 (1996).
291. O. Maier, V. Oberle, and D. Hoeckstra, *Chem. Phys. Lipids* 116, 3 (2002).
292. P. F. Devaux, P. Fellmann, and P. Hervé, *Chem. Phys. Lipids* 116, 115 (2002).
293. M. Fröhlich, V. Brecht, and R. Peschka-Süss, *Chem. Phys. Lipids* 109, 103 (2001).
294. D. D. Lasic, in "Handbook of Biological Physics" (R. Lipowsky and E. Sackmann, Eds.), Vol. 1, p. 491. Elsevier, Amsterdam, 1995.
295. D. D. Lasic and Y. Barenholz, Eds., "Handbook of Nonmedical Applications of Liposomes," Vols. 1–4. CRC Press, Boca Raton, FL, 1996.
296. D. D. Lasic and D. Papahadjopoulos, Eds., "Medical Applications of Liposomes." Elsevier, Amsterdam, 1998.
297. A. S. Janoff, Ed., "Liposomes: Rational Design." Dekker, New York, 1999.
298. J. H. Fendler, "Membrane Mimetic Chemistry." Wiley, New York, 1982.
299. J. H. Fendler, *Chem. Rev.* 87, 877 (1987).
300. R. B. Gennis, "Biomembranes: Molecular Structure and Function." Springer-Verlag, New York, 1989.
301. I. Ash, "Handbook of Industrial Surfactants." Synapse Information Resources, New York, 2000.
302. Y. Okahata, R. Ando, and T. Kunitake, *Ber. Bunsenges. Phys. Chem.* 85, 789 (1981).

303. M. J. Blandamer, B. Briggs, P. M. Cullis, S. D. Kirby, and J. B. F. N. Engberts, *J. Chem. Soc. Faraday Trans.* 93, 453 (1997).
304. T. Kunitake and Y. Okahata, *J. Am. Chem. Soc.* 99, 3860 (1977).
305. B. W. Ninham, D. F. Evans, and G. J. Wei, *J. Phys. Chem.* 87, 5020 (1983).
306. A. M. Carmona-Ribeiro and B. R. Midmore, *J. Phys. Chem.* 96, 3542 (1992).
307. R. G. Laughlin, R. L. Munyon, J. L. Burns, T. W. Coffindaffer, and Y. Talmon, *J. Phys. Chem.* 96, 374 (1992).
308. E. Feitosa, P. C. A. Barreleiro, and G. Olofsson, *Chem. Phys. Lipids* 105, 201 (2000).
309. E. Feitosa and W. Brown, *Langmuir* 13, 4810 (1997).
310. D. B. Nascimento, R. Rapuano, M. M. Lessa, and A. M. Carmona-Ribeiro, *Langmuir* 14, 7387 (1998).
311. H. Kawasaki, M. Souda, S. Tanaka, N. Nemoto, G. Karlsson, M. Almgren, and H. Maeda, *J. Phys. Chem. B* 106, 1524 (2002).
312. H. Hauser, H. H. Mantsch, and H. L. Casal, *Biochemistry* 29, 2321 (1990).
313. N. K. Subbarao, R. I. MacDonald, K. Takeshita, and R. C. MacDonald, *Biochim. Biophys. Acta* 1063, 147 (1991).
314. T. A. A. Fonteijn, D. Hoeckstra, and J. B. F. N. Engberts, *Langmuir* 8, 2437 (1992).
315. M. J. Blandamer, B. Briggs, P. M. Cullis, J. B. F. N. Engberts, A. Wagenaar, E. Smits, D. Hoeckstra, and A. Kacperska, *Langmuir* 10, 3507 (1994).
316. E. Smits, M. J. Blandamer, B. Briggs, P. M. Cullis, and J. B. F. N. Engberts, *Recl. Trav. Chim. Pays-Bas* 115, 37 (1996).
317. L. A. M. Rupert, J. F. L. van Breemen, D. Hoeckstra, and J. B. F. N. Engberts, *J. Phys. Chem.* 92, 4416 (1988).
318. Y.-M. Tricot, D. N. Furlong, and W. H. F. Sasse, *Aust. J. Chem.* 37, 1147 (1984).
319. A. M. Carmona-Ribeiro and S. Hix, *J. Phys. Chem.* 95, 1812 (1991).
320. A. M. Carmona-Ribeiro, C. E. Castuma, A. Sesso, and S. Schreier, *J. Phys. Chem.* 95, 5361 (1991).
321. L. Hammarström, I. Velikian, G. Karlsson, and K. Edwards, *Langmuir* 11, 408 (1995).
322. R. Humphry-Baker, D. H. Thompson, Y. Lei, M. J. Hope, and J. K. Hurst, *Langmuir* 7, 2592 (1991).
323. N. Plobeck, S. Eifler, A. Brisson, Y. Nakatani, and G. Ourisson, *Tetrahedron Lett.* 33, 5249 (1992).
324. V. Birault, G. Pozzi, N. Plobeck, S. Eifler, M. Schmutz, T. Palanché, J. Raya, A. Brisson, Y. Nakatani, and G. Ourisson, *Chem. Eur. J.* 2, 789 (1996).
325. G. Pozzi, V. Birault, B. Werner, O. Dannenmuller, Y. Nakatani, G. Ourisson, and S. Terekawa, *Angew. Chem. Int. Ed. Engl.* 35, 177 (1996).
326. P. Walde, M. Wessicken, U. Rädler, N. Berclaz, K. Conde-Frieboes, and P. L. Luisi, *J. Phys. Chem. B* 101, 7390 (1997).
327. B. J. Ravoo and J. B. F. N. Engberts, *Langmuir* 10, 1735 (1994).
328. U. Brinkmann, E. Neumann, and B. H. Robinson, *J. Chem. Soc., Faraday Trans.* 94, 1281 (1998).
329. W. R. Hargreaves and D. W. Deamer, *Biochemistry* 17, 3759 (1978).
330. D. P. Cistola, J. A. Hamilton, D. Jackson, and D. M. Small, *Biochemistry* 27, 1881 (1988).
331. K. Morigaki, P. Walde, M. Misran, and B. H. Robinson, *Colloids Surf. A: Physicochem. Eng. Aspects*, 213, 37 (2003).
332. J. M. Gebicki and M. Hicks, *Nature* 243, 232 (1973).
333. T. Hains, *Proc. Natl. Acad. Sci. USA* 80, 160 (1983).
334. H. Fukuda, A. Goto, H. Yoshioka, R. Goto, K. Morigaki, and P. Walde, *Langmuir* 17, 4223 (2001).
335. M. Bergström and J. C. Eriksson, *Langmuir* 12, 624 (1996).
336. C. L. Apel, D. W. Deamer, and M. N. Mautner, *Biochim. Biophys. Acta* 1559, 1 (2002).
337. M. Gradzielski, M. Bergmeier, M. Müller, and H. Hoffmann, *J. Phys. Chem. B* 101, 1719 (1997).
338. L. L. Brasher and E. W. Kaler, *Langmuir* 12, 6270 (1996).
339. M. T. Yatcilla, K. L. Herrington, L. L. Brasher, E. W. Kaler, S. Chiruvolu, and J. A. Zasadzinski, *J. Phys. Chem.* 100, 5874 (1996).
340. C. Caillet, M. Hebrant, and C. Tondre, *Langmuir* 16, 9099 (2000).
341. A. Shioi and T. A. Hatton, *Langmuir* 18, 7341 (2002).
342. Y. Xia, I. Goldmints, P. W. Johnson, T. A. Hatton, and A. Bose, *Langmuir* 18, 3822 (2002).
343. E. F. Marques, O. Regev, A. Khan, M. G. Miguel, and B. Lindman, *J. Phys. Chem. B* 102, 6746 (1998).
344. E. F. Marques, O. Regev, A. Khan, M. G. Miguel, and B. Lindman, *J. Phys. Chem. B* 103, 8353 (1999).
345. L. Cantù, M. Corti, E. Del Favero, and A. Raudino, *J. Phys. II France* 4, 1585 (1994).
346. L. Cantù, M. Corti, E. Del Favero, M. Dubois, and Th. N. Zemb, *J. Phys. Chem. B* 102, 5737 (1998).
347. S. Bonaccio, P. Walde, and P. L. Luisi, *J. Phys. Chem.* 98, 6661 (1994).
348. S. Bonaccio, D. Capitani, A. L. Segre, P. Walde, and P. L. Luisi, *Langmuir* 13, 1952 (1997).
349. J. K. Harris, G. D. Rose, and M. L. Bruening, *Langmuir* 18, 5337 (2002).
350. K. Schillén, K. Bryskhe, and Y. S. Mel'nikova, *Macromolecules* 32, 6885 (1999).
351. C. Nardin, T. Hirt, J. Leukel, and W. Meier, *Langmuir* 16, 1035 (2000).
352. S. L. Regen, in "Liposomes: From Biophysics to Therapeutics" (M. J. Ostro, Ed.), p. 73. Dekker, New York, 1987.
353. B. A. Armitage, D. E. Bennett, H. G. Lamparski, and D. F. O'Brien, *Adv. Polym. Sci.* 126, 53 (1996).
354. J. R. Riess, F. Frézard, J. Greiner, M. P. Krafft, C. Santaella, P. Vierling, and L. Zarif, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 3, p. 97. CRC Press, Boca Raton, FL, 1996.
355. M. P. Krafft, F. Giulieri, and J. G. Riess, *Angew. Chem. Int. Ed. Engl.* 32, 741 (1993).
356. M. P. Krafft, L. Schieldknecht, P. Marie, F. Giulieri, M. Schmutz, N. Poulain, and E. Nakache, *Langmuir* 17, 2872 (2001).
357. P. Lo Nostro and G. Gabrielli, *Colloids Surf.* 44, 119 (1990).
358. J.-H. Fuhrhop and J. Köning, "Membrane and Molecular Assemblies: The Synkinetic Approach." Royal Society of Chemistry, Cambridge, 1994.
359. G. B. Patel, B. J. Agnew, L. Deschatelets, L. P. Fleming, and G. D. Sprott, *Int. J. Pharm.* 194, 39 (2000).
360. A. Gliozzi, A. Relini, and P. L.-G. Chong, *J. Membr. Sci.* 206, 131 (2002).
361. N. A. J. M. Sommerdijk, T. H. L. Hoeks, M. Synak, M. C. Feiters, R. J. M. Nolte, and B. Zwanenburg, *J. Am. Chem. Soc.* 119, 4338 (1997).
362. M. Bergsma, M. L. Fielden, and J. B. F. N. Engberts, *J. Colloid Interface Sci.* 243, 491 (2001).
363. H. Hoffmann, C. Thunig, and D. Miller, *Colloids Surf. A: Physicochem. Eng. Aspects* 210, 147 (2002).
364. Y. Tanaka, M. Miyachi, and Y. Kobuke, *Angew. Chem. Int. Ed.* 38, 504 (1999).
365. P. Gosh, S. Sengupta, and P. K. Bharadwaj, *Langmuir* 14, 5712 (1998).
366. M. Sano, K. Oishi, T. Ishii, and S. Shinkai, *Langmuir* 15, 3773 (2000).
367. K. Katagiri, K. Ariga, and J. Kikuchi, *Chem. Lett.* 662 (1999).
368. Y. Sumida, A. Masuyama, H. Maekawa, M. Takasu, T. Kida, Y. Nakatsuji, I. Ikeda, and M. Nojima, *Chem. Commun.* 2385 (1998).
369. T. Kunitake, *Angew. Chem. Int. Ed.* 31, 709 (1992).
370. S. Chiruvolu, S. Walker, J. Israelachvili, F.-J. Schmitt, D. Leckband, and J. A. Zasadzinski, *Science* 264, 1753 (1994).
371. J. A. Zasadzinski, E. Kisak, and C. Evans, *Curr. Opin. Colloid Interface Sci.* 6, 85 (2001).

372. C. M. Paleos, Z. Sideratou, and D. Tsiourvas, *ChemBioChem.* 2, 305 (2001).
373. T. Sato and J. Sunamoto, *Prog. Lipid Res.* 31, 345 (1992).
374. A. Sharma and U. S. Sharma, *Int. J. Pharm.* 154, 123 (1997).
375. D. D. Lasic, *Trends Biotechnol.* 16, 307 (1998).
376. A. Bloom, in "Liposomes: Rational Design" (A. S. Janoff, Ed.), p. 397. Dekker, New York, 1999.
377. Y. Barenholz, *Curr. Opin. Colloid Interface Sci.* 6, 66 (2001).
378. R. Banerjee, *J. Biomater. Appl.* 16, 3 (2001).
379. K. Egbaria and N. Weiner, *Cosmetics Toiletries* 106, 79 (1991).
380. H. Schreier and J. Bouwstra, *J. Controlled Release* 30, 1 (1994).
381. T. Lian and R. J. Y. Ho, *J. Pharm. Sci.* 90, 667 (2001).
382. A. K. Singh and R. G. Carbonell, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 4, p. 209. CRC Press, Boca Raton, FL, 1996.
383. H. A. H. Rongen, A. Bult, and W. P. van Bennekom, *J. Immunol. Methods* 204, 105 (1997).
384. T. Wink, S. J. van Zuilen, A. Bult, and W. P. van Bennekom, *Anal. Chem.* 70, 827 (1998).
385. J.-A. A. Ho and R. D. Wauchope, *Anal. Chem.* 74, 1493 (2002).
386. P. Lundahl, C.-M. Zeng, C. Lagerquist Hägglund, I. Gottschalk, and E. Greijer, *J. Chromatogr. B* 720, 103 (1999).
387. M. Yoshimoto, R. Kuboi, Q. Yang, and J. Miyake, *J. Chromatogr. B* 712, 59 (1998).
388. O. Braun-Falco, H. C. Korting, and H. I. Maibach, Eds., "Liposome Dermatics." Springer-Verlag, Berlin, 1992.
389. R. M. Handjani-Vila, A. Ribier, and G. Vanlerberghe, in "Liposome Technology" (G. Gregoriadis, Ed.), Vol. 2, 2nd ed., p. 201. CRC Press, Boca Raton, FL, 1993.
390. G. Vanlerberghe, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 4, p. 77. CRC Press, Boca Raton, FL, 1996.
391. T. E. Redelmeier and N. Kitson, in "Liposomes: Rational Design" (A. S. Janoff, Ed.), p. 283. Dekker, New York, 1999.
392. E. Sackmann in "Handbook of Biological Physics" (R. Lipowsky and E. Sackmann, Eds.), Vol. 1, p. 213. Elsevier, Amsterdam, 1995.
393. M. J. Pregel, L. Jullien, and J.-M. Lehn, *Angew. Chem. Int. Ed.* 31, 1637 (1992).
394. M. R. Ghadiri, J. R. Granja, and L. K. Buehler, *Nature* 369, 301 (1994).
395. W. L. Erdahl, C. J. Chapman, R. W. Taylor, and D. R. Pfeiffer, *Biophys. J.* 66, 1678 (1994).
396. T. D. Clark, L. K. Buehler, and M. R. Ghadiri, *J. Am. Chem. Soc.* 120, 651 (1998).
397. M. Fritz, P. Walde, and D. Seebach, *Macromolecules* 32, 574 (1999).
398. P. H. Schlesinger, R. Ferdani, J. Liu, J. Pajewska, R. Pajewski, M. Saito, H. Shabany, and G. W. Gokel, *J. Am. Chem. Soc.* 124, 1848 (2002).
399. V. Sidorov, F. W. Kotch, G. Abdrakhmanova, R. Mizani, J. C. Fettinger, and J. T. Davis, *J. Am. Chem. Soc.* 124, 2267 (2002).
400. P. L. Luisi, P. Walde, and T. Oberholzer, *Curr. Opin. Colloid Interface Sci.* 4, 33 (1999).
401. J. W. Szostak, D. P. Bartel, and P. L. Luisi, *Nature* 409, 387 (2001).
402. A. Pohorille and D. Deamer, *Trends Biotechnol.* 20, 123 (2002).
403. G. A. Reineccius, in "Encapsulation and Controlled Release of Food Ingredients," ACS Symposium Series 590 (S. J. Risch and G. A. Reineccius, Eds.), p. 113. Am. Chem. Soc., Washington, DC, 1995.
404. P. Dufour, E. Laloy, J.-C. Vuillemard, and R. E. Simard, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 4, p. 153. CRC Press, Boca Raton, FL, 1996.
405. B. C. Keller, *Trends Food Sci. Technol.* 12, 25 (2001).
406. P. J. Quinn and S. F. Perrett, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 4, p. 127. CRC Press, Boca Raton, FL, 1996.
407. G. Mossa, M. C. Annesini, and L. Di Marzio, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 1, p. 273. CRC Press, Boca Raton, FL, 1996.
408. F. Cioci and R. Lavecchia, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 1, p. 287. CRC Press, Boca Raton, FL, 1996.
409. M. Blocher, P. Walde, and I. J. Dunn, *Biotechnol. Bioeng.* 62, 36 (1999).
410. M. Treyer, P. Walde, and T. Oberholzer, *Langmuir* 18, 1043 (2002).
411. M. Yoshimoto, S. Wang, K. Fukunaga, P. Walde, R. Kuboi, and K. Nakao, *Biotechnol. Bioeng.,* 81, 695 (2003).
412. B. A. Korgel and H. G. Monbouquette, *J. Phys. Chem. B* 100, 346 (1996).
413. M. T. Kennedy, B. A. Korgel, H. G. Monbouquette, and J. A. Zasadzinski, *Chem. Mater.* 10, 2116 (1998).
414. B. A. Korgel and H. G. Monbouquette, *Langmuir* 16, 3588 (2000).
415. T. Kunitake and S. Shinkai, *Adv. Phys. Org. Chem.* 17, 435 (1980).
416. I. M. Cuccovia and H. Chaimovich, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 3, p. 239. CRC Press, Boca Raton, FL, 1996.
417. P. Scrimin, P. Tecilla, and U. Tonellato, *J. Am. Chem. Soc.* 114, 5086 (1992).
418. P. Scrimin, in "Supramolecular Control of Structure and Reactivity," Perspectives in Supramolecular Chemistry (A. D. Hamilton, Ed.), Vol. 3, p. 101. Wiley, Chichester, 1996.
419. J. Pérez-Juste, F. Hollfelder, A. J. Kirby, and J. B. F. N. Engberts, *Org. Lett.* 2, 127 (2000).
420. T. Rispens and J. B. F. N. Engberts, *Org. Lett.* 3, 941 (2001).
421. I. Shin, T. Silman, C. Bon, and L. Weiner, *Biochemistry* 37, 4310 (1998).
422. R. Kuboi, M. Yoshimoto, P. Walde, and P. L. Luisi, *Biotechnol. Prog.* 13, 828 (1997).
423. S. Okada, S. Peng, W. Spevak, and D. Charych, *Acc. Chem. Res.* 31, 229 (1998).
424. R. Jelinek and S. Kolusheva, *Biotechnol. Adv.* 19, 109 (2001).
425. J. Song, Q. Cheng, S. Zhu, and R. C. Stevens, *Biomed. Microdev.* 4, 213 (2002).
426. J. T. Groves, N. Ulman, P. S. Cremer, and S. G. Boxer, *Langmuir* 14, 3347 (1998).
427. H. Mueller, H.-J. Butt, and E. Bamberg, *J. Phys. Chem. B* 104, 4552 (2000).
428. C. A. Keller, K. Glasmästar, V. P. Zhdanov, and B. Kasemo, *Phys. Rev. Lett.* 84, 5443 (2000).
429. E.-K. Sinner and W. Knoll, *Curr. Opin. Chem. Biol.* 5, 705 (2001).
430. I. Stanish and H. G. Monbouquette, *J. Membr. Sci.* 192, 99 (2001).
431. S. A. Miller, J. H. Ding, and D. L. Gin, *Curr. Opin. Colloid Interface Sci.* 4, 338 (1999).
432. S. Liu and D. F. O'Brian, *J. Am. Chem. Soc.* 124, 6037 (2002).
433. A. Mueller and D. F. O'Brian, *Chem. Rev.* 102, 727 (2002).
434. S. Mann, J. P. Hannington, and R. J. P. Williams, *Nature* 324, 565 (1991).
435. E. D. Eanes, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 2, p. 349. CRC Press, Boca Raton, 1996.
436. J. H. Collier and P. B. Messersmith, *Annu. Rev. Mater. Res.* 31, 237 (2001).
437. S. Mann and G. A. Ozin, *Nature* 382, 313 (1996).
438. G. J. de, A. A. Soler-Illia, C. Sanchez, B. Lebeau, and J. Patarin, *Chem. Rev.* 102, 4093 (2002).
439. J. D. Morgan, C. A. Johnson, and E. W. Kaler, *Langmuir* 13, 6447 (1997).
440. J. Hotz and W. Meier, *Langmuir* 14, 1031 (1998).
441. F. Caruso, *Chem. Eur. J.* 6, 413 (2000).
442. M. Blocher, D. J. Liu, P. Walde, and P. L. Luisi, *Macromolecules* 32, 7332 (1999).

443. M. Blocher, D. J. Liu, and P. L. Luisi, *Macromolecules* 33, 5787 (2000).
444. M. Jung, B. H. Robinson, D. C. Steytler, A. L. German, and R. K. Heenan, *Langmuir* 18, 2873 (2002).
445. J. H. Fuhrhop and J. Mathieu, *Angew. Chem. Int. Ed.* 23, 100 (1984).
446. G. Steinberg-Yfrach, J.-L. Rigaud, E. N. Durantini, A. L. Moore, D. Gust, and T. A. Moore, *Nature* 392, 479 (1998).
447. D. K. Lee and Y. S. Kang, *Mol. Cryst. Liq. Cryst.* 377, 261 (2002).
448. K. Fukuda, Y. Sasaki, K. Ariga, and J. Kikuchi, *J. Mol. Catal. B: Enzym.* 11, 971 (2001).
449. J. Kikuchi, K. Ariga, Y. Sasaki, and K. Ikeda, *J. Mol. Catal. B: Enzym.* 11, 977 (2001).
450. T. S. Arrhenius, M. Blanchard-Desce, M. Dvolaitzky, J. M. Lehn, and J. Malthete, *Proc. Natl. Acad. Sci. USA* 83, 5355 (1986).
451. N. Berclaz, M. Müller, P. Walde, and P. L. Luisi, *J. Phys. Chem. B* 105, 1056 (2001).
452. N. Berclaz, E. Blöchliger, M. Müller, and P. L. Luisi, *J. Phys. Chem. B* 105, 1065 (2001).
453. S. Engström, L. Lindahl, R. Wallin, and J. Engblom, *Int. J. Pharm.* 86, 137 (1992).
454. K. Larsson, *J. Phys. Chem.* 93, 7304 (1989).
455. J. C. Shah, Y. Sadhale, and D. M. Chilukuri, *Adv. Drug. Del. Rev.* 47, 229 (2001).
456. M. Nakano, A. Sugita, H. Matsuoka, and T. Handa, *Langmuir* 17, 3917 (2001).
457. M. Hodenius, M. De Cuyper, L. Desender, D. Müller-Schulte, A. Steigel, and H. Lueken, *Chem. Phys. Lipids* 120, 75 (2002).
458. J. M. Lehn and P. Ball, in "The New Chemistry" (N. Hall, Ed.), p. 300. Cambridge University Press, Cambridge, 2000.
459. J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* 99, 4763 (2002).
460. P. Durbut, in "Handbook of Detergents: Part A: Properties" (G. Broze, Ed.), p. 47. Dekker, New York, 1999.
461. A. Pinazo, M. R. Infante, S. Y. Park, and E. I. Franses, *Colloids Surf. B: Biointerfaces* 8, 1 (1996).
462. S. Lee, D. H. Kim, and D. Needham, *Langmuir* 17, 5544 (2001).
463. H. T. Jung, B. Coldren, J. A. Zasadzinski, D. J. Iampietro, and E. W. Kaler, *Proc. Natl. Acad. Sci. USA* 98, 1353 (2001).
464. D. D. Lasic, *J. Liposome Res.* 9, 43 (1999).
465. D. D. Lasic, R. Joannic, B. C. Keller, P. M. Frederik, and L. Auvray, *Adv. Colloid Interface Sci.* 89–90, 337 (2001).
466. S. A. Safran, P. A. Pincus, D. Andelman, and F. C. MacKintosh, *Phys. Rev. A* 43, 1071 (1991).
467. K. Horbaschek, H. Hoffmann, and J. Hao, *J. Phys. Chem. B* 104, 2781 (2000).
468. N. L. Gershfeld, W. F. Stevens, Jr., and R. J. Nossal, *Faraday Discuss. Chem. Soc.* 81, 19 (1986).
469. N. L. Gershfeld, *Biochim. Biophys. Acta* 988, 335 (1989).
470. N. L. Gershfeld and L. Ginsberg, *J. Membr. Biol.* 156, 279 (1997).
471. M. Koshinuma, K. Tajima, A. Nakamura, and N. L. Gershfeld, *Langmuir* 15, 3430 (1999).
472. K. Tajima, M. Koshinuma, A. Nakamura, and N. L. Gershfeld, *Langmuir* 16, 2576 (2000).
473. D. D. Lasic, *Biochem. J.* 256, 1 (1988).
474. T. E. Thompson, *Hepatology* 12, 51S (1990).
475. D. D. Lasic and F. J. Martin, *J. Membr. Sci.* 50, 215 (1990).
476. W. Helfrich, *Prog. Colloid Polym. Sci.* 95, 7 (1994).
477. D. D. Lasic, *J. Liposome Res.* 5, 431 (1995).
478. B. W. Ninham, Y. Talmon, and D. F. Evans, *Science* 221, 1047 (1983).
479. E. Z. Radlinska, B. W. Ninham, J.-P. Dalbiez, and T. N. Zemb, *Colloids Surf.* 46, 213 (1990).
480. P. André, A. Filankembo, I. Lisiecki, C. Petit, T. Gulik-Krzywicki, B. W. Ninham, and M. P. Pileni, *Adv. Mater.* 12, 119 (2000).
481. C. Tanford, "The Hydrophobic Effect." Wiley, New York, 1980.
482. M. Boström, D. R. M. Williams, and B. W. Ninham, *Phys. Rev. Lett.* 87, 168103 (2001).
483. B. W. Ninham, *Prog. Colloid Polym. Sci.* 120, 1 (2002).
484. M. Boström, D. R. M. Williams, and B. W. Ninham, *Langmuir* 18, 8609 (2002).
485. D. Gopinath, D. Ravi, B. R. Rao, S. S. Apte, and D. Rambhau, *Int. J. Pharm.* 246, 187 (2002).
486. G. B. Patel and G. D. Sprott, *Crit. Rev. Biotechnol.* 19, 317 (1999).
487. M. Conacher, J. Alexander, and J. M. Brewer, *Vaccine* 19, 2965 (2001).
488. C. Tondre and C. Caillet, *Adv. Colloid Interface Sci.* 93, 115 (2001).
489. K. Katagiri, R. Hamasaki, K. Ariga, and J. Kikuchi, *Langmuir* 18, 6709 (2002).
490. D. L. Melchior and A. Makriyannis, *Biotechnol. Tech.* 11, 885 (1997).
491. D. L. Melchior, *J. Pharm. Sci.* 91, 1075 (2002).
492. M. Fix and D. L. Melchior, *FEBS Lett.* 516, 109 (2002).
493. Y. C. Wang, C. J. Lee, W. K. Chen, C. I. Huang, W. F. Chen, G. J. Chen, and S. Z. Lin, *Artif. Cell Blood Sub.* 24, 35 (1996).
494. E. Mastrobattista, G. A. Koning, and G. Storm, *Adv. Drug Delivery Rev.* 40, 103 (1999).
495. M. De Cuyper and M. Joniau, *Eur. Biophys. J.* 15, 311 (1988).
496. M. De Cuyper and M. Joniau, *Langmuir* 7, 647 (1991).
497. M. De Cuyper, in "Handbook of Nonmedical Applications of Liposomes" (D. D. Lasic and Y. Barenholz, Eds.), Vol. 3, p. 325. CRC Press, Boca Raton, FL, 1996.
498. C. Sangregorio, J. K. Wiemann, C. J. O'Connor, and Z. Rosenzweig, *J. Appl. Phys.* 85, 5699 (1999).
499. N. Moussaoui, N. Cansell, and A. Denizot, *Int. J. Pharm.* 242, 361 (2002).
500. I. F. Uchegbu and S. P. Vyas, *Int. J. Pharm.* 172, 33 (1998).
501. P. Arunothayanun, I. F. Uchegbu, D. Q. M. Craig, J. A. Turton, and A. T. Florence, *Int. J. Pharm.* 183, 57 (1999).
502. P. Arunothayanun, I. F. Uchegbu, and A. T. Florence, *J. Pharm. Pharmacol.* 51, 651 (1999).
503. P. Arunothayanun, M.-S. Bernard, D. Q. M. Craig, I. F. Uchegbu, and A. T. Florence, *Int. J. Pharm.* 201, 7 (2000).
504. A. Kozubek, J. Gubernator, E. Przeworska, and M. Stasiuk, *Acta Biochim. Pol.* 47, 639 (2000).
505. H. M. Chen, V. Torchilin, and R. Langer, *J. Controlled Release* 42, 263 (1996).
506. K.-H. Song, S.-J. Chung, and C.-K. Shim, *J. Controlled Release* 84, 27 (2002).
507. C. Hu and D. G. Rhodes, *Int. J. Pharm.* 206, 110 (2000).
508. H. Kunieda and V. Rajagopalan, in "Vesicles," Surfactant Science Series (M. Rosoff, Ed.), Vol. 62, p. 79. Dekker, New York, 1996.
509. R. Brunke, *Manufacturing Chem.* 61, 36 (1990).
510. T. M. Allen, *Trends Pharm. Sci.* 15, 215 (1994).
511. D. D. Lasic and D. Needham, *Chem. Rev.* 95, 2601 (1995).
512. D. D. Lasic and F. J. Martin, Eds., "Stealth Liposomes." CRC Press, Boca Raton, FL, 1995.
513. J.-C. Bradley, M.-A. Guedeau-Boudeville, G. Jandeau, and J.-M. Lehn, *Langmuir* 13, 2457 (1997).
514. R. Glück and A. Wegmann, in "Medical Applications of Liposomes" (D. D. Lasic and D. Papahadjopoulos, Eds.), p. 75. Elsevier, Amsterdam, 1998.
515. T. Daemen, L. Burgener, A. Huckriede, and J. Wilschut, *J. Liposome Res.* 10, 329 (2000).
516. I. P. Hunziker, R. Zurbriggen, R. Glueck, O. B. Engler, J. Reichen, W. J. Dai, W. J. Pichler, and A. Cerny, *Mol. Immunol.* 38, 475 (2001).
517. L. T. Boni, S. R. Minchey, W. R. Perkins, P. L. Ahl, J. L. Slater, M. W. Tate, S. M. Gruner, and A. S. Janoff, *Biochim. Biophys. Acta* 1146, 247 (1993).

518. G. Cevc, in "Phospholipids Handbook" (G. Cevc, Ed.), p. 1. Dekker, New York, 1993.
519. C. Sun and M. Ueno, *Colloid Polym. Sci.* 278, 855 (2000).
520. T. Nakashima, T. Anno, H. Kanda, Y. Sato, T. Kuroi, H. Fujii, S. Nagadome, and G. Sugihara, *Colloids Surf., B: Biointerfaces* 24, 103 (2002).
521. K. Esumi, M. Miyazaki, T. Arai, and Y. Koide, *Colloids Surf., A: Physicochem. Eng. Aspects* 135, 117 (1998).
522. H. Heerklotz and J. Seelig, *Biophys. J.* 78, 2435 (2000).
523. C. E. Giacomelli, A. W. P. Vermeer, and W. Norde, *Langmuir* 16, 4853 (2000).
524. R. A. Schwendener, M. Asanger, and H. G. Weder, *Biochem. Biophys. Res. Commun.* 100, 1055 (1981).
525. L. T. Mimms, G. Zampighi, Y. Nozaki, C. Tanford, and J. A. Reynolds, *Biochemistry* 20, 833 (1981).
526. Y. Nozaki, D. D. Lasic, C. Tanford, and J. A. Reynolds, *Science* 217, 366 (1982).
527. D. Levy, A. Gulik, M. Seigneuret, and J.-L. Rigaud, *Biochemistry* 29, 9480 (1990).
528. N. Ostrowsky, *Chem. Phys. Lipids* 64, 45 (1993).
529. S. Kölchens, V. Ramaswami, J. Birgenheier, L. Nett, and D. F. O'Brien, *Chem. Phys. Lipids* 65, 1 (1993).
530. A. J. Jin, D. Huster, and K. Gawrisch, *Eur. Biophys. J.* 28, 187 (1999).
531. J. Pencer, G. F. White, and F. R. Hallett, *Biophys. J.* 81, 2716 (2001).
532. J. H. van Zanten and H. G. Monbouquette, *J. Colloid Interface Sci.* 146, 330 (1991).
533. J. H. van Zanten and H. G. Monbouquette, *J. Colloid Interface Sci.* 165, 512 (1994).
534. O. Stauch, R. Schubert, G. Savin, and W. Burchard, *Biomacromolecules* 3, 565 (2002).
535. J. P. P. Starink and A. J. Verkleij, *J. Microsc.* 167, 181 (1992).
536. D. D. Miller, J. R. Bellare, D. F. Evans, Y. Talmon, and B. W. Ninham, *J. Phys. Chem.* 91, 674 (1987).
537. S. Lesieur, C. Grabielle-Madelmont, M. Paternostre, and M. Ollivon, *Chem. Phys. Lipids* 64, 57 (1993).
538. D. Kardassi, D. Tsiourvas, and C. M. Paleos, *J. Colloid Interface Sci.* 186, 203 (1997).
539. O. Teschke and E. F. de Souza, *Langmuir* 18, 6513 (2002).
540. P. C. Mason, B. D. Gaulin, R. M. Epand, G. D. Wignall, and J. S. Lin, *Phys. Rev. E* 59, 3361 (1999).
541. M. A. Kiselev, P. Lesieur, A. M. Kisselev, D. Lombardo, M. Killany, and S. Lesieur, *J. Alloys Compd.* 328, 71 (2001).
542. A. Tardieu, V. Luzzati, and F. C. Reman, *J. Mol. Biol.* 75, 711 (1973).
543. Y. Tahara and Y. Fujiyoshi, *Micron* 25, 1412 (1994).
544. J. Pencer and F. R. Hallett, *Phys. Rev. E* 61, 3003 (2000).
545. P. Balgavy, M. Dubnicková, N. Kucerka, M. A. Kiselev, S. P. Yaradaikin, and D. Uhriková, *Biochim. Biophys. Acta* 1512, 40 (2001).
546. G. Ma, D. J. Barlow, M. J. Lawrence, R. K. Heenan, and P. Timmis, *J. Phys. Chem B* 104, 9081 (2000).
547. B. Kachar, D. F. Evans, and B. W. Ninham, *J. Colloid Interface Sci.* 100, 287 (1984).
548. J. Käs and E. Sackmann, *Biophys. J.* 60, 825 (1991).
549. E. Farge and P. F. Devaux, *Biophys. J.* 61, 347 (1992).
550. R. Lipowsky, *Curr. Opin. Struct. Biol.* 5, 531 (1995).
551. W. Wintz, H.-G. Döbereiner, and U. Seifert, *Europhys. Lett.* 33, 403 (1996).
552. F. M. Menger, S. J. Lee, and J. S. Keiper, *Langmuir* 12, 4479 (1996).
553. B. L.-S. Mui, H.-G. Döbereiner, T. D. Madden, and P. R. Cullis, *Biophys. J.* 69, 930 (1995).
554. L. Vander Elst, C. Piérart, S. L. Fossheim, J.-C. Raux, A. Roch, and R. N. Muller, *Supramol. Chem.* 14, 411 (2002).
555. J. C. McIntyre and R. G. Sleight, *Biochemistry* 30, 11819 (1991).
556. J. A. Bouwstra, G. S. Gooris, W. Bras, and H. Talsma, *Chem. Phys. Lipids* 64, 83 (1993).
557. L. A. Bagatolli and E. Gratton, *Biophys. J.* 78, 290 (2000).
558. K. Arnold, A. Lösche, and K. Garwrisch, *Biochim. Biophys. Acta* 645, 143 (1981).
559. A. Blume, R. J. Wittebort, S. K. Das Gupta, and R. G. Griffin, *Biochemistry* 21, 6243 (1982).
560. J. Korlach, P. Schwille, W. W. Webb, and G. W. Feigenson, *Proc. Natl. Acad. Sci. USA* 96, 8461 (1999).
561. V. Borenstain and Y. Barenholz, *Chem. Phys. Lipids* 64, 117 (1993).
562. M. Shinitzky and Y. Barenholz, *Biochim. Biophys. Acta* 515, 367 (1978).
563. P. Walde and E. Blöchliger, *Langmuir* 13, 1668 (1997).
564. J. Seelig and A. Seelig, *Q. Rev. Biophys.* 13, 19 (1980).
565. D. B. Fenske, *Chem. Phys. Lipids* 64, 143 (1993).
566. B.-S. Lee, S. A. Mabry, A. Jonas, and J. Jonas, *Chem. Phys. Lipids* 78, 103 (1995).
567. A. A. Nezorov and M. F. Brown, *J. Chem. Phys.* 107, 10288 (1997).
568. P. F. Devaux and H. M. McConnell, *J. Am. Chem. Soc.* 94, 4475 (1972).
569. J. Tabony and B. Perly, *Biochim. Biophys. Acta* 1063, 67 (1990).
570. S. König, W. Pfeiffer, T. Bayerl, D. Richter, and E. Sackmann, *J. Phys. II France* 2, 1589 (1992).
571. B. R. Lentz, *Chem. Phys. Lipids* 64, 99 (1993).
572. H. Hauser, W. Guyer, I. Pascher, P. Skrabal, and S. Sundell, *Biochemistry* 19, 366 (1980).
573. R. G. Snyder, G. L. Liang, H. L. Strauss, and R. Mendelsohn, *Biophys. J.* 71, 3186 (1996).
574. R. D. Kornberg and H. M. McConnell, *Biochemistry* 10, 1111 (1971).
575. N. Oku, D. A. Kendall, and R. C. MacDonald, *Biochim. Biophys. Acta* 691, 332 (1982).
576. W. R. Perkins, S. R. Minchey, P. L. Ahl, and A. S. Janoff, *Chem. Phys. Lipids* 64, 197 (1993).
577. K. Kawakami, Y. Nishihara, and K. Hirano, *Anal. Biochem.* 269, 139 (1999).
578. J.-H. Fuhrhop, C. Endisch, A. Schulz, J. Turner, and M. Eaton, *Langmuir* 15, 3707 (1999).
579. D. Huster, A. J. Jin, K. Arnold, and K. Gawrisch, *Biophys. J.* 73, 855 (1997).
580. R. G. Males, J. C. Nelson, P. S. Phillips, W. R. Cullen, and F. G. Herring, *Biophys. Chem.* 70, 75 (1998).
581. T.-X. Xiang and B. D. Anderson, *Biophys. J.* 75, 2658 (1998).
582. R. G. Males, J. C. Nelson, P. S. Phillips, W. R. Cullen, and F. G. Herring, *Biophys. Chem.* 70, 75 (1998).
583. S. Paula, A. G. Volkov, A. N. Van Hoeck, and D. W. Deamer, *Biophys. J.* 70, 339 (1996).
584. K. Matsuzaki, O. Murase, K. Sugishita, S. Yoneyama, K. Akada, M. Ueha, A. Nakamura, and S. Kobayashi, *Biochim. Biophys. Acta* 1467, 219 (2000).

Protein-Doped Nanoporous Silica Gels

Stefano Bettati, Barbara Pioselli, Barbara Campanini,
Cristiano Viappiani, Andrea Mozzarelli

University of Parma, Parma, Italy

CONTENTS

1. Introduction
2. Preparation and Physico-Chemical Properties of Silica Gels
3. Structure, Dynamics, and Stability of Proteins Encapsulated in Silica Gels
4. Applications of Protein-Doped Silica Gels
5. Conclusions

Glossary

References

1. INTRODUCTION

Bioactive materials are of increasing interest for technological and nanotechnological applications [1]. Biological macromolecules, and even whole cells, have been intimately linked to inorganic or organic matrices to exploit their specific functions that are otherwise extremely difficult to mimic only with inorganic–organic molecules. These nanobiocomposites are very frequently based on proteins. Proteins are biological macromolecules able to perform a wide range of functions: (i) antibodies selectively recognize and strongly bind antigenes; (ii) carriers selectively recognize and reversibly bind small molecules transferring them from one location to another in living organisms; (iii) receptors selectively recognize specific compounds and trigger intra- and intercellular signals; (iv) enzymes recognize specific substrates and catalyze their transformation into products. Proteins achieve their functions via a fine tuning of the three-dimensional structure and conformational dynamics. Proteins are labile molecules and their native active state is stabilized over inactive, denatured states by only a few tens of kcal/mole, the energy of a few hydrogen bonds. Versatility, specificity, and efficiency of proteins are attractive properties that boost interest in their usage in technological applications. Most of these applications rely on the nanobioencapsulation of proteins into inorganic or organic matrices in order to increase stability, confine in space, obtain fast separation, convert a chemical reaction in optical or electrical signals, and isolate defined protein conformations. Among the different inorganic materials, silicates are by far the most used to adsorb, covalently link, and encapsulate proteins. In particular, in the past 10 years there has been an explosion in the use of sol–gel technology to trap proteins [2–5]. The field was reviewed with particular emphasis on bioencapsulation within synthetic polymers [6–8], bioactive sol–gel coatings [9], functional properties of immobilized proteins [10], sol–gel glass as a matrix for chemical and biochemical sensing [11], and properties and applications of proteins within sol–gel derived materials [12]. The great advantages of sol–gel methods for the entrapment of proteins are the very mild conditions of pH, solvents, and temperature required to prepare stable protein-doped nanoporous gels, the possibility to fabricate materials of different size and geometry, such as monoliths, films, fibers, nanospheres, sandwiches, powders, either in the absence of water (xerogels) or in its presence (hydrogels), the gel optical quality that makes them suitable for the detection of signals by spectroscopic techniques, and the very reduced rate of protein leaching. Because the conformational flexibility and the biological activity of the entrapped proteins are usually maintained, protein-doped silica gels have been used for the development of bioreactors, biosensors, photonic devices, biocoating, and controlled protein release. Furthermore, silica gel encapsulation selects defined conformational states of proteins by decreasing the rate of their interconversion [10]. This feature is exploited to investigate functional properties of defined tertiary and quaternary states of allosteric proteins.

2. PREPARATION AND PHYSICO-CHEMICAL PROPERTIES OF SILICA GELS

2.1. General Properties

Bioimmobilization allows one to exploit functional properties of biological species, such as proteins, peptides, nucleic acids, and whole cells, for biotechnological applications.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (81–103)

In order to obtain a suitable material, it is necessary that the immobilized biological reagents are characterized by a high concentration, a high degree of activity, long term stability even in adverse environmental conditions, free accessibility to reagents or analytes, and absence of leaching [10, 12]. However, in specific cases, like protein drug delivery *in vivo*, encapsulation methods have been designed to entrap biomolecules and slowly release them as a consequence of the erosion of the matrix.

Typical techniques to immobilize biomolecules are based on physical adsorption [13], surface covalent binding [14], entrapment in semipermeable membranes [15], and micro-encapsulation into polymeric microspheres and gels [16, 17]. Most of these methods are not generic; thus they are often limited to a selected type of biomolecules. Encapsulation of biological components into silica-based matrices is an immobilization procedure that, for a long time, has only been exploited for organic polymers and small molecules. Labile biomolecules do not withstand high temperature processing of glasses. Therefore, initially, immobilization techniques were studied that linked biomolecules on the surface of porous glasses via chemical agents. However, chemical modifications might affect the bioactivity of proteins and cells. Therefore, entrapment methods based on the physical encapsulation in sol–gel glasses (sol–gel process) were developed. First reports on the sol–gel encapsulation of enzymes were published in the early 1970s [18]. The method was extended to a wide range of biological species, immobilized in a variety of silica-based biocomposite materials.

Several features make the silica-based sol–gel technique an ideal method to immobilize proteins. The polymeric network limits the mobility of the protein, preventing aggregation and/or leaching. Nevertheless, the entrapped proteins usually retain their native structure and function. The nanoporous structure makes the protein active sites accessible to small reagents that diffuse in the solvent phase, and the pore size and distribution can be tuned. Futhermore, the silica matrix is characterized by a high thermal, mechanical, and chemical stability [19, 20]. In addition to being a long-term storage site, it gives to encapsulated biomolecules a greater resistance toward thermal and chemical insults (see Section 3.2). Silica gels are optically transparent above 300 nm [4, 21–25]. This allows one to detect spectral changes induced by ligand–protein binding and thus to develop optical sensors. Silica gels are not able to conduct electrons, but the addition of electron transport mediators allows one to obtain electrochemical sensors (see [7] and references therein and also [21]). Beyond all these properties, simple reaction conditions are the basis for the great success of protein immobilization via sol–gel methods. The absence of coupling agents, the aqueous medium, and room temperature processing are conditions that favor the maintenance of biological structure and function. The polymerization reaction is compatible with the pH and ionic strength required for protein function (typical pH range 4–10; ionic strength range 0.01–1 M). The amount of protein loaded depends on both gelation conditions and the necessity to avoid intermolecular interactions which could cause aggregation and alteration of protein function. The size of the protein does not seem to be a limiting factor for the encapsulation [21]. The addition of low concentrations of organic molecules affects the network formation.

2.2. Materials and Protocols for Gel Preparation

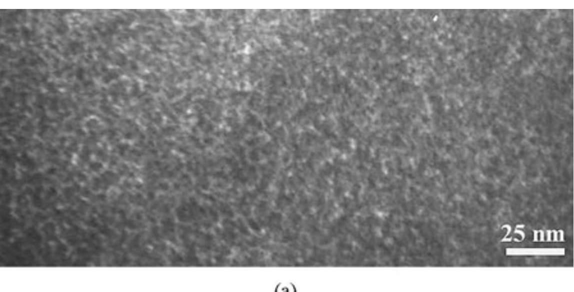
Sol–gel derived biocomposite materials are generally prepared via the partial or complete hydrolysis of a suitable silicium-containing precursor leading to the formation of an aqueous sol. The most frequently used precursors are tetraalkoxysilanes, mono-, di-, or tri-alkyl alkoxysilanes. Although we focus on silica derived glasses, other metal centers, such as Al, Ti, V, or Ce, can be used to modulate the material properties. The precursors may comprise functional groups ranging from alkenyl, aryl, amino, carboxyl, thiol, or other groups as well as redox-active moieties, flavins, or quinones. The hydrolysis reaction can be either acid or base catalyzed and forms a mixture of $R_xSi(OH)_y$ molecules (Fig. 1). Sonication of the mixture is frequently used to obtain a homogeneous sol. Hydrolysis may be followed by transesterification with additives, such as glycerol, to form polyglyceryl silicates which are more biocompatibile than alkoxysilane. The hydrolized precursor is mixed with an aqueous buffered solution containing the biomolecule and additives required to modify the properties of the final material (catalysts, drying control additives, polymers, templating agents, redox species). A polymeric network is achieved from the condensation between the silanol groups (Si–OH) obtained from hydrolysis, forming an oxo bridge (O–Si–O) with removal of a water molecule. As the condensation reaction proceeds, a wide range of linear, branched, and colloidal polysilicates is obtained, viscosity increases, and gelation occurs over a period of minutes to hours. The polymerization reaction

Tetraalkoxysilanes, $R_1=R_2=R_3=R_4$; TMOS, CH_3; TEOS, CH_3CH_2

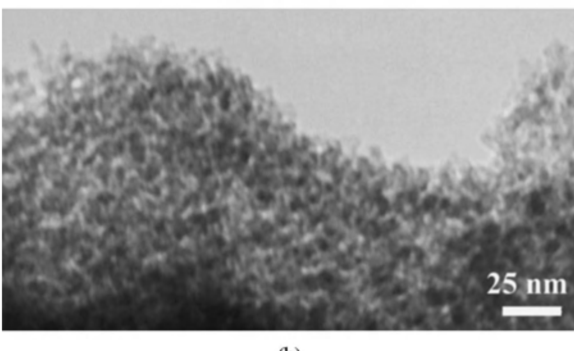
Mono-, di-, tri—alkoxysilanes, R_1, R_2, R_3, R_4 can be different groups

Figure 1. Sol–gel process. Reactions of the acid-catalyzed alkoxide hydrolysis and polycondensation.

occurs spontaneously and irreversibly but changes in pH, temperature, salts, and catalysts can promote it. The gel consists of an amorphous solid with an interstitial liquid phase. When a biomolecule is added, the silica network, growing around the biomolecule, entraps it within the porous gel. The initial gels are soft and have a high water content (50–80%) and large pores (up to 200 nm diameter). After gelation, gels are subjected to an *aging* process, during which the condensation reaction and the cross-linking continue, leading to gel shrinkage and strengthening. Aging typically causes a 10–30% shrinkage of the matrix, the average pore diameter decreases by about 25%, and the relative proportion of siloxane to silanol groups increases [19]. Hydrolysis and condensation by-products, such as methanol when the precursor is tetramethylorthosilicate (TMOS), are easily removed from the matrix by repeated washing procedures. These aged gels can be partially or completely dried at room temperature by slow removal of the liquid phase (solvents, water, and alcohol). During this process, the gel shrinks significantly and further cross-linking of the matrix occurs. The decrease of the pore size causes the reduction of the volume of the liquid phase, and, eventually, a collapse of the pores takes place. The resulting xerogels are more rigid and dimensionally stable than the corresponding hydrogels, with pore diameter ranging typically between 1 and 10 nm. Transmission electron microscopy of TMOS-derived dried gels prepared in the absence and presence of hemoglobin exhibited pore size of 3–5 nm (Fig. 2). Bulk glasses can be obtained upon densification around 1000 °C, but high temperatures destroy the entrapped biomolecules.



(a)



(b)

Figure 2. Transmission electron microscopy of silica gels prepared in the absence (a) and presence of hemoglobin (b). Reprinted with permission from [89], S. Abbruzzetti et al., *J. Nanosci. Nanotech.* 1, 407 (2001). © 2001, American Scientific Publishers.

Therefore, this stage of the sol–gel process is limited to the preparation of inorganic–organic-doped silica glasses.

According to the types of precursors and additives employed in sol–gel synthesis of hybrid materials, inorganic glasses, organically modified glasses, and organic–inorganic nanocomposite materials can be prepared [12]. The silica alkoxides most frequently used to obtain inorganic glasses are TMOS and tetraethylorthosilicate (TEOS). The conventional method involves acid-catalyzed hydrolysis and the use of methanol as a co-solvent. Indeed, the low water solubility and reactivity of these compounds requires the presence of co-solvents and catalysts that can be deleterius to the viability of entrapped biomolecules. Hydrolysis of alkoxides leads to the production of the corresponding alcohol, while the condensation reaction releases water and alcohols that are potentially protein denaturating agents. The method was improved by purging the sol with a nitrogen flow for 40 minutes at 4 °C prior to the addition of the protein, as in the case of hemoglobin-doped silica gels [26]. The gas flow allows one to remove at least part of the methanol produced by the sol–gel reaction. TMOS and TEOS are the most used precursors because methanol and ethanol are less detrimental for proteins than longer chain alcohols. Buffer systems, such as acetate and monobasic phosphate, were showed to stabilize cytochrome *c* against aggregation in a solution containing greater than 90 vol% methanol [27]. Currently, the sol–gel process with alkoxide silanes precursors is often performed in two steps. The first step is the hydrolysis of the alkoxide in acid medium in order to hydrolyze alkoxy groups. Sols are prepared by sonicating, in the absence of co-solvent, a mixture of TEOS or TMOS, water, and HCl until the mixture becomes clear and monophasic [28]. Then, upon mixing the sol and a buffer solution to adjust the pH to a value between 6 and 7, a solution containing a protein and a suitable buffer is added to the sol in ratios that vary from case to case. Condensation occurs at room temperature within a few minutes and without denaturation of most proteins. Following gelation, the hybrid material formed is allowed to age either in buffer (wet aging) or in air (dry aging).

An alcohol-free sol–gel process to encapsulate biological materials in inorganic glasses employs precursors such as sodium silicate or silicic acid to obtain aqueous silica gels [29, 30]. The method involves the mixing of a colloidal silica sol, commercially available in spherical particles at about pH 10.3, with a solution of acidified 0.1 M sodium silicate (Na_2SiO_3). After mixing, the resulting pH is close to physiological values (about 6.2–8.2). Then, a buffered protein-containing solution is mixed with an equal volume of the colloidal solution at the same pH value. An optically trasparent, solidlike, protein-containing gel is formed. After aging, the gel can be dried at 4 °C over a period of two to four weeks [29].

Poly(glyceryl silicate) (PGS) is a novel class of biocompatible precursors, namely, polyol esters of silicates, which allow one to overcome some complications encountered with other compounds, such as large-scale shrinkage and pore collapse during xerogel formation, and protein denaturation due to alcohol release. The important features of these precursors were summarized by Gill and Ballesteros [31]: high water solubility, autohydrolysis in aqueous media, and

hydrolysis reaction that liberates a nonvolatile, bioprotective alcohol, which also functions as a drying control additive [32–35]. Among various polyhydroxylated compounds, glycerol was selected for its high biocompatibility and functionality. PGS is a stable, water soluble solid formed by the partial hydrolysis and condensation of TMOS to poly(methylsilicate) followed by transesterification with glycerol, in a one-step process catalyzed by hydrocloridric acid or poly(antimony(III) ethylene glycoxide). In the absence of any catalyst and in the presence of an aqueous, buffered medium, PGS rapidly hydrolyzes and forms silica hydrogels in a few minutes. If polymerization occurs in the presence of proteins or cells, a very efficient entrapment of this species results, with little loss of activity. To yield bioactive silica glasses, after aging, the material is washed to remove glycerol and then dried to produce physically stable silica xerogels. The hydrolysis and polymerization rates of polyol silicates are higher than those of other precursors [36, 37]. This probably stems from the increased reactivity of the silicon center due to the electron-withdrawing effect of the polyol moieties and to the internally assisted hydrolysis by the hydroxyl function [32, 33]. Reactivity is controlled by adjusting the precursor composition.

The synthesis of inorganic materials does not offer the possibility to easily tune some fundamental characteristics of the hybrid material such as hydrophobic or hydrophilic properties, pore size, and mechanical stability. As outlined by Keeling-Tucker and Brennan [38], key problems with inorganic materials are (i) the relatively high polarity and surface charge due to high siloxide levels at neutral pH, (ii) the relatively rigid solvent phase due to adsorption of the solvent to the pore walls, and (iii) the heterogeneity in pore size, polarity, and mobility of the reactants. Therefore, other methods that allow better control over the final properties of a sol–gel derived glass have been developed. The addition of organically modified silanes to TEOS- or TMOS-based glasses provides organically modified silicates (Ormosils). A variety of organosilanes were used to form these materials [38]. Sol–gel methods involving these compounds were discussed [39, 40]. Although the organically modified silane is usually added to TEOS or TMOS, sometimes gels were obtained using the organosilane as the sole precursor [38]. With Ormosils it is possible to enhance the stability of hydrophobic proteins. Lipases provide a nice example showing how the chemical control of the gelation process could improve the enzymatic activity. Indeed, in aqueous media lipases catalyze hydrolysis of fats and oils to fatty acids and glycerol, whereas in organic media they catalyze esterification reactions. The active site is accessible to the substrate only when the displacement of an amphiphilic peptidic loop occurs. In aqueous solution this loop covers the active site, while at the lipid/water interface the loop undergoes the conformational rearrangement required to optimize function. Organic modified silicates offer a more lipophilic environment than hydrophylic silica matrices, and this type of encapsulation increases the activity of lipases [41] (see also Section 4). Ormosils are characterized by poor mechanical stability because of the presence of alkyl groups that reduce the degree of cross-linking. The same effect might contribute to increase the pore size and, thus, to improve the mobility of species in the glass. Phase separation occurs if several organic groups are added. In turn, this results in poor optical quality and cracking, perhaps owing to the existence of a critical point for the self-association of components. However, if the organosilane level increases over 50%, a homogeneous material is obtained [38]. Finally, nanocomposite materials are formed from inorganic or organic silicate precursors in combination with additives, such as hydrophobic or hydrophilic polymers [poly(ethylene glycol), poly(vinylimidazole), poly(ethyleneimine)], silicones, ionic or zwitterionic surfactants, micelles, liposomes, emulsions, liquid crystals, sugar, dyes, and redox species. These additives are usually dispersed into the silicate-containing solutions during the hydrolysis step. This might involve interaction with silica or segregation into independent phases, thus modifying the properties of the final material. Studies performed with fluorescent probes suggest that hydrophobic polymers tend to form a unique phase within silica derived nanocomposites, whereas hydrophilic polymers are characterized by a greater miscibility with aqueous solutions and, thus, are more frequently used. However, with high molecular weight hydrophilic polymers a segregation phase occurs [38].

2.3. Gel Structure and Interactions with Encapsulated Biomolecules

The design of useful biocomposites requires the optimization of the material features in order to adapt their properties to the performance required from the encapsulated protein. To this aim, the local environment around the active biomolecule has to be taken into account, as it may be affected by the physico-chemical properties of the silica matrix. Therefore, key sol–gel synthesis parameters are (i) the nature of the interactions between the gel matrix and the dopant protein and/or substrates/analytes and (ii) the influence of sol–gel glass on protein conformation and dynamics.

The surface chemistry of amorphous silica was extensively reviewed by Zhuravlev [42]. Here, we focus on the silica properties and synthesis parameters which affect protein bioactivity. One critical feature that controls the interaction between the biomolecule and the sol–gel silica matrix is the microstructure of the glass and, therefore, the formation and the characteristics of the pores that contain the protein molecules. In an enzyme doped silica matrix, pores should be large enough to unrestrict the free diffusion of solvent and ligand molecules, such as buffer ions, analytes, substrates, and products of enzymatic reactions. At the same time, they should be small enough to prevent leakage of the encapsulated macromolecules. The primary tool for characterizing the microstructure (i.e., the micropore volume and geometry) of xerogel samples is the measurements of nitrogen adsorption isotherms. Up to now, these experiments have showed that a unimodal pore size distribution does exist in protein-doped silica gels. The specific surface area and the pore size distribution of trypsin doped xerogels using Brunauer, Emmit, and Teller (BET) and Kelvin equations from nitrogen adsorption isotherms were calculated [43]. Glasses were characterized by an average distribution of pore diameter of 3–4 nm. The absence of autodigestion

by encapsulated trypsin, together with no detectable leakage from the gel, was interpreted as proof of the lack of diffusion within the matrix. Competion assays, performed to investigate the interaction of the substrate *N*-benzoyl-L-arginine-4-nitroanilide (BAPNA) with the active site, gave an indirect evaluation of the pore size. The inhibitory effect of polylysine polymers on BAPNA hydrolysis suggested that this effect is not stereospecific and is probably due to coating of the pore surface, interfering with BAPNA diffusion. More specific inhibitors were tested and found to be ineffective above 18,000 M_r, a cutoff value in agreement with the estimated pore size.

In aqueous silica xerogel doped with glucose-6-phosphate dehydrogenase and horseradish peroxidase, the pore distribution was determined and found to be relatively large [30]. Pores as small as 20–30 Å and as large as 800 Å were present. The mean pore size was approximately 200 Å. This finding is rather surprising because such large pores should allow the protein to escape from the gel (in the absence of interactions between protein surface and pore walls). Encapsulation of glucose-6-phosphate dehydrogenase and horseradish peroxidase in aqueous silica gel showed that, over the pH range 4–7, it is possible to tune the mean pore size and the volume fraction porosity: the lower the pH of the buffer used, the lower the mean pore size and volume fraction porosity [30]. Indeed, the pore size and porosity of a silica matrix are a strong function of reaction parameters as precursor ratio, pH, and temperature. Although the essential chemistry is simple, variation of one or more experimental parameters can dramatically affect the microstructure of the resulting gel. The dependence of pore structure and morphology on the relative rates of condensation and hydrolysis was investigated in the case of undoped silica gels [19, 44]. The measured value of times t_{clear} and t_{gel} provides information concerning the relative rates of hydrolysis and condensation [19]. t_{clear} is defined as the time after which two separate liquid phases are no longer discernible and is taken as an indication of the completion of the hydrolysis reaction, whereas t_{gel} is the time required for gelation. Studies performed on TEOS derived silica gels showed that t_{clear} decreases as the reactant ratio H_2O:TEOS increases at a fixed pH. This might reflect the faster hydrolysis rate due to a finer dispersion of the TEOS droplets in the aqueous phase. On the other hand, these results also suggested that the kinetics of the condensation reaction strongly depends on the reactant ratio, being more rapid as the ratio decreases at a fixed pH, due to more frequent interparticle collision [44]. The effect of pH on reaction kinetics and, thus, on gel microstructure, was also studied [19]. The hydrolysis and condensation reactions may be catalyzed by either an acid or a base. The condensation reaction proceeds via a base-catalyzed mechanism for pH values above the isoelectric point of silica (pH $\sim$ 2) and is acid-catalyzed for pH values below the isoelectric point. Gelation occurs rapidly at intermediate pH values (3–8) and relatively *mesoporous* gels are obtained. In these conditions, condensation is fast relative to hydrolysis and the formation of highly branched silica species is achieved. These species form loosely packed, clusterlike structures that coalesce leaving mesoporous regions between them. As the pH decreases and approaches the isoelectric point of silica, condensation becomes rate-limiting and randomly branched and relatively linear silicates result. The tightly interwoven gels, characterized by smaller pores, are defined as *microporous* structure.

Aging and, mainly, the hydration state affect the properties of the polymeric network microstructure. Xerogels possess an 8–10-fold smaller volume, reduced pore size, and higher mechanical strength than the corresponding wet gels. During drying, protein-containing pores behave differently from the empty ones. The former shrink but are adapted to the size of the protein, whereas the latter shrink and even collapse [27]. Although pore shrinkage is a common phenomenon in gel drying, the magnitude of the process is different in alkoxide based silica and in aqueous silica gel. Essentially, the shrinkage is much smaller in the latter, because some solid fraction already exists in the colloidal solution. The average pore size of the dried aqueous gel, determined using the BET adsorption method, was 5.3 nm in diameter, in contrast to the value of 2–3 nm found in dried alkoxide gels. It was also demonstrated that, in spite of the total shrinkage after complete drying being the same for a hemoglobin doped thin film and for bulk monoliths, the higher the rate of shrinkage (in a thin film pore contraction occurs rapidly), the higher the retention of native protein conformation, hence the protein stability [29].

PGS-derived silica gels form exceptionally porous matrices. Thermoporometric analysis of trypsin-doped sol–gels indicates that the mesoporous structure and pore volume of the hydrogels were substantially preserved upon formation of xerogels [31]. Gel shrinkage and pore collapse can be minimized by using additives acting as pore fillers or templating agents [45–47]. Nonsurfactant organic compounds as D-glucose, dibenzoyl-L-tartaric acid, or D-maltose might act as templating or pore-forming agents. A nonsurfactant pathway to mesoporous silica materials doped with acid phosphatase was reported [47]. D-glucose was added to HCl-catalyzed prehydrolyzed TEOS sol to affect the mesophase formation in polycondensation and gelation. After removal of the nonsurfactant molecule by solvent extraction at room temperature, a mesoporous silica matrix with high surface area and pore diameter in the range of 20–30 Å, with narrow pore size distribution, was obtained. The pore size, volume, and surface area were found to be tuneable by varying the concentration of the templating compounds in the preparation procedure.

The entrapment of the protein is achieved from the initial dispersion of an aqueous solution of the biomolecule into a silica sol. Since the protein is added prior to gelation, it is plausible that the dopant molecule could act as a template around which the gel network grows. Horseradish peroxidase and glucose 6-phosphate dehydrogenase are globular proteins with diameters of 64 and 88 Å, respectively. They were immobilized in wet silica gels with average pore size of 200 Å [30]. Hence, they should be able to diffuse within the pores and channels of the silica matrix, although at a slower rate than in solution. However, over time and during repeated washes, no significant leaching was observed, suggesting that the enzyme molecules could be sterically confined in smaller pores. Entrapped proteins could also be retained by electrostatic interactions or hydrogen bonds (directly or via hydration water molecules) between the silanols groups, which are negatively charged

above pH 3, and charged or polar groups on the protein surface. Polar interactions mainly depend on the isoelectric point of the protein. Strong adsorption was shown to occur in the case of cationic proteins [48, 49]. A high ionic strength of the buffer should minimize electrostatic interactions, as suggested by data on horseradish peroxidase and glucose 6-phosphate dehydrogenase silica gels in the presence of 1 M phosphate buffer [30]. Studies performed on silica gels doped with photochromic probes (azobenzene derivatives) indicated that the interaction between the host matrix and guest compounds affects the physico-chemical properties of the dopant and thus its reactivity [50]. The results suggest that a close interaction between large portions of the protein surface and the matrix does exists. In fact, it was found that the dye Comassie blue was not able to cover the complete surface of trypsin encapsulated in silica gel [50]. This tight binding could explain the thermal and chemical stability of the glass-encapsulated proteins. Indeed, experiments performed with fully acetylated trypsin showed that the encapsulated enzyme was not active, indicating the important role of ionic bonds between the lysine and arginine side chains and the negatively charged silicates for good dispersion and protein stability in the gel [43, 51].

Absorbance measurements of entrapped proteins such as cytocrome *c* have suggested that mixing the biomolecules into a silica sol prior to gelation leads to biomolecules dispersed homogeneously on a macroscopic scale [4]. At a molecular level, however, several absorbance and fluorescence studies using both small organic probes and fluorescent proteins indicated that dispersed species are actually distributed between several microenvironments within sol–gel derived materials.

Typical methods to study sol–gel materials, including ^{13}C and ^{29}Si nuclear magnetic resonance (NMR), infrared and Raman spectroscopies, cryogenic gas adsorption analysis, and small-angle X-ray scattering, report on the average properties of the bulk materials.

Electronic spectroscopy of specific probe molecules, especially emission spectroscopy, offers sensitive methods for studying the local environment around the probe molecule. If the dopant biomolecule itself cannot act as a probe, other optical probes are added during the sol stage. Studies probing the local microenvironments via electronic spectroscopy within biocomposite materials were the subject of several extensive reviews [23, 34, 38, 52]. An important issue arises from the macroscopic location within the gel of the dopant molecule, whose properties can also be affected by the microscopic heterogeneity of the pores. Depending on the nature of the protein and the specific precursors and additives used to obtain the biocomposite material, it is possible to identify four very general locations of the dopant molecules within the sol–gel matrix [52]: (i) the solvent region within the interior of a pore; (ii) the interface between the liquid solvent and the solid pore wall; (iii) the pore wall; (iv) the constraining region, where the distance between the pore walls is about the same as the dimension of the probe molecule itself. If a molecule is incorporated in the constraining region, it should be effectively confined in the solid state, because of the surrounding rigid solid matrix. Nevertheless, since proteins are added at the sol stage, they preferentially remain in the liquid phase of the gel. Thus, the most probable locations are the solvent filled pores and the pore walls. Indeed, frequent collisions between the molecule and the pore walls could favor the interactions, such as adhesion or hydrogen, electrostatic and covalent bonding between the protein and the pore walls.

The fluorescence properties of pyranine (8-hydroxy-1,3,6-trisulfonated pyrene) are sensitive to proton transfer phenomena and were employed to determine the solvent composition during the sol–gel process. The data also shed light on the nature of the hydrolysis and condensation reactions of the precursors. The first studies were carried out on TMOS-derived gels [53]. The results, also reviewed by Dunn and Zink [52], showed that hydrolysis and condensation processes concomitantly advance toward an equilibrium point, representative of the final solvent composition. pH influences the rate at which the equilibrium is reached, low pH allowing rapid hydrolysis and slow condensation rate. Since the condensation reaction proceeds slowly, these studies were able to ascertain that it occurs much later than the gelation point. Another important result is related to the possibility for the entrapped molecules to be affected by the external solvent and solutes. Although the processes are the same, the sol–gel dynamics differ significantly for bulk monoliths and thin films. The sensitivity of the luminescence spectrum of pyrene to changes in polarity allows one to investigate the characteristics of the solvent inside the pores of the matrix. In the sol and in the aged gel the liquid phase behaves similarly to an aqueous alcohol environment. A few studies were performed using pyrene as a photoprobe [54–57].

By affecting the microstructure of the gel network, drying produces changes in the polarity of the environment. When the xerogel stage is achieved, the residual solvent evaporates, and the pores collapse around the dopant molecule. Thus, especially in the case of silicate systems based on TMOS or TEOS, a more polar environment develops, since the hydrated surface of the pore comes in contact with the probe molecule. In organically modified silicates, experiments carried out with pyrene showed that the final stages of drying lead to a less polar environment, because of the introduction of organic groups into the matrix. Low levels of organosilanes tend to be well dispersed within the matrix and to reside at the solvent–silica interface, thereby reducing the number of silanol groups (and the amount of bound water), as suggested by several experiments performed on a considerable variety of Ormosils [38].

pH measurements within the solvent filled pores and at the interface region tackle another critical issue, especially for encapsulated proteins and organometallics sensitive to proton concentration. pH sensors, based on bromophenyl blue-doped sol–gel materials, showed a broad shape of the titration curve attributed to the entrapment of the probe in a series of chemical inequivalent microenvironments [58]. Titrations performed with acid–base indicators spanning a pH range from 3 to 10 suggested that the pH in the proximity of the pore walls might be one unit lower than that in the aqueous buffer solution [52]. These findings are supported by data on glucose-6-phosphate dehydrogenase and horseradish peroxidase-doped aqueous silica gels [30]. The effect of pH on the maximal catalytic activity (V_{max}) of both the enzymes was studied. In both cases, V_{max} versus pH

curves are shifted to higher pH values (about one unit) for the encapsulated enzymes with respect to the free form. This probably derives from the partitioning of hydrogen ions between the inside and the surface of the matrix. Because of the negatively charged silanol groups, an electrical double layer consisting of cationic buffer ions and hydrogen ions exists inside the silica matrix. Thus, the excess of protons makes the pH in the double layer lower than in the bulk solution. Since the average pore size is about 10 nm, a double-layer overlap into the pores might occur, leading to the presence of an excess of protons and, therefore, to the lowering of pH experienced by the embedded protein. However, it should be noted that the protonation of the solvent-shielded chromophore of green fluorescent protein mutants exhibits a pKa value essentially unchanged with respect to solution [59]. The same was observed for the alkaline Bohr effect of human hemoglobin [26, 60]. These apparently contradicting findings might be explained by the different location of the protein ionizable groups with respect to the proton gradient on the surface of the pore. The kinetics and the equilibrium behavior of the reaction between substrates and encapsulated proteins are determined by the accessibility of analytes to the entrapped protein. The latter is affected by both gel characteristics (such as thickness, hydration, pore size, and net charge), which are tuneable by controlling the synthesis conditions, and the physico-chemical properties of diffusing molecules. Even in the same sample, it could be that some of the protein molecules are encapsulated in pores that are completely accessible to analytes, some in partially accessible pores, and some in completely inaccessible pores. It is important to determine the fraction of the protein molecules actually accessible to the analytes. This appears to be highly dependent on the size of the protein [28]. Moreover, if repulsive or attractive interactions exist between the analyte and the silica matrix, a partial or a very high uptake of the reagent could occur. Thus, diffusion of charged species between silica glass and the surrounding solution results in different concentrations of species in the two phases, even when the system is in equilibrium. Neutral and charged quenchers, such as O_2, acrylamide, Co^{2+}, and I^-, were used to monitor the accessibility of small molecules to entrapped proteins and their partitioning within the gel [22, 50, 61, 62]. Studies performed on small proteins such as parvalbumin or oncomodulin (molar volume of 12–15 nm^3) showed that these proteins remain completely accessible to both neutral and charged analytes [63, 64]. A larger protein like monellin (molar volume 23 nm^3) showed an inaccessible fraction between 8 and 15%. Bovine serum albumin (molar volume 43 nm^3) may have as much as 50% of the protein in inaccessible locations [45, 65]. These measurements suggest that large proteins probably block the entrance of the pore within which they reside, preventing the entry of analytes.

A kinetic study of the photoinduced redox reaction of zinc cytocrome *c* by laser flash photolysis investigated the effects of pH and ionic strength on the diffusion of charged solutes ($[Fe(CN)_6]^{3-}$, $[Ru(NH_3)_6]^{3+}$, and $[Ru(NH_3)_5Cl]^{2+}$) and on the transport of electroneutral small species (*p*-benzoquinone, dioxygen) [22, 61]. The partitioning coefficient of charged quenchers strongly depends on ionic strength and pH; that is, high ionic strength and low pH values minimize the exclusion of the $[Fe(CN)_6]^{3-}$ from the interior of the gel, because the degree of interaction with silica surface is reduced. Experiments performed on azobenzene derivatives confirmed that sol–gel silica can act as a hydrogen donor to an acceptor molecule. This could induce an enormous excess in the uptake of hydrogen-bonding solutes from solution, a phenomenon called imbalanced uptake [50]. Neutral species are not excluded from the glass, but their diffusion is delayed by the matrix because of pore connectivity and tortuosity increasing the effective diffusion distance. Therefore, in the case of these species, it is possible to define a hindrance of diffusion in glasses. All these results suggest that incorporation of charged polymers, manipulation of pH and aging conditions, and the addition of templating agents to promote larger pores could affect and maximize the accessibility of the protein by analytes.

3. STRUCTURE, DYNAMICS, AND STABILITY OF PROTEINS ENCAPSULATED IN SILICA GELS

Proteins perform a variety of functions in living organisms, from enzymatic catalysis to metabolite transport and storage, from cell movement and mechanical support to signal transduction and cell regulation. Most of these activities require structural flexibility and the interconversion among conformational states and substates, controlled by ligand binding and/or variations of the physical and chemical environment. Encapsulation in a silica gel matrix can alter protein function by distorting the structure of functional conformations or, more likely, by posing thermodynamic and/or kinetic restraints to structural dynamics and conformational equilibria underlying protein function and regulation. In many cases, the stabilization of selected conformers with defined properties is not a shortcoming but a desired characteristic of encapsulated macromolecules, for example, in the fabrication of biosensors or whenever the isolation of stable or metastable conformers is required for biochemical and biophysical investigations (see Section 4). A large number of studies have been reported dealing with functional properties of proteins encapsulated in silica gels [10]. Less abundant is the information currently available on protein stability in the gel, and lesser yet the number of investigations that specifically tackled the problem of achieving detailed information on the effect of encapsulation on protein structure.

3.1. Structural and Dynamic Properties

The nature of the gel limits spectroscopic investigations on entrapped proteins to techniques that can probe the global structural properties but, apart from Raman and fluorescence spectroscopy, are characterized by poor spatial resolution. To date, a limited number of studies exploiting Raman or circular dichroism spectroscopy to gather direct structural information on encapsulated proteins have been published. Steady state fluorescence and emission decays of intrinsic and extrinsic probes have been shown to provide a wealth of useful information on structure, dynamics, and function of encapsulated proteins [28]. However, these techniques do not probe conformational events that do not perturb the

fluorophore microenvironment. Furthermore, the interpretation of data can be challenging when fluorescence signals arise from multiple probes (e.g., most proteins contain more than one tryptophan residue), or whenever the effects on fluorescence of solvent structure and polarity are not perfectly understood. Time-resolved fluorescence anisotropy investigations allow one to calculate the scope and speed of local motions and the global rotational diffusion times. The comparison of these parameters with those observed in solution also yields important information on the microenvironment experienced by the encapsulated proteins, since they are related to protein conformation and flexibility and to the nature and strength of protein–gel matrix interactions.

In many reports the observation of unperturbed absorbance spectra has been assumed as indirect evidence of conserved local structural features in the gel. Such an inference, originally expressed by Zink and co-workers [4], is reasonable at least for hemoglobin, myoglobin, cytochrome *c*, and other metalloproteins, whose specific spectral bands are extremely sensitive to the metal ion and active site geometry. Stronger evidence of native structure and dynamics in the gel arises from the quantitative conservation of functional properties and ligand regulation, as in the case of hemoglobin and several enzymes (see Section 4). In this regard, it is of paramount interest to relate the amplitude of conformational dynamics required by protein function and regulation, often known from X-ray studies, to the extent of function conservation in the gel. The emerging scenario, supported by an increasing amount of data, is one where silica gel encapsulation impedes large-scale quaternary conformational changes while allowing, though usually slowing down, local structural dynamics [6]. Whether the uncoupling of tertiary and quaternary conformational changes originates mainly from thermodynamic or kinetic restraints is not obvious and is likely to be case-dependent. The most extensive biophysical investigations on tertiary and quaternary structural dynamics in silica sol–gel systems have been carried out on hemoglobin and fructose-1,6-bisphosphatase. The results will be mainly reviewed in Section 4, together with other studies providing functional characterization of encapsulated proteins.

3.1.1. Structural Properties

Zink and co-workers investigated the structural properties of encapsulated bacteriorhodopsin by circular dichroism and Raman spectroscopy [25]. The results indicated that bacteriorhodopsin was encapsulated along with its membrane lipids and that neither the quaternary conformation nor the local structure of the retinal chromophore was significantly affected.

The sol–gel encapsulation of the complex cytochrome *c* : cytochrome *c* peroxidase allowed one, for the first time, to measure electron paramagnetic resonance signals of biomacromolecules at room temperature [66]. This made it possible to investigate the coordination and spin state, the charge transfer activity, and the structural orientation of the complex and its constituents.

The polypeptide backbone and the heme group conformation of native and Zn-cytochrome *c* have been investigated by ultraviolet(uv)-visible absorption, circular dichroism, and resonance Raman spectroscopy [22]. Far-UV circular dichroism spectra of native cytochrome *c* encapsulated in hydrogels indicated a very modest perturbation of the secondary structure. A more significant loss of α-helicity was observed for Zn-cytochrome *c*, which also exhibited a more evident perturbation of the near-UV spectrum, sensitive to the heme environment conformation. However, visible circular dichroism and resonance Raman spectra indicated that for both proteins the conformational changes of the heme pocket induced by encapsulation must be small. This result was confirmed by resonance Raman experiments carried out by Zink and co-workers [67], suggesting that neither sol–gel encapsulation, gel aging, nor drying significantly alters the heme site geometry.

The structural properties of three encapsulated metalloproteins, cytochrome *c*, horseradish peroxidase, and ascorbate oxidase, have been studied by circular dichroism, absorbance, and fluorescence spectroscopy [68]. Circular dichroism spectra of cytochrome *c* and horseradish peroxidase in the Soret region indicated that in both cases the heme microenvironment and the protein backbone to which the heme is covalently bound are unaffected by gel encapsulation. Absorbance and fluorescence spectra of tryptophan residues and the extrinsic probe ANS indicated a partial denaturation of ascorbate oxidase, which nevertheless retained the bound copper atoms. Partial denaturation has been reported for other encapsulated proteins, like apo-myoglobin [65, 69] and, upon aging of the gels, human [70, 71] and bovine serum albumin [70].

Encapsulation of the isozyme II of bovine carbonic anhydrase caused only minor changes of near- and far-UV circular dichroism spectra, which are sensitive to the aromatic side chains microenvironment and backbone conformation, respectively [62]. Thus, both the secondary and tertiary structure appeared to be generally conserved.

A series of elegant studies by Friedman and co-workers demonstrates that hemoglobin encapsulated in silica sol–gels retains the conformational markers of T and R quaternary conformations, as monitored by ultraviolet [72] and visible resonance Raman spectroscopy [73]. Ultraviolet Raman spectra are sensitive to the environment of aromatic aminoacids and allowed one to demonstrate that the quaternary structure-dependent $\alpha_1\beta_2$ interface of hemoglobin was unperturbed upon encapsulation [72]. Resonance Raman spectra of the Soret band can probe the perturbations of the heme iron-proximal histidine stretching mode, exhibiting frequencies that are characteristic of different tertiary and quaternary conformations [73]. The effect of mutations of cysteine β93 on T and R state conformation was also investigated [74]. The intensities and frequencies of Raman bands were found to be the same as in solution also for encapsulated deoxy-myoglobin [75].

Near- and far-UV circular dichroism spectra in the gel were found to be similar to those measured in solution also for hen egg-white lysozyme, bovine α-lactalbumin, and met-myoglobin [69]. Apo-myoglobin, however, was found to be largely unfolded upon encapsulation [69], as already observed by fluorescence spectroscopy [65]. The secondary structure was insensitive to ionic strength and pH values above the isoelectric point of the protein, ruling out the possibility that the partial unfolding was due to electrostatic

interactions between the positively charged protein and the negatively charged silica oxide groups. Instead, salts other than KCl induced a concentration-dependent increase in the α-helical content, and the ion effect was in accordance with the Hofmeister series, suggesting an important role of the "unusual" water structure in the gel pores on the conformation of encapsulated apo-myoglobin. The spectroscopic and kinetic heterogeneity of sol–gel encapsulated carbonmonoxy (CO)-myoglobin was investigated by resonance Raman and absorption spectroscopy [76]. The authors compared the visible Raman spectra and CO rebinding kinetics after photodissociation of myoglobin encapsulated in the deoxy or liganded state. Differences in the resonance Raman frequency of the iron-proximal histidine stretching mode were attributed to different positions of the F helix, which in turn account for differences in CO rebinding. The CO binding-induced conformational changes were reported to be qualitatively similar to the tertiary changes occurring within both the T and R quaternary states of gel-encapsulated hemoglobin: a clam-shell-type rotation of the E–F helices. Differences between the two proteins in the relaxation kinetics were attributed to different hydrogen bonding and salt bridging patterns.

3.1.2. Dynamic Properties

The rotational dynamics in silica sol–gels have been investigated for several proteins. The results span the whole range of rotational freedom, from unrestricted rotation to complete "freezing" of the entrapped molecules. The observed variability could in part originate from differences in the gel matrix, dependent on precursors, gelification protocols, aging, and drying. However, most likely the main source of diversity resides in the physical and chemical characteristics of the encapsulated proteins. For instance, the templating effect of the gel would hinder the global tumbling of macromolecules with a pronounced oblate or prolate shape. Moreover, one should take into account the possibility of chemical interactions between polar and/or charged groups on the protein exterior and the gel pore surface, carrying several functional groups which account for a net negative charge at neutral pH. Such interactions would be peculiar of each protein and are expected to be pH- and ionic strength-dependent.

Dielectric relaxation measurements on cytochrome *c* suggested that the rotational movement of the protein within a TMOS-derived matrix is only slightly restricted, as indicated by a small 1.1 kcal/mol increase of the activation energy for dipolar relaxation [77, 78].

Luminescence anisotropy decays of Zn-deutero-myoglobin, Zn-cytochrome *c* peroxidase, and [Zn, Fe^{3+}L] mixed-metal hemoglobin hybrids encapsulated in wet TMOS-derived gels indicated that a minor fraction of each of the three proteins is immobilized in the sol–gel, presumably by interactions with the silicate backbone [79]. A larger fraction, ~94%, ~70%, and ~85% respectively, appeared to rotate freely within the sol–gel, though more slowly than in solution (on a time scale of 20 μs or less).

Bright and co-workers reported that the global rotational mobility of bovine and human serum albumin, as monitored by time-resolved fluorescence anisotropy of the extrinsic probe acrylodan covalently attached to cysteine 34, was restricted by a factor of 2–3, while local motions of the probe were unperturbed in aged wet gels with respect to solution [70]. Upon gel drying, the data suggested an opening of the pocket hosting the acrylodan reporter. Doping the gel with polyethylene glycol resulted in slightly increased global rotational dynamics of bovine serum albumin, while the local acrylodan motion was relatively unaffected [46]. Time-resolved anisotropy of the coenzyme FAD indicated that the tumbling motion of glucose oxidase was slowed down by less than twofold upon encapsulation [80]. FAD local motions and, remarkably, the opening of the active site pocket upon binding of the substrate glucose were not significantly altered. Authors from the same group reported that the rotational reorientation dynamics of intact immunoglobulin G antidansyl antibodies are hampered in aged gels, despite roughly conserved equilibrium binding constants for the target hapten [81].

Brennan and co-workers carried out several studies in order to characterize the conformational flexibility and the rotational diffusion of gel encapsulated proteins, like monellin [82, 83] and human serum albumin [71]. The emission spectra of monellin upon excitation at 295 nm were identical in wet-aged and dry-aged gels and in solution, indicating the conservation of protein structure in the proximity of the single tryptophan residue, while steady-state anisotropy data and the changes in fluorescence properties upon denaturation were indicative of a reduced protein mobility [82] and conformational flexibility [83]. The acrylamide quenching indicated that the range of conformational motions is reduced for encapsulated monellin with respect to solution [83]. Entrapped human serum albumin exhibited a large residual anisotropy, consistent with a restriction of global rotational dynamics [71]. The higher fractional contribution of local motions to anisotropy decays, and the shortening of the longer rotational correlation time with respect to solution suggested a partial denaturation of encapsulated albumin, consistent with the observation of perturbed spectroscopic properties of the single tryptophan residue.

The anisotropy decays of magnesium protoporphyrin IX substituted myoglobin, which due to the chromophore rigidity are sensitive only to the global protein motion, indicated that the rotational dynamics in gels prepared with different protocols are almost completely hampered, with rotational correlation times of the order of 1 μs [84].

The structural and dynamic properties of a highly fluorescent mutant of the green fluorescent protein, GFPmut2, appeared to be conserved in silica sol–gels, for both the chromophore moiety and the whole protein molecule [59]. The pH-dependent emission spectra, fluorescence lifetimes, anisotropy, and rotational dynamics are essentially equal to those observed in solution. The impressive photostability of gel-encapsulated GFPmut2 allowed single-molecule imaging by confocal spectroscopy, a first-time experiment for biomolecules encapsulated in silica sol–gels. These results demonstrate that the system exhibits ideal properties for biochemical and biophysical studies at a single-molecule level.

Indirect evidence of structural dynamics in the gel arises from the observation of the interconversion between functionally distinct species. Once again, hemoglobin is the paradigm molecule for the investigation of allosteric transitions. Tertiary and quaternary conformational changes

appear to be kinetically well separated in gel-encapsulated hemoglobin, and for both the time course is slowed down by orders of magnitude with respect to solution, as demonstrated by oxygen binding [26, 60, 85–87], CO rebinding after photodissociation [88, 89], and resonance Raman studies [72, 73]. This allows one to trap nonequilibrium species of hemoglobin and probe both structural and functional properties of intermediate conformations.

The series of conformational changes following carbon monoxide binding to gel-encapsulated human adult hemoglobin has been characterized by ultraviolet resonance Raman spectroscopy [72]. The results indicated that ligand binding in the gel triggers, more than a single concerted switch, a sequence of conformational steps starting at the distal heme pocket with E helix motion and leading to shifts of the "switch" and the "hinge" region of $\alpha_1\beta_2$ interface which are characteristic of the T to R quaternary structure transition. Conformational dynamics in the gel were slower than ligand diffusion times and could be investigated in the absence of the traditional experimental limitation of geminate rebinding of the ligand leading to mixed populations of partially liganded species. Resonance Raman spectra of oxygenated hemoglobin at various times after oxygen scavenging with dithionite revealed a marked slowdown of tertiary and quaternary relaxations [73]. In the gel, these transitions are strongly temperature-dependent, both relaxations being essentially prevented at 4 °C. The carbon monoxide rebinding studies to horse myoglobin by Abbruzzetti et al. [89, 90] showed an enhanced amplitude of the geminate rebinding phase with respect to solution, indicating a decreased flexibility of the protein matrix that reduces the efficiency of CO escape to the solvent following photodissociation. The slower rate for the transition between conformational substates was tentatively attributed to high effective viscosity inside the gel matrix. Friedman and co-workers reported that upon bathing gel-encapsulated deoxy- and carbonmonoxy (CO)-myoglobin in 100% glycerol, not only are tertiary relaxations slowed down but also equilibrium and nonequilibrium conformations can be efficiently locked [76]. The authors attributed this observation, beyond possible viscosity effects, to the templating behavior of the gel. Conformational modifications requiring transition states with enhanced volume and hydration changes might be efficiently hampered by steric hindrance and solvent shell stabilization effects by the gel matrix [73, 76]. The latter hypothesis is supported by the observation of dramatically slowed relaxations of encapsulated hemoglobin below 10 °C, suggesting a relevant role in the immobilization of hydration shell water molecules [73].

Beyond the several studies on heme proteins, a few papers reported detailed investigation on tertiary and quaternary conformational dynamics accompanying function and regulation of proteins encapsulated in silica gels. Encapsulation of pig kidney fructose-1,6-bisphosphatase largely restrained the structural transition between the T and R quaternary conformations, while allowing the local conformational changes induced by allosteric regulators [91]. Interestingly, glutamate dehydrogenase, a large six-subunit enzyme, retains function and allosteric regulation in the gel, indicating that encapsulation does not hinder the large-scale conformational changes that have been structurally resolved, accompanying substrate and coenzyme binding, catalysis, and regulation [92].

The conformational changes induced by calcium binding in cod III parvalbumin were probed by measuring intrinsic tryptophan fluorescence and acrylamide binding [63], absorbance, and steady-state and time-resolved fluorescence of the extrinsic probes fluorescein, acrylodan, and nitrobenzoxadiazole attached at cysteine 18 [93]. Single tryptophan mutants of oncomodulin, another calcium-binding protein, retained in the gel the calcium concentration dependence of thermodynamic stability and terbium binding capacity [64].

The light-sensitive and proton-pumping function and the relative conformational changes are conserved in immobilized wild-type bacteriorhodopsin [25, 94, 95] and in the D96N mutant [14, 96–98].

An elegant work by Hoffman and co-workers [79] describes the effect of encapsulation on photoinitiated electron transfer within three protein–protein complexes: (i) [Zn, Fe^{3+}L] mixed-metal hemoglobin hybrids, (ii) the Zn-cytochrome *c* peroxidase complex with ferri-cytochrome *c* and (iii) the [Zn-deuteromyoglobin, ferri-cytochrome b_5] complex. In all cases transient absorption experiments allowed one to observe electron transfer in the encapsulated complexes. The effect of encapsulation on electron transfer was different for the three complexes that were selected for their largely different affinities and dynamic processes. No complex formation or dissociation was observed in the gel.

Brennan and co-workers observed reversible dissociation of a complex between bovine calmodulin and the peptide melittin, upon introduction in the gel of the denaturant guanidine hydrochloride or small antagonists of the complex [99]. The detection of reversible interprotein interactions in the gel is striking, provided that quaternary transitions in entrapped oligomeric proteins are normally hampered, and points out that silica sol–gels can be a suitable device to host biological systems more complex than single macromolecules, for both biophysical studies and biotechnological applications (see [12] and references therein).

3.2. Stability of Proteins Encapsulated in Silica Gels

Several authors observed thermal and/or chemical stabilization of gel-encapsulated proteins. Such effect could derive, at least in part, from a reduced conformational freedom in the limited space of gel pores (Fig. 3a), preventing a complete denaturation of polypeptide chains and the irreversible aggregation often observed for partially or fully denatured proteins [12]. For proteins existing as an equilibrium between different conformers, stabilization could arise from a lower probability to sample thermodynamically unfavorable conformations, when species interconversion occurs through conformationally expanded transition states. Steric effects could also facilitate refolding by virtue of an "easier" search of the native topology [100]. However, several authors reported either slower refolding kinetics, as in the case of acid-denatured myoglobin [101] or irreversible though incomplete unfolding, as for acid-denatured bovine serum albumin [65] and heat- or guanidinium chloride-denatured carbonic anhydrase II [62]. Depending on experimental protocols and recipes, gel aging, and hydration state,

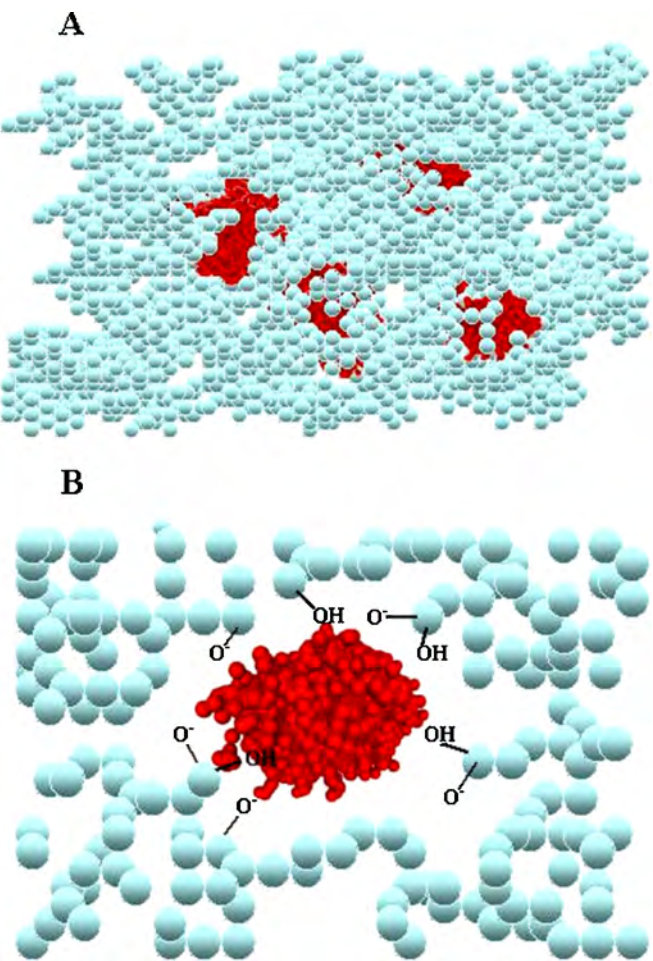


Figure 3. Schematic representation of protein-doped silica gels (A) and protein–gel matrix interactions (B).

variable amounts of functional groups can be present on the pore surface, like siloxane (Si–O–Si), silaketone (Si=O), silanol (Si–OH), and siloxide ($Si–O^-$) groups (Fig. 3b). Therefore, multiple polar and/or electrostatic interactions between surface aminoacids and the pore wall are likely to play a relevant role in protein kinetic and/or thermodynamic stabilization. Once again, general rules cannot be drawn from the data currently available, and one should take into account the relevant study by Heller and Heller [49], suggesting that electrostatic interactions between positively charged arginine surface residues and polysilicate anions could disrupt the function or structure of gel-encapsulated lactate oxidase and glycolate oxidase. When the enzymes were complexed with polycations prior to immobilization, they were stabilized instead of being deactivated. Moreover, the peculiar structure of the solvent and the high effective viscosity within gel pores could play a key role in limiting conformational dynamics, thus stabilizing the structure of entrapped proteins [69, 73, 76, 83]. Finally, it should be noted that the behavioral variability of encapsulated proteins does not only arise from the different properties of the macromolecules and of the hosting matrices, but even in the same system gel aging time and conditions can influence protein conformation, dynamics, stability, accessibility, and function (see [71] and references therein).

Increased thermal stability has been reported for alkaline phosphatase [102, 103], acid phosphatase [31, 43, 47], β-glucosidase [31, 104], lipases (see [31, 105, 106] and references therein), lipoxygenase [107], glucose oxidase [48], monellin [83], carbonic anhydrase II [62], urease [108], cytochrome *c* [67], lysozyme, α-lactalbumin and apomyoglobin [69], antibodies [109], pronase and α-chimotrypsin [110], and human serum albumin [71]. Stabilization with respect to pH, alcohols, or chemical denaturants has been reported, for instance, for bovine serum albumin [65], β-glucosidase [104], cytochrome *c* [27, 77, 78, 111], catalase [111], lipoxygenase [107], lipases (see [106] and references therein and also [112]), alcohol dehydrogenase [113], deoxy- [75] and CO-myoglobin [101], cod III parvalbumin [63], monellin [83], carbonic anhydrase II [62], horseradish peroxidase and myoglobin [114], and antibodies [109]. Enhanced stability toward photodegradation has been reported for a light-transducing protein, phycoerythrin [115], which, unlike the other phycobiliproteins phycocyanin and allophycocyanin, was shown to undergo only minor changes in its native structure.

Only in a few cases have the thermodynamic and kinetic aspects of thermal or chemical denaturation been investigated in detail. Such studies are of extreme interest for at least two reasons: (i) they pursue a qualitative and quantitative description of the behavior of gel-encapsulated proteins with respect to physical and chemical insults, which is fundamental for a rational design of biotechnological applications; (ii) they display to the fundamental research community the unique opportunity provided by sol–gel techniques to isolate "pure" conformational states out of complex conformational equilibria, slow down kinetics, or stabilize metastable intermediates, making this species accessible to biophysical investigations.

One of the earliest studies on unfolding/refolding of proteins encapsulated in silica gels was carried out on bovine serum albumin (BSA) and horse heart myoglobin by Saavedra and co-workers [65]. Based on the steady-state fluorescence properties of intrinsic (tryptophans) and extrinsic (acrylodan) probes, the authors concluded that complete and reversible denaturation of encapsulated BSA by guanidine hydrochloride could be achieved, while acid denaturation, involving a significant increase in the overall size of the molecule, was incomplete and irreversible.

Thermal unfolding of human serum albumin in gels at early aging times occurred at higher temperature with respect to solution [71]. However, the unfolding temperatures decreased upon aging. This observation, together with the spectroscopic evidence of incomplete unfolding and non-native conformation after entrapment, led the authors to raise two major concerns, which should be taken into account for any gel matrix-encapsulated protein system: (i) it is possible that the unfolding temperatures do not reflect denaturation from a fully native to the same fully unfolded conformations ensemble observed in solution, which limits their reliability as stability indicators; (ii) it is not obvious that encapsulation will lead to a higher long-term thermal stability.

The thermal and guanidinium chloride-induced unfolding of monellin has been investigated by measuring the fluorescence spectra, steady-state anisotropy, and acrylamide quenching of the single tryptophan residue [83]. Wet-aged gels did not show a significant thermal or chemical stabilization of the encapsulated monellin. In dry-aged gels, thermal unfolding was characterized by a 14 °C increase of T_m and a broader transition with respect to solution, suggesting a distribution of microenvironments endowed with different thermodynamic stability. Chemical denaturation also showed a broadened transition. The spectral changes upon unfolding and the calculated thermodynamic parameters were different with respect to solution and indicated conformational restrictions upon encapsulation leading to incomplete denaturation or unfolding to a different final state. The enhanced stability was attributed to the gel pores sterically restricting the mobility of the encapsulated protein, and thus limiting the sampling of thermodynamically unfavorable conformations, and to the highly structured nature of the solvent in the gel matrix. Complete unfolding of cod III parvalbumin by urea was achieved in the gel, as indicated by the denaturant dependence of tryptophan fluorescence intensity and emission wavelength [63]. As in solution, calcium binding provided a significant stabilization of the protein. Both in the absence and in the presence of calcium ions, higher denaturant concentrations were required to achieve denaturation in the gel with respect to the protein in solution. Reversible thermal denaturation of oncomodulin has been reported by Brennan and co-workers [64].

Resonance Raman studies by Friedman and co-workers [75] demonstrated that the native structure of deoxy-myoglobin is stabilized toward acid denaturation in the gel. At pH 2.6, the Raman bands assigned to the heme propionate and vinyl groups shifted to values similar to those observed in solution for the first acid intermediate, which occurs at pH between 4.5 and 3.5 and is characterized by the cleavage of the iron–proximal histidine bond, while the heme remains five-coordinate high spin, with water as the fifth ligand (see [75] and references therein). The cleavage of the iron–proximal histidine bond was markedly slowed down in the gel, even at pH values as low as 2. At pH 2.6, it took about 20 h to obtain spectroscopic indications of a small amount of the second acid intermediate, exhibiting a four-coordinate heme still bound in the pocket and occurring in solution at pH between 3.5 and 2.5. The authors suggested that the formation of acid denaturation intermediates of deoxy-myoglobin requires large amplitude conformational fluctuations, like an opening of the heme pocket to allow cleavage of the iron–proximal histidine bond, loss of the heme, and globin unfolding, which are strongly inhibited in the gel. The acid denaturation and refolding of CO-myoglobin have been studied by fluorescence, absorption, and visible resonance Raman spectroscopy and by geminate recombination in order to explore the heme conformation and environment, the globin, and the exposure to solvent of the A helix [101]. Both unfolding and refolding kinetics were dramatically slowed down by gel encapsulation. Different encapsulation and unfolding/refolding protocols allow one to stabilize several intermediate species, some of which have not been observed previously in solution studies. Moreover, the results suggested that the initial step in acid denaturation of CO-myoglobin is the unfolding or the increased exposure to the solvent of A helix residues and that the refolding of the A helix is a late process along the refolding pathway.

Both thermal and guanidinium chloride-induced denaturation of carbonic anhydrase II, as monitored by circular dichroism, appeared to be incomplete and largely irreversible [62]. The broad thermal unfolding transition has been attributed to some degree of heterogeneity in the microenvironment of the encapsulated molecules, which affects their stability toward unfolding. The incomplete unfolding and refolding were ascribed to steric effects of the gel pores, restricting the required protein mobility.

Complete and reversible denaturation of cytochrome *c* encapsulated in aged wet gels was induced by 3.5 M guanidinium chloride and monitored by circular dichroism spectra of the Soret region [68]. The unfolding reaction was dramatically slowed down with respect to solution. In xerogels, only partial denaturation was observed following 24 h incubation in the presence of 3.5 M guanidinium chloride.

Reversible thermal denaturation was observed for lysozyme, α-lactalbumin, and apomyoglobin [69]. The increased thermal stability is discussed in terms of theoretical models of crowding effects on protein stability [116], mimicked by molecular confinement in the gel pores, and perturbed water structure. Different from solution, calcium binding to encapsulated α-lactalbumin was unable to rescue the native structure upon reduction of one of the four intramolecular disulfide bonds. Furthermore, removal of calcium caused a loss in secondary and tertiary structure in the absence of reducing agents.

4. APPLICATIONS OF PROTEIN-DOPED SILICA GELS

4.1. Bioreactors

A biorector is a device in which one or more enzymes are physically confined. The purpose of a biorector is to obtain the selective and efficient transformation of substrates in products and an easy separation between catalysts and reagents. The requirements of a biorector are the maintenance of the catalytic activity of entrapped enzymes, long-term protein stability and reusability, and no limitation for reagents and products to diffuse in and out of the matrix. Immobilization of enzymes in silicic acid gels was first achieved by Dickey [117] and several years later by Johnson and Whateley [18], but only in the last 20 years or less have sol–gel methods became widely used to prepare biorectors and biosensors [3, 4, 34, 43, 102, 118, 119]. A list of enzyme-doped silica gels is reported in Table 1. Several classes of enzymes have been encapsulated in silica gels, ranging from lipases to oxidoreductases and proteases, without any specific limitations. Silicate precursors characterized by different hydrophobicities, and additives as glucose, polyethylene glycol, or polyvinylalcohols, were used to achieve better performances.

A critical issue that is not always properly addressed in the evaluation of the catalytic activity of encapsulated enzymes is the diffusion of substrates within the gel pores to reach

Table 1. Enzyme-doped silica gels.

Enzyme	Precursor	Ref.
C. cylindracea lipase	TMOS[a]–alkylmethoxisilanes	[202]
P. cepacia lipase	TMOS–alkylmethoxisilanes	[47]
Lipase	cellulose acetate–TiO_2 gel fiber	[203]
C. antartica lipase	cellulose acetate–TiO_2 gel fiber	[204]
C. cylindracea lipase	trimethoxypropylsilanes–TMOS	[128]
P. cepacia lipase	phyllosilicate	[205]
R. miehei lipase	cellulosae acetate–TiO_2 gel fiber	[206]
Lipase	TEOS[b]–PEG	[127]
Lipase	TMOS–alkylsiloxanes	[207]
P. cepacia, C. antarctica lipase	MTMOS[c]–alkylsiloxanes	[208]
Lipase	TMOS–alkylsiloxanes	[124]
Lipase	different matrices	[209]
Lipase		[210]
Lipase	MTMOS–TMOS–celite 545	[105]
HRP[d]	TMOS	[136]
HRP	different alkylsiloxanes precursors	[211]
Pronase, α-chymotripsin	TMOS	[110]
Serine proteases		[212]
Lactate oxidase/ glucose oxidase	different matrices	[213]
Urease	TEOS	[214]
Urease	TMOS	[215]
Urease, glucose oxidase, invertase	different matrices	[135]
HRP, GOD[e]	sodium silicate	[30]
HRP-cholesterol oxidase	TEOS	[216]
HRP, ascorbate oxidase	TMOS	[68]
Soybean lipoxygenase	TMOS	[142]
Lipoxygenase	alginate–TMOS	[107]
20 different lipases and proteases, enzyme array (solzymes)		[171]
HRP, GOD		
Acid phosphatase	TMOS + D-glucose	[47]
Alkaline phosphatase	TMOS	[103]
Carbonic anydrase II	TMOS	[62]
Glutamate dehydrogenase	TMOS	[92]
GOD	TEOS–veratryaldehyde	[217]
GOD, lactate oxidase, glycolate oxidase	TMOS	[48]
Glycolate oxidase, lactate oxidase	TMOS	[49]
Lactate dehydrogenase	TEOS	[218]
Bilirubin oxidase	alginate–TMOS	[219]
Atrazine-degrading enzymes	TMOS	[143]
Formaldehyde dehydrogenase/ formate dehydrogenase/ alcohol dehydrogenase	TMOS	[144]
Atrazine chlorohydrolase	TMOS–alkylsiloxane	[141]
Butyrylcholinesterase	TMOS	[220]
Acetylcholinesterase, butyrylcholinesterase	TMOS	[221]
α-amylase	chitosan–SiO_2 membrane	[222]
β-galactosidase	TMOS	[138]
Catalase	colloidal silica–sodium silicate	[223]
Parathion hydroxilase	TMOS	[224]
Cytochrome *c*, Myoglobin Hemoglobin, HRP	TMOS	[225]
Tryptophan synthase *O*-acetylserine sulfhydrylase	TMOS	[201]

[a] Tetramethylorthosilicate, TMOS.
[b] Tetraethylorthosilicate, TEOS.
[c] Methyltrimethoxysilane, MTMOS.
[d] Horseradish peroxidase, HRP.
[e] Glucose oxidase, GOD.

the active sites [120]. It is often assumed that small solutes can freely and rapidly diffuse within the gel and, therefore, there is no limitation to the enzyme activity by the size and geometry of the gel [3, 8, 94]. However, this might not be true, as was previously recognized in the investigation of the enzymatic activity of enzyme crystals [10, 121, 122]. The size of the crystals required to avoid effects on enzymatic parameters critically depends on the rate of the catalytic reaction. Only microcrystals with dimensions in the micrometer range can be used. In several investigations mm size silica gel monolythes were used, and, not surprisingly, the observed catalytic activities were much lower with respect to the soluble enzyme. On the other hand, alcohols released during the sol–gel process might lead to enzyme inactivation, and constraints placed by the Si–O–Si network to the protein might decrease the enzyme flexibility required for function.

Lipase, glucose oxidase, and horseradish peroxidase are among the enzymes more frequently encapsulated in silica gels and for which several studies were carried out to define the influence of the precursor alkylsiloxanes and additives on the catalytic activity and stability. The interest originated in their use for biotechnological applications, including biosensors.

4.1.1. Lipase

Lipases are used as hydrolases in detergent additives and as synthases in the formation of enantiomeric compounds [123–125]. Lipase silica gels are now commercially available. The most critical issue for enzymatic catalysis by lipases is the interfacial activation (i.e., the increase in activity at the interphase between polar and apolar solvents). This is due to an amphiphilic loop that covers the active site and undergoes a closed-to-open transition at a lipid–water interface, making the active site accessible to substrates. Several studies were carried out to produce active lipase-doped silica gels [105, 106, 112, 126]. TMOS-based silica gels led to encapsulated lipases characterized by very reduced activity, indicating a detrimental effect of the polar microenvironment of silica gels [106]. Different $RSi(OCH_3)_3$ precursors and mixtures of these and TMOS were used to encapsulate lipases from *Pseudomonas cepacia*. The use of 100% $CH_3Si(OCH_3)_3$ led to an immobilized lipase with 1300% of the activity of the commercial enzyme [106]. Higher activity was obtained using other R-groups, such as ethyl, *n*-propyl, *n*-buthyl, and *n*-$C_{18}H_{37}$, and additives such as polyethylene glycol and polyvinyl alcohol [112, 126, 127]. The immobilized enzyme is remarkably stable even after several catalytic cycles. Higher esterification activity was also observed when methyltrimethoxysilane and TMOS were used as precursors in a 3:1 ratio and the gel was casted on the surface of Celite 545 [105]. The enantioselective aminolysis of ethyl 2-chloropropionate was also found to increase by carrying out the reaction with *Candida cylindracea* lipase encapsulated in silica gels [128]. An alternative strategy to immobilize lipases was designed, making use of matrices derived from poly(hydroxymethylsiloxane) [129]. This method leads to a biocomposite silicone material with higher activity and stability with respect to the native enzyme. Encapsulation of enzymes within silicate, siloxane, and hybrid sol–gel polymers provides an efficient and generic approach to the

preparation of stable and active sol–gel bioceramics [31]. PGS was used to prepare transparent, mesoporous, and physically stable silica xerogels containing lipases as well as several other enzymes [31]. It was found that, because of the absence of released alcohols that might cause protein denaturation or aggregation, PGS-based biogels retain 83–98% of the activity of the native enzyme, with an 88–98% entrapping efficiency at loading concentration as high as 20% w/w of xerogel. The resulting materials exhibit high stability and low release of proteins after several washing cycles. This remarkably extensive study [31] provides a very careful and detailed investigation of the influence of (i) metallosilicate sol–gel precursors on the activity of subtilisin, aldolase, glycerol 3-phosphate oxidase, glucose oxidase, and pyruvate decarboxylase; (ii) alkylsiloxanes sol–gel precursors on the activity of thermolysin, alcohol dehydrogenase, lipoxygenase, esterase, and phospholipase D; (iii) polyol ester-derived functionalized siloxane on the activity of proteinase K, sialic acid aldolase, bacteriorhodopsin, aspartatate aminotransferase, almond oxynitrilase, β-glucosidase, tyrosinase, glucose oxidase, and carboxypeptidase; (iv) composite sol–gel polymers on the activity of β-glucoronidase, penicillin acylase, and luciferase. *Mucor javanicus* lipase was encapsulated in silica aerogels, retaining about half of the activity of the native enzyme even after incubation at 100 °C in the presence of organic solvents [130]. The comparison of the activity of *Psudomonas cepacia* lipase encapsulated in silica and aluminosilicate xerogels and aerogels indicates that the enzyme is more active in the latter ones. In aerogels the pore size does not shrink during CO_2 supercritical drying as, on the contrary, was observed with drying by evaporation [131]. Esterification rates were found to be higher when aerogels were obtained using a base catalyst as sodium fluoride in the presence of polyvinyl alcohol [132]. However, so far only lipase and myoglobin [133] have been entrapped in aerogels and demonstrated to remain active. The conditions for aerogel preparation, CO_2 at supercritical state, might not be favorable for most proteins.

4.1.2. Glucose Oxidase

Glucose oxidase catalyzes the formation of gluconic acid and hydrogen peroxide, using D-glucose and oxygen as substrates. This enzyme that contains flavin as a coenzyme is widely used in the development of biosensors for the detection of glucose, oxygen, and hydrogen peroxide (see Section 4.4, Table 2 [11, 134]). For example, co-immobilization of the enzyme with horseradish peroxidase and quinonamine dyes was exploited to detect spectrophotometrically glucose [24, 94]. Glucose oxidase activity of aged gels and xerogels was found to be 20% and 10% of that determined for the enzyme in solution. When only glucose oxidase was encapulated, the K_M was twofold higher than in solution and the turnover number was very similar to that in solution, suggesting that the lower activity of the enzyme in the coupled assays was due to slow diffusion of substrates and products within the gel. Enzyme activity in silica gels prepared in the presence of different additives, urotropin, veratraldehyde, glyoxal, glutaraldehyde, and aminocapronic acid, was measured using a calorimetric method [135]. The aminocapronic acid was found to be quite efficient in increasing enzyme activity.

4.1.3. Horseradish Peroxidase

Horseradish peroxidase catalyzes the oxygen peroxide-dependent oxidation of a variety of substrates via a multistep process involving the iron of bound heme. For the development of biosensors, the enzyme is co-immobilized with either glucose oxidase or oxalate oxidase (see Section 4.4 and Table 2). The rate of pyrogallol oxidation by horseradish peroxidase gels was found to be about 30-fold less than that of the soluble enzyme and to decrease as a function of gel aging, suggesting that the reaction was limited by the diffusion of the substrates through the pores of the gel [68]. This finding was confirmed by investigating the rate of formation of compounds I and II, two spectroscopically detectable reaction intermediates [68]. Sol–gel encapsulation was demonstrated to protect heme enzyme peroxidase activity of horseradish peroxidase and methemoglobin from inactivation by hydrogen peroxide, especially at high substrate concentrations [114]. Interestingly, in the sulfoxydation of thioanisole by hydrogen peroxide, catalyzed by horseradish peroxidase silica gels, a decrease of the enantiomeric excess and a similar yield of methyl phenyl sulfoxide enantiomers in the presence and absence of protein-doped gels were observed. These findings indicate that silica glasses are not necessarily inert with respect to particular reactions, thus requiring proper controls [136]. To avoid the formation of alcohols in gel preparation, sodium silicate was used as a precursor and gelation was achieved at neutral pH in the presence of either horseradish peroxidase or glucose-6-phosphate dehydrogenase [30]. The specific activity was 73% and 36% of that of the soluble enzyme, respectively, and the K_M values were higher than those of the enzyme in solution. A pH shift of 0.5–1 units was also observed in the pH dependence of the catalytic activity for both enzymes.

4.1.4. Other Enzymes

Other enzymes that have been encapsulated in silica gels include ascorbate oxidase, β-glucosidase, superoxide dismutase, catalase, and several proteases.

Ascorbate Oxidase is a homodimer containing four copper ions and catalyzing the oxidation of vitamin C. The enzyme, previously immobilized in calcium alginate gels [137], has been also encapsulated in silica gels [68]. The enzyme retains the catalytic activity and exhibits a K_M value similar to that observed in solution. However, the activity of fresh gels was about 2% of the soluble enzyme and further decreased as a function of aging in the first four days. Since no control of gel thickness was carried out it is likely that the reduced activity is due to a limited diffusion of substrates within the gels.

β-Glucosidase might be used for the improvement of the aroma of wine by the degradation of naturally present glycosides. The enzyme was encapsulated in an alginate–silicate sol–gel matrix [104]. The recovery of activity in the gels was about 46%, with a K_M for the substrate methylumbelliferyl-β-D-glucose sixfold higher than that of the soluble enzyme. The inhibition by glucose and ethanol was similar in solution and in gels. Increased stability was also detected for β-galactosidase in silica gels that exhibit kinetic parameters close to those observed in solution [138].

Superoxide dismutase, a copper-zinc dimeric enzyme, encapsulated in aged silica gels and xerogels, was investigated by studying the reaction with cyanide [4]. Spectral changes at 680 and 530 nm take place as in solution, suggesting that, at least qualitatively, aged enzyme gels and xerogels retain catalytic activity. Encapsulation of the enzyme does not perturb the spectral properties that are very sensitive to metal ion geometry. Furthermore, the removal of the metal from the encapsulated enzyme was achieved, leading to an apo-enzyme with spectral properties similar to the apo-enzyme in solution. However, metal rebinding on aged hydrogels led to an enzyme characterized by spectral properties different from those of the soluble enzyme, suggesting that encapsulation stabilizes a non-native conformation of the apo-enzyme.

Catalase was encapsulated either in TMOS-derived silica gels [111] or using a colloidal silica sol to avoid the formation of alcohol during gelation [29]. The rate of oxygen formation, catalyzed by the enzyme trapped in the latter gels, was found to be about fivefold higher than that in TMOS-derived gels, but still much lower than in the soluble enzyme.

Trypsin was one of the first enzymes to be immobilized using TMOS as a precursor [43], either in the absence or presence of polyethylene glycol or albumin. Activity was about 21% of the soluble enzyme in the absence of additives and 47% in the presence of polyethylene glycol. The activity of the encapsulated enzyme was found to be affected by protein concentration, the high enzyme concentration exhibiting lower activity likely due to aggregation. However, 100% activity was found when albumin was co-immobilized with trypsin, at a ratio of albumin/trypsin of 20:1. The enzyme in silica gels retained activity for several months upon storage at room temperature, without leaching and autodigestion. This latter finding suggests that a single enzyme molecule is present in each pore of the gel. Other proteolytic enzymes encapsulated in silica gels are pronase and chymotrypsin. Enzymes were prepared and casted into thin films or incorporated into an oil-based paint formulation, retaining catalytic activity and exhibiting higher thermostability than the soluble enzyme [110]. The protease-containing silicate was resistant to fouling by nonselective protein binding of human serum albumin [139].

Acid and alkaline phosphatases were encapsulated under a variety of protocols retaining most of their activity [43, 47, 103, 140]. For alkaline phosphatase it was found that the pore size (30 Å) of mesoporous gels, obtained in the presence of D-glucose template, is greater than that obtained using the classical method (15 Å) [47, 103]. This finding might correlate with the 2–10-fold increased enzymatic activity in the former gels, as diffusion of substrates within the gel is less restricted [47]. It might also correlate with a less tight inorganic network, leading to an increased protein flexibility. The activity of fine particles of enzyme-doped gels (60–100 mesh) was measured by monitoring the absorption changes at 405 nm caused by the cleavage of the substrate analog *p*-nitrophenye phosphate to *p*-nitrophenol [102]. Product formation was linear as a function of time for at least four hours. The enzyme kinetics did not follow the Michaelis–Menten behaviour, likely due to a heterogenous population of entrapped protein molecules.

Carbonic anhydrase in silica gels catalyzes the hydrolysis of *p*-nitrophenyl acetate with an apparent K_M similar to that in solution, but the k_{cat} is only 1–2% of the value determined for the soluble enzyme, due to restricted diffusion of substrates within silica monoliths sized $8 \times 8 \times 27$ mm [62]. No attempt was made to correlate the decreased activity with the size of the gel.

Atrazine chlorohydrolase was encapsulated in silica gels using different alkyloxysilanes as precursors. A 40% activity with respect to the free enzyme was obtained even after three weeks of aging on dried crushed gels prepared from mixtures of TMOS and the more hydrophobic methyltrimethoxysilane [141].

Lipoxygenase catalyzes the selective dioxygenation of polyunsaturated fatty acids to form hydroperoxy products. The enzyme entrapped in alginate–silicate gels, ground and vacuum-dried, exhibited an activity about 30% of the free enzyme and no decrease even after 25 days of storage in glycerol [107]. A more detailed comparison indicated that the optimal temperature for the soluble enzyme is 25 °C, whereas the immobilized enzyme has a broader optimal temperature range, between 25 and 35 °C. The K_M and v_{max} values of the enzyme in the two physical states were close, within a factor of two [142].

Co-immobilization of different enzymes was used to obtain coupled reactions leading to either colorimetric detections or production of valuable compounds. Silica gels containing glucose oxidase and horseradish peroxidase were prepared for the detection of D-glucose. In aged crushed gels and xerogels the activity was about 20% and 10% of that observed for the soluble enzyme, respectively [24, 94]. A similar approach was used to detect oxalate by encapsulating oxalate oxidase and horseradish peroxidase [21, 94]. An interesting application was the encapsulation of a cell-free crude extract of *Pseudomonas sp*. in TMOS derived gels [141, 143]. The extract was able to degrade atrazine both in solution and (to a much lesser extent) in gel. Silica gels doped by the three enzymes involved in the conversion of carbon dioxide to methanol, formate dehydrogenase, formaldehyde dehydrogenase, and alcohol dehydrogenase, were prepared [144]. The striking result was that methanol formation takes place at a rate 3–10-fold higher than in solution depending on the concentration of added NADH, indicating that enzyme confinement increases the catalytic efficiency. This biorector might be used for the decrease of the greenhouse effect and the concomitant production of methanol for energy technology. Remarkably, increased catalytic efficiency was also observed upon the co-entrapment of six enzymes involved in the formation of α-(2,6)sialyl-*N*-acetyllactosamine from *N*-acetyllactosamine and *N*-acetylmannosamine, sialic acid aldolase, myokinase, pyruvate kinase, pyrophosphatase, CMP-sialate synthetase, and α(2,6)sialyl transferase [31].

4.2. Biosensors and Biomaterials

A biosensor is a device that exploits the functional properties of a biomolecule to selectively recognize and quantitate an analyte in a complex medium (see [145, 146] and references therein). The recognition process is transduced in optical or electrochemical signals to obtain an analytical

information. Therefore, a biosensor is made by a biological recognition element and a transducer. Selectivity, sensitivity, detection limit, reversibility, response time, size, ruggedness, reliability, signal recovery, and cost are features that are addressed in the development of biosensors. The biological element, usually a protein, but also a whole cell, can be immobilized in a variety of matrices [147], including silica gels [7, 12, 148, 149], tightly coupled to optical, electric, and electronic devices. The most frequently used transducers are electrochemical, such as electrodes or potentiometers, and optical ones, such as absorbance and fluorescence detectors [150–154]. For *in-situ* monitoring, fiber-optic biosensors have been demonstrated to be very useful [21, 155, 156]. Low-cost screen-printed electrodes have been developed that allow both *in-situ* and real time monitoring [157]. In these systems, an important issue is the immobilization of a very thin layer of biomolecules on the electrode surface [31]. A cutting edge field of research is the development of nanobiosensors [158] and nanomaterials [159] based on protein-doped silica gels. A branch of increasing relevance in which such devices are strongly required is protein profiling via chip array [160, 161].

The analyte to be monitored diffuses in the silica gels and interacts with the proteins, triggering signals that are dependent on its concentration, or on the concentration of co-substrates or catalytic products. The analytes for which protein-doped silica gel biosensors have been developed are listed in Table 2. For the detection of glucose, biosensors were developed based on encapsulated glucose oxidase in silica gels, either in the absence or presence of additives and other enzymes to improve biosensor performance [147]. Some of the first biosensors for glucose were based on the enzyme immobilized on the surface of an oxygen electrode [134] or on an electrode in the presence of ferrocene as a mediator [162]. A screen-printed electrode containing immobilized glucose oxidase was obtained [163, 164], as well as immobilized horseradish peroxidase [163]. The comparison of three different procedures for glucose oxidase immobilization indicated that a sandwich configuration containing a sol–gel film and a layer of the enzyme was the best system to achieve fast response and high enzyme loading [165]. Co-immobilization of glucose oxidase and horseradish peroxidase was also achieved for glucose detection, exploiting hydrogen peroxide transformation by the latter enzyme [166]. By preparing a composite gel using polyol silicate/siloxane precursors, high activity and fast response were found for a sensor for glucose using immobilized glucose oxidase [31]. Galactose oxidase, lactate oxidase, and horseradish peroxidase were immobilized on the same plate, forming a multidetecting system for galactose, lactose, and lactate [31]. The same approach gave high performance biosensors for acetylcholine/choline, based on immobilized acetylcholine esterase and choline oxidase [31]. For the detection of hydrogen peroxide, horseradish peroxidase is usually encapsulated in biosensors exploiting amperometric detection methods. Hydrogen peroxide was also detected using a chemiluminescent [167, 168] flow-through sensor based on hemoglobin silica gels [169]. For the detection of a wealth of other analytes, such as pesticides, nitrate, nitrite, and urea, specific enzymes or proteins were encapsulated in silica gels (Table 2). A detailed comparison of three different

Table 2. Analytes detected by sol–gel based biosensors.

Analyte	Biological element	Transducer	Ref.
Glucose	GOD[a]	Amp[b]/photometer	[165]
Glucose	GOD	Amp	[226]
Glucose	GOD-HRP[c]	Amp	[227]
Glucose	GOD	Toxicity tests	[228]
Glucose	GOD	Thermistor	[229]
Glucose	GOD	Amp	[230]
Glucose	GOD	Amp/optical	[231]
Glucose	GOD	chemilumiscent (Ru(dpp))	[232]
Glucose	GOD	Amp	[233]
Glucose	GOD	chemilumiscent (Ru(dpp))	[234]
Glucose	GOD	fluorescence (fluoresceine)	[235]
Glucose	GOD	Amp	[223]
Glucose	GOD	Amp	[164]
Glucose	GOD	Amp	[230]
Glucose	GOD	Amp	[236]
Glucose	GOD	Amp	[237]
Glucose	GOD	Amp	[238]
Glucose	GOD	Amp	[239]
Glucose	GOD	Amp	[166]
Glucose	GOD	Amp	[240]
Glucose	GOD	Amp	[241]
Glucose	GOD	Amp	[242]
Glucose	GOD	Amp	[163]
Glucose	GOD	Amp	[243]
Glucose	GOD	Amp	[244]
Glucose	GOD	Amp	[245]
Glucose	GOD	Amp	[246]
Glucose	GOD	Amp	[242]
Glucose	GOD	Amp	[162]
Glucose	GOD	Amp	[134]
HP[d]	Hb	chemiluminescence (luminol)	[169]
HP	HRP	Amp	[247]
HP	HRP	Amp	[248]
HP	HRP/polypyrrole	Amp	[249]
HP	Catalase	Amp	[248]
HP	HRP	Amp	[250]
HP	HRP	Amp	[251]
HP	HRP	Amp	[252]
HP	HRP	chemiluminescence	[253]
HP	HRP	chemiluminescence	[167]
HP	HRP	chemiluminescence	[168]
HP	HRP	Amp	[254]
HP	HRP	Amp	[255]
HP	Soybean peroxidase	Amp	[256]
HP	HRP	Amp	[163]
HP	HRP	Amp	[257]
HP	HRP	Amp	[258]
Peroxides	HRP	Amp	[259]
Pesticides	CHE[e]	optodes	[260]
Pesticides	CHE	optodes	[261]
Pesticides	ACHE[f]/OPH[g]	pH-sensitive field effect transistor	[262, 263]
Pesticides	CHE	fluorescence	[264]
Pesticides	ACHE	screen-printed electrodes	[170]
Acetylcholine/ paraoxon	ACHE	optodes	[156]

continued

Table 2. Continued.

Analyte	Biological element	Transducer	Ref.
Acetylthiocholine /Acetylcholine	ACHE	Amp	[265]
Thiocoline/ choline	Choline oxidase	Amp	[265]
L-Lactate	LDH[h]	optical element	[266]
Oxalate	Oxalate oxidase, peroxidase	absorbance	[94]
Oxalate	Oxalate oxidase, peroxidase	absorbance	[172]
Ethanol	ADH[i]	Amp	[267]
Phenols	HRP	chemilumines cent (luminol)	[268]
Phenols	HRP	optical element	[269]
Phenols	HRP	Amp	[270]
Phenols	Tyrosinase	Amp	[271]
Phenols	Tyrosinase	Amp	[272]
Phenols	Phenol hydroxylase	Amp	[273]
Phenols	Tyrosinase	Amp	[274]
Nitrite	Cytochrome cd_1	optical element	[275]
Nitrate	Nitrate reductase	absorbance	[276]
NO	Cytochrome *c′*	absorbance	[277]
NO	Cytochrome *c′*	fluorescence	[184]
Camphor, pyrene	CytP450cam	Amp	[278]
NO, CO	Hb[j], MnMb[k]	absorbance	[67]
NO, CO	Hb, Mb, Cyt c	absorbance	[179] [276]
O_2	Hb	absorbance	[26, 60, 85, 180]
O_2	Mb		[181]
Hypoxantine	Xantine oxidase	Amp	[279]
Hypoxantine	Xantine oxidase	Amp	[280]
Urea	Urease	conductometer	[281]
Urea	Urease	conductometer	[282]
Urea	Urease	Amp	[108]
Urea	Urease	absorbance	[98]
Urea	Urease	Amp	[283]
Urea	Urease/albumin	conductometer	[282]
Urea, glyceryl tributyrate	Urease	fluorescence	[174]
Cyanide	HRP	Amp	[284]
2,6-dimethoxy-phenol laccase		absorbance	[285]
Calcium ions	Aequorin	fluorescence	[177]
Calcium ions	Parvalbumin	fluorescence	[63]
Calcium ions	Oncomodulin	fluorescence	[64]
Iron (III)	Pyoverdin peptide	fluorescence	[178]
Proteins	Human serum albumin	fluorescence	[176]
Fluorescein	Antibody	fluorescence	[186]

[a] GOD, glucose oxidase.
[b] Amp, amperometric detection.
[c] Horseradish peroxidase, HRP.
[d] Hydrogen peroxide, HP.
[e] Choline esterase, CHE.
[f] Acetylcholine esterase, ACHE.
[g] Organophosphate hydrolase, OPH.
[h] Lactate dehydrognase, LDH.
[i] Alcohol dehydrogenase, ADH.
[j] Hemoglobin, Hb.
[k] Myoglobin, Mb.

methods of acetylcholine esterase immobilization, including the sol–gel approach, was carried out for the preparation of screen-printed electrodes designed for the detection of pesticides [170]. Of particular interest is the preparation of an array of enzymes for the detection of multiple compounds [171] and of devices with a sandwich configuration [21, 24, 94], where two coupled reactions are catalyzed within a silica gel by distinct immobilized enzymes. This latter setup increases the speed of the detection and the overall biosensor performance.

Another approach for analyte detection makes use of fluorescent signals. It was demonstrated that the activity of the enzyme glucose-6-phosphate dehydrogenase could be monitored by the change both in absorbance and fluorescence emission of the coenzyme nicotinamide adenine dinucleotide phosphate, taking place upon oxidation of D-glucose-6-phosphate [172]. The entrapping of fluorescently labeled proteins [173] or the co-entrapping of fluorescently labelled dextran was also exploited, as in the case of urease or lipase for the detection of urea and glyceryl tributyrate [174]. Enzyme activity causes pH changes that, in turn, affect fluorescent properties of the dye. A similar approach was applied to detect ionic surfactants, such as cetyltrimethylammonium bromide [175]. Bovine serum albumin, labelled with the fluorescent dye acrylodan, was encapsulated in a silanized silica optical fiber. The interaction of the surfactant with the protein structure is signalled by a change of the fluorescence emission of acrylodan. By using the fluorescence changes of iodoacetoxy-nitrobenzoxadiazole bound to bovine serum albumin, proteins were detected [176]. Furthermore, by exploiting the fluorescence changes of intrinsic probes of aequorin [177], parvalbumin [63, 93], and oncomodulin [64] silica gels, calcium ions were detected. Similarly, ground gels of pyoverdin, a natural peptide pigment, were used to detect trace amounts of iron (III) in tap water and human serum by monitoring fluorescence quenching [178].

Gaseous species, such as oxygen, carbon monoxide, and nitric oxide, were detected by measuring the absorbance or fluorescence changes of the hemoproteins cytochrome *c*, cytochrome *c′*, Mn myoglobin, and hemoglobin, encapsulated in silica gels [4, 26, 60, 67, 85, 179–184]. The interaction of carbon dioxide with myoglobin was investigated in supercritical CO_2, the condition required to obtain highly porous aerogels [185].

Antibodies encapsulated in silica gels were used to detect specific antigens. Antifluorescein antibodies retain their ability to bind the antigen in the gel, but the affinity appeared to be decreased by two orders of magnitude with respect to free antibodies [186]. Antiatrazine monoclonal antibodies in gel were able to bind between 60 and 91% of atrazine, loaded on silica sol–gel columns doped with the antibody [187]. The leaching of the antibody was negligible and the gels were stable for at least two months [188]. Similarly, antibodies against trinitrotoluene [109, 189], isoprotrun [190], and dansyl [81] were encapsulated. In the latter case, binding affinity decreases only fivefold with respect to the antidansyl antibody in solution [81].

High-throughput compound screening is a key step in the discovery of new drugs. Immobilization on the same platform of several biological targets allows one to speed up the

search and to decrease the amount of naturally or chemically synthesized compounds. The sol–gel immobilization of the complex between the calcium binding protein calmodulin and mellitin was used for the screening of mellitin antagonists [99]. The change of luminescence of Tb(III) calmodulin, originated from an energy transfer from tryptophan emission of mellitin, provided a signal to monitor the competition between mellitin and antagonists. This study evidences that protein-doped gels might be a valuable tool for protein nano- and microarray technologies [191].

Optical memories and branched three-dimensional memories are electronic devices that have made use of biomolecules. Light-addressable systems were developed on the basis of the spectroscopic properties of immobilized native and D96N bacteriorhodopsin [14, 25, 96–98, 192, 193] and phycoerythrin [115, 194]. These proteins in the gel maintain their light-sensitive properties and exhibit different absorbance bands depending on the on–off cycle of the illuminating light.

By encapsulating in silica gels the iron storage protein ferritin, magnetic glasses can be prepared [195]. It was possible to release the iron from holo-ferritin but the reverse process was hampered.

Protein-doped silica gels have potential pharmaceutical technology applications. Controlled release of drugs from microcapsules is a therapeutic goal. When the drug is a protein the goal is even more challenging. Sol–gel derived cages were used to encapsulate the pharmaceutically transforming relevant growth factor–β1 [196] and the model protein trypsin inhibitor [197]. In both cases, a slow protein release was observed over several days. Since the growth factor is involved in eliciting bone tissue reactivity, this composite biomaterial might be useful for bone regeneration. Furthermore, by using as a precursor bis[3-(trimethoxysilyl)-propyl]ethylenediamine, a gel was fabricated that leads to different intakes and releases of cytochrome *c*, myoglobin, and hemoglobin [198]. Biocomposite layers of silica, entrapping collagen, gelatine, and collagen hydrolysate, were used to coat bone implants for the stimulation of hydroxyapatite deposition and bone regeneration [199]. The highest cell adhesion and proliferation of fibroblasts on the layer were observed using acidic biocomposite sols. The coatings were found to be not toxic and to exhibit good biocompatibility.

4.3. Biochemical Investigations

The discovery that proteins can be encapsulated in silica gels retaining biological activity was exploited by researchers to deepen the understanding of the functional properties of the allosteric proteins hemoglobin and fructose-1,6-bisphosphatase. In particular, it was found that encapsulation decreases by several orders of magnitude the rate of quaternary conformational changes and allows one to fix both proteins either in the T or R state, depending on the experimental conditions [60, 73, 85, 86, 91].

4.3.1. Hemoglobin

Hemoglobin is the paradigm of allosteric proteins. This protein has been investigated by a large variety of biophysical methods in order to understand the mechanism of oxygen binding and release and allostery. The fixation of the T and R state of hemoglobin by encapsulation in TMOS-derived gels has provided surprising insights on hemoglobin function and regulation. Trasmission electron microscopy of silica gels prepared in the absence and presence of T state Hb (Fig. 2) indicates that the pore/channels are about 3 and 4 nm, respectively. As hemoglobin diameter is about 6 nm, the protein cannot leach. The T and R states of hemoglobin were first characterized by measuring oxygen binding curves [26, 60, 85, 87, 89]. The oxygen binding curves of both the T and R states were characterized by a Hill coefficient very close to unity, indicating that the ligand binds noncooperatively, in agreement with the prediction of the Monod et al. model [200]. By varying the procedure of encapsulation, it was possible to isolate two distinct tertiary conformations of the T state of hemoglobin that exhibit a 10-fold difference in oxygen affinity [26]. The oxygen affinities of the two states closely correspond to those observed in solution for the binding of the first oxygen to T state hemoglobin under selected conditions, indicating that even tertiary transitions are slowed down in silica gels without altering ground state properties. The kinetics of carbon monoxide rebinding to R and T state hemoglobin gels was investigated (Fig. 4) [88, 90]. Carbon monoxide bound to the heme iron was laser-flash photolyzed and the time course of ligand rebinding was monitored by a nanosecond laser spectrometer. The T and R states exhibited a geminate phase and bimolecular phase similar to those observed in solution. When the kinetics was monitored on the high and low affinity T states striking differences were observed, indicating that encapsulation in silica gels strongly decreases the rate of not only quaternary changes but also tertiary transitions. This feature might be exploited to investigate the structure–dynamics–function relationships of complex conformational distributions. Functional information was correlated to the structural and dynamic characterization of T and R states hemoglobin gels, carrying out resonance Raman studies [72, 73] (see Sections 3.1.1 and 3.1.2). Hemoglobin gels exhibit the same marker bands as in solution. In particular, a band associated with the heme iron–proximal histidine stretching mode, diagnostic of the tertiary–quaternary

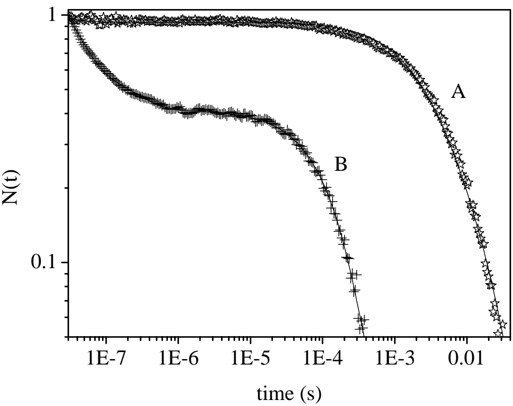


Figure 4. Transient absorption at 488 nm following 532 nm photolysis of T state (A) and R state (B) hemoglobin, encapsulated in silica gels, at 15 °C. Reprinted with permission from [89], S. Abbruzzetti et al., *J. Nanosci. Nanotech.* 1, 407 (2001). © 2001, American Scientific Publishers.

state of hemoglobin, was used to monitor the conformational transitions as a function of time upon exposure of T state hemoglobin gels to carbon monoxide.

4.3.2. Fructose-1-6-bisphosphatase

In the case of fructose-1,6-bisphosphatase, the enzyme was encapsulated either in the T or R state in the presence and absence of the allosteric effector AMP [91]. The k_{cat} of the R state was 4-fold less than that of the enzyme in solution; the K_M was 350-fold higher. Interestingly, the cooperative binding of magnesium ions, an enzyme activator, was observed at lower concentrations for the encapsulated enzyme than in solution. As for hemoglobin ligand-induced conformational changes, binding of AMP triggers the R to T quaternary transition at rates much smaller than in solution. The inactivation process is biphasic, suggesting a direct effect of the allosteric effector on the enzyme active site, followed by the quaternary switch.

4.3.3. Glutamate Dehydrogenase

The allosteric enzyme glutamate dehydrogenase encapsulated in TMOS gels was active, as shown by recording spectrophotometrically the formation of the coenzyme NADH [92]. The enzyme exhibited a 40-fold decrease in v_{max} with respect to the soluble enzyme, whereas the K_M was very similar. Binding of the positive and negative allosteric effectors GTP and ADP, respectively, parallels the effects observed in solution on the catalytic activity. The low activity was explained by a limited diffusion of substrates in the gel. However, it is also possible that the encapsulation conditions have entrapped a different conformation of the enzyme with respect to that predominant in solution.

4.3.4. Pyridoxal 5′-Phosphate-Dependent Enzymes

The catalytic activity of several proteins is associated with open-closed conformations of flexible lids or domains, covering or forming the active site. A class of enzymes in which these conformational changes have been crystallographically characterized is represented by pyridoxal 5′-phosphate-dependent enzymes. Silica gel protein encapsulation is a valuable method to isolate conformations or to dramatically decrease the rate of tertiary transitions. Taking advantage of the chromophoric properties of the bound coenzyme, investigations on the effect of silica gel encapsulation on the enzyme activity and the distribution of tertiary conformations of tryptophan synthase, *O*-acetylserine sulfhydrilase [201], tryptophanase, and tyrosine phenol lyase are being carried out in our laboratory.

5. CONCLUSIONS

Composite biomaterials are generated via either adsorption, covalent attachment of biomolecules to activated surfaces or encapsulation in organic, silicate, or organo-silicate matrices. The latter approach has led to the development of biorectors, biosensors, and biocoatings that are successfully used in industrial catalysis, the food industry, environmental protection, and biomedical diagnosis. The present technological challenge is to expand the use of protein-doped nanoporous gels for advanced electronic devices, and for nanodevices containing protein molecules arrays, able to simultaneously detect a large variety of analytes.

GLOSSARY

Absorbance The absorption of light by a substance, calculated as the $-\log(I/I_O)$, where I and I_O are the intensities of the light before and after passing through the sample, respectively. It is proportional to the concentration of the substance and the optical pathlength.

ANS 8-Anilino-1-naphthalenesulphonic acid, an apolar dye characterized by an enhanced fluorescence upon binding to hydrophobic patches of proteins. It is often used to detect unfolding intermediates.

Bioreactor A device in which one or more enzymes are physically confined, allowing one to obtain the selective and efficient transformation of substrates in products, and an easy separation between catalysts and reagents.

Biosensor A device exploiting the functional properties of a biomolecule for the selective recognition and quantitation of an analyte in a complex medium.

Circular dichroism Differential absorption of left and right circularly polarized light by optically active molecules.

Fluorescence The emission of light by a substance immediately after the absorption of energy from light of shorter wavelength.

Fluorescence anisotropy The measurement of the average angular displacement of the fluorophore which occurs between absorption and the subsequent emission of light by fluorescence. The angular displacement depends on the rate and extent of the rotational diffusion.

Protein A biological macromolecule formed by a sequence of aminoacids. Proteins perform a variety of functions, most of which require the recognition of other molecules.

Raman emission Discrete bands originating from molecular normal mode of vibrations, serving as a specific fingerprint of three-dimensional structure, intermolecular interactions, and conformations.

R- and T-state “Relaxed” and “tense” quaternary conformations of an allosteric protein. For hemoglobin R and T conformations are endowed with high and low oxygen affinity, and for enzymes with high and low catalytic efficiency, respectively.

Sol–gel process A versatile solution process for making ceramic and glass materials, involving the transition of a system from a liquid “sol” phase (mostly colloidal) into a solid “gel” phase.

ACKNOWLEDGMENTS

This work was supported by grants from the Italian Ministry of Instruction, University, and Research (COFIN 2002-2003 to A.M.) and the Italian Institute for the Physics of Matter. Support was also provided by NIH grant 1 P01GM58890 from the U.S. Public Health Service.

REFERENCES

1. E. Dujardin and S. Mann, *Adv. Mater.* 14, 775 (2002).
2. D. Avnir, S. Braun, and M. Ottolenghi, *ACS Symp. Ser.* 499, 384 (1992).
3. D. Avnir, S. Braun, O. Lev, and M. Ottolenghi, *Chem. Mater.* 6, 1605 (1994).
4. L. M. Ellerby, C. R. Nishida, F. Nishida, S. A. Yamanaka, B. Dunn, J. S. Valentine, and J. I. Zink, *Science* 255, 1113 (1992).
5. B. C. Dave, B. Dunn, J. S. Valentine, and J. I. Zink, in "Nanotechnology" (G.-M. Chow and K. E. Gonsalves, Eds.), p. 351. American Chemical Society, Washington, DC, 1996.
6. I. Gill and A. Ballesteros, *Trends Biotechnol.* 18, 282 (2000).
7. I. Gill, *Chem. Mater.* 13, 3404 (2001).
8. J. Livage, T. Coradin, and C. Roux, *J. Phys. Condens. Mat.* 13, R673 (2001).
9. H. Bottcher, *J. Prakt. Chem.* 342, 427 (2000).
10. A. Mozzarelli and S. Bettati, in "Advanced Functional Molecules and Polymers" (H. S. Nalwa, Ed.), Vol. 4, p. 55. Gordon and Breach, Singapore, 2001.
11. J. Lin and C. W. Brown, *Trend. Anal. Chem.* 16, 200 (1997).
12. W. Jin and J. D. Brennan, *Anal. Chim. Acta* 461, 1 (2002).
13. E. T. Vandenberg, R. S. Brown, and U. J. Krull, "Immobilized Biosystems in Theory and Practical Applications." Elsevier, Amsterdam, 1983.
14. H. H. Weetall, B. Robertson, D. Cullin, J. Brown, and M. Walch, *Biochim. Biophys. Acta* 1142, 211 (1993).
15. L. Doretti, D. Ferrara, and S. Lora, *Biosens. Bioelectron.* 8, 443 (1993).
16. K. F. O'Driscoll, *Method Enzymol.* 44, 169 (1976).
17. W. H. Scouten, *Method Enzymol.* 135, 30 (1987).
18. P. Johnson and T. L. Whateley, *J. Colloid Interf. Sci.* 37, 557 (1971).
19. C. J. Brinker and G. W. Scherer, "Sol–Gel Science: The Physics and Chemistry of Sol–Gel Processing." Academic Press, Boston, 1990.
20. L. C. Klein, *Annu. Rev. Mater. Res.* 15, 227 (1985).
21. B. C. Dave, B. Dunn, J. S. Valentine, and J. I. Zink, *Anal. Chem.* 66, A1120 (1994).
22. C. Y. Shen and N. M. Kostic, *J. Am. Chem. Soc.* 119, 1304 (1997).
23. B. Dunn and J. I. Zink, *J. Mater. Chem.* 1, 903 (1991).
24. S. A. Yamanaka, F. Nishida, L. M. Ellerby, C. R. Nishida, B. Dunn, J. S. Valentine, and J. I. Zink, *Chem. Mater.* 4, 495 (1992).
25. S. G. Wu, L. M. Ellerby, J. S. Cohan, B. Dunn, M. A. Elsayed, J. S. Valentine, and J. I. Zink, *Chem. Mater.* 5, 115 (1993).
26. S. Bruno, M. Bonaccio, S. Bettati, C. Rivetti, C. Viappiani, S. Abbruzzetti, and A. Mozzarelli, *Protein Sci.* 10, 2401 (2001).
27. B. Dunn, J. M. Miller, B. C. Dave, J. S. Valentine, and J. I. Zink, *Acta Mater.* 46, 737 (1998).
28. J. D. Brennan, *Appl. Spectrosc.* 53, 106A (1999).
29. D. M. Liu and I. W. Chen, *Acta Mater.* 47, 4535 (1999).
30. R. B. Bhatia, C. J. Brinker, A. K. Gupta, and A. K. Singh, *Chem. Mater.* 12, 2434 (2000).
31. I. Gill and A. Ballesteros, *J. Am. Chem. Soc.* 120, 8587 (1998).
32. L. L. Hench and J. K. West, *Chem. Rev.* 90, 33 (1990).
33. J. E. Mark, *Heterogen. Chem. Rev.* 3, 307 (1996).
34. D. Avnir, *Acc. Chem. Res.* 28, 328 (1995).
35. N. Husing and U. Schubert, *Angew. Chem. Int. Ed.* 37, 23 (1998).
36. R. J. Hook, *J. Non-Cryst. Solids* 195, 1 (1996).
37. J. J. Vanbeek, D. Seykens, and J. B. H. Jansen, *J. Non-Cryst. Solids* 146, 111 (1992).
38. T. Keeling-Tucker and J. D. Brennan, *Chem. Mater.* 13, 3331 (2001).
39. U. Schubert, N. Husing, and A. Lorenz, *Chem. Mater.* 7, 2010 (1995).
40. D. Avnir, L. C. Klein, D. Levy, U. Schubert, A. B. Wojcik, Z. Rappoport, and Y. Apeloig, "The Chemistry of Organic Compounds." New York, 1998.
41. R. D. Schmid and R. Verger, *Angew. Chem. Int. Ed.* 37, 1609 (1998).
42. L. T. Zhuravlev, *Colloid Surface A* 173, 1 (2000).
43. S. Shtelzer, S. Rappoport, D. Avnir, M. Ottolenghi, and S. Braun, *Biotechnol. Appl. Bioc.* 15, 227 (1992).
44. D. L. Meixner and P. N. Dyer, *J. Sol-Gel Sci. Techn.* 14, 223 (1999).
45. C. L. Wambolt and S. S. Saavedra, *J. Sol-Gel Sci. Techn.* 7, 53 (1996).
46. G. A. Baker, J. D. Jordan, and F. V. Bright, *J. Sol-Gel Sci. Techn.* 11, 43 (1998).
47. Y. Wei, J. G. Xu, Q. W. Feng, M. D. Lin, H. Dong, W. J. Zhang, and C. Wang, *J. Nanosci. Nanotech.* 1, 83 (2001).
48. Q. Chen, G. L. Kenausis, and A. Heller, *J. Am. Chem. Soc.* 120, 4582 (1998).
49. J. Heller and A. Heller, *J. Am. Chem. Soc.* 120, 4586 (1998).
50. J. D. Badjic and N. M. Kostic, *J. Mater. Chem.* 11, 408 (2001).
51. K. Martinek and V. V. Mozhaev, *Adv. Enzymol. RAMB* 57, 179 (1985).
52. B. Dunn and J. I. Zink, *Chem. Mater.* 9, 2280 (1997).
53. V. R. Kaufman, D. Avnir, D. Pines Rojanski, and D. Huppert, *J. Non-Cryst. Solids* 99, 2 (1988).
54. V. R. Kaufman and D. Avnir, *Langmuir* 2, 717 (1986).
55. K. Matsui and T. Nakazawa, *B. Chem. Soc. Jpn.* 63, 11 (1990).
56. K. Matsui, T. Nakazawa, and H. Morisaki, *J. Phys. Chem.* 95, 976 (1991).
57. R. Chambers, Y. Haruvy, and M. A. Fox, *Chem. Mater.* 6, 1351 (1994).
58. J. E. Lee and S. S. Saavedra, *Anal. Chim. Acta* 285, 265 (1994).
59. G. Chirico, F. Cannone, S. Beretta, A. Diaspro, B. Campanini, S. Bettati, R. Ruotolo, and A. Mozzarelli, *Protein Sci.* 11, 1152 (2002).
60. S. Bettati and A. Mozzarelli, *J. Biol. Chem.* 272, 32050 (1997).
61. C. Y. Shen and N. M. Kostic, *J. Electroanal. Chem.* 438, 61 (1997).
62. J. D. Badjic and N. M. Kostic, *Chem. Mater.* 11, 3671 (1999).
63. K. Flora and J. D. Brennan, *Anal. Chem.* 70, 4505 (1998).
64. L. L. Zheng, K. Flora, and J. D. Brennan, *Chem. Mater.* 10, 3974 (1998).
65. P. L. Edmiston, C. L. Wambolt, M. K. Smith, and S. S. Saavedra, *J. Colloid Interf. Sci.* 163, 395 (1994).
66. C. T. Lin, C. M. Catuara, J. E. Erman, K. C. Chen, S. F. Huang, W. J. Wang, and H. H. Wei, *J. Sol-Gel Sci. Techn.* 7, 19 (1996).
67. E. H. Lan, B. C. Dave, J. M. Fukuto, B. Dunn, J. I. Zink, and J. S. Valentine, *J. Mater. Chem.* 9, 45 (1999).
68. I. Savini, R. Santucci, A. Di Venere, N. Rosato, G. Strukul, F. Pinna, and L. Avigliano, *Appl. Biochem. Biotech.* 82, 227 (1999).
69. D. K. Eggers and J. S. Valentine, *Protein Sci.* 10, 250 (2001).
70. J. D. Jordan, R. A. Dunbar, and F. V. Bright, *Anal. Chem.* 67, 2436 (1995).
71. K. K. Flora and J. D. Brennan, *Chem. Mater.* 13, 4170 (2001).
72. L. J. Juszczak and J. M. Friedman, *J. Biol. Chem.* 274, 30357 (1999).
73. T. K. Das, I. Khan, D. L. Rousseau, and J. M. Friedman, *Biospectroscopy* 5, S64 (1999).
74. I. Khan, D. Dantsker, U. Samuni, A. J. Friedman, C. Bonaventura, B. Manjula, S. A. Acharya, and J. M. Friedman, *Biochemistry* 40, 7581 (2001).
75. T. K. Das, I. Khan, D. L. Rousseau, and J. M. Friedman, *J. Am. Chem. Soc.* 120, 10268 (1998).
76. U. Samuni, D. Dantsker, I. Khan, A. J. Friedman, E. Peterson, and J. M. Friedman, *J. Biol. Chem.* 277, 25783 (2002).
77. B. C. Dave, H. Soyez, J. M. Miller, B. Dunn, J. S. Valentine, and J. I. Zink, *Chem. Mater.* 7, 1431 (1995).
78. B. C. Dave, J. M. Miller, B. Dunn, J. S. Valentine, and J. I. Zink, *J. Sol-Gel Sci. Techn.* 8, 629 (1997).
79. J. M. Nocek, S. L. Hatch, J. L. Seifert, G. W. Hunter, D. D. Thomas, and B. M. Hoffman, *J. Am. Chem. Soc.* 124, 9404 (2002).

80. A. M. Hartnett, C. M. Ingersoll, G. A. Baker, and F. V. Bright, *Anal. Chem.* 71, 1215 (1999).
81. M. A. Doody, G. A. Baker, S. Pandey, and F. V. Bright, *Chem. Mater.* 12, 1142 (2000).
82. L. L. Zheng, W. R. Reid, and J. D. Brennan, *Anal. Chem.* 69, 3940 (1997).
83. L. Zheng and J. D. Brennan, *Analyst* 123, 1735 (1998).
84. D. S. Gottfried, A. Kagan, B. M. Hoffman, and J. M. Friedman, *J. Phys. Chem. B* 103, 2803 (1999).
85. N. Shibayama and S. Saigo, *J. Mol. Biol.* 251, 203 (1995).
86. N. Shibayama and S. Saigo, *J. Am. Chem. Soc.* 121, 444 (1999).
87. N. Shibayama, *J. Mol. Biol.* 285, 1383 (1999).
88. I. Khan, C. F. Shannon, D. Dantsker, A. J. Friedman, J. Perez-Gonzalez-de-Apodaca, and J. M. Friedman, *Biochemistry* 39, 16099 (2000).
89. S. Abbruzzetti, C. Viappiani, S. Bruno, S. Bettati, M. Bonaccio, and A. Mozzarelli, *J. Nanosci. Nanotechnol.* 1, 407 (2001).
90. S. Abbruzzetti, C. Viappiani, S. Bruno, and A. Mozzarelli, *Chem. Phys. Lett.* 346, 430 (2001).
91. J. K. McIninch and E. R. Kantrowitz, *BBA Biochim. Biophys. Acta Protein Struct. M* 1547, 320 (2001).
92. N. Husing, E. Reisler, and J. I. Zink, *J. Sol-Gel Sci. Techn.* 15, 57 (1999).
93. J. D. Brennan, K. K. Flora, G. N. Bendiak, G. A. Baker, M. A. Kane, S. Pandey, and F. V. Bright, *J. Phys. Chem. B* 104, 10100 (2000).
94. J. I. Zink, J. S. Valentine, and B. Dunn, *New J. Chem.* 18, 1109 (1994).
95. J. A. He, L. Samuelson, L. Li, J. Kumar, and S. K. Tripathy, *Adv. Mater.* 11, 435 (1999).
96. H. H. Weetall, *Appl. Biochem. Biotech.* 49, 241 (1994).
97. H. H. Weetall, *Biosens. Bioelectron.* 11, 327 (1996).
98. P. C. Pandey, S. Singh, B. Upadhyay, H. H. Weetall, and P. K. Chen, *Sensor. Actuat. B* 36, 470 (1996).
99. K. Flora, T. Keeling-Tucker, C. Hogue, and J. Brennan, *Anal. Chim. Acta* (2002).
100. H. X. Zhou and K. A. Dill, *Biochemistry* 40, 11289 (2001).
101. U. Samuni, M. S. Navati, L. J. Juszczak, D. Dantsker, M. Yang, and J. M. Friedman, *J. Phys. Chem. B* 104, 10802 (2000).
102. S. Braun, S. Rappoport, R. Zusman, D. Avnir, and M. Ottolenghi, *Mater. Lett.* 10, 1 (1990).
103. Y. Wei, J. G. Xu, Q. W. Feng, H. Dong, and M. D. Lin, *Mater. Lett.* 44, 6 (2000).
104. O. Heichalsegal, S. Rappoport, and S. Braun, *Bio-Technol.* 13, 798 (1995).
105. K. Kawakami and S. Yoshida, *J. Ferment. Bioeng.* 82, 239 (1996).
106. M. T. Reetz and K. E. Jaeger, *Chem. Phys. Lipids* 93, 3 (1998).
107. A. F. Hsu, T. A. Foglia, and G. J. Piazza, *Biotechnol. Lett.* 19, 71 (1997).
108. K. Ogura, K. Nakaoka, M. Nakayama, M. Kobayashi, and A. Fujii, *Anal. Chim. Acta* 384, 219 (1999).
109. E. H. Lan, B. Dunn, and J. I. Zink, *Chem. Mater.* 12, 1874 (2000).
110. Y. Kim, J. Dordick, and D. Clark, *Biotechnol. Bioeng.* 72, 475 (2001).
111. J. M. Miller, B. Dunn, J. S. Valentine, and J. I. Zink, *J. Non-Cryst. Solids* 202, 279 (1996).
112. M. T. Reetz, A. Zonta, and J. Simpelkamp, *Biotechnol. Bioeng.* 49, 527 (1996).
113. A. K. Williams and J. T. Hupp, *J. Am. Chem. Soc.* 120, 4366 (1998).
114. C. R. Lloyd and E. M. Eyring, *Langmuir* 16, 9092 (2000).
115. Z. P. Chen, L. A. Samuelson, J. Akkara, D. L. Kaplan, H. Gao, J. Kumar, K. A. Marx, and S. K. Tripathy, *Chem. Mater.* 7, 1779 (1995).
116. A. P. Minton, *Biophys. J.* 78, 101 (2000).
117. F. H. Dickey, *J. Phys. Chem.* 58, 695 (1955).
118. D. Venton, K. Cheesman, R. Chatterton, and T. Anderson, *Biochim. Biophys. Acta* 797, 343 (1984).
119. M. Glad, O. Norrlow, B. Sellergren, N. Siegbahn, and K. Mosbach, *J. Chromatogr. A* 347, 11 (1985).
120. W. Tischer and V. Kasche, *Trends Biotechnol.* 17, 326 (1999).
121. M. Makinen and A. Fink, *Annu. Rev. Biophys. Bioeng.* 6, 301 (1977).
122. A. Mozzarelli and G. L. Rossi, *Annu. Rev. Biophys. Biom.* 25, 343 (1996).
123. B. Rubin and E. A. Dennis, *Methods Enzymol.* 284 (1997).
124. K. E. Jaeger and M. T. Reetz, *Trends Biotechnol.* 16, 396 (1998).
125. U. Bornscheuer, C. Bessler, R. Srinivas, and S. Krishna, *Trends Biotechnol.* 20, 433 (2002).
126. M. T. Reetz, A. Zonta, and J. Simpelkamp, *Angew. Chem. Int. Ed.* 34, 301 (1995).
127. T. Keeling-Tucker, M. Rakic, C. Spong, and J. D. Brennan, *Chem. Mater.* 12, 3695 (2000).
128. J. D. Badjic, E. N. Kadnikova, and N. M. Kostic, *Org. Lett.* 3, 2025 (2001).
129. I. Gill, E. Pastor, and A. Ballesteros, *J. Am. Chem. Soc.* 121, 9487 (1999).
130. T. Antczak, J. MrowiecBialon, S. Bielecki, A. B. Jarzebski, J. J. Malinowski, A. I. Lachowski, and E. Galas, *Biotechnol. Tech.* 11, 9 (1997).
131. P. Buisson, C. Hernandez, M. Pierre, and A. C. Pierre, *J. Non-Cryst. Solids* 285, 295 (2001).
132. A. Pierre and P. Buisson, *J. Mol. Catal. B* 11, 639 (2001).
133. Q. Ji, C. R. Lloyd, B. C. Dunn, and E. M. Eyring, *Instrum. Sci. Technol.* 27, 23 (1999).
134. Y. Tatsu, K. Yamashita, M. Yamaguchi, S. Yamamura, H. Yamamoto, and S. Yoshikawa, *Chem. Lett.* 8, 1615 (1992).
135. H. Graebner, U. Georgi, R. Huttl, and G. Wolf, *Thermochim. Acta* 310, 101 (1998).
136. E. N. Kadnikova and N. M. Kostic, *J. Non-Cryst. Solids* 283, 63 (2001).
137. M. Esaka, K. Suzuki, and K. Kubota, *Agric. Biol. Chem.* 49, 2955 (1985).
138. O. Ariga, T. Suzuki, Y. Sano, and Y. Murakami, *J. Ferment. Bioeng.* 82, 341 (1996).
139. J. Kim, R. Delio, and J. S. Dordick, *Biotechnol. Progr.* 18, 551 (2002).
140. S. Braun, S. Shtelzer, S. Rappoport, D. Avnir, and M. Ottolenghi, *J. Non-Cryst. Solids* 147, 739 (1992).
141. C. Kauffmann and R. T. Mandelbaum, *J. Biotechnol.* 62, 169 (1998).
142. A. F. Hsu, E. Wu, T. A. Foglia, and G. J. Piazza, *J. Food Biochem.* 24, 21 (2000).
143. C. G. Kauffmann and R. T. Mandelbaum, *J. Biotechnol.* 51, 219 (1996).
144. R. Obert and B. C. Dave, *J. Am. Chem. Soc.* 121, 12192 (1999).
145. H. A. Fishman, D. R. Greenwald, and R. N. Zare, *Annu. Rev. Biophys. Biom.* 27, 165 (1998).
146. F. W. Scheller, U. Wollenberger, A. Warsinke, and F. Lisdat, *Curr. Opin. Biotech.* 12, 35 (2001).
147. J. F. Liang, Y. T. Li, and V. C. Yang, *J. Pharm. Sci.* 89, 979 (2000).
148. B. D. MacCraith, C. McDonagh, A. K. McEvoy, T. Butler, G. Okeeffe, and V. Murphy, *J. Sol-Gel Sci. Techn.* 8, 1053 (1997).
149. Y. Tang and B. C. Dave, *Adv. Mater.* 10, 1536 (1998).
150. G. Gilardi and A. Fantuzzi, *Trends Biotechnol.* 19, 468 (2001).
151. I. Willner and B. Willner, *Trends Biotechnol.* 19, 222 (2001).
152. S. D'Auria and J. R. Lakowicz, *Curr. Opin. Biotech.* 12, 99 (2001).
153. J. Wang, *Anal. Chim. Acta* 399, 21 (1999).
154. A. Walcarius, *Electroanalysis* 10, 1217 (1998).
155. J. Lin, *Trac-Trend Anal. Chem.* 19, 541 (2000).
156. R. A. Doong and H. C. Tsai, *Anal. Chim. Acta* 434, 239 (2001).
157. C. C. Liu and Z. H. Jin, *Trends Biotechnol.* 15, 213 (1997).
158. A. Curtis and C. Wilkinson, *Trends Biotechnol.* 19, 97 (2001).
159. A. Huczko, *Appl. Phys. A* 70, 365 (2000).

160. H. H. Zhou, S. Roy, H. Schulman, and M. J. Natan, *Trends Biotechnol.* 19, S34 (2001).
161. K. H. Lee, *Trends Biotechnol.* 19, 217 (2001).
162. P. Audebert, C. Demaille, and C. Sanchez, *Chem. Mater.* 5, 911 (1993).
163. J. Wang, P. V. A. Pamidi, and D. S. Park, *Anal. Chem.* 68, 2705 (1996).
164. Y. Z. Guo and A. R. Guadalupe, *Sensor. Actuat. B* 46, 213 (1998).
165. U. Narang, P. N. Prasad, F. V. Bright, K. Ramanathan, N. D. Kumar, B. D. Malhotra, M. N. Kamalasanan, and S. Chandra, *Anal. Chem.* 66, 3139 (1994).
166. L. CocheGuerente, S. Cosnier, and L. Labbe, *Chem. Mater.* 9, 1348 (1997).
167. A. N. Diaz, M. C. R. Peinado, and M. C. T. Minguez, *Anal. Chim. Acta* 363, 221 (1998).
168. J. Li, K. M. Wang, X. H. Yang, and D. Xiao, *Anal. Commun.* 36, 195 (1999).
169. B. X. Li, Z. J. Zhang, and L. X. Zhao, *Anal. Chim. Acta* 445, 161 (2001).
170. S. Andreescu, L. Barthelmebs, and J. L. Marty, *Anal. Chim. Acta* 464, 171 (2002).
171. C. B. Park and D. S. Clark, *Biotechnol. Bioeng.* 78, 229 (2002).
172. S. A. Yamanaka, B. Dunn, J. S. Valentine, and J. I. Zink, *J. Am. Chem. Soc.* 117, 9095 (1995).
173. J. D. Brennan, *J. Fluoresc.* 9, 295 (1999).
174. M. D. Gulcev, G. L. G. Goring, M. Rakic, and J. D. Brennan, *Anal. Chim. Acta* 457, 47 (2002).
175. J. S. Lundgren and F. V. Bright, *Anal. Chem.* 68, 3377 (1996).
176. K. Flora and J. D. Brennan, *Analyst* 124, 1455 (1999).
177. D. J. Blyth, S. J. Poynter, and D. A. Russell, *Analyst* 121, 1975 (1996).
178. J. M. Barrero, C. Camara, M. C. Perezconde, C. Sanjose, and L. Fernandez, *Analyst* 120, 431 (1995).
179. D. J. Blyth, J. W. Aylott, D. J. Richardson, and D. A. Russell, *Analyst* 120, 2725 (1995).
180. N. Shibayama and S. Saigo, *FEBS Lett.* 492, 50 (2001).
181. K. E. Chung, E. H. Lan, M. S. Davidson, B. S. Dunn, J. S. Valentine, and J. I. Zink, *Anal. Chem.* 67, 1505 (1995).
182. M. F. McCurley, G. J. Bayer, and S. A. Glazier, *Sensor. Actuat. B* 36, 491 (1996).
183. J. W. Aylott, D. J. Richardson, and D. A. Russell, *Chem. Mater.* 9, 2261 (1997).
184. S. L. R. Barker, R. Kopelman, T. E. Meyer, and M. A. Cusanovich, *Anal. Chem.* 70, 971 (1998).
185. Q. Ji, C. R. Lloyd, W. R. Ellis, and E. M. Eyring, *J. Am. Chem. Soc.* 120, 221 (1998).
186. R. Wang, U. Narang, P. N. Prasad, and F. V. Bright, *Anal. Chem.* 65, 2671 (1993).
187. A. Turniansky, D. Avnir, A. Bronshtein, N. Aharonson, and M. Altstein, *J. Sol-Gel Sci. Techn.* 7, 135 (1996).
188. A. Bronshtein, N. Aharonson, D. Avnir, A. Turniansky, and M. Altstein, *Chem. Mater.* 9, 2632 (1997).
189. M. Altstein, A. Bronshtein, B. Glattstein, A. Zeichner, T. Tamiri, and J. Almog, *Anal. Chem.* 73, 2461 (2001).
190. P. Pulido-Tofino, J. M. Barrero-Moreno, and M. C. Perez-Conde, *Anal. Chim. Acta* 429, 337 (2001).
191. G. Walter, K. Bussow, D. Cahill, A. Lueking, and H. Lehrach, *Curr. Opin. Microbiol.* 3, 298 (2000).
192. L. M. Shamansky, K. M. Luong, D. Han, and E. L. Chronister, *Biosens. Bioelectron.* 17, 227 (2002).
193. K. J. Wise, N. B. Gillespie, J. A. Stuart, M. P. Krebs, and R. R. Birge, *Trends Biotechnol.* 20, 387 (2002).
194. Z. P. Chen, D. L. Kaplan, K. Yang, J. Kumar, K. A. Marx, and S. K. Tripathy, *J. Sol-Gel Sci. Techn.* 7, 99 (1996).
195. E. H. Lan, B. Dunn, J. S. Valentine, and J. I. Zink, *J. Sol-Gel Sci. Techn.* 7, 109 (1996).
196. S. B. Nicoll, S. Radin, E. M. Santos, R. S. Tuan, and P. Ducheyne, *Biomaterials* 18, 853 (1997).
197. E. M. Santos, S. Radin, and P. Ducheyne, *Biomaterials* 20, 1695 (1999).
198. M. S. Rao and B. C. Dave, *J. Am. Chem. Soc.* 120, 13270 (1998).
199. I. Brasack, H. Bottcher, and U. Hempel, *J. Sol-Gel Sci. Techn.* 19, 479 (2000).
200. J. Monod, J. Wyman, and J. Changeux, *J. Mol. Biol.* 12, 88 (1965).
201. A. Mozzarelli, B. Campanini, S. Bettati, and A. Peracchi, in "Biochemistry and Molecular Biology of Vitamin B6 and PQQ-Dependent Proteins" (A. Iriarte, H. Kagan, and M. Martinez-Carrion, Eds.), p. 349. Birkhauser-Verlag, Basel, 2000.
202. J. P. Chen, W. S. Lin, and M. F. Chang, *J. Am. Oil Chem. Soc.* 79, 309 (2002).
203. Y. Ikeda, R. Nozaki, and Y. Kurokawa, *J. Chem. Technol. Biot.* 77, 86 (2002).
204. Y. Ikeda and Y. Kurokawa, *J. Am. Oil Chem. Soc.* 78, 1099 (2001).
205. A. F. Hsu, K. Jones, W. N. Marmer, and T. A. Foglia, *J. Am. Oil Chem. Soc.* 78, 585 (2001).
206. Y. Ikeda and Y. Kurokawa, *J. Sol-Gel Sci. Techn.* 21, 221 (2001).
207. M. Pierre, P. Buisson, F. Fache, and A. Pierre, *Biocatal. Biotransfor.* 18, 237 (2000).
208. M. T. Reetz, A. Zonta, V. Vijayakrishnan, and K. Schimossek, *J. Mol. Catal. A* 134, 251 (1998).
209. M. T. Reetz, *Adv. Mater.* 9, 943 (1997).
210. M. Heidt, U. Bornscheuer, and R. D. Schmid, *Biotechnol. Tech.* 10, 25 (1996).
211. F. A. ElEssi, A. Z. AbuZuhri, S. I. AlKhalil, and M. S. AbdelLatif, *Talanta* 44, 2051 (1997).
212. D. J. van Unen, J. F. Engbersen, and D. N. Reinhoudt, *Biotechnol. Bioeng.* 75, 154 (2001).
213. B. Lillis, C. Grogan, H. Berney, and W. A. Lane, *Sensor. Actuat. B* 68, 109 (2000).
214. U. Narang, P. N. Prasad, F. V. Bright, A. Kumar, N. D. Kumar, B. D. Malhotra, M. N. Kamalasanan, and S. Chandra, *Chem. Mater.* 6, 1596 (1994).
215. A. Ulatowska and H. Podbielska, *Opt. Appl.* 30, 193 (2000).
216. A. Kumar, R. Malhotra, B. D. Malhotra, and S. K. Grover, *Anal. Chim. Acta* 414, 43 (2000).
217. U. Georgi, H. Graebner, G. Roewer, and G. Wolf, *J. Sol-Gel Sci. Techn.* 13, 295 (1998).
218. K. Ramanathan, M. N. Kamalasanan, D. B. Malhotra, D. R. Pradhan, and S. Chandra, *J. Sol-Gel Sci. Techn.* 10, 309 (1997).
219. J. P. Chen and H. Y. Wang, *Biotechnol. Tech.* 12, 851 (1998).
220. F. Akbarian, A. Lin, B. S. Dunn, J. S. Valentin, and J. I. Zink, *J. Sol-Gel Sci. Techn.* 8, 1067 (1997).
221. M. Altstein, G. Segev, N. Aharonson, O. Ben-Aziz, A. Turniansky, and D. Avnir, *J. Agr. Food Chem.* 46, 3318 (1998).
222. G. J. Cho, I. S. Moon, and J. S. Lee, *Chem. Lett.* 6, 577 (1997).
223. Z. J. Liu, B. H. Liu, M. Zhang, J. L. Kong, and J. Q. Deng, *Anal. Chim. Acta* 392, 135 (1999).
224. C. Dosoretz, R. Armon, J. Starosvetzky, and N. Rothschild, *J. Sol-Gel Sci. Techn.* 7, 7 (1996).
225. S. G. Wu, J. Lin, and S. I. Chan, *Appl. Biochem. Biotech.* 47, 11 (1994).
226. J. J. Niu and J. Y. Lee, *Sensor. Actuat. B* 82, 250 (2002).
227. F. M. Tian and G. Y. Zhu, *Anal. Chim. Acta* 451, 251 (2002).
228. A. Kros, M. Gerritsen, V. S. I. Sprakel, N. A. J. M. Sommerdijk, J. A. Jansen, and R. J. M. Nolte, *Sensor. Actuat. B* 81, 68 (2001).
229. K. Ramanathan, B. R. Jonsson, and B. Danielsson, *Anal. Chim. Acta* 427, 1 (2001).
230. S. Bharathi and O. Lev, *Anal. Commun.* 35, 29 (1998).
231. S. Bharathi and O. Lev, *Appl. Biochem. Biotech.* 89, 209 (2000).
232. O. S. Wolfbeis, I. Oehme, N. Papkovskaya, and I. Klimant, *Biosens. Bioelectron.* 15, 69 (2000).
233. P. C. Pandey, S. Upadhyay, and H. C. Pathak, *Sensor. Actuat. B* 60, 83 (1999).

234. X. J. Wu, M. M. F. Choi, and D. Xiao, *Analyst* 125, 157 (1999).
235. S. de Marcos, J. Galindo, J. F. Sierra, J. Galban, and J. R. Castillo, *Sensor. Actuat. B* 57, 227 (1999).
236. F. Q. Tang, J. F. Shen, J. F. Zhang, and G. L. Zhang, *Chem. J. Chinese U.* 20, 634 (1999).
237. J. Li, L. S. Chia, N. K. Goh, and S. N. Tan, *J. Electroanal. Chem.* 460, 234 (1999).
238. S. P. Yang, Y. F. Lu, P. Atanassov, E. Wilkins, and X. C. Long, *Talanta* 47, 735 (1998).
239. B. Q. Wang, B. Li, Q. Deng, and S. J. Dong, *Anal. Chem.* 70, 3170 (1998).
240. T. Yao and K. Takashima, *Biosens. Bioelectron.* 13, 67 (1998).
241. J. Wang, D. S. Park, and P. V. A. Pamidi, *J. Electroanal. Chem.* 434, 185 (1997).
242. J. Gun and O. Lev, *Anal. Chim. Acta* 336, 95 (1996).
243. J. Li, L. S. Chia, N. K. Goh, S. N. Tan, and H. Ge, *Sensor. Actuat. B* 40, 135 (1997).
244. U. Kunzelmann and H. Bottcher, *Sensor. Actuat. B* 39, 222 (1997).
245. I. Pankratov and O. Lev, *J. Electroanal. Chem.* 393, 35 (1995).
246. S. Sampath and O. Lev, *Electroanalysis* 8, 1112 (1996).
247. B. Q. Wang, J. Z. Zhang, G. J. Cheng, and S. J. Dong, *Anal. Chim. Acta* 407, 111 (2000).
248. D. D. Chen, B. H. Liu, Z. J. Liu, and J. L. Kong, *Anal. Lett.* 34, 687 (2001).
249. F. M. Tian, B. Xu, L. D. Zhu, and G. Y. Zhu, *Anal. Chim. Acta* 443, 9 (2001).
250. Y. Miao and S. N. Tan, *Anal. Chim. Acta* 437, 87 (2001).
251. X. Chen, B. Q. Wang, and S. J. Dong, *Electroanalysis* 13, 1149 (2001).
252. P. C. Pandey, S. Upadhyay, I. Tiwari, and V. S. Tripathi, *Sensor. Actuat. B* 72, 224 (2001).
253. B. X. Li, Z. J. Zhang, and Y. Jin, *Sensor. Actuat. B* 72, 115 (2001).
254. B. Q. Wang and S. J. Dong, *Talanta* 51, 565 (2000).
255. P. Wang, X. P. Wang, L. H. Bi, and G. Y. Zhu, *J. Electroanal. Chem.* 495, 51 (2000).
256. B. Wang, B. Li, Z. Wang, G. Xu, Q. Wang, and S. Dong, *Anal. Chem.* 71, 1935 (1999).
257. S. L. Chut, J. Li, and S. N. Tan, *Analyst* 122, 1431 (1997).
258. J. Li, S. N. Tan, and H. L. Ge, *Anal. Chim. Acta* 335, 137 (1996).
259. J. Li, S. N. Tan, and J. T. Oh, *J. Electroanal. Chem.* 448, 69 (1998).
260. V. G. Andreou and Y. D. Clonis, *Biosens. Bioelectron.* 17, 61 (2002).
261. H. C. Tsai and R. A. Doong, *Water Sci. Technol.* 42, 283 (2000).
262. A. K. Singh, A. W. Flounders, J. V. Volponi, C. S. Ashley, K. Wally, and J. S. Schoeniger, *Biosens. Bioelectron.* 14, 703 (1999).
263. A. W. Flounders, A. K. Singh, J. V. Volponi, S. C. Carichner, K. Wally, A. S. Simonian, J. R. Wild, and J. S. Schoeniger, *Biosens. Bioelectron.* 14, 715 (1999).
264. A. N. Diaz and M. C. R. Peinado, *Sensor. Actuat. B* 39, 426 (1997).
265. P. C. Pandey, S. Upadhyay, H. C. Pathak, C. M. D. Pandey, and I. Tiwari, *Sensor. Actuat. B* 62, 109 (2000).
266. C. I. Li, Y. H. Lin, C. L. Shih, J. P. Tsaur, and L. K. Chau, *Biosens. Bioelectron.* 17, 323 (2002).
267. P. C. Pandey, S. Upadhyay, I. Tiwari, and V. S. Tripathi, *Anal. Biochem.* 288, 39 (2001).
268. M. C. Ramos, M. C. Torijas, and A. N. Diaz, *Sensor. Actuat. B* 73, 71 (2001).
269. J. Li, K. M. Wang, D. Xiao, and X. H. Yang, *Chem. J. Chinese U.* 21, 1018 (2000).
270. S. A. Kane, E. I. Iwuoba, and M. R. Smyth, *Analyst* 123, 2001 (1998).
271. B. Wang, J. Zhang, and S. Dong, *Biosens. Bioelectron.* 15, 397 (2000).
272. Z. Liu, B. Liu, J. Kong, and J. Deng, *Anal. Chem.* 72, 4707 (2000).
273. J. Metzger, M. Reiss, and W. Hartmeier, *Biosens. Bioelectron.* 13, 1077 (1998).
274. J. Li, L. S. Chia, N. K. Goh, and S. N. Tan, *Anal. Chim. Acta* 362, 203 (1998).
275. S. Ferretti, S. K. Lee, B. D. MacCraith, A. G. Oliva, D. J. Richardson, D. A. Russell, K. E. Sapsford, and M. Vidal, *Analyst* 125, 1993 (2000).
276. J. W. Aylott, D. J. Richardson, and D. A. Russell, *Analyst* 122, 77 (1997).
277. D. J. Blyth, J. W. Aylott, J. W. B. Moir, D. J. Richardson, and D. A. Russell, *Analyst* 124, 129 (1999).
278. E. I. Iwuoha, S. Kane, C. O. Ania, M. R. Smyth, P. R. O. de Montellano, and U. Fuhr, *Electroanalysis* 12, 980 (2000).
279. J. J. Niu and J. Y. Lee, *Sensor. Actuat. B* 62, 190 (2000).
280. J. J. Niu and J. Y. Lee, *Anal. Commun.* 36, 81 (1999).
281. W. Y. Lee, S. R. Kim, T. H. Kim, K. S. Lee, M. C. Shin, and J. K. Park, *Anal. Chim. Acta* 404, 195 (2000).
282. W. Y. Lee, K. S. Lee, T. H. Kim, M. C. Shin, and J. K. Park, *Electroanalysis* 12, 78 (2000).
283. P. C. Pandey and G. Singh, *Talanta* 55, 773 (2001).
284. T. M. Park, E. I. Iwuoha, and M. R. Smyth, *Electroanalysis* 9, 1120 (1997).
285. R. A. Simkus, V. Laurinavicius, L. Boguslavsky, T. Skotheim, S. W. Tanenbaum, J. P. Nakas, and D. J. Slomczynski, *Anal. Lett.* 29, 1907 (1996).

www.aspbs.com/enn

Quantum Dot Atoms, Molecules, and Superlattices

Hiroyuki Tamura, Hideaki Takayanagi
NTT Basic Research Laboratories, NTT Corporation, Atsugi, Kanagawa, Japan

Kenji Shiraishi
University of Tsukuba, Tsukuba, Ibaraki, Japan

CONTENTS

1. Introduction
2. Single Quantum Dot (Artificial Atom)
3. Coupled Quantum Dots (Artificial Molecule)
4. Quantum Dot Arrays
5. Conclusions
Glossary
References

1. INTRODUCTION

Rapid progress in nanoscale fabrication technology ("nanotechnology") has enabled us to make various types of semiconductor devices using quantum dots (QDs). Quantum dots are small regions defined by tailor-made confining potentials in semiconductor materials, in which the number of electrons can vary between one and several hundred and whose size is comparable to the Fermi wavelength of the electrons. Because QDs are analogous to real atoms, they are often referred to as artificial atoms. Unlike real atoms, though, current and voltage leads can be attached to probe a QD's electronic state. The interplay between quantum confinement and charging effects manifests itself in a wide range of physical phenomena [1–3]. The single-electron charging effect is one of the more important of these and will be addressed in this chapter. When the charging energy of a small quantum dot is larger than the thermal energy, electrons in the lead cannot transfer into the dot ("Coulomb blockade effect"). By coupling several quantum dots, we can create artificial molecules. An important feature of these artificial molecules is that the couplings between different dots can be tuned by changing gate voltages or interdot distances. This tunability allows us to achieve various interesting phenomena, ranging from the formation of a "chemical bond" between two coupled dots to the manipulation of a single electron in "turnstiles" and "pumps."

If we consider a QD artificial atom as a building block, we can create QD arrays. Electrons in QD arrays can move around and interact with each other through the Coulomb interaction, and many interesting effects manifest themselves in magnetic, transport, and optical properties. When QDs are arranged on a periodic lattice and are coupled to each other coherently, a band structure is obtained. This type of dot lattice or artificial crystal was first proposed by Sakaki [4] and it is called quantum dot superlattice (QDSL). It is analogous to quantum well superlattices [5], in which a sequence of semiconductor layers with different bandgaps produces new materials and devices. The energy spectrum of the superlattice is determined by the artificial periodicity and the coupling between quantum wells (dots) rather than by the properties of the individual semiconductor materials.

This chapter is organized as follows. Section 2 surveys the studies of the single-electron effect in a single QD. Starting from a classical single-charge theory as a basis, we extend it to include the quantized single-particle-energy and the exchange interaction and discuss the physics of single QD systems. Section 3 describes multiple-dot systems or QD molecules. Section 4 covers the physics in quantum dot arrays, stressing theoretical predictions and proposals that will stimulate future study. Finally, Section 5 makes some concluding remarks and suggests future directions.

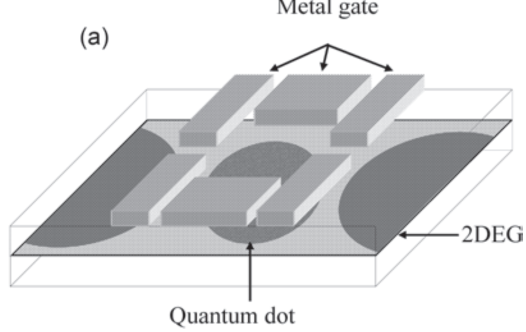
2. SINGLE QUANTUM DOT (ARTIFICIAL ATOM)

2.1. Single-Electron Tunneling in Semiconductor Structures

In this section, we focus on two QD structures that are commonly used in the study of single-electron phenomena. One is the lateral QD shown in Figure 1a. Metal electrodes are

*This chapter first appeared in *Quantum Dots and Nanowires*, Edited by S. Bandyopadhyay and H. S. Nalwa, ©2003, American Scientific Publishers.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.



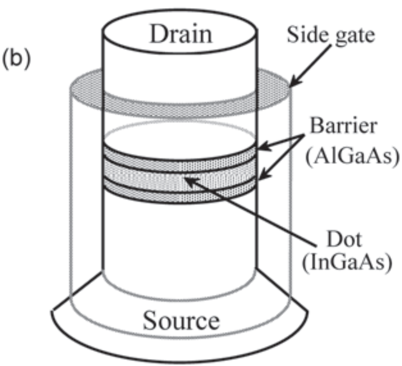


Figure 1. (a) Schematic of a lateral QD. Negatively biased metal electrodes patterned on the surface of a 2DEG heterostructure of GaAs depletes the underlying 2DEG in the vicinity of the gates. (b) Schematic of a vertical QD fabricated in a heterostructure where a disk of GaAs is sandwiched by two thin barriers of AlGaAs. A negatively biased gate-electrode surrounding the pillar depletes electrons in the outer region of the pillar.

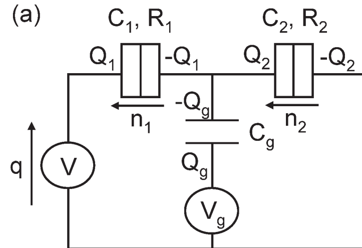
patterned on the surface of a two-dimensional electron gas (2DEG) heterostructure of GaAs/AlGaAs. A negative voltage applied to the gates raises the electrostatic potential and depletes the underlying 2DEG in the vicinity of the gates. Under suitable biasing conditions, a small region of 2DEG remains at the center of the structure and is separated from the rest of the 2DEG by two narrow constrictions. When electrostatic potential at the constriction is higher than the Fermi energy of the 2DEG, there arises an energy barrier under which electrons can tunnel. In many cases, an additional gate is capacitively coupled to the QD to modify the electrostatic potential inside the QD. This additional gate can be a separate surface electrode near the QD or a back-gate on the other side of the 2DEG.

The other structure is obtained in the vertical QD illustrated in Figure 1b. A narrow pillar is fabricated in a heterostructure where a disk of GaAs is sandwiched between two thin AlGaAs barriers. A negatively biased gate-electrode surrounding the pillar depletes electrons in the outer region of the pillar and shrinks the quantum disk. Consequently, the electrons are confined at a very narrow region in the center of the pillar. Current flowing along the pillar is measured between the source and drain electrodes on the top and bottom of the pillar. The lateral confining potential has a circular symmetry approximated by a two-dimensional parabolic potential. This allows us to observe a shell structure having highly degenerate levels, as will be described in the following section. Another type of vertical dot structure has been proposed by Ashoori et al. [6, 7]. The dot is located in the GaAs quantum well between AlGaAs barriers and its size is controlled by the top gate voltage as in the lateral QD. Electrons tunnel vertically to the bottom electrode. The addition spectra are measured in a unique ac capacitance measurement by varying the gate voltage. Singlet–triplet transition and Landau level transition have also been observed in these kinds of structures [7].

2.2. Classical Model of the Single-Electron Charging Effect

If one considers a quantum dot to be a classical conductor having capacitance C, the charging energy needed to add one electron to it is expressed by $E_c = e^2/C$. For a sphere, $C_{\text{sphere}} = 4\pi\varepsilon R$ and, for a disk, $C_{\text{disk}} = 8\varepsilon R$, where ε is the dielectric constant around the dot. At low temperatures, the charging energy E_c can be larger than the thermal energy $k_B T$ in the QD, where the number of electrons does not fluctuate.

The quantum dot is coupled to external electrodes to add or remove electrons and modify the electrostatic potential in the dot. Figure 2a illustrates the equivalent circuit of the single-electron transistor (SET), which consists of two tunnel junctions in series that form a QD capacitively coupled to the gate-electrode with voltage V_g, capacitance C_g, and accumulated charge Q_g. Each tunnel junction has capacitance C_i, resistance R_i, and accumulated charge Q_i. The



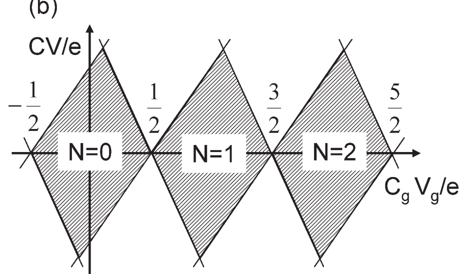


Figure 2. (a) Equivalent circuit of the single-electron transistor. Two tunnel junctions with capacitance C_i, resistance R_i, and accumulated charge Q_i $(i = 1, 2)$ form a QD capacitively coupled to the gate-electrode with voltage V_g, capacitance C_g, and accumulated charge Q_g. The transport voltage V induces a net flow of charge through the device. (b) Stability diagram of a SET. Inside the shaded regions, tunneling events are suppressed by the Coulomb blockade effect and the number of electrons N is well defined. For low V, the current oscillates periodically when $V_g \simeq (n + 1/2)e/C_g$.

transport voltage V induces a net flow of charge through the device. Consider the situation where R_1 and R_2 are much larger than the resistance quantum h/e^2 and electrons n_i tunnel through tunnel junction i. The electrostatic energy of the whole system is given by

$$E(n_1, n_2) = \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} + \frac{Q_g^2}{2C_g} - (en_1 + Q_1)V - Q_g V_g \tag{1}$$

where the charges must satisfy the relations

$$\frac{Q_1}{C_1} + \frac{Q_2}{C_2} = V \quad \frac{Q_g}{C_g} + \frac{Q_2}{C_2} = V_g \quad -Q_1 + Q_2 - Q_g = eN \tag{2}$$

where $N = n_1 - n_2$ is the number of electrons in the QD. Substituting Eq. (2) into Eq. (1), we obtain

$$E(n_1, n_2) = \frac{1}{2C}(C_1 V + C_g V_g + eN)^2 - en_1 V \tag{3}$$

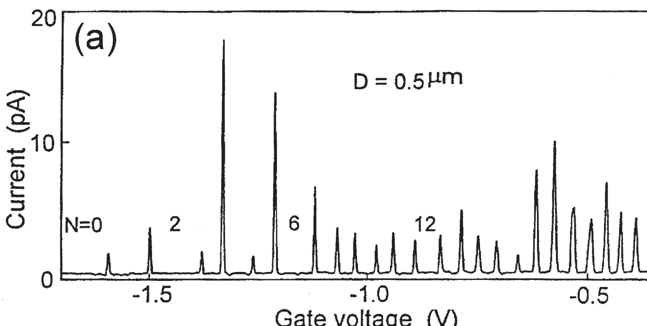
Tunneling is prohibited at junction 1 if $E(n_1 \pm 1, n_2) > E(n_1, n_2)$ and at junction 2 if $E(n_1, n_2 \pm 1) > E(n_1, n_2)$. Under these conditions, the Coulomb blockade occurs if

$$\frac{C_g V_g}{e} - \left(N + \frac{1}{2}\right) < \frac{(C_2 + C_g)V}{e} < \frac{C_g V_g}{e} - \left(N - \frac{1}{2}\right)$$
$$-\frac{C_g V_g}{e} + \left(N - \frac{1}{2}\right) < \frac{C_1 V}{e} < -\frac{C_g V_g}{e} + \left(N + \frac{1}{2}\right) \tag{4}$$

where we define the total capacitance as $C = C_1 + C_2 + C_g$ and assume $E_c \gg k_B T$. In the $V - V_g$ plane shown in Figure 2b, rhombus-shaped regions ("Coulomb diamonds") appear along the axis of V_g, within which the QD is charged with a fixed number of electrons. Inside these regions, tunneling events are suppressed and no current flows through the device ("Coulomb blockade effect"). For low V, the current flows only when $C_g V_g = e(N + 1/2)$, and the gate capacitance C_g can be determined from the current peak separations. C_1 and C_2 can be deduced from the slopes of the two edges of the rhombus.

When a QD having N electrons is in equilibrium with reservoirs with the chemical potential μ, the thermodynamic potential is given by $\Omega(N) = E(N) - N\mu$. When $\Omega(N') > \Omega(N)$ for all $N'(\neq N)$, electron tunneling is prohibited by the Coulomb blockade effect. When the chemical potential satisfies the condition $\Omega(N+1) = \Omega(N)$ at some μ, a small bias between the source and drain reservoirs induces a current through the dot via a transition between states N and $N+1$. This condition can be simply expressed by $\mu(N) = E(N+1) - E(N)$. This quantity is called an addition energy and represents the energy necessary to add an electron to the dot. By changing the external gate-electrode V_g, the chemical potential μ in the reservoirs is also modulated, where these two quantities are related via the relation $\mu(N) = \alpha e V_g(N)$. The factor α can be determined by fitting the conductance peak to a $\cosh^{-2}$ line shape [8].

The classical model described here gives the current oscillations with an equal spacing $\Delta V_g = e/C_g$ as a function of the gate voltage. Measured current oscillations in the vertical QD are shown in Figure 3. The separation of the current peaks is not exactly periodic. In the vertical QD shown in Figure 1b, the actual QD is much smaller than the diameter of the pillar, since the negatively biased side-gate depletes the electrons and the electrostatic potential strongly confines them in the center of the quantum disk. When the side-gate voltage is just above the threshold voltage, a small number of electrons remain in the QD. When the actual diameter of the GaAs QD is $d = 20$ nm, for example, the single-particle level spacing is as large as $\hbar^2/m^*d^2 \simeq 6$ meV ($m^* = 0.07$), which is comparable to the Coulomb energy $e^2/4\pi\epsilon d \simeq 12$ meV ($\epsilon = 13$). In this situation, it is obvious that the classical charging model is insufficient and we have to take into account the quantum size effect. Coulomb interactions also become important, as we will see in the following sections.



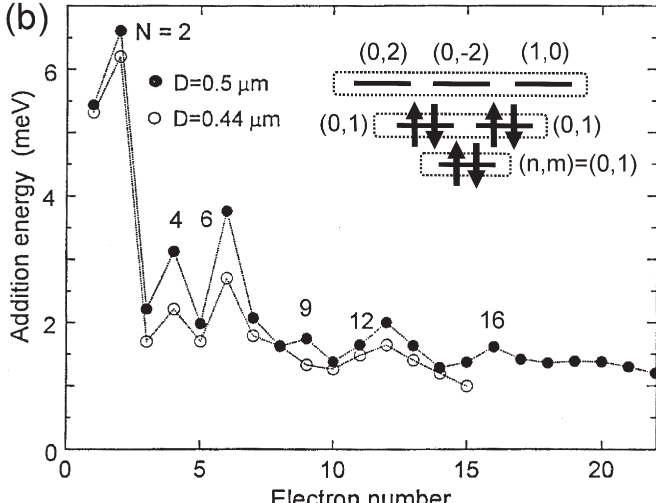


Figure 3. (a) Current oscillations measured in a vertical QD. (b) Peak spacings as a function of the electron number in quantum dots having pillar diameters of $D = 0.5$ and 0.44 μm. Inset: Schematic of energy levels of a QD in the parabolic confining potential. Reprinted with permission from [10], S. Tarucha et al., *Phys. Rev. Lett.* 77, 3613 (1996). © 1996, American Physical Society.

2.3. Constant Interaction Model with Exchange Term

We consider the Hamiltonian for a QD expressed by

$$\mathscr{H} = \sum_{\alpha\sigma} \varepsilon_\alpha c_{\alpha\sigma}^+ c_{\alpha\sigma} + \frac{1}{2} \sum_{\alpha\alpha'\beta\beta'} \sum_{\sigma\sigma'} U_{\alpha\beta\beta'\alpha'} c_{\alpha\sigma}^+ c_{\beta\sigma'}^+ c_{\beta'\sigma'} c_{\alpha'\sigma} \tag{5}$$

where $c^+_{\alpha\sigma}(c_{\alpha\sigma})$ is the creation (annihilation) operator for the state α with energy ε_α and spin σ, and $U_{\alpha\beta\beta'\alpha'}$ is the matrix element of Coulomb interaction. To go beyond the classical charging model, let us start from the reduced Hamiltonian including the single-particle, direct, and exchange terms [9]. Neglecting the off-diagonal terms, the interacting Hamiltonian $\mathscr{H}_I$ is reduced to

$$\mathscr{H}_I = \mathscr{H}_{\text{Coulomb}} + \mathscr{H}_{\text{exchange}} \tag{6}$$

$$\mathscr{H}_{\text{Coulomb}} = \frac{1}{2}\sum_{\alpha\beta}\sum_{\sigma\sigma'} U_{\alpha\beta\beta\alpha} c^+_{\alpha\sigma} c_{\alpha\sigma} c^+_{\beta\sigma'} c_{\beta\sigma'} \tag{7}$$

$$\mathscr{H}_{\text{exchange}} = -\frac{1}{2}\sum_{\alpha\beta}\sum_{\sigma\sigma'} U_{\alpha\beta\alpha\beta} c^+_{\alpha\sigma} c_{\alpha\sigma'} c^+_{\beta\sigma'} c_{\beta\sigma} \tag{8}$$

It is convenient to introduce operators of the number of electrons $\hat{n}_\alpha$ and the spin $\widehat{\mathbf{S}}_\alpha$ on the QD such that

$$\hat{n}_\alpha = \sum_\sigma c^+_{\alpha\sigma} c_{\alpha\sigma} \qquad \widehat{\mathbf{S}}_\alpha = \frac{1}{2}\sum_{\sigma\sigma'} c^+_{\alpha\sigma} \vec{\sigma}_{\sigma\sigma'} c_{\alpha\sigma'} \tag{9}$$

where $\vec{\sigma}$ is the Pauli matrix. When one assumes the matrix elements in the Coulomb term and the exchange term are given by $U_{\alpha\beta\beta\alpha} = U$ and $U_{\alpha\beta\alpha\beta} = J$, one can express $\mathscr{H}_{\text{Coulomb}}$ and $\mathscr{H}_{\text{exchange}}$ in a simple form as

$$\mathscr{H}_{\text{Coulomb}} = \frac{1}{2}\sum_{\alpha\beta}\sum_{\sigma\sigma'} U_{\alpha\beta\beta\alpha} \hat{n}_{\alpha\sigma} \hat{n}_{\beta\sigma'} = \frac{U}{2}\widehat{N}^2 \tag{10}$$

$$\begin{aligned}\mathscr{H}_{\text{exchange}} &= -\sum_{\alpha\beta} U_{\alpha\beta\alpha\beta}\left[\widehat{\mathbf{S}}_\alpha \cdot \widehat{\mathbf{S}}_\beta + \frac{1}{4}\sum_{\sigma\sigma'} \hat{n}_{\alpha\sigma}\hat{n}_{\beta\sigma'}\right] \\ &= -J\left[\widehat{\mathbf{S}}^2 + \frac{1}{4}\widehat{N}^2\right]\end{aligned} \tag{11}$$

where we define $\widehat{N} = \sum_\alpha \hat{n}_\alpha$, $\widehat{\mathbf{S}} = \sum_\alpha \widehat{S}_\alpha$, and use the identity $\vec{\sigma}_{\sigma\sigma'}\vec{\sigma}_{\sigma_1\sigma'_1} = 2\delta_{\sigma\sigma'_1}\delta_{\sigma'\sigma_1} - \delta_{\sigma\sigma'}\delta_{\sigma_1\sigma'_1}$. Then, by adding the single-particle term in Eq. (5), the total Hamiltonian is simply expressed by

$$\mathscr{H} = \sum_{\alpha\sigma} \varepsilon_\alpha \hat{n}_{\alpha\sigma} + E_c\widehat{N}^2 - J\widehat{\mathbf{S}}^2 \tag{12}$$

where $E_c = U/2 - J/4$ is the charging energy of the QD. The last two terms in Eq. (12) represent the dependence of the energy of the QD on the total number of electrons and the total spin respectively. They commute with each other and with the first term in Eq. (12). Therefore, all states of the QD can be classified by n_α and $\mathbf{S}_\alpha$. Choosing the direction of the total spin of the system in the z axis, we can express the energy of this system in terms of the occupation numbers $n_{\alpha\sigma}$ as

$$E(N, S) = \sum_{\alpha\sigma} \varepsilon_\alpha n_{\alpha\sigma} + E_c N^2 - JS(S+1) \tag{13}$$

where $N = \sum_{\alpha\sigma} n_{\alpha\sigma}$ and $S = |\sum_\alpha (n_{\alpha\uparrow} - n_{\alpha\downarrow})|/2$. This Hamiltonian represents the constant interaction model with the exchange term (CIE model). Without the last (exchange) term, it is simply called the constant interaction (CI) model.

Let us consider the situation where the number of electrons is fixed by the Coulomb blockade effect. The lowest energy is obtained by minimizing E with respect to the occupation numbers $n_{\alpha\sigma}$ satisfying Pauli's principle. When two levels at the Fermi energy are nearly degenerate with a small energy splitting Δ, the energy difference between the lowest energy states with the singlet $S = 0$ and the triplet $S = 1$ state is given by $E(N, 1) - E(N, 0) = \Delta - 2J$. The triplet state becomes energetically more favorable than the singlet one when $J > \Delta/2$. In general, when we fill $2p(\geq 2)$ electrons in $q(\geq 2p)$ degenerate levels with a splitting Δ near the Fermi energy, the energy difference between spin $S = k - 1$ and $S = k$ $(1 \leq k \leq p)$ states is given by $E(N, k) - E(N, k-1) = (2k-1)\Delta - 2kJ$. By doing a similar algebra for the odd number of electrons $2p + 1$, the energy difference between spin states $S = k + 1/2$ and $S = k - 1/2$ is given by $E(N, k+1/2) - E(N, k-1/2) = 2k\Delta - (2k-1)J$. The largest spin state $S = p$ (or $S = p + 1/2$) becomes the lowest state when the exchange energy is larger than the energy splitting such that $J > (2p-1)\Delta/2p$ (or $J > 2p\Delta/(2p-1)$). This is known as Hund's first rule. This rule generally holds when the separation between two (or more) single-particle levels is smaller than the exchange energy, for example, in the presence of magnetic fields.

The CIE model described here is quite simplified. The Coulomb and exchange energies $U_{\alpha\beta\beta\alpha}$ and $U_{\alpha\beta\alpha\beta}$ differ from state to state, and it is not obvious that the off-diagonal terms can be neglected. Nevertheless, it does qualitatively explain many features observed in experiment for a single QD, as we will see in the next subsection.

2.4. Quantum Dot in Parabolic Potential

Here, let us consider the single-electron charging effect in a vertical QD. Figure 3a shows the gate-voltage dependence of the current through QDs with diameters of $D = 0.5$ and $0.44\ \mu$m [10]. In these devices, the number of electrons in the QD starts from one. It can be seen that the peaks of Coulomb blockade oscillations are not equally spaced. The peak spacings versus the number of electrons are plotted in Figure 3b. The 2nd, 6th, and 12th peak spacings are quite large. The 4th, 9th, and 16th ones are also large. As mentioned in the previous subsection, the confinement of the vertical QD can be represented by a two-dimensional isotropic parabolic potential. Let us examine the single-particle energies for this potential. In the presence of a magnetic field threaded perpendicularly to the 2D plane, the single-particle Hamiltonian is given by

$$H_0 = \frac{(\mathbf{p} - e\mathbf{A})^2}{2m^*} + \frac{1}{2}m^*\omega_0^2(x^2 + y^2) \tag{14}$$

where $\mathbf{A} = (-By/2, Bx/2)$, m^* is the effective mass of an electron, and ω_0 is the frequency of the parabolic confinement. The eigenvalue of the Hamiltonian H_0 is given by

$$\varepsilon_{nm} = (2n + |m| + 1)\hbar\omega - \frac{1}{2}m\hbar\omega_c \qquad (n = 0, 1, 2, \ldots \quad m = 0, \pm 1, \pm 2, \ldots) \tag{15}$$

where $\omega = \sqrt{\omega_0^2 + (\omega_c/2)^2}$ and $\omega_c = eB/m^*$. In the absence of the magnetic field $B = 0$, the single-particle levels are classified as shell structures; that is, $\varepsilon_l = l\hbar\omega_0$ $(l = 1, 2, \ldots)$,

where the energy shells are separated in an equal spacing $\hbar\omega_0$ and the lth shell has l degenerate levels distinguished by the magnetic quantum number $m = -(l-1), -(l-3), -(l-5), \ldots, l-3, l-1$ as shown in Figure 3. When the magnetic field is applied, the degeneracy of the levels in each shell is lifted and the energy levels have a magnetic field dependence called the Darwin–Fock spectrum [11, 12].

When we consider the single electron phenomena in the presence of highly degenerate levels, the CIE model serves as a starting point in examining what happens to the Coulomb blockade. As described in the previous section, the CIE model has degenerate single-particle levels in the topmost shell and shows the ground state with the largest total spin obeying Hund's first rule: In a zero magnetic field, first l electrons with up-spin fill the levels in the lth shell, and then the next l electrons with down-spin occupy the shell, satisfying the Pauli's principle. The configuration of electrons having the largest spin on (or above) the third shell is not always unique. For example, the third shell has three levels with the magnetic momenta $m = 2, 0, -2$ and there are three possibilities for filling two electrons with up-spins. The CIE model cannot predict how the up-spin electrons are filled in the degenerate levels with different magnetic momentum. The shell filling depends on the distribution of wavefunctions in each orbital, which results in different matrix elements of the Coulomb interaction $U_{\alpha\beta\beta\alpha}$. Many theoretical studies have been performed to take this orbital effect properly into account. These include Hartree–Fock calculations [13–17], numerical diagonalizations of the many-electron Hamiltonian [18–22], density-functional methods [23–30], Monte Carlo calculations [31, 32], and other techniques [33–35]. Figure 4 shows the chemical potential difference versus the number of electrons calculated by a self-consistent Hartree–Fock approximation [14]. The large peaks at $N = 2, 6, 12, 20$ reflect the shell structure and the peaks at $N = 4, 9, 16$ reflect high-spin states based on Hund's rule. These qualitative features are

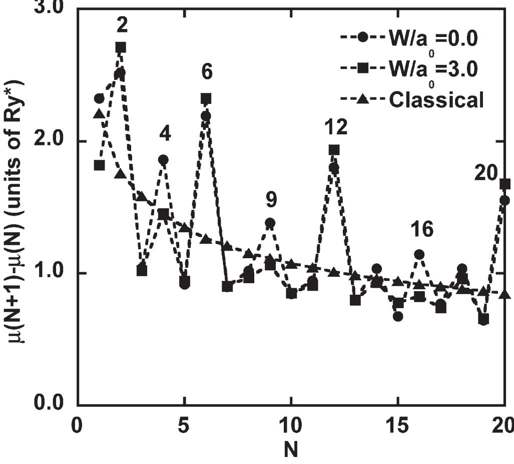


Figure 4. Calculated difference of adjacent chemical potentials as a function of the number of electrons in quantum disks having thickness W of $W/a_0 = 0.0$ and 3.0. Here, $a_0 = \sqrt{\hbar/2m^*\omega_0}$ and Ry^* is the effective Rydberg constant. Classical dependence is obtained at high temperature $k_B T/\hbar\omega_0 = 0.3$. Reprinted with permission from [14], H. Tamura, *Phys. B* 249–251, 210 (1998). © 1998, Elsevier Science.

in good agreement with the experiment. Experimental [36] and theoretical [36, 37] investigations on ellipsoidal deformation of vertical quantum dots have also been done. Theoretical calculations in the three-dimensional QD have been performed [19, 26, 38, 39].

When magnetic fields are applied, the degenerate levels in each shell are lifted depending on the magnetic momentum and form the Darwin–Fock spectrum. The magnetic field dependence of the peak positions measured in the vertical QD is plotted in Figure 5 [40]. Figure 6 shows the calculated addition spectra $\mu(N)$ in the CIE model as a function of $\omega_c = eB/m^*$. Another interesting feature in Figures 5 and 6 is that cusps appear when the two single-particle levels cross. If we neglect the exchange term in Eq. (13) in the CIE model as plotted by dotted lines in Figure 6, two curves of chemical potential simply approach and depart with the magnetic field, whereas, when the exchange term is included as plotted by solid lines, cusps appear when two single-particle levels cross. These cusps are clear evidence of the appearance of the high-spin state ("generalized Hund's rule") when the two single-particle levels cross by applying a magnetic field [14, 41, 42].

In magnetic fields $\omega_c \gg \omega_0$, Landau levels of filling factor $\nu = 2$ are formed, where N electrons occupy the lowest Landau level with the lowest total spin. As the magnetic field becomes stronger, the adjacent level spacing in the Landau level decreases. When the difference between the highest unoccupied level for up-spin and the highest occupied level for down-spin becomes smaller than the exchange energy, one electron flips its spin and the total spin increases by one. This spin-flip event continues until there are no more down-spin electrons and the total spin becomes maximum $S_{\max} = N/2$. This transition region is called the spin-flip region.

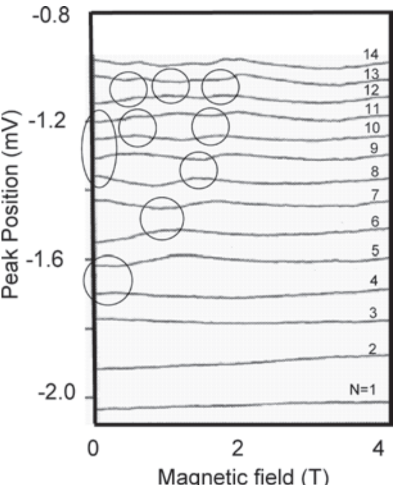


Figure 5. Peak positions of Coulomb oscillations in the vertical QD ($D = 0.54$ μm) having electron number $N = 1$ to 14 as a function of magnetic field. Circles at $B = 0$ indicate the region where Hund's rule holds in the shell structure and circles for $B > 0$ indicate the region where the generalized Hund's rule holds when two single-particle levels cross. Reprinted with permission from [40], D. G. Austing et al., *Japan. J. Appl. Phys.* 38, 372 (1999), ©1999, Institute of Pure and Applied Physics.

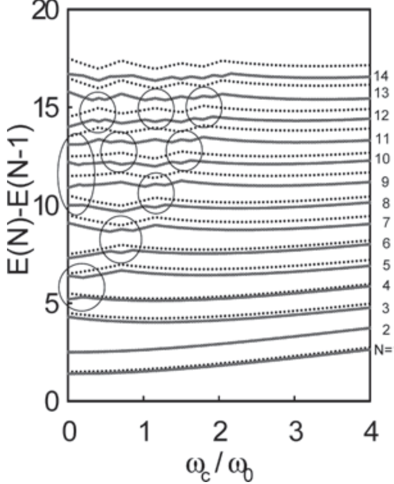


Figure 6. Addition spectra in the constant interaction model with (solid lines) and without (dotted lines) exchange term as a function of magnetic field. Here, $\hbar\omega_0 = U = 1$ and $J = 0.1$. The features in Figure 5 are well reproduced.

Spin-polarized electrons occupy adjacent orbitals with consecutive angular momenta. This compact occupation of states maximizes the electron density, and so it is called the maximum density droplet [43]. The maximum density droplet state has been observed in experiments [7, 44–46] and has been theoretically analyzed [47–49]. The root-mean-squared extent of orbital $|m\rangle$ with angular momentum m is expressed by $\langle \delta r\rangle = \langle m|r^2|m\rangle^{1/2} = \sqrt{2(m+1)(\hbar e/B)}$, where orbitals having smaller m are distributed in the inner area. A further increase of the magnetic field squeezes the QD and Coulomb repulsion becomes much stronger. The electrons jump from the lower levels with lower angular momentum to the higher ones distributed in the outer area of the QD to relax the Coulomb repulsive energy. This is called the reconstruction region [46]. The CIE model simply predicts that electrons in the innermost levels move to the outermost levels [43] and a hole of electron density is created in the center of the QD. Chamon and Wen [50] found that Hartree–Fock calculations predict a ring of high-density electrons that breaks off from the edge of the maximum density droplet. This has been confirmed by other authors [51–53]. Another interesting possibility in edge reconstruction is the formation of spin textures at the edges of the QD when the Zeeman splitting is included [54, 55]. These textures are driven by the same physics that leads to "skymions" in bulk quantum Hall system [56–59].

Transport and capacitance spectroscopy analyses of QDs have revealed many other interesting properties of electrons in artificial atoms. In the nonlinear transport regime, the current is measured as a function of source–drain voltage eV_{sd} as well as gate voltage. One can probe the level distribution in the energy window eV_{sd} by measuring the dependence of the current with respect to this energy window. For small bias $eV_{sd} \ll \Delta$ (Δ is the single-particle-level spacing), usual Coulomb blockade oscillations are observed. For large bias $eV_{sd} > \Delta$, the source–drain voltage gives not only broadened peaks but also additional structures caused by the excited states. From the structure in the energy diagram, a lot of valuable information on the ground and excited states has been obtained [60, 61]. Transport spectra of lateral [62–64] and self-assembled QDs [65–67] have been measured and the shell structure has been observed. The Coulomb oscillations in a vertical dot containing a very large number of electrons have exhibited a surprising behavior in magnetic fields [68]. The addition spectra as observed in capacitance spectroscopy measurements group into bunches. A theoretical explanation has been proposed [69], but further study is needed in order to understand this phenomenon.

2.5. Spin-Blockade Effect

In the previous sections, we have described the Coulomb blockade effect based on the classical and CIE charging models, where the large charging energy makes the number of electrons well defined in the isolated QD. In these models, the coupling Γ between electrons in the QD and the leads through the tunneling barriers is not taken into account. When the thermal energy $k_BT(\gg\Gamma)$ is much smaller than the single-particle-level spacing Δ, tunneling occurs through only one level. The conductance peak has a lineshape of the derivative of the Fermi–Dirac distribution, where its peak height scales as $1/T$ and its peak width as T. On the other hand, multiple levels can contribute to the tunnel current when the thermal energy exceeds the level spacing. The temperature dependence of conductance oscillations becomes complicated, since the elastic coupling to leads differs from level to level depending the spatial or spin distribution and the symmetry of wavefunction. Meir et al. [70] have employed the Anderson model and the Landauer conductance formula to investigate the transport properties through multiple levels in a QD having very different elastic couplings. They predicted that participation of the multiple levels can give rise to an increase of current with increasing temperatures. This anomalous temperature dependence explained experimental observations [71].

It is not always true that a tunneling event occurs when the Coulomb blockade condition is broken and the current should flow in the classical model. The transition amplitude is closely related to the matrix elements between two states of N and $N+1$ electrons in the QD, and the tunneling probability is suppressed when the matrix element vanishes. To analyze this possibility quantum mechanically, we have to deduce the conductance formula including the transition amplitudes explicitly [72, 73]. When we define $|N, i\rangle$ and $E(N, i)$ as the ith state and energy for N electrons in the QD and $P_{\mathrm{eq}}(N, i) = \exp(-E(N, i)/T)/\sum_{M,j}\exp(-E(M, j)/T)$ as the equilibrium probability of the state, the tunnel current at low bias ($V \ll T$) is expressed by

$$I = V\frac{e^2}{T}\sum_{i,j} P_{\mathrm{eq}}(N, i) f(\epsilon_{ij})\frac{\Gamma_L\Gamma_R}{\Gamma_L+\Gamma_R} \tag{16}$$

$$\Gamma_{L(R)} = 2\pi N(0)|t_{L(R)}|^2\left|\sum_k\langle N+1, j|c_k^+|N, i\rangle\right|^2 \tag{17}$$

where $f(\epsilon)$ is the Fermi–Dirac distribution, $N(0)$ is the density of states in the leads, $t_{L(R)}$ is the tunnel amplitude from the left (right) lead, and $\epsilon_{ij} = E(N+1, j) - E(N, i)$. If we

assume $2\pi N(0)|t_L|^2 = 2\pi N(0)|t_R|^2 = \Gamma$ for simplicity, the above expression becomes

$$I = V\frac{e^2}{2T}\Gamma\sum_{i,j}P_{\rm eq}(N,i)f(\epsilon_{ij})\left|\sum_k \langle N+1, j|c_k^+|N,i\rangle\right|^2 \quad (18)$$

in which all the contributions of transition from $|N, i\rangle$ to $|N+1, j\rangle$ are summed by adding one electron on every possible level k. At low temperatures, $T \ll \epsilon = E(N+1,0) - E(N,0)$, only the ground states $|N,0\rangle$ and $|N+1,0\rangle$ can contribute and the conductance is expressed simply by

$$I = V\frac{e^2}{8T}\Gamma\frac{1}{\cosh^2(\epsilon/2T)}\left|\sum_k \langle N+1, 0|c_k^+|N,0\rangle\right|^2 \quad (19)$$

It is noted that the expression of the tunnel current involves the matrix elements between $c_k^+|N,i\rangle$ and $|N+1,j\rangle$. This means the current is suppressed at low temperature if transition from $|N,i\rangle$ to $|N+1,j\rangle$ does not occur just by adding one electron.

What suppresses the tunneling current? One possibility is the so-called spin-blockade effect proposed by Weinmann et al. [74–77]. If the ground states $|N,i\rangle$ and $|N+1,j\rangle$ respectively have spin S and S' and if $|S'-S| > 1/2$, one cannot change the total spin S into S' by adding spin 1/2. This results in a zero matrix element and the suppression of the current. This situation can occur when the single-particle levels are degenerate like in the shell structure in the circular QD or in the level crossings under magnetic field [78, 79]. As we considered previously, when p levels are almost degenerate near the Fermi energy, the total spin can be $S = p/2$ at most. Transition is prohibited from a state with three electrons having minimum total spin of 1/2 to a state with four electrons having maximum spin of 2. The spin of the ground state depends on the competition between the single-particle energy and the exchange energy, as discussed in Section 2.3, and total spin of the ground state does not always change by 1/2.

How can the spin-blockade effect be observed in experiment? One method is to measure the temperature dependence of the Coulomb blockade peaks. As temperatures increase, the excited states having nonzero matrix elements between $|N,0\rangle$ participate in the tunneling events and the current increases rapidly with temperature. This is in a sharp contrast to the normal $1/T$ dependence of the Coulomb blockade peak as seen in Eq. (19). Another method is to perform transport measurements in the nonlinear regime. Applying large source–drain voltage involves multilevels in the tunnel current through excited states. Evidence of the spin blockade has been obtained in the non-linear transport through GaAs QDs [80, 81]. Magnetotransport measurements have also provided indication of the spin-blockade effect in GaAs dots [81] and Si dots [82] having degenerate energy levels slightly split by the Zeeman effect. The spin-blockade effect at low temperatures is not always observed even when the condition $\Delta S > 1/2$ seems satisfied, as reported in ref. [82]. This problem deserves future study.

2.6. Kondo Effect in Quantum Dots

In the previous sections, we have treated source and drain electrodes as classical reservoirs that add electrons to or remove them from the QD incoherently, where electrons loose their quantum-mechanical phase memory during tunneling events. At extremely low temperatures, this treatment cannot be justified and a coherent coupling between electrons in the QD and leads should be taken into account. One consequence is the so-called Kondo effect.

In the 1930s, a peculiar phenomenon was found in the temperature dependence of the resistance of normal metals, such as copper, containing a dilute amount (below 0.1%) of magnetic impurities of iron or manganese [83–85]. The resistance decreased with temperature due to the reduction of phonon scatterings but increased upon cooling to very low temperatures. This anomalous behavior in resistance had been unresolved for a long time, but a crucial key to understanding this phenomenon was introduced by Kondo in 1964 [86].

The system is described by the Anderson model given by

$$\begin{aligned}\mathcal{H} &= \sum_{\mathbf{k}\sigma}\varepsilon_{\mathbf{k}}c_{\mathbf{k}\sigma}^+c_{\mathbf{k}\sigma} + \sum_{\sigma}\varepsilon_d n_{d\sigma} + Un_{d\uparrow}n_{d\downarrow} \\ &\quad + \sum_{\mathbf{k}\sigma}V\left(c_{\mathbf{k}\sigma}^+d_\sigma + d_\sigma^+c_{\mathbf{k}\sigma}\right)\end{aligned} \quad (20)$$

where $c_{\mathbf{k}\sigma}$ ($c_{\mathbf{k}\sigma}^+$) and d_σ (d_σ^+) are the annihilation (creation) operators of a conduction electron and a d-orbital electron with spin σ respectively, $n_\sigma = d_\sigma^+ d_\sigma$ is the number operator of the localized level with an energy ε_d measured from the Fermi energy, and V is the matrix element between states of the conduction electron and the d-orbital. In the strong coupling limit ($U \ll V$), the system is spin singlet $S = 0$. In the weak coupling limit ($U \gg V$), the number N and spin S on the localized level in the ground state are simply given by $(N,S) = (0,0)$ for $\varepsilon_d > 0$, $(N,S) = (1,1/2)$ for $\varepsilon_d < 0 < \varepsilon_d + U$, and $(N,S) = (2,0)$ for $\varepsilon_d + U < 0$.

Let us consider the weakly coupled regime $U \gg V$ and assume $\varepsilon_d < 0 < \varepsilon_d + U$. We can apply perturbation theory with respect to V, starting from the ground state of $d_\uparrow^+|F\rangle$ and $d_\downarrow^+|F\rangle$, where $|F\rangle$ is the state of conduction electrons filled up to the Fermi energy. The first-order process, which simply puts (removes) an electron in (from) the localized level, costs a large energy and can be neglected. The second-order process can be expressed by

$$\mathcal{H}'\frac{1}{E-\mathcal{H}_0}\mathcal{H}' \quad (21)$$

where $\mathcal{H}'$ is the fourth term in Eq. (20) and $\mathcal{H}_0 = \mathcal{H} - \mathcal{H}'$. By operating Eq. (21) to $d_\uparrow^+|F\rangle$ and $d_\downarrow^+|F\rangle$, one can obtain the effective interaction [87]

$$\begin{aligned}\mathcal{H}_{\rm eff} &= \sum_{\mathbf{k}\mathbf{k}'\sigma}\frac{V^2}{2}\left(-\frac{1}{\varepsilon_d} - \frac{1}{\varepsilon_d + U}\right)c_{\mathbf{k}\sigma}^+c_{\mathbf{k}'\sigma} \\ &\quad + \sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'}2V^2\left(-\frac{1}{\varepsilon_d} + \frac{1}{\varepsilon_d + U}\right)c_{\mathbf{k}\sigma}^+c_{\mathbf{k}'\sigma'}\vec{\sigma}_{\sigma\sigma'}\cdot\mathbf{S}\end{aligned} \quad (22)$$

where $\vec{\sigma}$ is the Pauli matrix and $\mathbf{S}$ is the spin operator of the localized electron. The first term in Eq. (22) expresses normal scatterings independent of spins. The spin-dependent

scatterings in the second term represent so-called *s–d* exchange interactions with antiferromagnetic coupling when $\varepsilon_d < 0 < \varepsilon_d + U$. Kondo pointed out that higher Born scatterings between conduction electrons and the localized spin via antiferromagnetic interaction in Eq. (22) increase the scattering probability (or the inverse of relaxation time) below a characteristic temperature, resulting in a resistance increase. A sea of conduction electrons and the localized moment of spin in a magnetic impurity are coupled by the effective exchange interaction, forming a spin-singlet bound state. The localized moment of spin is screened by the so-called "Kondo cloud" of conduction electrons as if the localized spin disappeared. This many-body effect is called the Kondo effect and the characteristic temperature is the Kondo temperature.

The Kondo effect is a universal phenomenon which manifests itself when a localized state with a net spin couples with conduction electrons in the Fermi sea. A QD confining an unpaired electron of spin 1/2 coupled to leads, schematically shown in Figure 7, can be considered a Kondo system. This "mesoscopic Kondo effect" in a QD has been theoretically predicted [88–91]. First-order tunneling is prohibited by the Coulomb blockade effect as illustrated in Figure 7a because the addition of a second electron costs an energy of $\varepsilon_0 + U$, which exceeds the Fermi energies of the leads μ_L and μ_R. On the other hand, higher-order processes by co-tunneling allow a spin-flip on the dot, as shown in Figure 7c, via the intermediate virtual state depicted in Figure 7b. Successive spin-flip processes effectively screen the local spin on the dot leading to a macroscopically correlated, spin-singlet state. The Kondo temperature in this dot system is given by [92–94]

$$T_K = \sqrt{\frac{\Gamma U}{4}} \exp\left(\frac{\pi \varepsilon_0(\varepsilon_0 + U)}{\Gamma U}\right) \tag{23}$$

where ε_0 is the single-particle energy, U is the on-site Coulomb energy, and $\Gamma = \Gamma_L + \Gamma_R$ is the level broadening caused by the coupling between the dot and the leads. Below T_K, a resonant level ("Kondo peak") is induced near the Fermi energy in the density of states in the dot, and this level allows electrons to tunnel from the left to the right leads resonantly, even when the Coulomb blockade condition is satisfied.

Figure 7d schematically shows how normal Coulomb blockade oscillations change when the Kondo effect takes place. For $T > T_K$, the conductance oscillates with an equal spacing as in Figure 2 and its height is much smaller than $2e^2/h$, reflecting tunneling probabilities less than unity. Below T_K, an unpaired electron in the dot is screened by the conduction electron in the leads and the conductance approaches $2e^2/h$ in the so-called unitary limit where the transmission probability is unity due to a resonant tunneling. For an even number of electrons in the dot, when all the electrons have a pairing partner with an opposite spin, the Kondo effect does not occur and the usual Coulomb blockade effect suppresses the current through the dot.

Recently, the Kondo effect in QDs has been observed in many experiments [95–102]. As described in Section 2.1, lateral QDs have a tunable barrier formed by the gate-electrodes, and tunneling couplings between the dot and the leads are controllable. The conductance increase at valleys predicted in Figure 7d has been experimentally demonstrated in the lateral QDs [95–97]. To obtain ideal characteristics in the unitary limit like in Figure 7d, the barrier height must be low and symmetric ($\Gamma_L = \Gamma_R$). Van der Wiel et al. [99, 102] have observed $2e^2/h$ conductance in their lateral QD as shown in Figure 8. In Figure 8a, the valleys reach the maximum conductance value of $2e^2/h$. Two separate Coulomb peaks develop with narrowing peak spacing. The even–odd asymmetry of the temperature dependence of neighboring peaks is clearly seen. The zero-bias peak in the differential conductance around $V_{SD} = 0$ shown in Figure 8b reflects the Kondo resonance in the density of states at the Fermi energy.

The detailed aspects of the Kondo effect predicted in a QD have been investigated by theorists [103–114]. Here, we show examples calculated by a numerical renormalization group method [114], which gives a unified description of the temperature dependence of the Kondo effect in a single-site QD of an energy ε_d. The single-site approximation is justified when the single-particle energy spacing is much larger than the level broadening Γ. Figure 9 plots the conductance as a function of the single-particle level ε_d in the dot for different temperatures. When temperatures are higher than Γ, normal Coulomb blockade peaks can be seen near $\varepsilon_d \sim -U$ and 0. Decreasing temperatures $T \gtrsim \varepsilon_d$ makes the two peaks rapidly sharper and come closer together, while the valley between peaks has a weak temperature dependence and remains low. A further decrease of temperatures raises the valley up to the unitary value. This temperature dependence in the conductance at the peaks

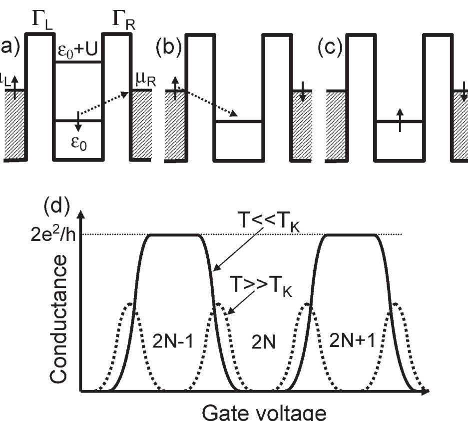


Figure 7. (a) Schematic of a quantum dot having single-site energy ε_0 and on-site Coulomb energy U. The dot is connected to the left and right leads by the tunneling probabilities Γ_L and Γ_R. The series (a)–(c) depicts the spin-flip of the electron at energy ε_0 due to a co-tunneling process via a virtual state shown in (b). (d) Schematic behavior of conductance oscillations as a function of the gate voltage. Above the Kondo temperature T_K, the Coulomb blockade oscillation has a periodicity of e/C_g with the peak heights smaller than e^2/h. In temperatures much lower than T_K, the conductance peak height reaches $2e^2/h$ when the number of electrons in the QD is odd, whereas the conductance is still suppressed for an even number of electrons.

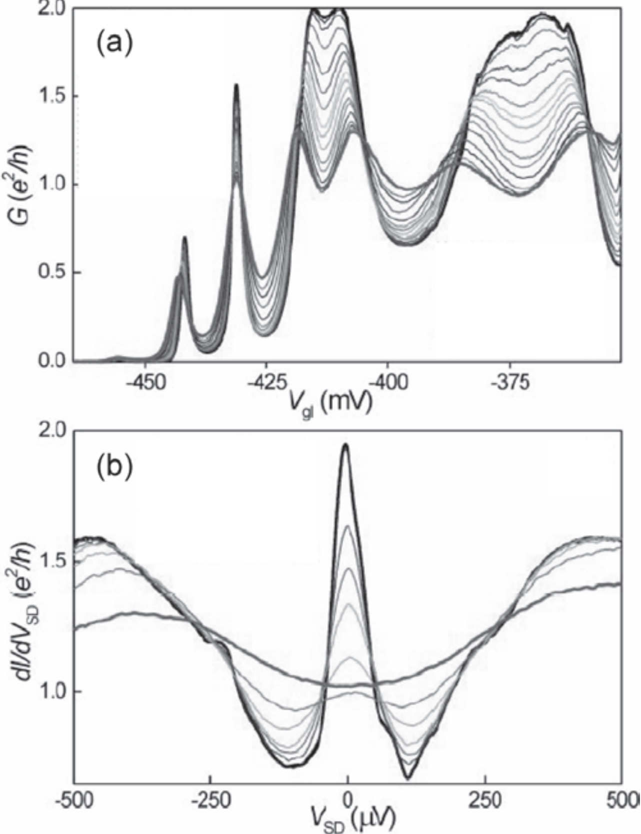


Figure 8. (a) Conductance oscillations as a function of the gate voltage at $B = 0.4$ T in the temperature range from 15 to 800 mK. (b) Differential conductance as a function of dc bias voltage between source and drain leads for temperatures ranging from 15 to 900 mK at $B = 0.4$ T. Reprinted with permission from [102], W. G. van der Wiel, Ph.D. thesis, Delft University of Technology, 2002.

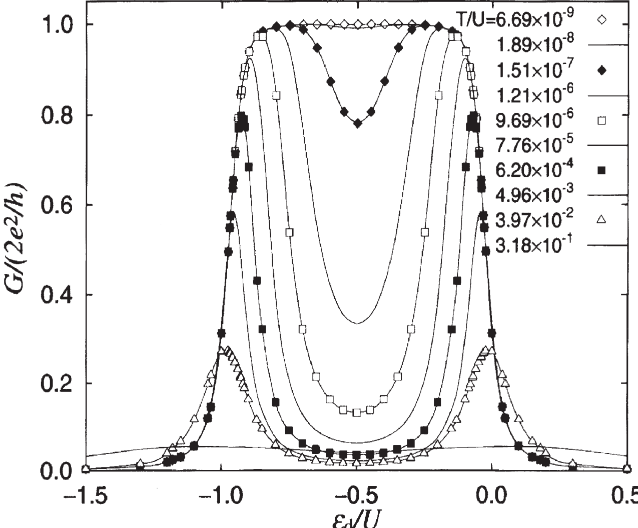


Figure 9. Calculated conductance as a function of the energy of the orbital ε_d in the dot for different temperatures. Reprinted with permission from [114], W. Izumida et al., *J. Phys. Soc. Japan* 70, 1045 (2001). ©2001, The Physical Society of Japan.

and the valley comes from the characteristic temperatures $\widetilde{T}_K$ dependent on ε_d; $\widetilde{T}_K \sim \Gamma$ near the peaks ($\varepsilon_d \sim -U$ and 0), whereas, at the valley ($\varepsilon_d \sim -U/2$), $\widetilde{T}_K$ is comparable to $T_K = \sqrt{\Gamma U/4}\exp(-\pi\varepsilon_0(\varepsilon_0 + U)/\Gamma U)$, which is typically smaller than Γ. The calculation in Figure 9 not only reproduces the experimental observation in Figure 8 but also fits almost perfectly to the conductance curves measured in the experiment of Goldhaber-Gordon et al. [97].

In Figure 7d, we have assumed nondegenerate levels in the dot, where the total spin alternates between 0 and 1/2 according to even and odd numbers of electrons. The conductance oscillations have an even–odd parity below T_K. In real dots, this assumption is not always applied, and the higher levels can affect the Kondo effect. In general, the single-particle levels can be energetically degenerate, where the total spin is not necessarily the lowest due to Hund's rule, as we have seen in the previous subsection. Theorists have investigated the effect of multilevels in a Kondo system of QDs [108, 112, 115–121]. Inoshita et al. [108] have pointed out that the Kondo temperature is significantly enhanced in a multilevel QD compared to a single-level QD. The Kondo effect has been studied at the singlet–triplet transition in Zeeman splitting [117, 119] or the orbital effect [118, 120], where the level degeneracy enhances the Kondo temperature [118, 120] and the conductance peak for a two-level system approaches $4e^2/h$ in the unitary limit [112, 120, 122]. The Kondo effect without the even–odd parity has been observed in experiments [98, 100]. Sasaki et al. [98] have used a vertical QD having tunnel barriers thinner than in the circular QD shown in the previous subsection, in order to raise the Kondo temperature. They have observed the Kondo effect caused by high-spin states in the singlet–triplet transition where two adjacent levels cross in the presence of magnetic fields.

In coupled QDs, the splitting of Kondo resonant peaks has been observed, which reflects the energy difference between the bonding and antibonding states formed by the coherent superposition of the Kondo states of each dot [101]. The Kondo physics in the coupled QDs has been extensively studied by theorists [123–127]. The Kondo effect in QDs, as compared to that in metals, has revealed many fascinating aspects by means of tunable gate voltages or magnetic fields. Recently, the Kondo effect has been found in carbon nanotubes, where the unitary limit is achieved and the Kondo resonance for even numbers of electrons has also been observed [128]. The theoretical and experimental findings reviewed in this section have clearly proven the universality of the Kondo effect. Kondo physics will continue to be an important field in materials science.

2.7. Statistics of Coulomb Blockade Peak Spacings

So far, we have focused mainly on circular QDs with neither disorder nor anisotropic boundaries, where the Coulomb peak spacings have a regular distribution. However, Coulomb oscillations in typical QDs have random spacings, which reflect unintentional irregularities caused by disorder as well as the chaotic nature of electron trajectories. In QDs having a large number of electrons, very complicated mutual interactions also causes energy levels to fluctuate

randomly. This reminds us of the level statistics in the random matrix theory (RMT) used in analysis of the spectrum of complicated nuclei or metal clusters. For a review of RMT, see ref. [129]. In mesoscopic systems, RMT has been used to examine, for example, the universal conductance fluctuations in chaotic and diffusive metals, as reviewed in ref. [130]. In general, the energy distribution in the presence of randomness tends to obey certain statistics. In this subsection, we review the level statistics in the Coulomb blockade oscillations in QDs.

Let us consider the level statistics based on the CI model described in Section 2.3. We neglect exchange terms for simplicity. Without the spin degree of freedom, the CI model gives peak spacings $\Delta\mu = E_c + \Delta\varepsilon$ with the constant charging energy E_c and the single-particle-level spacing $\Delta\varepsilon$ characterizing the distribution. If one assumes the single-particle Hamiltonian is represented by a random matrix, the level distribution $P(s)$ of $s = \Delta\varepsilon$ is described by the Wigner–Dyson distribution: $P(s) = (\pi/2)s\exp(-\pi s^2/4)$ for the Gaussian orthogonal ensemble (GOE) in the presence of time-reversal symmetry and $P(s) = (32/\pi^2)s^2\exp(-4s^2/\pi)$ for the Gaussian unitary ensemble (GUE) in the presence of magnetic fields or magnetic impurities breaking the time-reversal symmetry. The level-spacing fluctuations are given by $\delta s = (\langle s^2\rangle - \langle s\rangle^2)^{1/2} = 0.53\langle s\rangle$ for the GOE, and $\delta s = 0.42\langle s\rangle$ for the GUE. This simplest model is called the CI-plus-RMT model (CI + RMT). When the spin degree of freedom is included in the CI model, two electrons with opposite spins occupy each level, and the spacing distribution has a bimodal behavior according to $\Delta\mu = E_c + \Delta\varepsilon$ for even numbers of electrons and $\Delta\mu = E_c$ for odd numbers of electrons. This model is the CI-plus-spin-degenerate-RMT (CI + SDRMT) model.

Experiments on the level statistics in Coulomb blockade oscillations have been carried out [131–135]. Sivan et al. measured the peak distribution in disordered GaAs and In_2O_{3-x} QDs and found unimodal Gaussian distributions with width $\delta s \sim (0.1 - 0.15)e^2/C \sim (2 - 3)\langle s\rangle$ which is larger than expected from RMT. By performing an exact diagonalization of the Anderson Hamiltonian, they showed that, as the on-site Coulomb interaction increases, the distribution approaches the Gaussian with width $\delta s \sim (0.1 - 0.2)e^2/C$, which is consistent with their experimental observation. Simmel et al. [132] have obtained similar results for lateral GaAs dots. Coulomb oscillations have also been measured in Si dots formed by the modulated gate technique in the metal–oxide–semiconductor field-effect transistor (MOSFET) geometry and the distribution with $\delta s \sim 0.06e^2/C$ has been obtained [134]. Patel et al. [133] have observed hundreds of clear Coulomb peaks in lateral QDs, as shown in Figure 10a. Like other experimental observations, these peaks have unimodal distributions fitted to Gaussian curves as shown in Figure 10b. However, the width of fluctuations is estimated as $\delta s \sim 0.019e^2/C$, which is much smaller than obtained in other experiments and is rather comparable to the single-particle-level spacing. The width becomes about 15% narrower by applying magnetic fields as seen in Figure 10c. The parameters obtained in four experiments are summarized in Table 1. Interaction parameter $r_s \sim 1$, where $r_s = \sqrt{1/(\pi a_B^{*2} n_{2DEG})}$ (n_{2DEG} is the electron

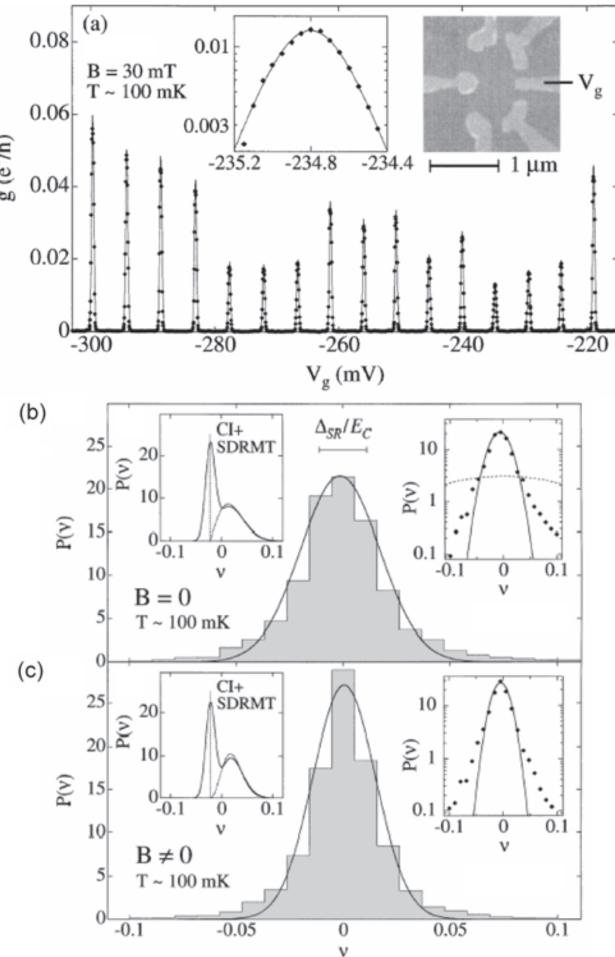


Figure 10. (a) Conductance peaks (diamonds) at $B = 30$ mT as a function of the gate voltage V_g for a lateral QD having a single-particle-level spacing of $\Delta = 14$ μV and charging energy $E_c = 460$ μV. Solid curve shows fits to a $\cosh^{-2}$ lineshape. Left inset: Detailed view of data and fit on a log-linear scale. Right inset: scanning electron microscope image of the device. (b), (c) Histograms of normalized peak spacing ν (bars) for (b) $B = 0$ and (c) $B \neq 0$ for three devices. Solid curves show fits to normalized Gaussian of width 0.019 (0.015) for $B = 0$ ($B \neq 0$). Data represent 4300 (10800) Coulomb blockade peaks with ~4300 (1600) statistically independent for $B = 0$ ($B \neq 0$). Horizontal bar indicates (spin-resolved) mean level spacing Δ_{SR}/E_c of three devices. Right inset: Plot of histogram (diamonds) and best fit Gaussian of width 0.13. Left inset: Dotted curves are CI + SDRMT peak spacing distributions; solid curves correspond to CI + SDRMT distributions convolved with Gaussian of width 0.009 averaged over three dots. Reprinted with permission from [133], S. R. Patel et al., *Phys. Rev. Lett.* 80, 4522 (1998). ©1998, American Physical Society.

density in 2D, and a_B^* is the effective Bohr radius) representing the strength of the Coulomb interaction at an average distance between the electrons relative to their kinetic energy. Besides the difference in the width of the distributions, all the experiments have shown unimodal distributions nicely fitted to Gaussian curves and no correlation between peak spacings and heights.

The experiments mentioned above have clearly indicated that the simple CI-plus-RMT model or its modified versions cannot properly describe the peak distributions. A lot

Table 1. Summary of parameters obtained in four experiments for Coulomb blockade oscillations in QDs.

	Sivan et al. [131]	Simmel et al. [132]	Patel et al. [133]	Simmel et al. [134]
Material	GaAs, In_2O_{3-x}	GaAs	GaAs	Si MOSFET
Level spacing Δ (μeV)	30–60	50	9–11	15
Charging energy E_c (meV)	0.5–0.8	0.3	0.38–0.58	1.9
Measured temperature (mK)	50	25	$\sim$100	320
Interaction parameter r_s	>1	$\sim$1	$\sim$1	2.1
Peak width (E_c)	0.1–0.15	0.1	0.019	0.06

of theoretical calculations that take the electron–electron interactions into account have been performed [136–151]. Blanter et al. [136] have used a random phase approximation (RPA) to construct an effective screened Coulomb potential in finite geometries. They found that peak-spacing fluctuations are enhanced with respect to the RMT but still of the order of the mean level spacing Δ. This result cannot be directly compared with the experiments ($r_s \sim 1$), because the RPA is valid only in the high-density limit $r_s \ll 1$. A self-consistent Hartree–Fock approximation [137] has also obtained a slight enhancement $\delta s \sim 0.6\Delta$ for $r_s \lesssim 1$. Self-consistent Hartree–Fock methods have been applied in the low-density limit $r_s \gtrsim 1$ [143, 145, 146] and have shown that the peak distribution rapidly approaches Gaussian as r_s increases and its width is significantly enhanced compared to the noninteracting case. On the other hand, it has been pointed out that the effect of screened Coulomb interactions treated by the density functional theory [151] enhances the level fluctuations only slightly and the width is comparable to the single-electron-level spacing, which reproduces the experimental finding by Patel et al.

Why does the width of distributions differ in experiments? There are several possible reasons for that: First, the electrostatic coupling to external electrodes could affect the peak fluctuations. It has been pointed out on basis of a self-consistent calculations that spacings in addition spectra in QDs are reduced when QDs are coupled to the leads [78]. This indicates that we have to consider the screening effect of external electrodes to explain different experimental results. The thermal effect is also an important consideration. In the experiment of Simmel et al. [134], thermal energy was comparable to the single-particle-level spacing, where the thermal broadening reduces the peak spacings [133, 144]. Third, the dots used in the experiments are rather ballistic than diffusive, which requires an estimate of fluctuations based on correlation between wavefunctions in chaotic systems. Several calculations have shown that scars along periodic orbits can enhance peak spacing distributions [152, 153].

Recently, bimodal distribution of peak spacings has been obtained in a chaotic QD of GaAs, where interactions are weak ($r_s = 0.72$) [154]. The Coulomb blockade peaks show a pair correlation in both position and amplitude due to spin pairing. The distribution can be approximated by a modified bimodal Wignar–Dyson distribution composed of two components in the presence of spin splitting.

3. COUPLED QUANTUM DOTS (ARTIFICIAL MOLECULE)

In the previous section, we have seen that QDs act as artificial atoms and that Hund's high-spin states or Kondo resonance states, which have been observed in real atoms or metals, also emerge in the QD system in a tunable manner. In this section, we will consider two (or several) QDs coupled with each other electrostatically and/or by tunneling barriers. Such coupled QDs are called artificial molecules. When the tunneling coupling is small, the orthodox Coulomb blockade theory treated in Section 2.1 can be applied to electrostatically coupled QDs. Single-electron turnstiles [155, 156] and pumps [157, 158] in metallic dots are interesting devices having electrostatic interdot couplings. Lateral-type QDs have been the most widely used in the study of artificial molecules, in which gate voltages can tune the coupling strength between separate dots. It has been proposed that a coherently coupled QD can be used for a *qubit* (quantum bit, or two-level quantum system) in quantum computation [159]. Computation proceeds by a succession of coherent interactions between specific pairs of qubits. Quantum computation can be viewed as a coherent superposition of digital computations proceeding in parallel. This parallelism enables prime factoring, which has been viewed as intractable. Successful experimental coupling and entangling of quantum states in QD molecules has been reported [160], in which the coherent tunneling of an electron–hole pair (exciton) causes an energy splitting as a function of the separation between the dots.

3.1. Quantum Dot Molecules

In quantum dot artificial molecules, electrons can move between spatially separated dots, and electrostatic interaction strongly influences the distribution of charges. Transport spectroscopy in laterally coupled QDs has been studied in experiments [161–166] and theoretical calculations [167–171]. Waugh et al. [161] have reported on low-temperature tunneling measurements through coupled QDs by adjusting interdot coupling tuned by gates. When the tunneling coupling is small, the Coulomb blockade oscillations have the same peak spacing as in a single QD. As the conductance through the tunneling barrier increases, the conductance peaks split. When the tunnel conductance approaches e^2/h, the peak spacing is half as large as for a single QD, indicating that the two QDs have combined into a single

large QD with twice the gate capacitance. Similar peak splittings have been observed in vertically coupled QDs [166, 172, 173], which are separated by a thin barrier. The strength of coupling varies inversely with the barrier thickness and height, preserving the circular symmetry. By applying magnetic fields perpendicularly to the pillar in vertical QDs, the coupling strength can be modulated [174].

Multiple QDs can also be coupled by coherent tunneling. Coherent tunneling between coupled QDs has been experimentally observed [164, 173, 175–178], and many theories have investigated the coherent properties in coupled QDs [14, 17, 30, 179–197]. When the barrier thickness (or barrier height) is small and the coupling is strong, the QD molecule is regarded as a single QD; the shell structure and Hund's state have been observed in addition spectra of vertical double-dot systems with circular symmetry [172], which are in qualitative agreement with Hartree–Fock calculations [14]. Weak coherent coupling between two QDs recombines states into bonding and antibonding states, resulting in peak splitting of Coulomb oscillations.

Magnetic field effects have been observed in vertically coupled QDs [166, 198]. It has been found that the stability of the spin-polarized maximum-density droplet is considerably reduced compared to the single-dot system. Weak tunneling coupling splits the lowest Landau levels into bonding and antibonding levels, which are occupied by spin-polarized electrons in strong magnetic fields. The bonding (antibonding) state can be represented as a pseudospin-up (-down) state [199, 200]. This pseudospin (or isospin) representation has been used to describe double-layer systems in the quantum Hall regime under strong magnetic fields [201]. In this representation, a pair of electrons occupying the bonding and antibonding levels has zero pseudospin. As the magnetic field is increased further, electrons in the outer pseudospin-down orbital move to the pseudospin-up orbital one by one to relax Coulomb repulsion in strong magnetic confinement. This phenomenon is called the pseudospin-flip effect. An experimental attempt to observe this effect in vertical QDs has been made [166]. In QD molecules, the spin-blockade effect [188, 189, 191, 202] and pseudospin-blockade effect [195] have been discussed theoretically.

3.2. Electron Correlation in the Two-Site Hubbard Model

We have seen that a single-particle model, such as the CIE model or Hartree–Fock approximation, works very well in describing the addition spectra of single QDs having multilevels. When multiple QDs are coupled and electrons can move between spatially separated QDs, the electrons tend to avoid each other spatially and correlation effects become important. To examine this effect, let us consider a two-site Hubbard model where two electrons occupy one of two sites. Each site has a single-particle level $\varepsilon = 0$ and on-site Coulomb repulsion U. The transfer integral between two sites is given by t ($t > 0$) and the Hubbard Hamiltonian by $\mathscr{H} = -t\sum_\sigma (c^+_{1\sigma}c_{2\sigma} + c^+_{2\sigma}c_{1\sigma}) + U(n_{1\uparrow}n_{1\downarrow} + n_{2\uparrow}n_{2\downarrow})$. The possible six states for two electrons are expressed by $|\uparrow,\uparrow\rangle = c^+_{1\uparrow}c^+_{2\uparrow}|0\rangle$, $|\downarrow,\downarrow\rangle = c^+_{1\downarrow}c^+_{2\downarrow}|0\rangle$, $|\uparrow,\downarrow\rangle = c^+_{1\uparrow}c^+_{2\downarrow}|0\rangle$, $|\downarrow,\uparrow\rangle = c^+_{1\downarrow}c^+_{2\uparrow}|0\rangle$, $|\uparrow\downarrow,0\rangle = c^+_{1\uparrow}c^+_{1\downarrow}|0\rangle$, $|0,\uparrow\downarrow\rangle = c^+_{2\uparrow}c^+_{2\downarrow}|0\rangle$. It is easy to verify that the nonbonding states $|\uparrow,\uparrow\rangle$, $|\downarrow,\downarrow\rangle$ are eigenstates with eigenenergy $E = 0$ and total spin $S = 1$. The Hamiltonian on the basis of the remaining four states is given in the matrix form by

$$\begin{array}{c} \quad |\uparrow,\downarrow\rangle \ |\downarrow,\uparrow\rangle \ |\uparrow\downarrow,0\rangle \ |0,\uparrow\downarrow\rangle \\ \mathscr{H} = \begin{bmatrix} 0 & 0 & -t & -t \\ 0 & 0 & t & t \\ -t & t & U & 0 \\ -t & t & 0 & U \end{bmatrix} \begin{array}{l} \langle\uparrow,\downarrow| \\ \langle\downarrow,\uparrow| \\ \langle\uparrow\downarrow,0| \\ \langle 0,\uparrow\downarrow| \end{array} \end{array} \tag{24}$$

Diagonalizing this Hamiltonian, we obtain four solutions comprising eigenenergy, eigenstate, and total spin given by

$$\begin{array}{lll} E_{\rm GS} = \frac{U-\sqrt{U^2+16t^2}}{2} & \phi_{\rm GS} = \frac{1}{\sqrt{1+\alpha^2}}(\phi_{\rm covalent} + \alpha\phi_{\rm ionic}) & S=0 \\ E_1 = 0 & \phi_1 = \frac{1}{\sqrt{2}}(c^+_{1\uparrow}c^+_{2\downarrow} + c^+_{1\downarrow}c^+_{2\uparrow})|0\rangle & S=1 \\ E_2 = U & \phi_2 = \frac{1}{\sqrt{2}}(c^+_{1\uparrow}c^+_{1\downarrow} - c^+_{2\uparrow}c^+_{2\downarrow})|0\rangle & S=0 \\ E_3 = \frac{U+\sqrt{U^2+16t^2}}{2} & \phi_3 = \frac{1}{\sqrt{1+\alpha^2}}(\phi_{\rm covalent} - \alpha\phi_{\rm ionic}) & S=0 \end{array} \tag{25}$$

Here, $\alpha = (-U + \sqrt{U^2+16t^2})/4t$, the covalent state $\phi_{\rm covalent} = (c^+_{1\uparrow}c^+_{2\downarrow} + c^+_{2\uparrow}c^+_{1\downarrow})|0\rangle/\sqrt{2}$, and the ionic state $\phi_{\rm ionic} = (c^+_{1\uparrow}c^+_{1\downarrow} + c^+_{2\uparrow}c^+_{2\downarrow})|0\rangle/\sqrt{2}$. The $\phi_{\rm covalent}$ is also called the Heitler–London orbital. The ground-state wavefunctions and energies for $U/t > 0$ are denoted by $\phi_{\rm GS}, E_{\rm GS}$, and for the nth excited state by ϕ_n, E_n. The first excited state ϕ_1 is one of three nonbonding states with $E = 0$ and $S = 1$ having z-component of the spin $S_z = 0$. The parameter α represents the proportion of the ionic state with respect to the covalent state. For $U/t = 0, \alpha = 1$ and the covalent and ionic states contribute equally to the ground state. For $U/t \gg 1$, $\alpha \simeq 2t/U \ll 1$ and the proportion of the ionic state decreases with increasing Coulomb repulsion, which raises the total energy of the two electrons occupying the same site. The total energy of the ground state for $U/t \gg 1$ is given by $E_{\rm GS} \simeq -4t^2/U$. This dependence can also be deduced by performing a perturbation expansion with respect to t when $U \gg t$. In the unperturbed state for t=0, each site has one electron. When we turn on small $t > 0$ between two sites, the electron on one site can move to the other and go back when two electrons have opposite spins. This second-order process gives an energy decrease of about $-t^2/U$. This effective interaction is called kinetic exchange interaction. The nonbonding states $|\uparrow,\uparrow\rangle$, $|\downarrow,\downarrow\rangle$, $(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)/\sqrt{2}$ at $E = 0$ are equivalent to the state of two uncoupled sites, each having one electron with the same spin because, by the Pauli principle, an electron on one site cannot transfer to the other. As a result, the energy is independent of t and U.

The ground state properties of the two-site Hubbard model cannot be explained by a Hartree–Fock (HF) approximation. The thermal average of the number of electrons for this system is given by $\langle n_{i\sigma}\rangle = n/2$, where n is the average number of electrons per site. The HF approximation gives $\langle n_{i\uparrow}n_{i\downarrow}\rangle \to \langle n_{i\uparrow}\rangle\langle n_{i\downarrow}\rangle = n^2/4$, so that the Coulomb term $Un^2/2 = U/2$ is independent of t. The ground state is given by the bonding state filled by up- and down-spin electrons; that is, $\frac{1}{\sqrt{2}}(\phi_{\rm covalent} + \phi_{\rm ionic})$. This state corresponds to the weak interaction limit ($U = 0$ or $\alpha = 1$) in the exact ground state. The mean-field approximation fails to include

the correlation effect between electrons with opposite spin. This simple example clearly shows the essential role of correlation effects in coupled QDs, as will be discussed in the following section.

3.3. Ferromagnetic State in Quantum Dot Molecules

To see that electron correlations and magnetic properties are closely related, we consider two types of simple QD molecules, which we will call a Lieb molecule (Fig. 11a) and a triangular molecule (Fig. 11b). Omitting the Coulomb repulsion ($U = 0$) in the Hubbard model, we obtain the single-particle Hamiltonian

$$\begin{bmatrix} 0 & -t & -t & -t \\ -t & 0 & 0 & 0 \\ -t & 0 & 0 & 0 \\ -t & 0 & 0 & 0 \end{bmatrix} \mathbf{c} = E\mathbf{c} \quad \text{(Lieb molecule)} \tag{26}$$

$$\begin{bmatrix} 0 & -t & -t \\ -t & 0 & -t \\ -t & -t & 0 \end{bmatrix} \mathbf{c} = E\mathbf{c} \quad \text{(triangular molecule)} \tag{27}$$

It can be easily verified by substitution that the eigenvalues and eigenstates are given by $E = 0, 0, \pm\sqrt{3}t$, $\mathbf{c} = (0, 1, 0, -1)$, $(0, 0, 1, -1)$, $(\pm\sqrt{3}, 1, 1, 1)$ for the Lieb molecule and $E = -2t, t, t$, $\mathbf{c} = (1, 1, 1), (1, -1, 0)$, $(1, 0, -1)$ for the triangular molecule. The second and third lowest levels are doubly degenerate and $N = 4$ corresponds to half filling of these degenerate levels. Let four electrons interact with each other by the on-site Coulomb repulsion in these molecules. Diagonalizing the Hubbard Hamiltonian numerically, the ground states in both molecules have $S = 1$ high spin. Figure 12 plots the energy difference $\Delta E = E(S = 0) - E(S = 1)$ between the ground state ($S = 1$) and the first excited state ($S = 0$) as a function of U/t. This energy difference represents the effective exchange energy between electrons in the degenerate levels. For $U/t \lesssim 1$, $\Delta E \propto U$ in both types of molecules, because, from the Pauli principle, the on-site Coulomb repulsion between electrons in the low-spin state $S = 0$ raises the energy of order U, whereas it does not for the high-spin state $S = 1$. For $U/t \gg 1$, the energy difference behaves as $\Delta E \sim t^2/U$ for the Lieb molecule, and $\Delta E \sim t$ for the triangular molecule. The mechanism of this dependence is schematically shown in Figure 13. The effective interaction for the Lieb molecule is the antiferromagnetic kinetic exchange interaction, and the spins on the center site and on the end sites are in the opposite direction, which leads to the net spin of $S = 1$. On the other hand, the high spin in the triangular molecule is caused by a ferromagnetic ring exchange interaction of third order. In this process two electrons with the opposite spins cycle within the triangular sites, which gives an energy loss of order t and favors a parallel spin [203].

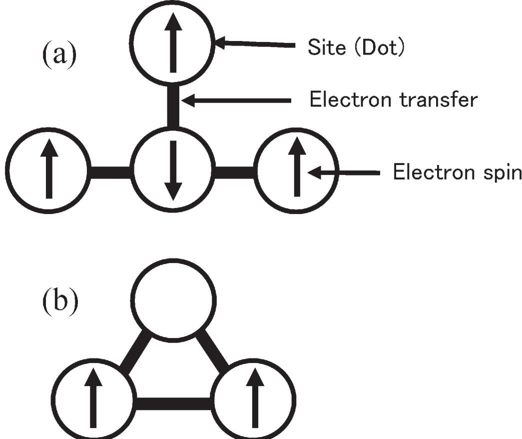


Figure 11. Schematics of quantum dot molecules. (a) A Lieb molecule and (b) a triangular molecule.

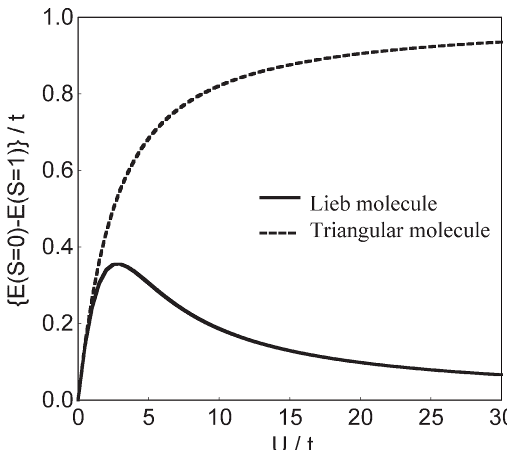
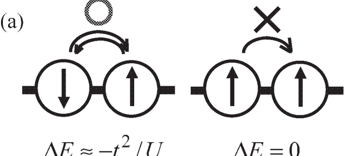


Figure 12. Calculated energy difference between the lowest excited state with spin $S = 0$ and the ground state with spin $S = 1$. Solid (dashed) line represents the Lieb (triangular) molecule.



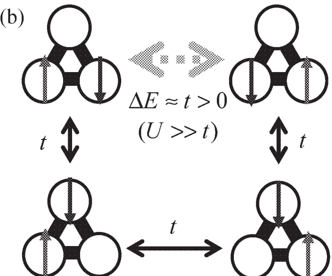


Figure 13. Effective exchange interaction for $U \gg t$. (a) Antiferromagnetic exchange interaction based on the second-order process of two electrons with opposite spins moving back and forth between two sites. (b) Ferromagnetic interaction based on the third-order process of two electrons cycling in triangular sites. When two electrons have opposite spins, this process raises their energy of order t, whereas it does not do so for two electrons with the same spin.

3.4. Magnetic Field Effect in Quantum Dot Molecules

In real molecules, the interatom distance a is very small and an extremely high magnetic field is necessary to induce magnetic field effects. In QD molecules, the magnitude of magnetic fields necessary to observe such effects is quite reasonable, because the interdot distance is typically $a = 10–1000$ nm. For example, one flux quantum h/e threading inside the triangular area $S = \sqrt{3}a^2/4$ in Figure 11b of $a = 10$ Å (100 nm) corresponds to $B \approx 10000$ T (1 T). Let us see what happens in the triangular QD molecule when it is threaded by an Aharonov–Bohm magnetic flux $\phi = BS/(h/e)$ in the triangle. The magnetic flux changes the phase of the transfer integrals as shown in Figure 14a. For $U = 0$, the tight-binding Hamiltonian is given by

$$\begin{bmatrix} 0 & -te^{\pi i\phi} & -t \\ -te^{-\pi i\phi} & 0 & -te^{\pi i\phi} \\ -t & -te^{-\pi i\phi} & 0 \end{bmatrix} \mathbf{c} = E\mathbf{c} \tag{28}$$

By solving this equation, the single-particle levels are given by $E = -2t\cos(2\pi\phi/3), 2t\cos\{2\pi(\phi \pm 1)/3\}$, where the doubly degenerate levels $E = t$ for $\phi = 0$ are lifted for $\phi > 0$. By taking into account the on-site Coulomb repulsion U and solving the Hubbard model for this molecule, we can show that the magnetic field induces a singlet–triplet transition for $N = 4$, as depicted in Figure 14b. The applied magnetic field splits the degenerate levels and the high-spin state is broken when the level splitting exceeds the exchange energy.

Magnetic field effects in several types of QD molecules have been discussed in refs. [185, 190, 204]. Yu et al. [185] used an extended Hubbard model to calculate addition spectra of linear chains and ring-shaped arrays of QDs and showed the doubly degenerate peak is split by the Aharonov–Bohm magnetic flux threaded inside the ring. Stafford et al. [190] discussed coherent resonant tunneling through a linear array of QDs in inhomogeneous magnetic fields and found a giant magnetoresistance caused by magnetic disorder. Kolehmainen et al. [204] employed the current spin-density-functional theory and discussed spin ordering in coupled QDs. They showed that four coupled dots with 40 electrons in a square or row have a ferromagnetic ground state.

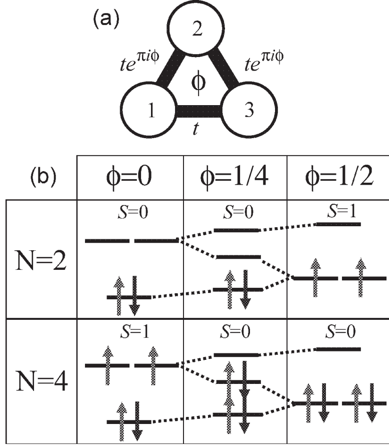


Figure 14. Magnetic field effect in triangular molecules. (a) ϕ is the magnetic flux threading the triangular area. (b) Schematic of energy levels as a function of the magnetic flux for $N = 2$ and 4. Applied magnetic flux lifts the degenerate levels, resulting in a spin-singlet–triplet transition.

4. QUANTUM DOT ARRAYS

In this section, we review the study of arrays of quantum dots. Here, we consider two types of dot arrays. One is an array of electrostatically coupled dots, where electron tunneling between dots occurs incoherently. This type of dot array is represented by a quantum dot cellular automaton. This device implementing Boolean logic functions can process and carry analog information. The other type of array is represented by a periodic lattice of dots in which coherent coupling between dots makes a band structure in the energy spectrum of electrons. This type is called a quantum dot superlattice. A wide range of lattice patterns is possible with tunable carrier concentration and coupling strength between dots. This enables us to achieve various correlation effects, such as ferromagnetism, in semiconducting nanomaterials.

4.1. Quantum Dot Cellular Automaton

Lent et al. have proposed a new device called a quantum dot cellular automaton (QCA) [205–209]. The QCA device consists of arrays of electrostatically coupled cells, each containing several coupled QDs. The electrons can tunnel between dots in the same cell but not between different cells. A basic cell consists of four dots in a square and is charged with two excess electrons. Coulomb repulsion makes these electrons arrange in one of two diagonal opposite positions, representing a binary logic state, 0 or 1. Long-range electrostatic Coulomb interaction provides an intercellular coupling between different cells. The overall state of the QCA device is determined by the boundary condition on edge cells, which act as the input channels. The QCA scheme can implement Boolean logic functions and perform general computational tasks [208]. When coupled cells interact with neighbors within a certain range, they can be viewed as an analog nonlinear (or neural) network [210].

Although the idea of QCAs seemed quite concrete and attractive, some of the proposals have been criticized by Bandyopadhyay [211–213] and Landauer [214]. Bandyopadhyay [213] argued that logic signals in the QCA device cannot propagate unidirectionally from the input to the output channel and the input bit cannot uniquely determine the output bit. To overcome this problem, Bandyopadhyay proposed an alternative type of logic architecture based on single-electron spin effects in QDs [213].

Experimental realization of QCA devices has been reported [215–222]. Orlov et al. [215] built a basic cell of QCAs consisting four metallic dots in Al/AlO_x/Al tunnel junctions and successfully achieved the polarization of two electrons by controlling an external gate voltage. Amlani et al. [220] demonstrated QCA majority gate operations in metallic QDs, where the inputs are driven by biased external electrodes and the output is detected by two single-dot electrometers.

In these devices, the polarization of bit cells is retained by the biased input electrodes and the electrostatic forces of neighboring cells. When the input polarization is switched, unidirectionality prevents the signal propagation. Lent et al. [223, 224] proposed a new type of QCAs which ensures unidirectionality. In the basic cell, three QDs are connected in series, where the middle dot acts as an interdot barrier between the top and bottom dots. By lowering or raising the bias to the middle dot, the tunneling between the two dots is allowed or suppressed. Whether the electron moves to the top or bottom dot depends on the input signal. Once the polarization is determined, it is retained (latched) after the input bias is removed. Orlov et al. [225, 226] experimentally demonstrated the clocked switching of a single electron in this type of QCA (see Fig. 15). They formed a QCA half-cell consisting of two QDs with an extra dot between them (Fig. 15a and b). Each dot of the device is capacitively coupled to a corresponding gate electrode (Fig. 15c), and the input signal is applied to the top and bottom gates. As demonstrated in Figures 15d–f, the clocking bias drives the charge state of the QCA half-cell. Although the clocked type of QCA can retain bit information much better than the original QCA scheme, a drawback arises in constructing the logic design: the number of clocked gates must increase with the number of cells, which may restrict scalability of the QCA logic architecture.

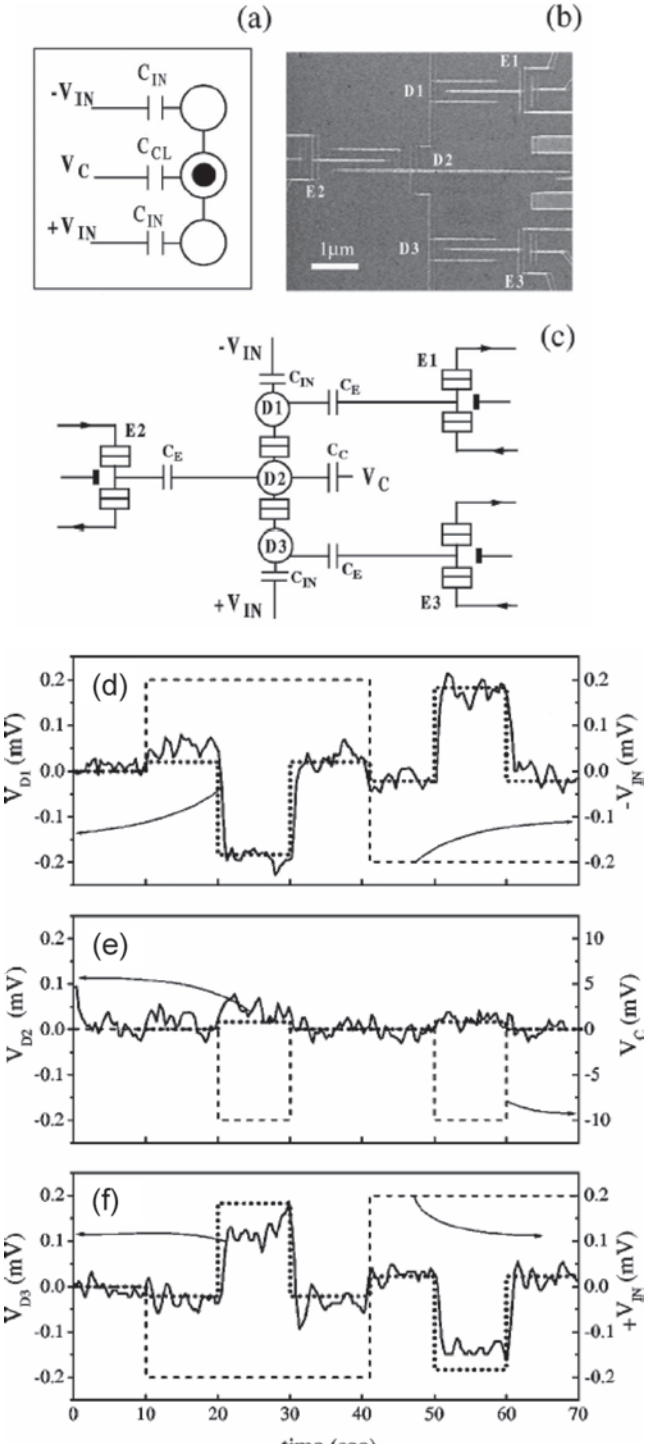


Figure 15. (a) Schematic of a QCA half cell. (b) Scanning electron microscopy image of the device utilizing $Al/AlO_x/Al$ tunnel junctions. (c) Schematic diagram of the experimental circuit. A single electron moves between the top (D1) and the bottom (D3) quantum dots through the middle dot (D2), which acts as a barrier and locks the electron in either D1 or D3 when it is biased. (d)–(f) Measured (solid lines) and calculated (dotted lines) responses to the input and clocking signals together with applied input signals (dashed lines) of the top (d), middle (e), and bottom (f) dot. Reprinted with permission from [225], A. O. Orlov et al., *Appl. Phys. Lett.* 77, 295 (2000). ©2002, American Institute of Physics.

4.2. Quantum Dot Superlattices

Here, we focus on another type of QD array in which the coherent coupling between QDs forms a band structure in the energy spectrum. Kouwenhoven et al. studied the magnetotransport properties of an artificial one-dimensional (1D) crystal of 15 QDs defined in the 2DEG of GaAs/AlGaAs heterostructures by means of a split-gate technique [227]. They observed large oscillations with two deep dips enclosing 15 small oscillations. They argued that the larger oscillations correspond to large energy gaps of the 1D crystal of QDs and the smaller ones to the discrete levels in the miniband. Haug et al. [228] built linear chains of QDs consisting of one to four dots. Transport measurements of these devices show sharp resonant conductance peaks around threshold, which are attributed to the Coulomb charging and quantum confinement effects.

Phase transition [229–235], correlated electron transport [184, 190, 236–238], and addition spectra [185, 229, 239] in linear chains of coupled QDs have been theoretically studied. Stafford and Das Sarma [181] studied the addition spectra in a 1D array of QDs and obtained quantum phase transitions, such as the Mott–Hubbard metal–insulator transition. Ugajin [230–232] predicted a Mott–Hubbard metal–insulator transition driven by an external electric field. This effect provides a method of modulating collective excitations locally and can be applied to new field-effect devices.

Engineering of electronic and magnetic properties of a lattice of QDs has also been investigated [240–247]. Sugimura [240] theoretically proposed the possibility of magnetic ordered states in a semiconductor quantum dot array. Describing the GaAs/AlGaAs QD system by the one-band Hubbard model, he showed that magnetic orderings, such as ferromagnetism and antiferromagnetism, can be expected in

QD arrays. Khurgin et al. showed that exchange interactions between holes confined in arrays of strained $In_{1-x}Ga_xAs$ QDs (see Fig. 16) can lead to *ferrimagnetic* arrangement of magnetic moments. The ferrimagnetism is induced by coupling two neighboring QDs having different values of orbital angular momenta. The Curie temperature for QDs with thickness of $t_1 = 7$ nm and $t_2 = 3$ nm is of the order of 25 K (see Fig. 16d). Khurgin et al. also proposed a method of engineering ferroelectric arrays of coupled strained QDs [242].

Lattice structure and carrier concentration are essential factors determining the electronic and magnetic properties in solids. Carrier concentration in solids can be controlled by chemical or gate-induced doping [248, 249]. On the other hand, building a desired lattice of atoms is very hard, because the lattice structure is strictly determined by the atomic nature of constituent elements. The design of materials with desired lattice is possible by putting QD atoms on the lattice points and coupling the QDs with each other to form artificial crystals called quantum dot superlattices (QDSLs). Semiconductor fabrication technology enables us to form QDSLs in various patterns as we wish, including lattices not existing in nature. The electron concentration can be tuned by the gate voltages. Therefore, QDSLs are suitable for achieving various interesting effects, such as flat-band ferromagnetism [244–247], which have been considered unlikely to occur in semiconductors.

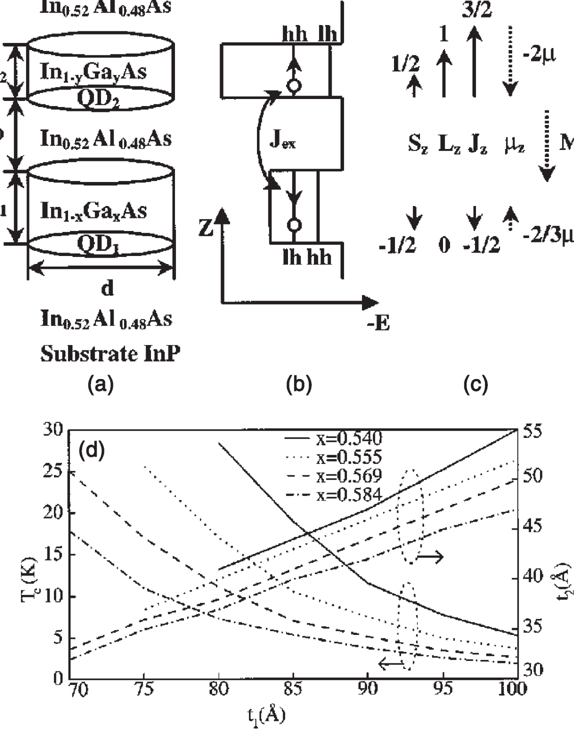


Figure 16. (a) One period of a QD array of $In_{1-x}Ga_xAs$ and $In_{1-y}Ga_yAs$ buried in $In_{0.52}Al_{0.48}As$. (b) Schematic of the valence-band alignment of the QD array and (c) directions of spin and orbital angular momenta. (d) Curie temperature T_c and the thickness of unstrained QD_2-t_2 as functions of the thickness of strained QD_1-t_1 for different compositions x of strained $In_{1-x}Ga_xAs$. Lateral size $d = 120$ Å. Barrier thickness $t_b = 20$ Å. Reprinted with permission from [241], J. B. Khurgin et al., *Appl. Phys. Lett.* 73, 3944 (1998). ©1998, American Institute of Physics.

4.3. Flat-Band Ferromagnetism

It has been proved that the Hubbard model on specific types of lattice having a flat band exhibits ferromagnetism in the presence of on-site Coulomb interaction [203, 250–255]. The flat band is a dispersionless subband in the single-particle band structure. Figure 17 shows two examples of lattices having flat bands, where (a) is a Lieb lattice and (b) is a Kagome lattice. The Hubbard Hamiltonian for these lattices is

$$H = -t \sum_{(i,j)\sigma} c^{+}_{i\sigma} c_{j\sigma} - t' \sum_{\langle i,k\rangle\sigma} c^{+}_{i\sigma} c_{k\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow} \tag{29}$$

where t (t') is the transfer integral between the (next) nearest neighbor sites (i, j) ($\langle i, k\rangle$), U is the on-site Coulomb energy, $c^{+}_{i\sigma}(c_{i\sigma})$ is the creation (annihilation) operator of an electron with spin $\sigma = \uparrow$ or $\downarrow$ on site i, and $n_\sigma = c^{+}_{i\sigma} c_{i\sigma}$. The single-particle band diagrams calculated for $t' = U = 0$ are shown in Figure 18, where flat subbands can be seen. In the presence of on-site Coulomb repulsion, it has been mathematically proven that the Hubbard model exhibits flat-band ferromagnetism on the Lieb [250] and Kagome lattice [251] at zero temperatures. The mechanism of ferromagnetism in these lattices has already been explained in Figure 13. In the bipartite Lieb lattice, kinetic exchange interactions make the neighboring spins align in opposite directions (see Fig. 13a) and the different numbers of up- and down-spins result in *ferri*magnetism. The ferromagnetism in the Kagome lattice is caused by a ring exchange interaction of the third-order process cycling two electrons with the opposite spins within the triangular lattice of three sites (see Fig. 13b). When the next nearest neighbor transfer t' is included, the flat band is broken as shown in Figure 18. It is very important to know how t' affects the flat-band ferromagnetism, because the next nearest neighbor transfer t' is not negligible in QDSLs [247]. To include the effect of t', we consider a realistic potential in QDSL. The transfer and on-site Coulomb energies for QDSL are evaluated by assuming that electrons are

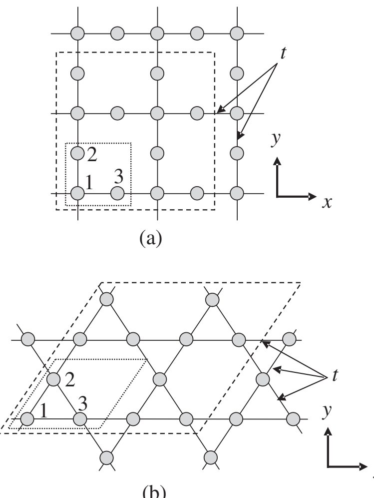


Figure 17. The Lieb lattice (a) and Kagome lattice (b). Solid lines represent the nearest neighbor transfer t. Dotted and dashed lines indicate one and 2×2 unit cell(s). A unit cell contains three sites $i = 1, 2, 3$.

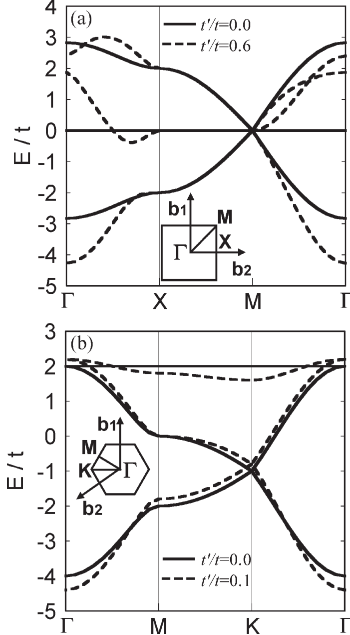


Figure 18. Single-particle band energy E for the (a) Lieb lattice and (b) Kagome lattice for the next nearest neighbor transfer $t' = 0$ (solid line) and $t' > 0$ (dashed line) calculated using the tight-binding approximation, where E and t' are normalized in units of the nearest neighbor transfer t. Insets: Brillouin zone. $\mathbf{b}_1$ and $\mathbf{b}_2$ are the reciprocal lattice vectors. Reprinted with permission from [247], H. Tamura et al., *Phys. Rev. B* 65, 085324 (2002). ©2002, American Physical Society.

confined in a two-dimensional potential [256] given by

$$V(\mathbf{r}) = \sum_i v(\mathbf{r} - \mathbf{R}_i),$$

$$v(\mathbf{r}) = \begin{cases} -\dfrac{1}{16} m^* \omega^2 a^2 [\cos(\pi x/a)\cos(\pi y/a)]^2 \\ \qquad \text{for } r < a/2 \\ 0 \qquad \text{for } r \geq a/2 \end{cases} \tag{30}$$

where m^* is the effective mass of an electron, ω is the confining oscillator frequency, $\mathbf{R}_i$ is the position of the ith dot, and a is the interdot distance. Noting that $[\cos(\pi x/a)\cos(\pi y/a)]^2 \simeq [(2r/a)^2 - 1]^2 \simeq 1 - 2(2r/a)^2$ for $r \ll a/2$, the "atomic" wave function localized in the potential $v(\mathbf{r})$ is given in a good approximation by

$$\phi(\mathbf{r}) = \frac{2}{\sqrt{\pi} d} \exp\left(-\frac{2r^2}{d^2}\right), \tag{31}$$

where $d = 2\sqrt{\hbar/m^*\omega}$ is the dot diameter.

The transfer and on-site Coulomb energies are given by

$$t(\mathbf{R}_i, \mathbf{R}_j) = -\int d\mathbf{r}\, \phi(\mathbf{r} - \mathbf{R}_i)\{-\hbar^2\nabla^2/2m^* + V(\mathbf{r})\} \cdot \phi(\mathbf{r} - \mathbf{R}_j) \tag{32}$$

$$U = \iint d\mathbf{r}_1\, d\mathbf{r}_2 \frac{e^2 |\phi(\mathbf{r}_1)|^2 |\phi(\mathbf{r}_2)|^2}{4\pi\varepsilon |\mathbf{r}_1 - \mathbf{r}_2|} = \frac{\sqrt{2\pi} e^2}{4\pi\varepsilon d} \tag{33}$$

where ε is the dielectric constant. Here, the nearest neighbor transfer t is calculated for $\mathbf{R}_i = (0, 0)$ and $\mathbf{R}_j = (a, 0)$ and the next nearest neighbor transfer t' for $\mathbf{R}_i = (a, 0)$ and $\mathbf{R}_j = (0, a)$ in the Lieb lattice and for $\mathbf{R}_i = (a, 0)$ and $\mathbf{R}_j = (a/2, \sqrt{3}a/2)$ in the Kagome lattice. For $d/a \lesssim 0.5$, t' is much smaller than t and negligible.

An exact diagonalization of the Hubbard Hamiltonian of 2×2 unit cells has been done using Eqs. (32) and (33) [247]. In realistic dot arrays formed by a negatively biased gate-electrode depleting the underneath 2DEG, the interdot distance is usually fixed and cannot be changed. By modifying the gate voltage, the dot diameter can be changed. To simulate this, we evaluate t, t', and U from Eqs. (32) and (33) for a fixed interdot distance and diagonalize the Hubbard Hamiltonian in Eq. (29) for a half filled flat band. Figure 19 plots the energy difference $\Delta E = E(S = 0) - E(S = 2)$ between the ground state with $S = 2$ and the lowest excited state with $S = 0$ as a function of dot diameter. This energy difference can be regarded as a qualitative estimate of the transition temperature of ferromagnetic and paramagnetic transition in a macroscopic sample. ΔE takes a peak around $d/a = 0.5 - 0.6$, where the effect of t' is negligible and ΔE is not affected by t'. Table 2 lists estimated ΔE's in units

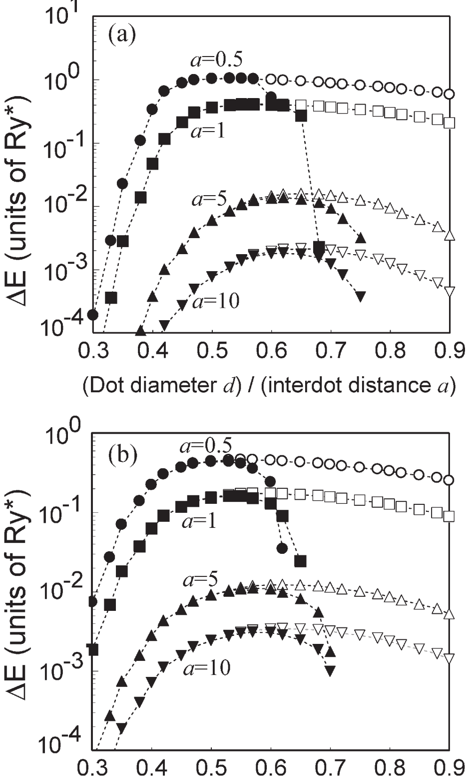


Figure 19. Energy difference $\Delta E = E(S = 0) - E(S = 2)$ as a function of dot diameter d for the (a) Lieb and (b) Kagome lattices for different interdot distances a in the dot model given in Eq. (30) represented by solid symbols. Open symbols represent the energy difference when we neglect the next nearest neighbor transfer t'. Energy and length are scaled in units of the effective Rydberg constant $Ry^* = 13.6\ \text{eV} \times (m^*/m_0)/(\varepsilon/\varepsilon_0)^2$ and the effective Bohr radius $a_B^* = 0.53\ \text{Å} \times (\varepsilon/\varepsilon_0)/(m^*/m_0)$, where m_0 and ε_0 are the mass of a bare electron and the dielectric constant in vacuum respectively. Reprinted with permission from [247], H. Tamura et al., *Phys. Rev. B* 65, 085324 (2002). ©2002, American Physical Society.

Table 2. Estimated energy difference $\Delta E = E(S=0) - E(S=2)$ (Kelvin) between the ground state with spin 2 and the lowest excited state with spin 0 for dot arrays of the kagome lattice of typical interdot distance $a = 5$, 10, 50, 100 nm and dot diameter $a/2$.

	ΔE (Kelvin) for Lieb Lattice				ΔE (Kelvin) for Kagome Lattice			
Distance	5 nm	10 nm	50 nm	100 nm	5 nm	10 nm	50 nm	100 nm
GaAs	76	26	0.5	0.06	31	11	0.6	0.2
InAs	90	43	3.5	0.6	45	19	1.7	0.5
Si	27	4.5	0.04	0.005	15	4.3	0.2	0.05

Note: We consider GaAs dots ($m^*/m_0 = 0.067$, $\varepsilon/\varepsilon_0 = 12.4$, $Ry^* = 6$ meV, $a_B^* = 10$ nm), InAs dots ($m^*/m_0 = 0.02$, $\varepsilon/\varepsilon_0 = 12.4$, $Ry^* = 1.8$ meV, $a_B^* = 34$ nm), and Si dots ($m^*/m_0 = 0.2$, $\varepsilon/\varepsilon_0 = 12$, $Ry^* = 19$ meV, $a_B^* = 3$ nm). *Source:* Reprinted with permission from [247], H. Tamura et al., *Phys. Rev. B* 65, 085324 (2002). © 2002, American Physical Society.

of Kelvin for dot arrays of various sizes and materials. We consider GaAs, InAs, and Si dots and assume $a = 2d$. For dots with a distance of 100 nm which is available within the present fabrication technology, ΔE is of the order of several hundreds milli-Kelvin and we can expect that ferromagnetism can be observed in the low temperature region. For dots of distance 5 nm, ΔE is as high as a few ten Kelvin.

Ideas for achieving flat-band ferromagnetism based on carbon networks [257, 258], graphite ribbons [259, 260], and Ga and As atomic wires [261, 262] have also been proposed. However, there has been no clear evidence for flat-band ferromagnetism in these materials. This is supposedly due to lattice distortion effects that would destabilize the ferromagnetism when the flat band is half filled [263]. In real materials, the number of valence electrons is determined in such a way that the crystal structure is stable. It follows then that unrestricted material design is difficult in general. The QDSL does not have such disadvantages and is suitable for achieving the flat-band ferromagnetism with tunable parameters.

One may think that a finite magnetization could not appear at finite temperatures in two dimensions because of the absence of long-range order described in the Mermin–Wagner theorem [264]. However, it should be noted that this theorem is applied in the thermodynamic limit. When a sample has a finite size smaller than the spin–spin correlation length, a finite magnetization can appear [247].

4.4. Quantum Wire Geometry for QDSLs

We discuss another implementation of the QDSL system. Schedelbeck et al. [265] have fabricated coupled QDs, each of which is formed in the intersection of three quantum wells, by using cleaved edge overgrowth [266] and demonstrated a coherent peak-splitting between the bonding and antibonding states in the photoluminescence spectrum. The area at the intersection is larger than the normal width of the wire, and the potential energy at each intersection in which a QD is formed becomes lower. In a similar way, an array of QDs is formed at the intersections in a network of quantum wires. Such a network geometry has been fabricated in GaAs wires by selective area growth techniques [267–270] and in Si wires by combining nanolithography and nanoetching techniques [271]. In the latter case, the widths of the Si quantum wires are less than 10 nm, which would provide much smaller QDs at intersections.

Recently, a QDSL implemented in a two-dimensional Kagome quantum wire network has been suggested and the possibility of flat-band ferromagnetism has been discussed [245]. InAs quantum wires are surrounded by $In_{0.72}Ga_{0.28}As$ barriers with a band offset of 0.21 eV, as schematically shown in Figure 20. The electronic band structure was calculated using first-principles calculations based on a local density functional as shown in Figure 21a, which is very similar to the band diagram calculated by the tight-binding approximation in Figure 18b. At 5/3 filling where the flat band is half filled, the charge density is well localized at the intersections of wires which form the Kagome lattice as seen in Figure 21b. A local spin-density-functional calculation reveals the ferromagnetic spin density at the intersections, as shown in Figure 22. This ferromagnetism can be turned off by changing the filling from 5/3 to 6/3 (or to 4/3). This wire network geometry is suitable for achieving strong potential modulations needed for creation of an effective Hubbard system in the QDSL.

4.5. Magnetic Field Effects in QDSLs

In Section 3.4, we explained that transfer integrals acquire the Peierls phase factor in magnetic fields, which splits degenerate levels in the triangular QD molecule. Magnetic fields significantly affect the band structure in QDSLs as well. The energy spectrum of the tight-binding model in magnetic fields has a self-similar (fractal) structure called the Hofstadter butterfly [272], which is obtained by solving the Harper equation [273]. Figure 23 shows the Hofstadter butterfly for the square lattice [274]. The Hofstadter butterfly of other lattices having a flat band has been studied

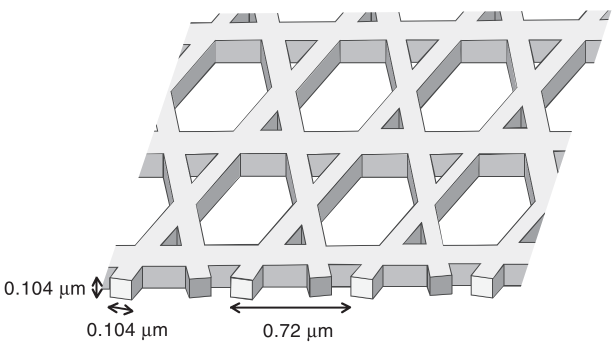


Figure 20. Schematic of a quantum-wire network effectively acting as a quantum dot superlattice. InAs wires are buried in barrier regions of $In_{0.72}Ga_{0.28}As$ with a height of 0.21 eV. Reprinted with permission from [245], K. Shiraishi et al., *Appl. Phys. Lett.* 78, 3702 (2001). ©2001, American Institute of Physics.

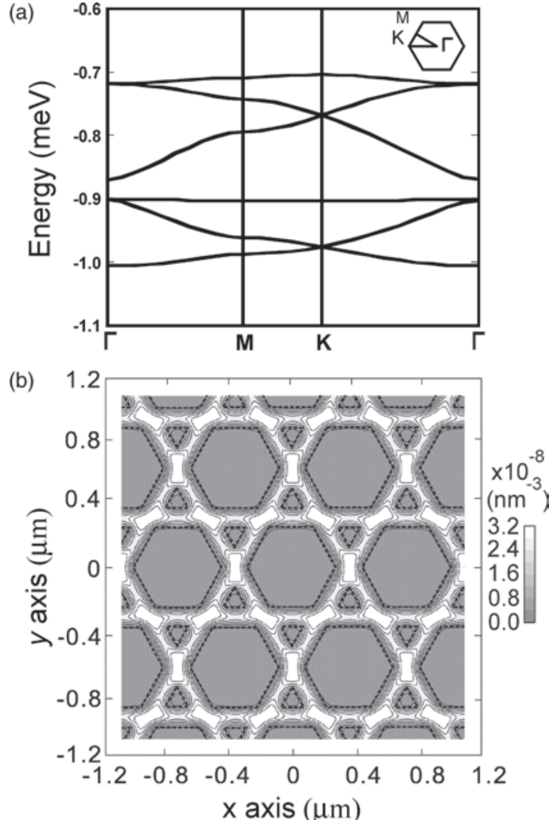


Figure 21. (a) Electronic band diagram of paramagnetic states for the Kagome wire network calculated using the local density-functional approximation which neglects the spin degree of freedom. (b) A cross-sectional view of the total charge density for filling 5/3. Reprinted with permission from [245], K. Shiraishi et al., *Appl. Phys. Lett.* 78, 3702 (2001). ©2001, American Institute of Physics.

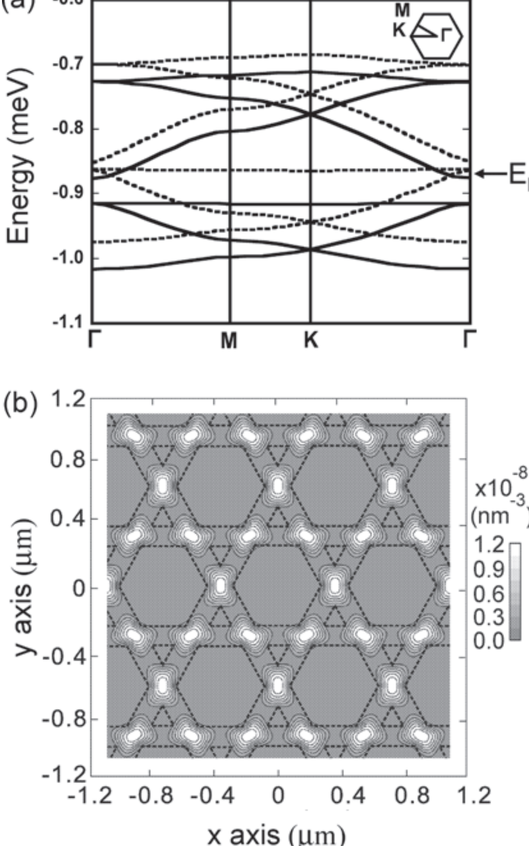


Figure 22. (a) Electronic band diagram of ferromagnetic states for the Kagome wire network calculated using the local spin-density-functional approximation which takes the spin degree of freedom into account. (b) A cross-sectional view of the total spin density of ferromagnetic states for filling 5/3. Reprinted with permission from [245], K. Shiraishi et al., *Appl. Phys. Lett.* 78, 3702 (2001). ©2001, American Institute of Physics.

in ref. [275]. It is well known that the Hall conductance is in integer multiples of e^2/h when the Fermi level lies in a gap of the butterfly spectrum [276]. In Figure 23, the quantized values of the lowest Landau level are indicated in units of e^2/h. Experimental attempts have been done over the last couple of decades to find signatures of the Hofstadter butterfly [277–280]. Albrecht et al. [280] measured the quantized Hall effect of a 2DEG in the presence of a perpendicular magnetic field and a weak lateral periodic potential with square symmetry. They observed the expected quantized values when the Fermi level is in one of the minigaps in the Hofstadter butterfly diagram as shown in Figure 24. In real QDSLs, only primary gaps separating clusters of subbands will survive due to broadening induced by disorders and lattice imperfections.

Magnetic fields also affect the transport property of QDSLs, since they influence the many-body wavefunction of electrons in them. Stafford et al. [190] considered coherent tunneling through a 3×3 array of QDs and demonstrated a giant magnetoconductance effect. In finite magnetic fields, some Coulomb blockade peaks are strongly suppressed compared with those in a zero magnetic field, as shown in Figure 25. The interdot kinetic exchange interaction causes an antiferromagnetic spin-density-wave ground state in a zero magnetic field. When an external magnetic field is applied, the electron spins tend to align with the field to minimize the Zeeman energy and a transition occurs from the spin-density-wave ground state at $B = 0$ to a ferromagnetic state at some critical magnetic field. This spin-polarization transition causes the conductance suppression of the sixth peak at $B = 1.3$ T. The seventh peak at $B = 1.3$ T is caused by the spin-blockade effect explained in Section 2.5.

The effect of magnetic fields is especially interesting in the Kagome lattice, because the flat band in the Kagome lattice originates from the interference of wavefunctions, which is disturbed by the Peierls phase in magnetic fields. It has been shown that application of a perpendicular magnetic field destroys the flat-band ferromagnetism [246] and induces a ferromagnetic–paramagnetic transition when the Coulomb interaction is strong. When the Fermi level is in the flat band, the Kagome lattice acts as a band insulator in a zero magnetic field, because the dispersionless flat band has zero group velocity. When magnetic fields are applied, the flat band is destroyed and has dispersion, giving rise to a metal–insulator transition (or "a negative magnetoresistance"). When the flat band is half filled, the ferromagnetic–paramagnetic transition and the metal–insulator one occur

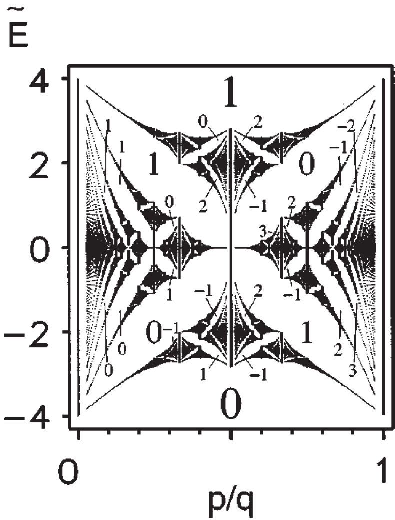


Figure 23. The scaled energy $\widetilde{E}$ versus inverse magnetic flux p/q (Hofstadter butterfly). The Landau band coupling is neglected. The numbers in the energy gaps are the quantized Hall conductances in units of e^2/h. For the Hall conductance in higher Landau bands, one has to add the Landau band index ν to these values. Reprinted with permission from [274], D. Springsguth et al., *Phys. Rev. B* 56, 2036 (1997). ©1997, American Physical Society.

simultaneously at a magnetic field when the Coulomb interaction is strong. The magnetic-field-induced metal–insulator transition in a 2D square lattice has also been discussed by An et al. [281]. These interesting phenomena can be observed in experiment under reasonable magnetic fields in QDSLs.

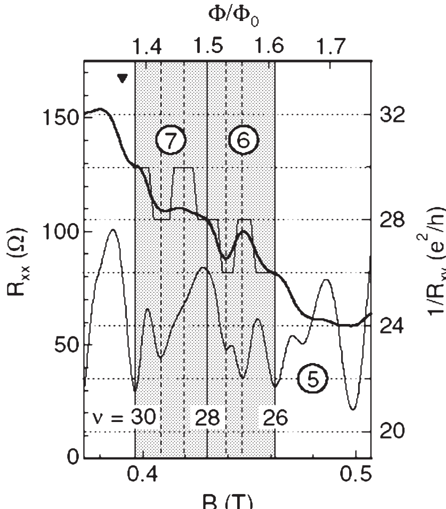


Figure 24. The longitudinal resistance R_{xx} and inverse Hall resistance, $1/R_{xy}$, for a 120 nm lattice at 50 mK near filling $\nu = 26$ to 30. Horizontal dotted lines indicate the quantized values of R_{xy}. The Hall conductance exhibits the 0–1–0–1 sequence in units of e^2/h when it is extracted by the contribution from the filled Landau levels $\nu e^2/h$. This is as expected when the Fermi level is in the right side of Hofstadter's butterfly (see Fig. 23). Reprinted with permission from [280], C. Albrecht et al., *Phys. Rev. Lett.* 86, 147 (2001). ©2001, American Physical Society.

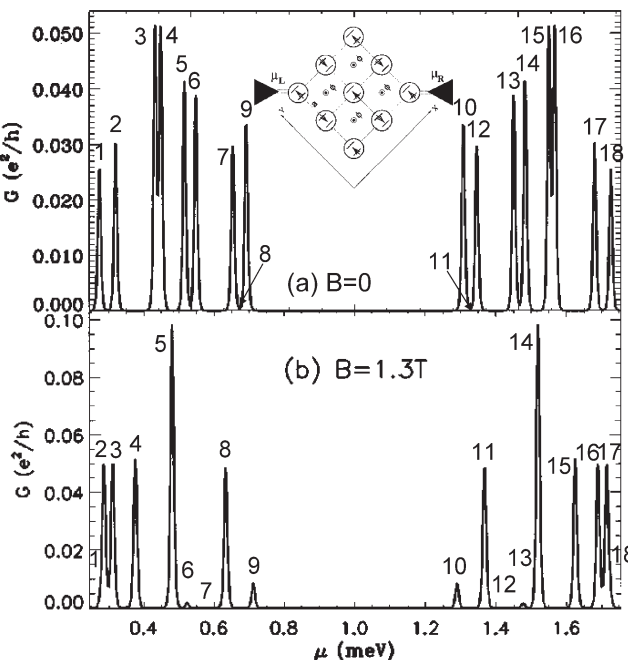


Figure 25. Conductance versus chemical potential μ through the 3×3 array of QDs at (a) $B = 0$ and (b) $B = 1.3$ T. Each peak is labeled by the number of electrons. The ground state (N, S, S_z) sequence for $B = 0$ is given by $(1, 1/2, 1/2) \rightarrow (2, 0, 0) \rightarrow (3, 3/2, 1/2) \rightarrow (4, 1, 1) \rightarrow (5, 1/2, 1/2) \rightarrow (6, 0, 0) \rightarrow (7, 1/2, 1/2) \rightarrow (8, 0, 0) \rightarrow (9, 1/2, 1/2)$, and for $B = 1.3$ T by $(1, 1/2, 1/2) \rightarrow (2, 1, 1) \rightarrow (3, 3/2, 3/2) \rightarrow (4, 2, 2) \rightarrow (5, 5/2, 5/2) \rightarrow (6, 2, 2) \rightarrow (7, 1/2, 1/2) \rightarrow (8, 1, 1) \rightarrow (9, 1/2, 1/2)$. The sixth peak at $B = 1.3$ T is strongly suppressed by the spin-polarization transition. The suppression of the seventh peak at $B = 1.3$ T is caused by the spin-blockade effect. Reprinted with permission from [190], C. A. Stafford et al., *Phys. Rev. B* 58, 7091 (1998). ©1998, American Physical Society.

5. CONCLUSIONS

The advantage of using semiconductors to achieve ferromagnetism exists not only in making semiconducting "permanent magnets." As we mentioned, the controllability of the magnetic property by changing the electron filling and magnetic field will make dot lattices useful in electronic devices, such as memories, sensors, and magnetic heads, since they contain only semiconductor materials, such as Si and GaAs. Magnetic devices can be fabricated without using any magnetic elements, such as iron and manganese, which are incompatible with the conventional large-scale-integrated-circuit fabrication process.

Regarding QDSLs, one can expect research pursuing various correlation effects that have been predicted in mathematical models of lattice systems, other than the flat-band ferromagnetism. One of the most fascinating examples is high-temperature (high-T_c) superconductivity. High-T_c d-wave superconductivity in a repulsive Hubbard model has been predicted [282, 283]. Since the energy scale is a hundred times smaller than in conventional CuO_2 high-T_c superconductors ($T_c \sim 100$ K), the transition temperature of the superconductivity in semiconductor dot arrays should be quite low: the estimated transition temperature using the

prediction is $T_c \simeq 0.01t \sim 1$ K for $a = 2d = 10$ nm GaAs dots. It has been shown that, in other types of lattices, the transition temperature becomes much higher [284]. It would be very interesting if superconductivity (or at least the Kosterlitz-Thouless-Berezinskii transition) could be achieved in semiconductors.

Other types of lattice structures are also fascinating. One is the ladder structure constructed with oxide materials. This structure exhibits various interesting properties, such as the spin gap and superconductivity [285–287]. Optical properties of dot arrays deserve future study, since the large density of states in the flat band will significantly affect photoluminescence and laser characteristics [288].

GLOSSARY

Addition energy Energy necessary to add an extra electron to a quantum dot. It is the same as the chemical potential within the dot.

Artificial atom (molecule, crystal) Single (double, or lattice of) quantum dot showing the properties of real atom (molecule or crystal). Artificial crystal is also called quantum dot superlattice.

Constant interaction model Simple model to describe the Coulomb charging effect in a quantum dot. It is usually expressed by a Hamiltonian consisting of a single-particle energy and a classical charging energy.

Flat band A subband which does not have dispersion in a band diagram. Flat band is degenerate in energy, so the electron correlation becomes strong if the Fermi level is at the flat band.

Hofstadter butterfly Fractal band diagram when the tight-binding model is threaded by magnetic fields.

Spin blockade effect When the total spins of the initial (N) and the final ($N+1$) states in a quantum dot having N electrons are different by more than 1/2, the transition is forbidden and the tunneling current is suppressed.

ACKNOWLEDGMENTS

We have benefitted from extensive interactions with numerous colleagues and collaborators over the past years. We thank Tatsushi Akazaki, Hideo Aoki, Ryotaro Arita, David G. Austing, Yuichi Harada, Wataru Izumida, Takashi Kimura, Kazuhiko Kuroki, Kenichi Matsuda, Seigo Tarucha, and Masahito Ueda for collaborations and valuable discussions. We also thank Andreas Richter for critical reading of the manuscript. Some of the work reviewed here was supported by JPSJ under Contract RTF96P00203 and the NEDO International Joint Research Grant.

REFERENCES

1. H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in "Single Charge Tunneling." (H. Grabert and M. H. Devoret, Eds.), Vol. 294 of NATO ASI Series, p. 167. Plenum Press, New York, 1992.
2. L. P. Kouwenhoven, C. M. Markus, P. L. McEuen, S. Tarucha, R. M. Westervelt, and N. S. Wingreen, in "Mesoscopic Electron Transport." (L. L. Sohn, L. P. Kouwenhoven, and G. Schön, Eds.), Vol. 345 of NATO ASI Series, p. 105. Kluwer Academic, Dordrecht, 1997.
3. U. Meirav and E. B. Foxman, *Semicond. Sci. Technol.* 10, 255 (1995).
4. H. Sakaki, *Japan J. Appl. Phys.* 28, L314 (1989).
5. L. Esaki and T. Tsu, *IBM J. Res. Develop.* 14, 61 (1970).
6. R. C. Ashoori, H. L. Stormer, J. S. Weiner, L. N. Pfeifer, S. J. Pearton, K. W. Baldwin, and K. W. West, *Phys. Rev. Lett.* 68, 3088 (1992).
7. R. C. Ashoori, H. L. Stormer, J. S. Weiner, L. N. Pfeifer, K. W. Baldwin, and K. W. West, *Phys. Rev. Lett.* 71, 613 (1993).
8. E. B. Foxman, P. L. McEuen, U. Meirav, N. S. Wingreen, Y. Meir, P. A. Belk, N. R. Belk, M. A. Kastner, and S. J. Wind, *Phys. Rev. B* 47, 10020 (1993).
9. I. K. Kurland, I. L. Aleiner, and B. L. Altshuler, *Phys. Rev. B* 62, 14886 (2000).
10. S. Tarucha, D. G. Austing, T. Honda, R. J. van der Hage, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 77, 3613 (1996).
11. V. Fock, *Z. Phys.* 47, 446 (1928).
12. C. G. Darwin, *Proc. Cambridge Philos. Soc.* 27, 86 (1930).
13. M. Macucci, K. Hess, and G. J. Iafrate, *Phys. Rev. B* 48, 17354 (1993).
14. H. Tamura, *Phys. B* 249–251, 210 (1998).
15. S. Bednarek, B. Szafran, and J. Adamowski, *Phys. Rev. B* 61, 4461 (2000).
16. B. Szafran, J. Adamowski, and S. Bednarek, *Phys. Rev. B* 61, 1971 (2000).
17. C. Yannouleas and U. Landman, *Phys. Rev. Lett.* 82, 5325 (1999).
18. T. Ezaki, N. Mori, and C. Hamaguchi, *Phys. Rev. B* 56, 6428 (1997).
19. M. Rontani, F. Rossi, F. Manghi, and E. Molinari, *Phys. Rev. B* 59, 10165 (1999).
20. B. Jouault, G. Santoro, and A. Tagliacozzo, *Phys. Rev. B* 61, 10242 (2000).
21. S. M. Reimann, M. Koskinen, and M. Manninen, *Phys. Rev. B* 62, 8108 (2000).
22. N. A. Bruce and P. A. Maksym, *Phys. Rev. B* 61, 4718 (2000).
23. M. Ferconi and G. Vignale, *Phys. Rev. B* 50, 14722 (1994).
24. S. Nagaraja, P. Matagne, V.-Y. Thean, J.-P. Leburton, Y.-H. Kim, and R. M. Martin, *Phys. Rev. B* 56, 15752 (1997).
25. M. Koskinen, M. Manninen, and S. M. Reimann, *Phys. Rev. Lett.* 79, 1389 (1997).
26. I.-H. Lee, V. Rao, R. M. Martin, and J.-P. Leburton, *Phys. Rev. B* 57, 9035 (1998).
27. K. Hirose and N. S. Wingreen, *Phys. Rev. B* 59, 4604 (1999).
28. I.-H. Lee, K.-H. Ahn, Y.-H. Kim, R. M. Martin, and J.-P. Leburton, *Phys. Rev. B* 60, 13720 (1999).
29. T. F. Jiang, X.-M. Tong, and S.-I. Chu, *Phys. Rev. B* 63, 045317 (2001).
30. I. I. Yakimenko, A. M. Bychkov, and K.-F. Berggren, *Phys. Rev. B* 63, 165309 (2001).
31. J. Harting, O. Mülken, and P. Borrmann, *Phys. Rev. B* 62, 10207 (2000).
32. F. Pederiva, C. J. Umrigar, and E. Lipparini, *Phys. Rev. B* 62, 8120 (2000).
33. A. Angelucci and A. Tagliacozzo, *Phys. Rev. B* 56, R7088 (1997).
34. S. Sinha, R. Shankar, and M. V. N. Murthy, *Phys. Rev. B* 62, 10896 (2000).
35. K. Varga, P. Navratil, J. Usukura, and Y. Suzuki, *Phys. Rev. B* 63, 205308 (2001).
36. D. G. Austing, S. Sasaki, S. Tarucha, S. M. Reimann, M. Koskinen, and M. Manninen, *Phys. Rev. B* 60, 11514 (1999).
37. S. Akbar and I.-H. Lee, *Phys. Rev. B* 63, 165301 (2001).
38. L. R. C. Fonseca, J. L. Jimenez, J. P. Leburton, and R. M. Martin, *Phys. Rev. B* 57, 4017 (1998).
39. S. Bednarek, B. Szafran, and J. Adamowski, *Phys. Rev. B* 59, 13036 (1999).

40. D. G. Austing, Y. Tokura, T. Honda, S. Tarucha, M. Danoesastro, J. Janssen, T. Oosterkamp, and L. P. Kouwenhoven, *Japan J. Appl. Phys.* 38, 372 (1999).
41. M. Eto, *Japan J. Appl. Phys.* 36, 3924 (1997).
42. A. Natori, Y. Sugimoto, and M. Fujito, *Japan J. Appl. Phys.* 36, 3960 (1997).
43. A. H. MacDonald, S. T. E. Yang, and M. D. Johnson, *Austral. J. Phys.* 46, 245 (1993).
44. O. Klein, C. de C. Chamon, D. Tang, D. M. Abusch-Magder, U. Meirav, X.-G. Wen, M. A. Kastner, and S. J. Wind, *Phys. Rev. Lett.* 74, 785 (1995).
45. R. C. Ashoori, *Nature* 379, 419 (1996).
46. T. H. Oosterkamp, J. W. Janssen, L. P. Kouwenhoven, D. G. Austing, T. Honda, and S. Tarucha, *Phys. Rev. Lett.* 82, 2931 (1999).
47. A. H. MacDonald and M. D. Johnson, *Phys. Rev. Lett.* 70, 3107 (1993).
48. P. Hawrylak, *Phys. Rev. Lett.* 71, 3347 (1993).
49. J. J. Palacios, L. Martín-Moreno, G. Chiappe, E. Louis, and C. Tejedor, *Phys. Rev. B* 50, 5760 (1994).
50. C. d. C. Chamon and X. G. Wen, *Phys. Rev. B* 49, 8227 (1994).
51. K. Ahn, J. H. Oh, and K. J. Chang, *Phys. Rev. B* 52, 13757 (1995).
52. S. M. Reimann, M. Koskinen, M. Manninen, and B. R. Mottelson, *Phys. Rev. Lett.* 83, 3270 (1999).
53. M. Ferconi and G. Vignale, *Phys. Rev. B* 56, 12108 (1997).
54. A. Karlhede, S. A. Kivelson, K. Lejnell, and S. L. Sondhi, *Phys. Rev. Lett.* 77, 2061 (1996).
55. M. Franco and L. Brey, *Phys. Rev. B* 56, 10383 (1997).
56. E. H. Rezayi, *Phys. Rev. B* 36, 5454 (1987).
57. E. H. Rezayi, *Phys. Rev. B* 43, 5944 (1991).
58. D. H. Lee and C. L. Kane, *Phys. Rev. Lett.* 64, 1313 (1990).
59. S. L. Sondhi, A. Karlhede, S. A. Kivelson, and E. H. Rezayi, *Phys. Rev. B* 47, 16419 (1993).
60. D. R. Stewart, D. Sprinzak, C. M. Marcus, C. I. Duruöz, and J. S. Harris, *Science* 278, 1784 (1997).
61. L. P. Kouwenhoven, T. H. Oosterkamp, M. W. S. Danoesastro, M. Eto, D. G. Austing, T. Honda, and S. Tarucha, *Science* 278, 1788 (1997).
62. I. V. Zozoulenko, A. S. Sachrajda, C. Gould, K.-F. Berggren, P. Zawadzki, Y. Feng, and Z. Wasilewski, *Phys. Rev. Lett.* 83, 1838 (1999).
63. M. Ciorga, A. S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, S. Jullian, Y. Feng, and Z. Wasilewski, *Phys. Rev. B* 61, R16315 (2000).
64. A. Fuhrer, S. Lüscher, T. Heinzel, K. Ensslin, W. Wegscheider, and M. Bichler, *Phys. Rev. B* 63, 125309 (2001).
65. B. T. Miller, W. Hansen, S. Manus, R. J. Luyken, A. Lorke, J. P. Kotthaus, S. Huant, G. Medeiros-Ribeiro, and P. M. Petroff, *Phys. Rev. B* 56, 6764 (1997).
66. R. J. Warburton, B. T. Miller, C. S. Dürr, C. Bödefeld, K. Karrai, J. P. Kotthaus, G. Medeiros-Ribeiro, P. M. Petroff, and S. Huant, *Phys. Rev. B* 58, 16221 (1998).
67. E. E. Vdovin, A. Levin, A. Patane, L. Eaves, P. C. Main, Y. N. Khanin, Y. V. Dubrovskii, M. Henini, and G. Hill, *Science* 290, 122 (2000).
68. N. B. Zhitenev, R. C. Ashoori, L. N. Pfeifer, and K. W. West, *Phys. Rev. Lett.* 79, 2308 (1997).
69. V. Moldoveanu, A. Aldea, A. Manolescu, and M. Nita, *Phys. Rev. B* 63, 045301 (2001).
70. Y. Meir, N. S. Wingreen, and P. A. Lee, *Phys. Rev. Lett.* 66, 3048 (1991).
71. E. B. Foxman, U. Meirav, P. L. McEuen, M. A. Kastner, O. Klein, P. A. Belk, D. M. Abusch, and S. J. Wind, *Phys. Rev. B* 50, 14193 (1994).
72. C. W. J. Beenakker, *Phys. Rev. B* 44, 1646 (1991).
73. H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in NATO ASI Series, Vol. 294, p. 167 (1992).
74. D. Weinmann, W. Häusler, W. Pfaff, B. Kramer, and U. Weiss, *Europhys. Lett.* 26, 467 (1994).
75. W. Pfaff, D. Weinmann, W. Häusler, B. Kramer, and U. Weiss, *Z. Phys. B* 96, 201 (1994).
76. D. Weinmann, W. Häusler, and B. Kramer, *Phys. Rev. Lett.* 74, 984 (1995).
77. L. Jauregui, W. Häusler, D. Weinmann, and B. Kramer, *Phys. Rev. B* 53, R1713 (1996).
78. Y. Tanaka and H. Akera, *Phys. Rev. B* 53, 3901 (1996).
79. M. Eto, *J. Phys. Soc. Japan* 66, 2244 (1997).
80. A. T. Johnson, L. P. Kouwenhoven, W. de Jong, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon, *Phys. Rev. Lett.* 69, 1592 (1992).
81. J. Weis, R. J. Haug, K. v. Klitzing, and K. Ploog, *Phys. Rev. Lett.* 71, 4019 (1993).
82. L. P. Rokhinson, L. J. Guo, S. Y. Chou, and D. C. Tsui, *Phys. Rev. B* 63, 035321 (2001).
83. K. Yosida and A. Yoshimori, "Magnetism," Vol. V. Academic Press, San Diego, 1973.
84. A. C. Hewson, "The Kondo Problem to Heavy Fermions." Cambridge Univ. Press, Cambridge, UK, 1993.
85. D. L. Cox and A. Zawadowski, *Adv. Phys.* 47, 599 (1998).
86. J. Kondo, *Prog. Theor. Phys.* 32, 37 (1964).
87. J. R. Schrieffer and P. A. Wolff, *Phys. Rev.* 149, 491 (1966).
88. T. K. Ng and P. A. Lee, *Phys. Rev. Lett.* 61, 1768 (1988).
89. L. I. Glazman and M. E. Raikh, *JETP Lett.* 47, 452 (1988).
90. A. Kawabata, *J. Phys. Soc. Japan* 60, 3222 (1991).
91. N. S. Wingreen, Y. Meir, and P. A. Lee, *Phys. Rev. Lett.* 70, 2601 (1993).
92. F. D. Haldane, *Phys. Rev. Lett.* 40, 416 (1978).
93. A. Okiji and N. Kawakami, *Phys. Rev. Lett.* 50, 1157 (1983).
94. N. Kawakami and A. Okiji, *Phys. Rev. Lett.* 51, 2011 (1983).
95. D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, and M. A. Kastner, *Nature* 391, 156 (1998).
96. S. M. Cronenwett, T. H. Oosterkamp, and L. P. Kouwenhoven, *Science* 281, 540 (1998).
97. D. Goldhaber-Gordon, J. Göres, and M. A. Kastner, *Phys. Rev. Lett.* 81, 5225 (1998).
98. S. Sasaki, S. D. Franceschi, J. M. Elzerman, W. G. van der Wiel, M. Eto, S. Tarucha, and L. P. Kouwenhoven, *Nature* 405, 764 (2000).
99. W. G. van der Wiel, S. D. Franceschi, T. Fujisawa, J. M. Elzerman, S. Tarucha, and L. P. Kouwenhoven, *Science* 289, 2105 (2000).
100. J. Schmid, J. Weis, K. Eberl, and K. v. Klitzing, *Phys. Rev. Lett.* 84, 5824 (2000).
101. H. Jeong, A. M. Chang, and M. R. Melloch, *Science* 293, 2221 (2001).
102. W. G. van der Wiel, Ph.D. Thesis, Delft University of Technology, 2002.
103. S. Hershfeld, J. H. Davies, and J. W. Wilkins, *Phys. Rev. Lett.* 67, 3720 (1991).
104. S. Hershfeld, J. H. Davies, and J. W. Wilkins, *Phys. Rev. B* 46, 7046 (1992).
105. A. L. Yeyati, A. Martin-Rodero, and F. Flores, *Phys. Rev. Lett.* 71, 2991 (1993).
106. Y. Meir, N. S. Wingreen, and P. A. Lee, *Phys. Rev. Lett.* 70, 2601 (1993).
107. N. S. Wingreen and Y. Meir, *Phys. Rev. B* 49, 11040 (1994).
108. T. Inoshita, A. Shimizu, Y. Kuramoto, and H. Sasaki, *Phys. Rev. B* 48, 14 725 (1993).
109. A. Oguri, H. Ishii, and T. Saso, *Phys. Rev. B* 51, 4715 (1995).
110. A. Oguri, H. Ishii, and T. Saso, *Phys. Rev. B* 56, 13422 (1997).
111. W. Izumida, O. Sakai, and Y. Shimizu, *J. Phys. Soc. Japan* 66, 717 (1997).
112. W. Izumida, O. Sakai, and Y. Shimizu, *J. Phys. Soc. Japan* 67, 2444 (1998).
113. W. Izumida and O. Sakai, *Phys. B* 281–282, 32 (1998).

114. W. Izumida, O. Sakai, and S. Suzuki, *J. Phys. Soc. Japan* 70, 1045 (2001).
115. T. Pohjola, J. König, M. M. Salomaa, J. Schmid, H. Schoeller, and G. Schön, *Europhys. Lett.* 40, 189 (1997).
116. A. L. Yeyati, F. Flores, and A. Martin-Rodero, *Phys. Rev. Lett.* 83, 600 (1999).
117. M. Pustilnik, Y. Avishai, and K. Kikoin, *Phys. Rev. Lett.* 84, 1756 (2000).
118. M. Eto and Y. V. Nazarov, *Phys. Rev. Lett.* 85, 1306 (2000).
119. D. Giuliano and A. Tagliacozzo, *Phys. Rev. Lett.* 84, 4677 (2000).
120. M. Pustilnik and L. I. Glazman, *Phys. Rev. Lett.* 85, 2993 (2000).
121. D. Giuliano, B. Jouault, and A. Tagliacozzo, *Phys. Rev. B* 63, 125318 (2001).
122. W. Izumida, O. Sakai, and S. Tarucha, *Phys. Rev. Lett.* 87, 216803 (2001).
123. A. Georges and Y. Meir, *Phys. Rev. Lett.* 82, 3508 (1999).
124. R. Aguado and D. C. Langreth, *Phys. Rev. Lett.* 85, 1946 (2000).
125. C. A. Büsser, E. B. Anda, A. L. Lima, M. A. Davidovich, and G. Chiappe, *Phys. Rev. B* 62, 9907 (2000).
126. W. Izumida and O. Sakai, *Phys. Rev. B* 62, 10260 (2000).
127. T. Aono and M. Eto, *Phys. Rev. B* 63, 125327 (2001).
128. J. Nygård, D. H. Cobden, and P. E. Lindelof, *Nature* 408, 342 (2000).
129. T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong, *Rev. Mod. Phys.* 53, 385 (1981).
130. C. W. J. Beenakker, *Rev. Mod. Phys.* 69, 731 (1997).
131. U. Sivan, R. Berkovits, Y. Aloni, O. Prus, A. Auerbach, and G. Ben-Yoseph, *Phys. Rev. Lett.* 77, 1123 (1996).
132. F. Simmel, T. Heinzel, and D. A. Wharam, *Europhys. Lett.* 38, 123 (1997).
133. S. R. Patel, D. R. Stewart, C. M. Marcus, M. Gökçedag, Y. Alhassid, A. D. Stone, C. I. Duruöz, and J. S. Harris, Jr., *Phys. Rev. Lett.* 81, 5900 (1998).
134. F. Simmel, D. Abusch-Magder, D. A. Wharam, M. A. Kastner, and J. P. Kotthaus, *Phys. Rev. B* 59, R10441 (1999).
135. O. Millo, D. Katz, Y. Cao, and U. Banin, *Phys. Rev. B* 61, 16773 (2000).
136. Y. M. Blanter, A. D. Mirlin, and B. A. Muzykantskii, *Phys. Rev. Lett.* 78, 2449 (1997).
137. H. Tamura and M. Ueda, *Phys. Rev. Lett.* 79, 1345 (1997).
138. R. Berkovits and B. L. Altshuler, *Phys. Rev. B* 55, 5297 (1997).
139. A. A. Koulakov, F. G. Pikus, and B. I. Shklovskii, *Phys. Rev. B* 55, 9223 (1997).
140. J. A. Vergés, E. Cuevas, E. Louis, and M. Ortuno, *Phys. Rev. B* 56, R7045 (1997).
141. R. O. Vallejos, C. H. Lewenkopf, and E. R. Mucciolo, *Phys. Rev. Lett.* 81, 677 (1998).
142. R. Berkovits, *Phys. Rev. Lett.* 81, 2128 (1998).
143. P. N. Walker, G. Montambaux, and Y. Gefen, *Phys. Rev. B* 60, 2541 (1999).
144. Y. Alhassid and S. Malhotra, *Phys. Rev. B* 60, R16315 (1999).
145. S. Levit and D. Orgad, *Phys. Rev. B* 60, 5549 (1999).
146. A. Cohen, K. Richter, and R. Berkovits, *Phys. Rev. B* 60, 2536 (1999).
147. Y. Avishai, D. Berend, and R. Berkovits, *Phys. Rev. B* 59, 10707 (1999).
148. G. S. Jeon, S. Wu, M. Y. Choi, and H.-W. Lee, *Phys. Rev. B* 59, 2841 (1999).
149. K.-H. Ahn, K. Richter, and I.-H. Lee, *Phys. Rev. Lett.* 83, 4144 (1999).
150. Y. Alhassid, *Rev. Mod. Phys.* 72, 895 (2000).
151. K. Hirose, F. Zhou, and N. S. Wingreen, *Phys. Rev. B* 63, 075301 (2001).
152. M. Stopa, *Phys. B* 251, 228 (1998).
153. S. Hortikar and M. Srednicki, *Phys. Rev. Lett.* 80, 1646 (1998).
154. S. Lüscher, T. Heinzel, K. Ensslin, W. Wegscheider, and M. Bichler, *Phys. Rev. Lett.* 86, 2118 (2001).
155. L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 64, 2691 (1990).
156. L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon, *Phys. Rev. Lett.* 67, 1626 (1991).
157. H. Pothier, P. Lafarge, P. F. Orfla, C. Urbina, D. Esteve, and M. H. Devoret, *Phys. B* 169, 573 (1991).
158. H. Pothier, P. Lafarge, C. Urbina, D. Esteve, and M. H. Devoret, *Europhys. Lett.* 17, 249 (1992).
159. D. Loss and D. P. DiVincenzo, *Phys. Rev. A* 57, 120 (1998).
160. M. Bayer, P. Hawrylak, K. Hinzer, S. Fafard, M. Korkusinski, Z. R. Wasilewski, O. Stern, and A. Forchel, *Science* 291, 451 (2001).
161. F. R. Waugh, M. J. Berry, D. J. Mar, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Phys. Rev. Lett.* 75, 705 (1995).
162. C. Livermore, C. H. Crouch, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Science* 274, 1332 (1996).
163. C. H. Crouch, C. Livermore, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Appl. Phys. Lett.* 71, 817 (1997).
164. R. H. Blick, D. Pfannkuche, R. H. Haug, K. von Klizing, and K. Eberl, *Phys. Rev. Lett.* 80, 4032 (1998).
165. A. S. Adourian, C. Livermore, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Appl. Phys. Lett.* 75, 424 (1999).
166. D. G. Austing, Y. Tokura, S. Tarucha, P. Matagne, and J. P. Leburton, *Phys. E* 11, 63 (2001).
167. I. M. Ruzin, B. Chandrasekhar, E. I. Levin, and L. I. Glazman, *Phys. Rev. B* 45, 13469 (1992).
168. J. M. Golden and B. I. Halperin, *Phys. Rev. B* 53, 3893 (1996).
169. K. A. Matveev, L. I. Glazman, and H. U. Baranger, *Phys. Rev. B* 53, 1034 (1996).
170. J. M. Golden and B. I. Halperin, *Phys. Rev. B* 54, 16757 (1996).
171. B. Partoens, V. A. Schweigert, and F. M. Peeters, *Phys. Rev. Lett.* 79, 3990 (1997).
172. D. G. Austing, T. Honda, K. Muraki, Y. Tokura, and S. Tarucha, *Phys. B* 251, 206 (1998).
173. T. Schmidt, R. J. Haug, K. v. Klizing, A. Förster, and H. Lüth, *Phys. Rev. Lett.* 78, 1544 (1997).
174. S. Sasaki, D. G. Austing, and S. Tarucha, *Phys. B* 256–258, 157 (1998).
175. N. C. van der Vaart, S. F. Godijn, Y. V. Nazarov, C. J. P. M. Harmans, J. E. Mooij, L. W. Molenkamp, and C. T. Foxon, *Phys. Rev. Lett.* 74, 4702 (1995).
176. F. R. Waugh, M. J. Berry, C. H. Crouch, C. Livermore, D. J. Mar, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Phys. Rev. B* 53, 1413 (1996).
177. D. Dixon, L. P. Kouwenhoven, P. L. McEuen, Y. Nagamune, J. Motohisa, and H. Sakaki, *Phys. Rev. B* 53, 12625 (1996).
178. R. H. Blick, R. J. Haug, J. Weis, D. Pfannkuche, K. v. Klitzing, and K. Eberl, *Phys. Rev. B* 53, 7899 (1996).
179. C. Y. Fong, J. S. Nelson, L. A. Hemstreet, R. F. Gallup, L. L. Chang, and L. Esaki, *Phys. Rev. B* 46, 9538 (1992).
180. G. W. Bryant, *Phys. Rev. B* 48, 8024 (1993).
181. C. A. Stafford and S. D. Sarma, *Phys. Rev. Lett.* 72, 3590 (1994).
182. G. Klimeck, G. Chen, and S. Datta, *Phys. Rev. B* 50, 2316 (1994).
183. R. Kotlyar and S. D. Sarma, *Phys. Rev. B* 56, 13235 (1997).
184. Z. Yu, T. Heinzel, and A. T. Johnson, *Phys. Rev. B* 55, 13697 (1997).
185. Z. Yu, A. T. Johnson, and T. Heinzel, *Phys. Rev. B* 58, 13830 (1998).
186. R. Kotlyar, C. A. Stafford, and S. Das Sarma, *Phys. Rev. B* 58, 3989 (1998).
187. R. Kotlyar, C. A. Stafford, and S. Das Sarma, *Phys. Rev. B* 58, R1746 (1998).
188. Y. Asano, *Phys. Rev. B* 58, 1414 (1998).
189. H. Imamura, H. Aoki, and P. A. Maksym, *Phys. Rev. B* 57, R4257 (1998).
190. C. A. Sta.ord, R. Kotlyar, and D. D. Sarma, *Phys. Rev. B* 58, 7091 (1998).

191. H. Imamura, P. A. Maksym, and H. Aoki, *Phys. Rev. B* 59, 5817 (1999).
192. F. Ramírez, E. Cota, and S. E. Ulloa, *Phys. Rev. B* 59, 5717 (1999).
193. S. Nagaraja, L. R. C. Fonseca, and J. P. Leburton, *Phys. Rev. B* 59, 14880 (1999).
194. M. Rontani, F. Rossi, F. Manghi, and E. Molinari, *Solid State Comm.* 112, 151 (1999).
195. Y. Tokura, D. G. Austing, and S. Tarucha, *J. Phys. Condens. Matter* 11, 6023 (1999).
196. A. Wensauer, O. Steffens, M. Suhrke, and U. Rössler, *Phys. Rev. B* 62, 2605 (2000).
197. M. Pi, A. Emperador, M. Barranco, and F. Garcias, *Phys. Rev. B* 63, 115316 (2001).
198. S. Amaha, D. G. Austing, Y. Tokura, K. Muraki, K. Ono, and S. Tarucha, *Solid State Comm.* 119, 183 (2001).
199. J. J. Paracios and P. Hawrylak, *Phys. Rev. B* 51, 1769 (1995).
200. J. Hu, E. Dagotto, and A. H. MacDonald, *Phys. Rev. B* 54, 8616 (1996).
201. S. M. Girvin and A. H. MacDonald, “Novel Quantum Liquids in Low-Dimensional Semiconductor Structures” Wiley, New York, 1995.
202. Y. Tokura, S. Sasaki, D. G. Austing, and S. Tarucha, *Phys. E* 6, 676 (2000).
203. K. Penc, H. Shiba, F. Mila, and T. Tsukagoshi, *Phys. Rev. B* 54, 4056 (1996).
204. J. Kolehmainen, S. M. Reimann, M. Koskinen, and M. Manninen, *European. Phys. J. B* 13, 731 (2000).
205. C. G. Lent, P. D. Tougaw, and W. Porod, *Appl. Phys. Lett.* 62, 714 (1993).
206. P. D. Tougaw, C. G. Lent, and W. Porod, *J. Appl. Phys.* 74, 3558 (1993).
207. C. G. Lent and P. D. Tougaw, *J. Appl. Phys.* 74, 6227 (1994).
208. P. D. Tougaw and C. G. Lent, *J. Appl. Phys.* 75, 1818 (1994).
209. P. D. Tougaw and C. G. Lent, *J. Appl. Phys.* 80, 4722 (1996).
210. G. Toth, C. S. Lent, P. D. Tougaw, Y. Brazhnik, W. W. Weng, W. Porod, R. W. Liu, and Y. F. Huang, *Superlatt. Microstruct.* 20, 473 (1996).
211. S. Bandyopadhyay, B. Das, and A. E. Miller, *Nanotechnology* 5, 113 (1994).
212. S. Bandyopadhyay and V. P. Roychowdhury, *Phys. Low Dimension. Syst.* 8/9, 29 (1995).
213. S. Bandyopadhyay, *Japan J. Appl. Phys.* 1 35, 3350 (1996).
214. R. Landauer, *Philos. Trans. R. Soc. London Ser. A* 353, 367 (1995).
215. A. O. Orlov, I. Amlani, G. H. Bernstein, C. S. Lent, and G. L. Snider, *Science* 277, 928 (1997).
216. I. Amlani, A. O. Orlov, G. L. Snider, C. S. Lent, and G. H. Bernstein, *Appl. Phys. Lett.* 71, 1730 (1997).
217. I. Amlani, A. O. Orlov, G. L. Snider, C. S. Lent, and G. H. Bernstein, *J. Vac. Sci. Technol. B* 16, 3795 (1998).
218. I. Amlani, A. O. Orlov, G. L. Snider, C. S. Lent, and G. H. Bernstein, *Appl. Phys. Lett.* 72, 2179 (1998).
219. G. L. Snider, A. O. Orlov, I. Amlani, X. Zou, G. H. Bernstein, C. S. Lent, J. L. Merz, and W. Porod, *J. Vac. Sci. Technol. A* 17, 1394 (1999).
220. I. Amlani, A. O. Orlov, G. Toth, G. H. Bernstein, C. S. Lent, and G. L. Snider, *Science* 284, 289 (1999).
221. I. Amlani, A. O. Orlov, R. K. Kummamuru, G. H. Bernstein, C. S. Lent, and G. L. Snider, *Appl. Phys. Lett.* 77, 738 (2000).
222. C. Single, F. E. Prins, and D. P. Kern, *Appl. Phys. Lett.* 78, 1421 (2001).
223. C. S. Lent and P. D. Tougau, *Proc. IEEE* 85, 541 (1997).
224. G. Toth and C. S. Lent, *J. Appl. Phys.* 85, 2977 (1999).
225. A. O. Orlov, I. Amlani, R. K. Kummamuru, R. Ramasubramaniam, G. Toth, C. S. Lent, G. H. Bernstein, and G. L. Snider, *Appl. Phys. Lett.* 77, 295 (2000).
226. A. O. Orlov, R. K. Kummamuru, R. Ramasubramaniam, G. Toth, C. S. Lent, G. H. Bernstein, and G. L. Snider, *Appl. Phys. Lett.* 78, 1625 (2001).
227. L. P. Kouwenhoven, F. W. J. Hekking, B. J. van Wees, C. J. P. M. Harmans, C. E. Timmering, and C. T. Foxon, *Phys. Rev. Lett.* 65, 361 (1990).
228. R. J. Haug, J. M. Hong, and K. Y. Lee, *Surf. Sci.* 263, 415 (1992).
229. C. A. Stafford and S. D. Sarma, *Phys. Rev. Lett.* 72, 3590 (1994).
230. R. Ugajin, *J. Appl. Phys.* 76, 2833 (1994).
231. R. Ugajin, *Appl. Phys. Lett.* 68, 2657 (1996).
232. R. Ugajin, *Phys. Rev. B* 53, 10141 (1996).
233. H. Chen, J. Wu, Z. Q. Li, and Y. Kawazoe, *Phys. Rev. B* 55, 1578 (1007).
234. R. Ugajin, *Phys. Rev. Lett.* 80, 572 (1998).
235. R. Kotlyar, C. A. Stafford, and S. D. Sarma, *Phys. Rev. B* 58, R1746 (1998).
236. G. Chen, G. Klimeck, S. Datta, G. Chen, and W. A. Goddard, *Phys. Rev. B* 50, 8035 (1994).
237. M. R. Wegewijs and Y. V. Nazarov, *Phys. Rev. B* 60, 14318 (1999).
238. W. Z. Shangguan, T. C. A. Yeung, Y. B. Yu, and C. H. Kam, *Phys. Rev. B* 63, 235323 (2001).
239. K. Kang, M.-C. Cha, and S.-R. E. Yang, *Phys. Rev. B* 56, R4344 (1997).
240. A. Sugimura, *Japan J. Appl. Phys.* 29, L2463 (1990).
241. J. B. Khurgin, F. Jin, and A. Obeidat, *Appl. Phys. Lett.* 73, 3944 (1998).
242. J. B. Khurgin and F. Jin, *Appl. Phys. Lett.* 73, 3102 (1998).
243. S. S. Li and J. B. Xia, *J. Appl. Phys.* 84, 3710 (1998).
244. H. Tamura, K. Shiraishi, and H. Takayanagi, *Japan. J. Appl. Phys.* 39, L241 (2000).
245. K. Shiraishi, H. Tamura, and H. Takayanagi, *Appl. Phys. Lett.* 78, 3702 (2001).
246. T. Kimura, H. Tamura, K. Shiraishi, and H. Takayanagi, *Phys. Rev. B* 65, 081307 (2002).
247. H. Tamura, K. Shiraishi, T. Kimura, and H. Takayanagi, *Phys. Rev. B* 65, 085324 (2002).
248. J. H. Schon, C. Kloc, R. C. Haddon, and B. Batlogg, *Science* 288, 656 (2000).
249. J. H. Schon, C. Kloc, and B. Batlogg, *Nature* 406, 702 (2000).
250. E. H. Lieb, *Phys. Rev. Lett.* 62, 1201 (1989).
251. A. Mielke, *J. Phys. A* 24, L73 (1991).
252. H. Tasaki, *Phys. Rev. Lett.* 69, 1608 (1992).
253. H. Sakamoto and K. Kubo, *J. Phys. Soc. Japan* 65, 3732 (1996).
254. Y. Watanabe and S. Miyashita, *J. Phys. Soc. Japan* 66, 2123 (1997).
255. Y.Watanabe and S. Miyashita, *J. Phys. Soc. Japan* 66, 3981 (1997).
256. R. Fleischmann, T. Geisel, and R. Ketzmerick, *Phys. Rev. Lett.* 68, 1367 (1992).
257. N. Shima and H. Aoki, *Phys. Rev. Lett.* 71, 4389 (1993).
258. M. Fujita, T. Umeda, and M. Yoshida, *Phys. Rev. B* 51, 13778 (1995).
259. M. Fujita, K. Wakabayashi, K. Nakada, and K. Kusakabe, *J. Phys. Soc. Japan* 65, 1920 (1996).
260. K. Kusakabe, K. Wakabayashi, M. Igami, K. Nakada, and M. Fujita, *Mol. Cryst. Liq. Cryst.* 305, 445 (1997).
261. R. Arita, K. Kuroki, H. Aoki, A. Yajima, M. Tsukada, S. Watanabe, M. Ichimaru, T. Onogi, and T. Hashizume, *Phys. Rev. B* 57, R6854 (1998).
262. A. Yajima, M. Tsukada, S. Watanabe, M. Ichimura, Y. Sawa, T. Onogi, and T. Hashizume, *Phys. Rev. B* 60, 1456 (1999).
263. S. Okada and A. Oshiyama, *Phys. Rev. B* 62, R13286 (2000).
264. N. D. Mermin and H. Wagner, *Phys. Rev. Lett.* 17, 1133 (1966).
265. G. Schedelbeck, W. Wegscheider, M. Bichler, and G. Abstreiter, *Science* 278, 1792 (1997).
266. L. Pfeiffer, K. W. West, H. L. Stormer, J. P. Eisenstein, K. W. Baldwin, D. Gershoni, and J. Spector, *Appl. Phys. Lett.* 56, 1697 (1990).

267. T. Fukui, S. Ando, T. Tokura, and T. Toriyama, *Appl. Phys. Lett.* 58, 2018 (1991).
268. Y. Nagamune, S. Tsukamoto, M. Nishioka, and Y. Arakawa, *J. Crys. Growth* 126, 707 (1993).
269. K. Kumakura, K. Nakakoshi, J. Motohisa, T. Fukui, and H. Hasegawa, *Japan J. Appl. Phys.* 34, 4387 (1995).
270. K. Kumakura, J. Motohisa, and T. Fukui, *J. Cryst. Growth* 170, 700 (1997).
271. H. Namatsu, S. Horiguchi, Y. Takahashi, M. Nagase, and K. Kurihara, *Japan J. Appl. Phys.*. 36, 3669 (1997).
272. D. T. Hofstadter, *Phys. Rev. B* 14, 2239 (1976).
273. P. G. Harper, *Proc. R. Soc. London Ser. A* 68, 874 (1955).
274. D. Springsguth, R. Ketzmerick, and T. Geisel, *Phys. Rev. B* 56, 2036 (1997).
275. H. Aoki, M. Ando, and H. Matsumura, *Phys. Rev. B* 54, R17296 (1996).
276. D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* 49, 405 (1982).
277. R. R. Gerhardts, D.Weiss, and U. Wulf, *Phys. Rev. B* 43, 5192 (1991).
278. D. Weiss, A. Mensching, K. von Klitzing, and G. Weimann, *Surf. Sci.* 263, 314 (1992).
279. T. Schlösser, K. Ensslin, J. P. Kotthaus, and M. Holland, *Europhys. Lett.* 33, 683 (1996).
280. C. Albrecht, J. H. Smet, K. von Klitzing, D. Weiss, V. Umansky, and H. Schweizer, *Phys. Rev. Lett.* 86, 147 (2001).
281. J. An, C. D. Gong, and H. Q. Lin, *Phys. Rev. B* 63, 174434 (2001).
282. T. Dahm and L. Tewordt, *Phys. Rev. B* 52, 1297 (1995).
283. J. Deisz, H. Hess, and J. Serene, *Phys. Rev. Lett.* 76, 1312 (1996).
284. K. Kuroki and R. Arita, *Phys. Rev. B* 64, 024501 (2001).
285. E. Dagotto, J. Riera, and D. Scalapino, *Phys. Rev. B* 45, 5744 (1992).
286. E. Dagotto and T. M. Rice, *Science* 271, 618 (1996).
287. M. Uehara, T. Nagata, J. Akimitsu, H. Takahashi, and T. M. K. Kinoshita, *J. Phys. Soc. Japan* 65, 2764 (1996).
288. Y. Arakawa and H. Sakaki, *Appl. Phys. Lett.* 40, 939 (1992).

www.aspbs.com/enn

Quantum Dot Infrared Photodetector

Jamie Phillips, Adrienne Stiff-Roberts, Pallab Bhattacharya

University of Michigan, Ann Arbor, Michigan, USA

CONTENTS

1. Introduction
2. Overview of Intraband Detector Operation
3. Epitaxial Growth and Electronic Properties of Quantum Dot Ensembles
4. Device Fabrication and Characterization
5. Evaluation of QDIP Performance
6. Conclusion
Glossary
References

1. INTRODUCTION

Infrared (IR) detectors are important for numerous defense, commercial, and scientific applications. Infrared systems continue to demand detectors with the highest performance, highest operating temperature, and lowest cost. The current leading IR detector technology is HgCdTe [1–5]. However, this II–VI semiconductor material is extremely soft, leading to difficulties in materials growth and detector processing, which results in low yield and high cost. As a result, alternative detection schemes using bandgap engineering and more mature III–V semiconductor materials technologies have been pursued. One of these approaches is the strained-layer superlattice with a type-II band lineup [6–8] (e.g., InAs/InGaSb), which is predicted to have reduced Auger scattering rates that may lead to increased operating temperatures in comparison to HgCdTe. This promising technology is still in the developmental stage and requires complex material heterostructures. A second approach is the use of quantum-confined structures to detect IR radiation through optical absorption across intraband electronic states. The quantum well infrared photodetector (QWIP) [9] has shown success with this technique as demonstrated by its use in IR camera systems [10]. Unfortunately, QWIPs have not demonstrated performance comparable to HgCdTe due to fundamental performance limitations. An extension of the QWIP is the quantum dot infrared photodetector (QDIP), utilizing advantages of quantum confinement in three dimensions. QDIPs are predicted to outperform QWIPs due to their inherent sensitivity to normal incidence radiation and reduced phonon scattering. This review describes the operation of the QDIP; charts progress in materials growth, device design, fabrication, and testing; and outlines future directions for the detector technology.

2. OVERVIEW OF INTRABAND DETECTOR OPERATION

The concept of using quantum confinement in semiconductor heterostructures for infrared detection may be explained using basic principles of quantum mechanics. A quantum dot is analogous to the basic particle-in-a-box problem in three dimensions. For an infinite potential well, the solution to the time-independent Schrödinger equation yields the allowed energy states or eigenvalues

$$E_{n_x, n_y, n_z} = \frac{\pi^2\hbar^2}{2m^*}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right) \tag{1}$$

where n_i are quantum numbers and L_i are the dimensions of the potential well in the x, y, and z directions. The energy separation between the ground state and the first excited state for a cubic potential well with $L = L_x = L_y = L_z$ is given by

$$E_{1,1,2} - E_{1,1,1} = \frac{3\pi^2\hbar^2}{2m^*L^2} \tag{2}$$

The potential barriers in semiconductor heterostructures are finite and determined by the bandgap energy and alignment of the bands. For common III–V semiconductor material systems such as GaAs/AlGaAs, InGaAs/GaAs, and InGaAs/InP, the barrier height and well dimensions may be designed to produce an energy separation between confined states in the range of 0–1 eV, corresponding to wavelengths of interest for infrared detection: short-wave infrared (SWIR, 1–3 μm), mid-wave infrared (MWIR, 3–5 μm), long-wave infrared (LWIR, 8–12 μm), and very-long-wave infrared (VLWIR, >12 μm).

Bulk semiconductor photon detectors are based on interband optical absorption, where a photon is absorbed to create electron–hole pairs in the conduction and valence

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

bands. The photoexcited carriers are then extracted to create a current or voltage as the basis for the detection signal. Quantum-confined nanostructures may be utilized to detect photons using *intraband* optical absorption, where an electron (hole) in a confined conduction (valence) band state is promoted to a higher energy state. The photoexcited carrier may then be extracted as a photocurrent. In the case of both interband and intraband infrared photodetectors, these devices must operate at reduced temperatures since the energy level spacing corresponding to the desired infrared wavelength range is comparable to the thermal energy of the carrier at room temperature ($k_B T = 26$ meV). Thus, carriers may be thermally excited out of the quantum dot ground state in the absence of incident infrared radiation, contributing to large dark currents. The detection mechanism for QDIPs is illustrated in Figure 1. It should be noted that this detection scheme requires the presence of an electron (hole) in the conduction (valence) band, and doping is typically required to provide these carriers. The selection rules for electron–photon interactions are determined by the matrix element

$$|a \cdot p_{\mathrm{if}}|^2 = \left| \int \psi_{\mathrm{f}}^* p_a \psi_{\mathrm{i}} e^{ik \cdot r} \, d^3 r \right|^2 \quad (3)$$

where ψ_{i} and ψ_{f} are the wave functions of the initial and final electronic states, p_a is the momentum operator, and $e^{ik \cdot r}$ represents the incident photon. An optical transition is forbidden if this matrix element is 0. Therefore, in order for an optical transition to occur, a polarization component of the optical electric field must correspond to a direction of quantum confinement. For example, a semiconductor quantum well with confinement in the growth direction requires an electric field component in the growth direction. This implies that quantum wells are not intrinsically sensitive to normal-incidence photoexcitation. Despite this problem, QWIPs have demonstrated success using reflectors and gratings to couple infrared radiation with intraband transitions, though this intrinsically limits the quantum efficiency to 50%. Quantum dots provide quantum confinement in all directions, thus providing intrinsic sensitivity to normal-incidence photoexcitation and an inherent advantage over QWIPs.

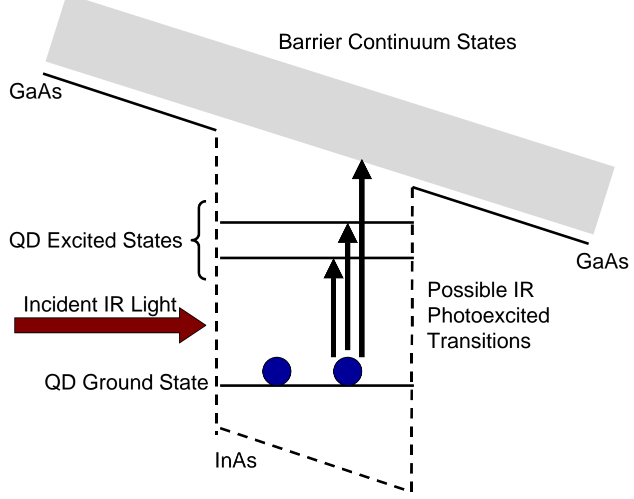


Figure 1. Illustration of intraband optical transitions in InAs/GaAs quantum dots that form the basis for QDIP detection.

Detectors based on intraband optical transitions are typically photoconductive in nature, where photoexcited carriers provide an increase in conductivity. These devices face challenges in extracting photoexcited carriers due to the inherent presence of the energy barriers used for quantum confinement. Ideally, the excited state for the optical transition should coincide with the top of the barrier for optimal carrier extraction. In practice, the excited state is significantly below the barrier and photoexcited carriers must be extracted by tunneling through the barrier or through thermionic emission. Detectors may also be designed using a single bound state, where absorption transitions are between the confined state and the continuum of states above the barrier energy.

3. EPITAXIAL GROWTH AND ELECTRONIC PROPERTIES OF QUANTUM DOT ENSEMBLES

Achieving nanoscale semiconductor heterostructures with excellent optical and electrical properties is a challenge. The realization of semiconductor quantum dots has been pursued for decades due to their potential applications in electronic and optoelectronic devices, including single-electron transistors, diode lasers, and infrared detectors and emitters. The development of epitaxial growth techniques, such as molecular beam epitaxy (MBE) and metal–organic chemical vapor deposition (MOCVD), has resulted in the routine ability to grow excellent quantum well materials with quantum confinement in the growth direction. Obtaining quantum confinement in the lateral direction is much more challenging. The most straightforward approach has been to pattern quantum well structures using lithographic and etching techniques [11, 12]. This method has not proven successful for device applications due to the inability to achieve high-quality interfaces after etching and regrowth. Growth on patterned substrates has also been investigated to provide lateral confinement using preferential growth on particular crystallographic planes [13–15]. This technique has been highly successful in producing quantum wire arrays and has proven to be an important technology for semiconductor diode lasers [16–18]. The preferential growth technique is limited in its ability to provide multiple layers of quantum dots, and both techniques are unable to provide high-density ensembles of quantum dots. The most successful approach to achieve multiple-layer, high-density, semiconductor quantum dot ensembles is the use of self-assembly techniques during epitaxial growth [19–25]. Strained-layer epitaxy in the Stranski–Krastanow growth mode has emerged as the premier method of forming quantum dot ensembles and has demonstrated success in a number of electronic and optoelectronic devices. All of the successful reports of QDIP operation use this technique to fabricate quantum dots. The epitaxial growth and electronic properties of self-assembled quantum dots using strained-layer epitaxy are described in the following discussion.

3.1. Epitaxial Growth

The surface morphology that evolves during epitaxial growth is determined by the relationship between the growth rate and the surface relaxation as the system is driven toward equilibrium by diffusion and kinetic processes. The surface profile is determined by the extent to which the surface atoms are mobile before incorporation into the underlying crystal lattice, which is under a driving growth flux. For unstrained systems, or systems with little strain, epitaxial growth is pseudomorphic, and growth occurs layer by layer. This growth mode is commonly referred to as the Frank–Van der Merwe growth mode. In a strained system, the minimum free-energy surface is not flat in all cases, but can have a three-dimensional form. Three-dimensional island growth may commence immediately (Volmer–Weber growth mode) or after initial layer-by-layer growth (Stranski–Krastanow growth mode [26]). The material deposited layer by layer in the Stranski–Krastanow growth mode is referred to as the wetting layer. The three-dimensional islands may have nanoscale dimensions where quantum confinement effects are observed. Several researchers have discovered that the highly strained growth of InAs on GaAs results in the Stranski–Krastanow growth mode [19–25], and three-dimensional islands of dimension less than 40 nm are formed. The driving forces for these self-assembled, three-dimensional quantum dots are elastic relaxation at the facet edges, renormalization of the surface energy of the facets, and interaction between neighboring islands via the substrate. This technique of self-assembling semiconductor nanostructures has proven to be a highly effective technique for fabricating dense ensembles of quantum dots with excellent electrical and optical properties. Self-assembled quantum dots have been achieved using MBE and MOCVD growth techniques in a number of semiconductor material combinations, including InGaAs/GaAs, InGaAs/InP, (In, Ga, Al)Sb/GaAs, InP/InGaP, and SiGe/Si. Self-assembled quantum dots also have been incorporated into a number of electronic and optoelectronic devices, including transistors, lasers, detectors, and modulators. A vast majority of QDIPs have used In(Ga)As/GaAs self-assembled quantum dots. Accordingly, we will primarily discuss the characteristics of QDIPs in this material system, though results from other materials are also discussed. For typical growth parameters used in MBE or MOCVD, an ensemble of dots with lateral sizes from 15 to 25 nm and heights from 5 to 8 nm are formed. The density of quantum dots is in the range of 10^9–10^{11} cm^{-2}. The typical size, shape, and density of self-assembled quantum dots used in QDIPs are shown in Figure 2.

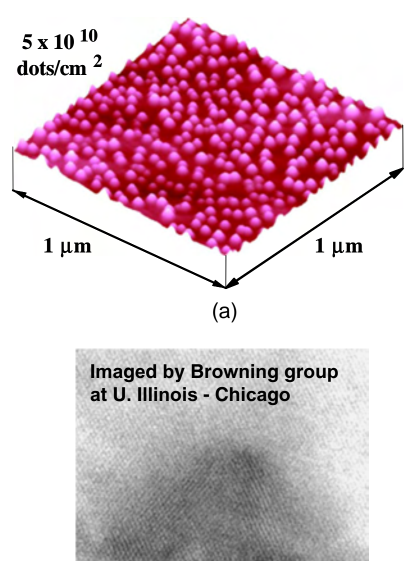


Figure 2. (a) Atomic force microscope and (b) cross-sectional transmission electron microscope image of self-assembled InAs/GaAs quantum dots illustrating characteristic size, shape, and density for QDIP applications.

3.2. Electronic Band Structure

The unique shape and strain distribution of self-assembled quantum dots leads to complexities in understanding the electronic band structure. The self-assembly technique results in an inhomogeneous distribution of quantum dot sizes. Each quantum dot in the ensemble demonstrates sharp atomic-like electronic states [27]. In the ensemble as a whole, however, there is an inhomogeneous distribution of electronic states that typically resembles a Gaussian distribution. Researchers have used a variety of theoretical and experimental techniques to gain an understanding of the band structure in individual and ensembles of quantum dots (QDs). The band structure of InAs/GaAs QDs has been calculated through modeling of the strain distribution and using an eight-band k.p model [28]. These calculations reveal a number of near-degenerate hole states and a few electron states for typical InAs/GaAs self-assembled QD dimensions. The photoluminescence spectra of self-assembled quantum dots reveal a number of excited-state interband transitions and agree with theoretical calculations [29–31]. For intraband detectors, it is desirable to experimentally verify the energy separation between electronic states within the conduction or valence band; however, it is difficult to distinguish these energy levels from the interband optical transitions probed in photoluminescence experiments. Electrical measurements such as capacitance spectroscopy [32] and deep-level transient spectroscopy [33] have been used to probe electronic states within the conduction or valence band. The fusion of the theoretical and experimental data provides a general picture of the band structure of self-assembled InAs/GaAs QDs, where a number of hole states with small energy separation, as well as single or multiple electron states, are observed. The energy separation between the ground electron state and an excited electron state (or the barrier) ranges from 50 to 500 meV [34–38], corresponding to the wavelengths of interest for infrared detection.

3.3. Intraband Optical Absorption

The absorption process requires the presence of a ground-state electron, where dopants have been incorporated into the heterostructure either by directly doping the quantum dots [39] or by modulation doping [40]. QDIP performance

has a distinct tradeoff between, and dependence on, quantum dot density and dopant concentration. As one would expect, the optimal dopant density corresponds to twice the QD areal density (i.e., just enough dopants to fill the ground states of the QDs). The dark current increases at higher dopant densities as dopants outnumber the electron states available in the QD ground state, providing additional "noise" carriers. At reduced dopant densities, there are fewer electrons in the QD ground state for absorption. The reduced absorption requires an increase in the number of QD layers for adequate detection. The optimal level for electron occupation is therefore the level required to fully occupy the ground state.

The primary wavelengths of interest for infrared detectors are less than 20 μm, corresponding to energy near 60 meV. The energy separation between electronic states in self-assembled quantum dots is on this order or larger for the conduction band, and much less than 60 meV for the valence band. As a result, devices based on intraband optical transitions in InAs/GaAs QDs thus far have exclusively used confined electron states in the conduction band. An understanding of intraband absorption is necessary to ascertain the suitability of self-assembled QDs for intraband detectors. The first report of intraband absorption in quantum dots was made by Pan et al. [41], where absorption was observed in the spectral range of 13–15 μm, corresponding to ground-state-to-excited-state optical absorption. Since then, intraband absorption has been reported for infrared wavelengths ranging from the MWIR to the VLWIR [42–44], where absorption peaks may be tuned by varying QD size through growth conditions or through modification of the barrier layer. The optical absorption spectral linewidths of the QDs are broadband and are determined by the size distribution of the QDs in the ensemble, where a typical spectrum is shown in Figure 3. QDIPs therefore have a tradeoff in absorption strength and broadband response for a given net QD volume. To illustrate, a perfectly uniform QD ensemble will have large peak absorption and a narrow absorption spectrum. A nonuniform QD ensemble will have reduced peak absorption, where the absorption response will be distributed in wavelength according to the electronic structure variation for differing QD sizes.

The intraband optical absorption shows the predicted intrinsic sensitivity to normal incidence excitation. The shape of self-assembled QDs resembles flattened pyramids or lenses, where quantum confinement is more significant in the vertical direction in comparison to the lateral direction. As a result, the intraband optical absorption is predicted to be more sensitive to in-plane polarization in comparison to polarization in the growth direction [45]. The significant decrease in intraband absorption for polarization in the growth direction has been reported [46, 47]. This polarization dependence has also been confirmed experimentally by measuring the change in peak absorbance at room temperature as a function of annealing time for in-plane ($\theta_{inc} = 0°$ to surface normal) and perpendicularly polarized ($\theta_{inc} = 65°$ to surface normal) light. The absorbance of a 70-layer InAs/GaAs quantum dot stack was obtained for various annealing times at a constant anneal temperature of 700 °C. As shown in Figure 4, in the unannealed case (anneal time = 0 min), absorbance is greater in the in-plane polarized case than in the perpendicularly polarized case, as predicted by theory. As the anneal time is increased, the in-plane polarized light absorbance decreases, while the perpendicularly polarized light absorbance increases. This result agrees with theoretical predictions since the dots become flatter and more like quantum wells as the anneal time is increased, allowing perpendicularly polarized absorbance to become dominant. This polarization dependence emphasizes the importance of realizing QDs with small lateral dimensions to obtain strong normal incidence sensitivity for IR detectors.

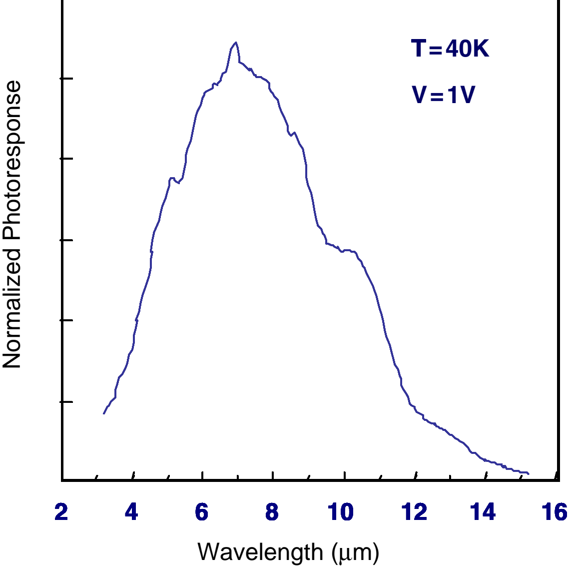


Figure 3. Absorption spectra in InAs/GaAs quantum dots for LWIR detection.

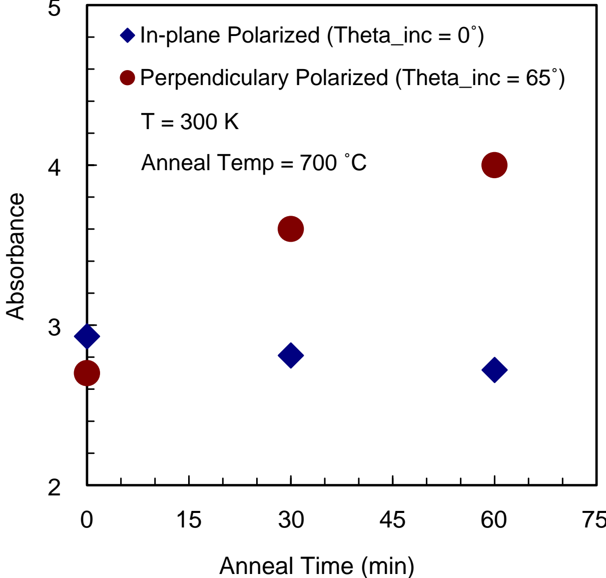


Figure 4. Room-temperature absorbance in 70-layer InAs/GaAs quantum dot stack as a function of anneal time for in-plane and perpendicularly polarized light.

3.4. Intraband Carrier Dynamics

Carrier excitation and relaxation processes play an integral role in intraband devices. Thermal excitation (generation and thermionic emission) is the principal mechanism governing dark current in an intraband detector. Relaxation of carriers reduces the efficiency of the detector, since a photoexcited carrier may relax into the ground state of a quantum dot before being collected as photocurrent. Once a carrier has been photoexcited by incident infrared radiation and has escaped a quantum dot, it can undergo several processes. The photoexcited carrier may drift in the electric field, be collected at a device contact as photocurrent, or relax into the ground state of the same quantum dot or a different one. The relaxation processes in an intraband device are analogous to minority carrier recombination in a photodiode; however, the carrier dynamics in quantum dots differ significantly from quantum wells.

The three-dimensional discrete nature of the QD electronic states suggests the existence of a "phonon bottleneck," where electron relaxation by a single phonon emission is suppressed. In contrast, the phonon process dominates carrier scattering and relaxation in quantum wells. Multiphonon events require a resonant condition, and therefore are typically slow for QDs (>1 ns). The suppression of phonon scattering in QDs suggests that carrier relaxation times will be significantly longer than in quantum wells. Recent investigations have demonstrated that Auger scattering [48] and electron–hole scattering [49] are responsible for interband carrier relaxation in QDs. Hole relaxation times measured by differential transmission spectroscopy (DTS) for In(Ga)As/GaAs QDs can be fast, less than 1 ps [50, 51], due to valence-band mixing, anisotropy, and high density of states. Electron relaxation times measured using the DTS technique show two time constants for relaxation from the excited state to the ground state [50, 51]. The short time constant ($\sim$7 ps) is believed to be governed by Auger, electron–hole, and multiphonon scattering, and the long time constant ($\sim$15–100 ps) by phonon-mediated processes. In quantum wells, electron relaxation processes are typically about 3 ps [52]. The DTS measurements described for QDs are performed under conditions where significant electron and hole densities are present. However, since the intraband detector is a unipolar device, there is a significantly reduced density of holes, which suppresses the Auger and electron–hole scattering processes and lengthens the electron relaxation time. The lengthened carrier relaxation times in self-assembled QDs are a desirable property for realizing efficient intraband detectors.

4. DEVICE FABRICATION AND CHARACTERIZATION

There are many possible heterostructure and device designs that have been used to fabricate QDIPs [53–64]. Therefore, the following section will discuss the most basic QDIP in order to understand the fundamental characteristics of the device and then briefly review various heterostructure and device designs to indicate the breadth of device possibilities.

The primary figures of merit for a detector are the dark current, spectral response, detectivity, quantum efficiency, and photoconductive gain. The dark current in a detector is the current flowing in the device in the absence of illumination. Dark current is a primary limitation in IR detectors where low incidence and corresponding low photocurrent are typical. The spectral response of a detector measures the photocurrent generated as a function of wavelength. The detectivity in a detector is a measure of the signal-to-noise ratio and represents a convolution of dark current and quantum efficiency, where the quantum efficiency represents the conversion of incident photons into an electrical signal. In a photoconductor, an analogous figure of merit to the quantum efficiency is the responsivity, given in amperes per watt (A/W) and representing the measured current for a given photon incidence. The photoconductive gain, also closely related to the quantum efficiency, is a measure of how many photoexcited carriers are generated for a single incident photon, and its value is based on the ratio of the carrier lifetime to the carrier transit time. After reviewing general QDIP characterization techniques, further description of these figures of merit and measured QDIP characteristics follow.

4.1. Device Heterostructure and Fabrication

QDIP devices are photoconductive in nature, where photoexcited carriers within the quantum dots provide a photocurrent and apparent increase in conductivity. QDIPs clearly have a resemblance to QWIPs in general operation and device structure; however, there are additional parameters unique to quantum dots that affect device performance. Essentially, QDIPs have more design parameters or "dials to turn," such as quantum dot size, shape, and density.

The common QDIP device heterostructure is an extension of QWIP devices, where a series of doped QD layers is placed between undoped barrier regions with doped contact layers at the device boundaries. Usually, InAs quantum dots are surrounded by GaAs barriers. IR detection requires a significant absorption volume to convert photons to electrons. Accordingly, QWIPs typically incorporate 10–50 quantum well layers. The small absorption volume of QDs also necessitates a large number of layers for adequate detection; however, the number of layers is typically limited by the ability to incorporate strain in the structure without generating dislocations. By calibrating the barrier region thickness to minimize the effect of strain from one dot layer to the next, QD layers can be repeated anywhere from 10 to 70 times without the propagation of dislocations.

In general, there are two device structures that have been studied, the lateral QDIP [56, 58, 59, 62] and the vertical QDIP [53–55, 57, 60, 61, 63, 64]. The device conduction band and heterostructure for the lateral and vertical devices are illustrated in Figure 5a and b, respectively. The lateral QDIP, which operates much like a field-effect transistor, conducts photocurrent through lateral transport of carriers across a high-mobility channel [62, 65]. AlGaAs barriers, which provide this high-mobility channel, are also necessary to modulation-dope the quantum dots. Since the major contributions to the dark current in lateral QDIPs are due to interdot tunneling and hopping conduction, these devices have demonstrated very low dark currents and high operating temperatures [56]. The vertical QDIP conducts photocurrent through vertical transport of carriers. In this case,

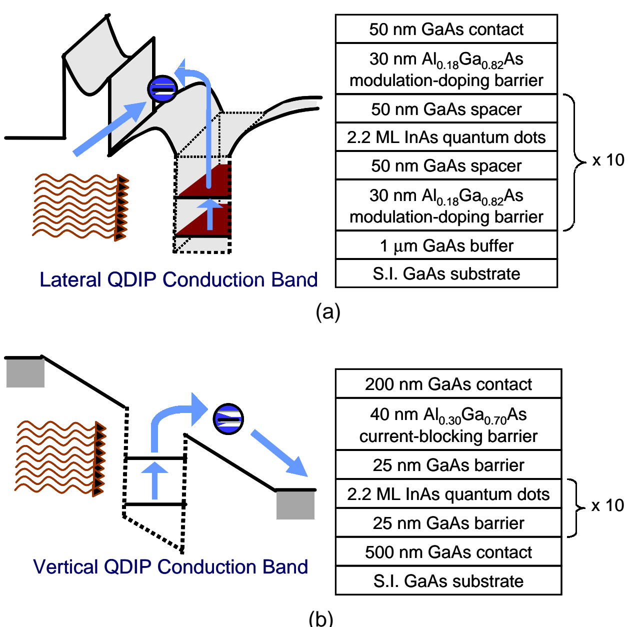


Figure 5. Schematics of QDIP conduction band and heterostructure designed for (a) lateral and (b) vertical detection.

the quantum dots are directly doped to provide free carriers during photoexcitation, and an AlGaAs barrier can be included in the vertical device to block dark current created by thermionic emission [66, 67].

For either type of device, fabrication typically consists of standard photolithography and wet-etching techniques. The lateral device requires a minimum of two lithography steps, the first for metal evaporation of two top contacts and the second for device definition by a mesa etch. The vertical device requires a minimum of three lithography steps: (1) metal evaporation of the top contact, (2) device definition by mesa etch around the top contact, and (3) metal evaporation of the bottom contact around the device mesa. An important advantage of QDIPs over QWIPs is the simpler device fabrication. QWIPs require the fabrication of random reflectors or gratings at the top of the detector in order to scatter normal-incidence light so that some portion will be laterally incident and thereby detectable by the QWIP. As discussed earlier, since QDIPs are inherently sensitive to normal-incidence light, no such processing is required. Most infrared detection applications require a two-dimensional format for imaging applications. Therefore, the vertical detection scheme in which arrays are illuminated from the top or backside and electrical contacts are placed correspondingly is typically used as opposed to the lateral detection scheme.

In addition to the typical InAs/GaAs QDIP, other III–V dot systems and heterostructures are being investigated for use as infrared detectors. In contrast to a typical QDIP with a series of symmetric and uniform layers, alternative material heterostructures may be designed to realize multicolor detectors [53]. In these structures, asymmetric barriers alter the response wavelength in differing bias directions to provide the capability for independent MWIR and LWIR response, depending on the bias conditions. As an example, InAs quantum dots embedded in strain-relieving InGaAs quantum wells are known as dot-in-a-well (DWELL) heterostructures. QDIPs using DWELLs not only permit greater control over detection wavelength tunability [68], but they have also demonstrated excellent device performance [69]. Devices containing an undoped active region of InAs quantum dots grown directly on AlGaAs with a GaAs cap layer have also demonstrated promising device performance [70]. Other III–V material systems under consideration include InAs/InAlAs QDs grown on InP [71] and InGaAs/InGaP QDs grown on GaAs [55]. Intraband transitions in the infrared have also been observed for the elemental semiconductor material system of Ge/Si quantum dots [72–74]. These boron-doped devices operate by the intraband absorption of infrared light by holes trapped in the quantum dot valence band. The device characterization of these Ge/Si QDIPs indicates they also may be an important technology for infrared detection [75–77]. A listing of reported QDIP results for various material systems and technologies is given in Table 1.

4.2. QDIP Characterization

Due to the large dark current present in all infrared photodetectors, QDIP properties must be measured at reduced temperatures (<300 K) by using a dewar or cryostat. Thus, QDIPs are usually mounted on a chip carrier, where contact is made with individual devices through gold wire bonds. The most straightforward detector measurement is the dark current as a function of bias voltage for various temperatures.

Spectral response measurements in QDIPs are often conducted at normal-incidence using a Fourier transform infrared (FTIR) spectrometer with a broadband (1–20 μm),

Table 1. Listing of QDIP results.

Quantum dot material	Barrier material	Growth technology	Device configuration	Ref.
InGaAs	GaAs	MBE	Vertical	[53, 54, 57, 58, 61, 64]
InGaAs	InGaP	MOCVD	Vertical	[55]
InGaAs	GaAs	MBE	Lateral	[56, 59, 62, 65]
InAs	InAlAs	MBE	Vertical	[60]
InGaAs	InGaAs/GaAs	MBE	Vertical	[63, 68, 69]
InGaAs	AlGaAs/GaAs	MBE	Vertical	[66, 67, 70]
InAs	InAlAs	MBE	Vertical	[71]
Ge	Si	MBE	Vertical	[75–77]

high-intensity source. Blackbody response measurements of QDIPs typically use a calibrated infrared photon source, such as a blackbody at a temperature anywhere from 500 to 800 K, to determine the absolute responsivity of the QDIP to normal-incidence IR radiation. In such a blackbody response measurement, the QDIP photocurrent signal and noise under illumination are measured with a fast Fourier transform (FFT) analyzer or a lock-in amplifier. A flat-band noise spectrum is desired because it indicates that the dominant noise mechanism is generation–recombination (GR) noise, as assumed in most theoretical calculations. The blackbody response measurement does not consider the wavelength of IR light that is detected; thus, spectral and blackbody responses are considered simultaneously to determine the detectivity and responsivity (quantum efficiency) at the peak wavelength response of the device.

The absolute responsivity may be determined by measuring the photocurrent induced by a calibrated blackbody source by

$$I_p = R_p \int R(\lambda) W(\lambda)\, d\lambda \tag{4}$$

where I_p is the photocurrent, R_p is the peak responsivity, and W is the blackbody power per unit wavelength at a given temperature. The responsivity shows a broadband response similar to the absorption spectra, corresponding to the inhomogeneous QD size distribution. In general, the spectral response, as shown in Figure 6, does not change significantly with bias. The peak responsivity, on the other hand, shows a clear dependence on bias, indicating the extraction process for photoexcited carriers. Typical values for peak responsivity at various temperatures are shown in Figure 7.

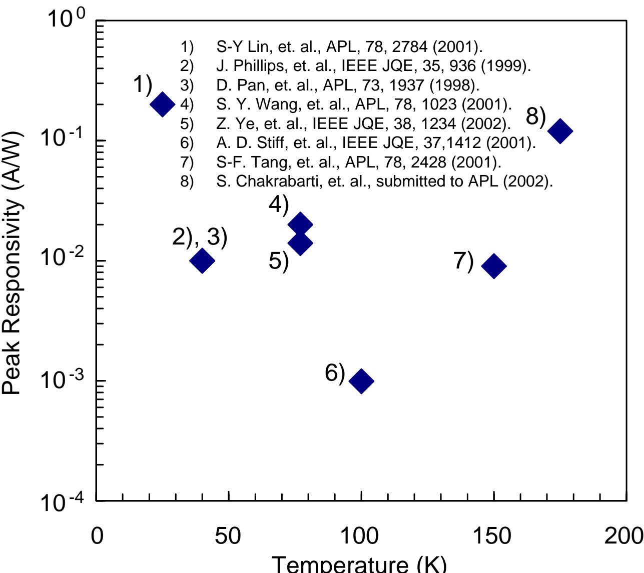


Figure 7. Peak responsivity as a function of temperature for several reported QDIP results.

4.3. Dark Current

Possible mechanisms for dark current in QDIPs include thermal generation, thermionic emission from QDs, interdot tunneling, and impact ionization [78]. Similar to QWIPs, the dominant mechanisms for dark current are suspected to be thermal generation and thermionic emission processes. To reduce dark current to sufficient levels, the devices are operated at cryogenic temperatures, as are all high-performance infrared photon detectors. A primary advantage predicted for QDIPs over QWIPs is the reduction of dark current due to reduced thermal generation in QDs versus QWs, as described in the above carrier dynamics section. A characteristic dark-current density–voltage curve for varying temperature in a vertical QDIP is shown in Figure 8. The asymmetry in the current–voltage curve is indicative of asymmetry in the device structure, which may stem from device structure (incorporation of barriers, nonuniform spacing

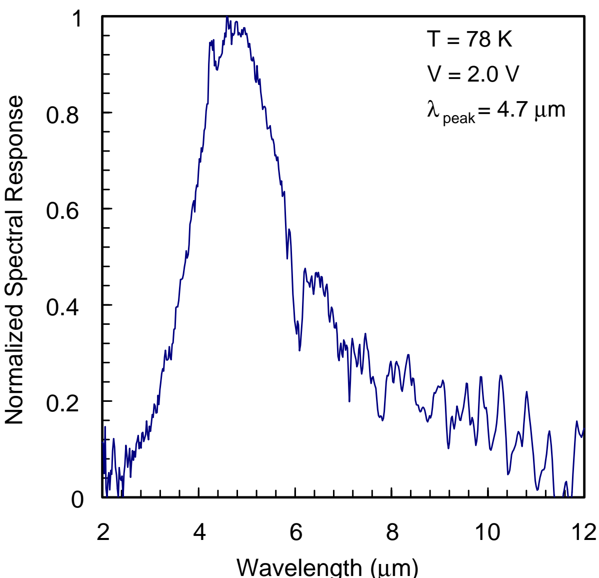


Figure 6. Spectral response of a vertical InAs/GaAs QDIP.

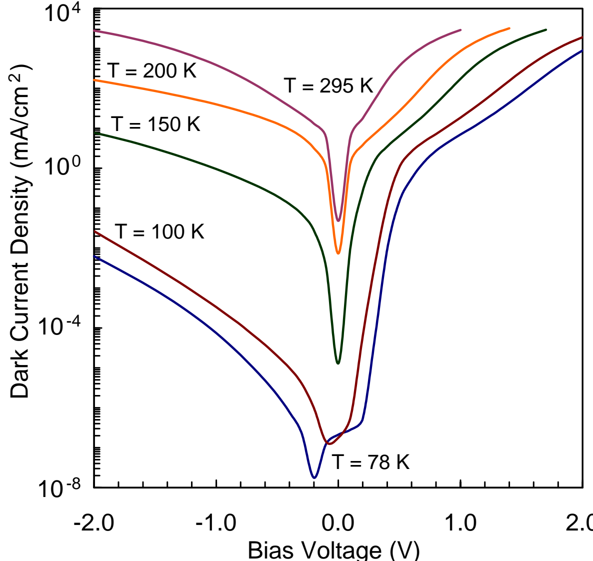


Figure 8. Temperature-dependent dark-current density–voltage characteristics for a vertical InAs/GaAs QDIP.

between QD layers), the QD shape, or doping nonuniformities. In addition to temperature dependence, the dark current shows a distinct dependence on energy separation or response wavelength correlating with the expected thermal generation processes. One of the inherent drawbacks with the typical photoconductor design is the larger dark current in comparison to a photodiode. As a result, QDIPs have been designed to incorporate higher bandgap barrier layers in the active region of the device and have demonstrated significant reductions in dark current [64].

4.4. Detectivity

A common figure of merit for detectors is the detectivity D^*, which reflects the signal-to-noise ratio independent of detector geometry and operation and is defined as

$$D^* = \frac{\text{Signal} \times \sqrt{\text{Bandwidth}}}{\text{Noise} \times \text{Incidance} \times \sqrt{\text{Area}}} \tag{5}$$

The value for detectivity is determined using calibrated blackbody response and dark-current noise measurements. Typical values for the peak detectivity are shown in Figure 9 for various temperatures.

4.5. Quantum Efficiency and Photoconductive Gain

The photocurrent in a photoconductive detector is given by

$$I_p = q\phi\eta g \tag{6}$$

where q is the charge of an electron, ϕ is the photon incidence, η is the quantum efficiency, and g is the photoconductive gain. The photoconductive gain is defined as the number of electrons flowing through the external circuit for each carrier generated in the device. This gain may be greater than 1 for the case where the carrier lifetime exceeds the transit time through the device. A direct determination of photoconductive gain or quantum efficiency requires a measurement of the product $\phi\eta$. This measurement is extremely difficult to do in practice due to optical reflection and the requirement to accurately determine absolute power levels. In practice, noise current measurements are typically used to determine photoconductive gain. In turn, the quantum efficiency is then inferred from photocurrent measurements.

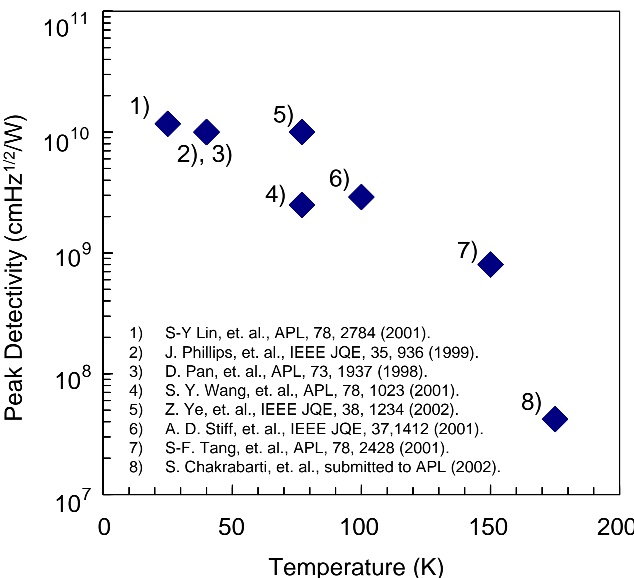


Figure 9. Peak detectivity as a function of temperature for several reported QDIP results.

In a QWIP, the commonly used expressions between gain and noise current were developed in the following manner. The mean carrier density flowing through a QWIP under steady state during time τ is given by n, where the mean density of captured and ejected carriers in a quantum well is given by

$$n_{ci} = n_{ei} = np_c \tag{7}$$

and p_c is the capture probability for a quantum well. This expression neglects tunneling between adjacent quantum wells. By the definition of g, the photoconductive gain is given by

$$g = \frac{n}{\sum n_{ei}} = \frac{1}{Np_c} \tag{8}$$

where N is the number of quantum wells in the device. Assuming that carrier generation and capture are statistically independent, the noise current may then be related to gain with the following expression

$$I_n^2 = 4qeIgB\left(1 - \frac{p_c}{2}\right) \tag{9}$$

Substitution of the relationship between gain and capture probability results in the following practical relationship

$$g = \frac{I_n^2}{4qIB} + \frac{1}{2N} \tag{10}$$

This relationship for gain in a photoconductor is commonly used to infer gain from noise current and dark current, where photoconductive gain is assumed to be equivalent to dark-current gain. This has been shown to be a good approximation for QWIPs.

For QDIPs, however, there are important fundamental differences, precluding the direct use of the QWIP photoconductive gain expression. In a quantum well, the presence of subbands provides a continuum of energy states for the capture and emission process, where state filling may be neglected under typical QWIP operation. In a QDIP, electron occupation in QD states governs the capture and emission process. Electron occupation in QD states in QDIPs with and without illumination differs significantly, and dark-current gain *does not* directly relate to photoconductive gain. The following simplified picture may be used to provide a revised expression for photoconductive gain in QDIPs. The capture and emission processes in steady state for a carrier density n are given by

$$n_{ei} = n_{ci} = n[p_{c1i}(1-m) + p_{c2i}] \tag{11}$$

where p_{c1i} and p_{c2i} are capture probabilities into the ground and excited QD states, respectively, and m represents the occupation probability for the QD ground state. Note that carrier relaxation processes between the QD ground and excited states are not explicitly represented in the previous expression, where these processes do not directly contribute to photocurrent or GR noise. These are also neglected in the QWIP analysis, and in the QDIP case will affect electron occupation in the QD states. The factor m is used to represent this relationship. The revised photoconductive gain expression is then given by

$$g = \frac{1}{N[p_{c1i}(1-m) + p_{c2i}]} \tag{12}$$

and the following relationship with noise current

$$g = \frac{I_n^2}{2qIB[2 + p_{c1}p_{c2}(1-m)]} + \frac{1}{N[2 + p_{c1}p_{c2}(1-m)]} \tag{13}$$

Note that gain cannot directly be determined without knowledge of the capture probability or occupation of the QD ground state. This analysis only accounts for state filling in the QD ground state, though filling of excited states may also significantly affect gain. For dark-current measurements, occupation of QD ground states should be significant, with $m \sim 1$. For illumination, occupation of QD ground states will be reduced due to photoexcitation, $m < 1$. The capture probabilities will be affected by the "fill factor" in the quantum dots and will change with electron occupation. The actual photoconductive gain predicted by the preceding analysis can be lower than the dark-current gain by approximately 30%. This analysis further suggests that the determination of the quantum efficiency (QE) of QDIPs in the current method of measurement is lower than the actual quantum efficiency (the real QE should be about 30% larger for the simulation given later). Note that the QDIP expressions reduce to the QWIP expressions when returning to a one-level versus two-level model for gain. Despite the differences between photoconductive gain in QWIPs and QDIPs, the expressions for QWIPs are currently being used to estimate gain and quantum efficiency from blackbody response and noise current measurements.

The quantum efficiency of QDIPs determined using the above technique has been generally low, approximately 1%, and has not come to equal that of QWIPs (~20%), suggesting that the size uniformity needs to be improved and/or the ability to increase the number of quantum dots in a detector is needed. The reduced quantum efficiency for QDIPs in comparison to QWIPs is consistent with the reduced QD peak absorption compared to quantum wells.

5. EVALUATION OF QDIP PERFORMANCE

The fundamental limitation for an IR detector is commonly referred to as the background limited performance (BLIP). This is the condition where the density of photogenerated carriers is equal to the density of carriers generated thermally. Kinch has made an excellent comparison for fundamental limitations in IR detector materials through analysis of the BLIP condition [79]. This analysis allows one to evaluate a materials technology independent of device configuration and examines the BLIP condition where photogenerated carriers are greater than thermally generated carriers in the material, given by

$$\eta\Phi\tau/t > n_{th} \tag{14}$$

where η is the absorption quantum efficiency, Φ is the photon flux, τ is the carrier lifetime, t is the material thickness in the direction of the incident photon flux, and n_{th} is the thermally generated carrier density. Using $\eta = \alpha t$, where α is the absorption coefficient in the material, we get the BLIP requirement $\Phi > n_{th}/\alpha\tau$ and obtain a normalized thermal generation rate [79]

$$G_{th} = n_{th}/\alpha\tau \tag{15}$$

This provides a simple technique for comparing differences between materials technologies for infrared detectors. The normalized thermal generation rate for a QDIP may be calculated using this technique, as described in [80]. A comparison of the normalized dark current for HgCdTe, QWIP, and QDIP, as obtained from the above equations is shown in Figure 10 for a bandgap energy of $E_g = 0.124$ eV corresponding to a cutoff wavelength of $\lambda_c \sim 10$ μm. Two cases for QDIPs are given in the figure: (1) The nonuniform case represents the typical QD size uniformity observed in current self-assembled QD ensembles, and (2) the uniform case represents a perfectly uniform QD ensemble. As can be seen from Figure 10, the performance of QDIPs is predicted to surpass that of QWIPs. In the case of dense QD ensembles with high size uniformity, QDIPs are predicted to rival and perhaps outperform HgCdTe.

These predictions agree with the theory reported by Ryzhii [81], where QDIPs were first proposed due to the following anticipated advantages over QWIPs: inherent sensitivity to normal-incidence radiation, low dark current,

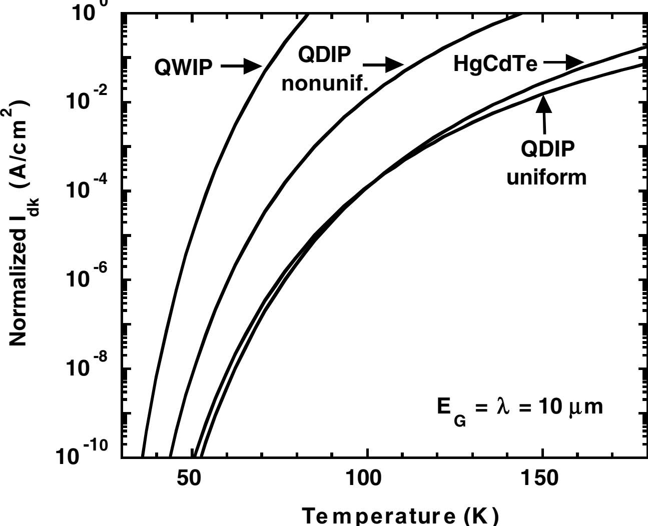


Figure 10. Normalized dark current for HgCdTe, QWIP, and QDIP detectors as a function of temperature for a peak response wavelength of 10 μm, using the analysis at BLIP outlined in the text.

inhibited carrier capture rates, and high photoconductive gain. This theory, however, considers somewhat idealized quantum dots in that the effect of inhomogeneous broadening resulting from the random size distribution in a QD ensemble is not considered. Such inhomogeneous broadening adversely affects actual device behavior. While the device characterization of QDIPs has demonstrated excellent, high-temperature performance, as discussed in Section 4, these devices have not yet yielded the same responsivity, quantum efficiency, or detectivity as QWIPs or as expected from theoretical calculations. Comprehensive theories have been developed to better understand the behavior of real devices [82–84]. In general, these theoretical models indicate that QDIP performance, in particular, the dark current and detectivity, can become comparable or superior to QWIP performance given the controlled growth of high-density, uniform QD ensembles composed of small dots in both the lateral and the growth directions.

6. CONCLUSION

III–V-based quantum dot infrared photodetectors are positioned to become an important technology in the field of infrared detection. Their inherent sensitivity to normal-incidence radiation, lower dark currents, higher operating temperatures, multispectral tunability, and mature processing techniques indicate that these devices could provide the type of high-resolution, low-cost, and high-yield imaging components required by the military and other highly sophisticated applications.

Excellent device performance has been demonstrated using QDIPs. Theoretical models predict that QDIPs should provide comparable performance to HgCdTe detectors and better performance than QWIPs, especially at elevated temperatures. However, QDIPs are not without disadvantages, and there are some challenges that must be overcome in order to take full advantage of this technology. The random growth process of self-assembled quantum dots adversely affects device performance due to the inhomogeneous broadening associated with the random distribution in quantum dot size. However, this problem can be resolved by the growth of dense, small quantum dot ensembles. Additionally, the quantum efficiency of these devices must be increased in order to improve the performance of QDIPs. Further study of MBE growth of self-assembled quantum dots, as well as more detailed modeling of realistic QDIP devices should aid in this development.

GLOSSARY

Dark current The external current that, under specified biasing conditions, flows in a photoconductive detector when there is no incident radiation.

Detectivity Detector figure of merit commonly used to characterize and compare the signal-to-noise ratio.

Epitaxy The growth of a thin layer on the surface of a crystal so that the layer has the same structure as the underlying crystal.

Long-wavelength infrared (LWIR) Typically referring to infrared wavelengths between 8 and 12 μm.

Mid-wavelength infrared (MWIR) Typically referring to infrared wavelengths between 3 and 5 μm.

Quantum dot infrared photodetector (QDIP) An infrared photon detector utilizing optical transitions between bound electronic states in the conduction or valence band of quantum dots.

Quantum efficiency In an optical detector, the ratio of the number of output quanta (electrons) to the number of input quanta (photons).

Quantum well infrared photodetector (QWIP) An infrared photon detector utilizing optical transitions between bound electronic states in the conduction or valence band of quantum wells.

Responsivity In a photodetector, the ratio of the electrical output to the optical input.

Short-wavelength infrared (SWIR) Typically referring to infrared wavelengths between 1 and 3 μm.

Very-long-wavelength infrared (VLWIR) Typically referring to infrared wavelengths longer than 12 μm.

REFERENCES

1. P. R. Norton, *Proc. SPIE* 3379, 102 (1998).
2. J. Bajaj, *Proc. SPIE* 3948, 42 (2000).
3. M. B. Reine, *Proc. SPIE* 4028, 320 (2000).
4. I. M. Baker and C. D. Maxey, *J. Electron. Mater.* 30, 682 (2001).
5. J. B. Varesi, R. E. Bornfreund, A. C. Childs, W. A. Radford, K. D. Maranowski, J. M. Peterson, S. M. Johnson, L. M. Giegerich, T. J. deLyon, and J. E. Jensen, *J. Electron. Mater.* 30, 566 (2001).
6. F. Fuchs, U. Weimar, E. Ahlswede, W. Pletschen, J. Schmitz, and M. Walther, *Proc. SPIE* 3287, 14 (1998).
7. M. H. Young, D. H. Chow, A. T. Hunter, and R. H. Miles, *Appl. Surf. Sci.* 123/124, 395 (1998).
8. H. Mohseni, M. Razeghi, G. J. Brown, and Y. S. Park, *Appl. Phys. Lett.* 78, 2107 (2001).
9. S. D. Gunapala and K. Bandara, *Thin Films* 21, 113 (1995), and references therein.
10. S. D. Gunapala, S. V. Bandara, J. K. Liu, W. Hong, M. Sundaram, R. Carralejo, C. A. Chott, P. D. Maker, and R. E. Miller, *Proc. SPIE* 3061, 722 (1997).
11. P. D. Wang, C. M. S. Torres, H. Benisty, C. Weisbuch, and S. P. Beaumont, *Appl. Phys. Lett.* 61, 946 (1992).
12. R. Steffen, T. Koch, J. Oshinowo, F. Faller, and A. Forchel, *Appl. Phys. Lett.* 68, 223 (1996).
13. S. Nilsson, E. Van Gieson, D. J. Arent, H. P. Meier, W. Walter, and T. Forster, *Appl. Phys. Lett.* 55, 972 (1989).
14. F. S. Turco, M. C. Tamargo, D. M. Hwang, R. E. Nahory, J. Werner, K. Kash, and E. Kapon, *Appl. Phys. Lett.* 56, 72 (1989).
15. S. Guha and A. Madhukar, *J. Appl. Phys.* 73, 8662 (1993).
16. S. Simhony, E. Kapon, E. Colas, D. M. Hwang, N. G. Stoffel, and P. Worland, *Appl. Phys. Lett.* 59, 2225 (1991).
17. M. Walther, E. Kapon, J. Christen, D. M. Hwang, and R. Bhat, *Appl. Phys. Lett.* 60, 521 (1992).
18. S. Tiwari, G. D. Pettit, K. R. Milkove, F. Legoues, R. J. Davis, and J. M. Woodall, *Appl. Phys. Lett.* 64, 3536 (1994).
19. L. Goldstein, F. Glas, J. Y. Marzin, M. N. Charasse, and G. LeRoux, *Appl. Phys. Lett.* 47, 1099 (1985).
20. S. Fujita, Y. Matsuda, and A. Sasaki, *Appl. Phys. Lett.* 47, 955 (1985).
21. P. R. Berger, K. Chang, P. K. Bhattacharya, and J. Singh, *J. Vac. Sci. Technol., B* 5, 1162 (1987).
22. D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, and P. M. Petroff, *Appl. Phys. Lett.* 63, 3202 (1993).

23. Q. Xie, P. Chen, A. Kalburge, T. R. Ramachandran, A. Nayfonov, A. Konkar, and A. Madhukar, *J. Cryst. Growth* 150, 357 (1995).
24. D. Bimberg, M. Grundmann, and N. N. Ledentsov, *Mater. Res. Soc. Bull.* 23, 31 (1998).
25. P. Bhattacharya, K. Kamath, J. Phillips, and D. Klotzkin, *Bull. Mater. Sci.* 22, 519 (1999).
26. I. N. Stranski and L. Krastanow, *Abt. IIb*, 146, 146 (1937).
27. J. L. Spithoven, J. Lorbacher, I. Manke, F. Heinrichsdorff, A. Krost, D. Bimberg, and M. Dahne-Prietsch, *J. Vac. Sci. Technol., B* 17, 1632 (1999).
28. H. Jiang and J. Singh, *Phys. Rev. B* 56, 4696 (1996).
29. M. Grundmann, N. N. Ledentsov, O. Stier, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, and Zh. I. Alferov, *Appl. Phys. Lett.* 68 (1996).
30. M. J. Steer, D. J. Mowbray, W. R. Tribe, M. S. Skolnick, M. D. Sturge, M. Hopkinson, A. G. Cullis, C. R. Whitehouse, and R. Murray, *Phys. Rev. B* 54, 17738 (1996).
31. R. Heitz, A. Kalburge, Q. Xie, M. Grundmann, P. Chen, A. Hoffman, A. Madhukar, and D. Bimberg, *Phys. Rev. B* 57, 9050 (1998).
32. G. Medeiros-Ribeiro, D. Leonard, and P. M. Petroff, *Appl. Phys. Lett.* 66, 1767 (1995).
33. S. Ghosh, B. Kochman, J. Singh, and P. Bhattacharya, *Appl. Phys. Lett.* 76, 2571 (2000).
34. S. Sauvage, P. Boucaud, F. H. Julien, J.-M. Gérard, and J.-Y. Marzin, *J. Appl. Phys.* 82, 3396 (1997).
35. S. Maimon, E. Finkman, G. Bahir, S. E. Schacham, J. M. Garcia, and P. M. Petroff, *Appl. Phys. Lett.* 73, 2003 (1998).
36. L. Chu, A. Zrenner, G. Böhm, and G. Abstreiter, *Appl. Phys. Lett.* 75, 3599 (1999).
37. Z. Chen, E.-T. Kim, and A. Madhukar, *Appl. Phys. Lett.* 80, 2490 (2002).
38. I. Mukhametzhanov, Z. H. Chen, O. Baklenov, E. T. Kim, and A. Madhukar, *Phys. Status Solidi B* 224, 697 (2001).
39. J. Phillips, K. Kamath, and P. Bhattacharya, *Appl. Phys. Lett.* 72, 2020 (1998).
40. N. Horiguchi, T. Futatsugi, Y. Nakata, N. Yokoyama, T. Mankad, and P. M. Petroff, *Jpn. J. Appl. Phys.* 38, 2559 (1999).
41. D. Pan, Y. P. Zeng, M. Y. Kong, J. Wu, Y. Q. Zhu, C. H. Zhang, J. M. Li, and C. Y. Wang, *Electron. Lett.* 32, 1726 (1996).
42. K. W. Berryman, S. A. Lyon, and M. Segev, *J. Vac. Sci. Technol.* 15, 1045 (1997).
43. J. Phillips, K. Kamath, X. Zhou, N. Chervela, and P. Bhattacharya, *Appl. Phys. Lett.* 71, 2079 (1997).
44. S. Sauvage, P. Boucaud, F. H. Julien, J. M. Gerard, and J. Y. Marzin, *J. Appl. Phys.* 82, 3396 (1997).
45. H. Jiang and J. Singh, *Appl. Phys. Lett.* 71, 3239 (1997).
46. S. Sauvage, P. Boucaud, F. H. Julien, J. M. Gerard, and V. Thierry-Mieg, *Appl. Phys. Lett.* 71, 2785 (1997).
47. S. J. Chua, S. J. Xu, X. H. Zhang, X. C. Wang, T. Mei, W. J. Fan, C. H. Wang, J. Jiang, and X. G. Xie, *Appl. Phys. Lett.* 73, 1997 (1998).
48. B. Ohnesorge, M. Albrecht, J. Oshinowo, A. Forchel, and Y. Arakawa, *Phys. Rev. B* 54, 11532 (1996).
49. I. Vurgaftman and J. Singh, *Appl. Phys. Lett.* 64, 232 (1994).
50. T. Sosnowski, T. Norris, H. Jiang, J. Singh, K. Kamath, and P. Bhattacharya, *Phys. Rev. B* 57, R9423 (1998).
51. J. Urayama, T. B. Norris, J. Singh, and P. Bhattacharya, *Phys. Rev. Lett.* 86, 4930 (2001).
52. D. Klotzkin and P. Bhattacharya, *J. Lightwave Technol.* 17, 1634 (1999).
53. D. Pan, E. Towe, and S. Kennerly, *Electron. Lett.* 34, 1019 (1998).
54. J. Phillips, K. Kamath, and P. Bhattacharya, *Appl. Phys. Lett.* 72, 2020 (1998).
55. S. Kim, H. Mohseni, M. Erdtmann, E. Michel, C. Jelen, and M. Razeghi, *Appl. Phys. Lett.* 73, 963 (1998).
56. S. W. Lee, K. Hirakawa, and Y. Shimada, *Appl. Phys. Lett.* 75, 1428 (1999).
57. J. Phillips, P. Bhattacharya, S. Kennerly, D. Beekman, and M. Dutta, *IEEE J. Quantum Electron.* 36, 936 (1999).
58. N. Horiguchi, T. Futatsugi, Y. Nakata, N. Yokoyama, T. Mankad, and P. Petroff, *Jpn. J. Appl. Phys.* 38, 2559 (1999).
59. J. Kim, J. Oh, S. Hong, C. Park, and T. Yoo, *IEEE Electron. Device Lett.* 21, 329 (2000).
60. E. Finkman, S. Maimon, V. Immer, G. Bahir, S. E. Schacham, O. Gauthier-Lafaye, S. Herriot, F. H. Julien, M. Gendry, and J. Brault, *Physica E* 7, 139 (2000).
61. H. C. Liu, M. Gao, J. McCaffrey, Z. R. Wasilewski, and S. Fafard, *Appl. Phys. Lett.* 78, 79 (2001).
62. L. Chu, A. Zrenner, M. Bichler, and G. Abstreiter, *Appl. Phys. Lett.* 79, 2249 (2001).
63. E. Kim, Z. Chen, M. Ho, and A. Madhukar, *J. Vac. Sci. Technol., B* 20, 1188 (2002).
64. A. D. Stiff, S. Krishna, P. Bhattacharya, and S. Kennerly, *IEEE J. Quantum Electron.* 37, 1412 (2001).
65. T. Cho, J.-W. Kim, J.-E. Oh, J.-W. Choe, and S. Hong, *Jpn. J. Appl. Phys.* 38, 2442 (1999).
66. S. Y. Wang, S. D. Lin, H. W. Wu, and C. P. Lee, *Appl. Phys. Lett.* 78, 1023 (2001).
67. S.-F. Tang, S.-Y. Lin, and S.-C. Lee, *Appl. Phys. Lett.* 78, 2428 (2001).
68. E.-T. Kim, Z. Chen, and A. Madhukar, *Appl. Phys. Lett.* 79, 3341 (2001).
69. S. Raghavan, P. Rotella, A. Stintz, B. Fuchs, S. Krishna, C. Morath, D. A. Cardimona, and S. W. Kennerly, *Appl. Phys. Lett.* 81, 1369 (2002).
70. Z. Chen, O. Baklenov, E. T. Kim, I. Mukhametzhanov, J. Tie, A. Madhukar, Z. Ye, and J. C. Campbell, *J. Appl. Phys.* 89, 4558 (2001).
71. A. Weber, O. Gauthier-Lafaye, F. H. Julien, J. Brault, M. Gendry, Y. Désieres, and T. Benyattou, *Appl. Phys. Lett.* 74, 413 (1999).
72. J. L. Liu, W. G. Wu, A. Balandin, G. L. Jin, and K. L. Wang, *Appl. Phys. Lett.* 74, 185 (1999).
73. P. Boucaud, V. Le Thanh, S. Sauvage, D. Dé Barre, and D. Bouchier, *Appl. Phys. Lett.* 74, 401 (1999).
74. J. L. Liu, W. G. Wu, A. Balandin, G. Jin, Y. H. Luo, S. G. Thomas, Y. Lu, and K. L. Wang, *Appl. Phys. Lett.* 75, 1745 (1999).
75. A. I. Yakimov, A. V. Dvurechenskii, Yu. Yu. Proskuryakov, A. I. Nikiforov, O. P. Pchelyakov, S. A. Teys, and A. K. Gutakovskii, *Appl. Phys. Lett.* 75, 1413 (1999).
76. C. Miesner, O. Röthig, K. Brunner, and G. Abstreiter, *Physica E* 7, 146 (2000).
77. A. I. Yakimov, A. V. Dvurechenskii, A. I. Nikiforov, and Yu. Yu. Proskuryakov, *J. Appl. Phys.* 89, 5676 (2001).
78. J. Allam, F. Capasso, K. Alavi, and A. Y. Cho, *Inst. Phys. Conf. Ser.* 83, 405 (1986).
79. M. A. Kinch, *J. Electron. Mater.* 29, 809 (2000).
80. J. Phillips, *J. Appl. Phys.* 91, 4590 (2002).
81. V. Ryzhii, *Semicond. Sci. Technol.* 11, 759 (1996).
82. V. Ryzhii, *J. Appl. Phys.* 89, 5117 (2001).
83. V. Ryzhii, I. Khmyrova, V. Pipa, V. Mitin, and M. Willander, *Semicond. Sci. Technol.* 16, 331 (2001).
84. V. Ryzhii, I. Khmyrova, V. Mitin, M. Stroscio, and M. Willander, *Appl. Phys. Lett.* 78, 3523 (2001).

www.aspbs.com/enn

Quantum Dots: Artificial Atoms and Molecules

Philippe Matagne, Jean-Pierre Leburton

Beckman Institute for Advanced Science and Technology, Urbana, Illinois, USA

CONTENTS

1. Introduction
2. Quantum Dots
3. Coulomb Blockade Effect and Single-Electron Tunneling
4. Tarucha's Experiment
5. The Many-Body Hamiltonian of Artificial Atoms
6. Quantum Dots as Electronic Devices
7. Full Scale Simulation of Quantum Dots
8. Asymmetrical Single Quantum Dots
9. Double Quantum Dots—Artificial Molecules
Glossary
References

1. INTRODUCTION

Recent years have witnessed the advancement of miniaturization techniques to such a sophisticated level that semiconductor structures with dimensions of a few hundred angstroms can be fabricated routinely. While this technology led to a reduction in feature sizes of conventional devices—for instance, deep submicron metal–oxide–semiconductor field effect transistors (MOSFETs)—it also made possible the realization of a whole new class of devices based on novel physical phenomena [1–3]. With the length scale of these devices approaching the de Broglie wavelength of carriers, electron motion is confined. This may be in one, two, or three dimensions with the corresponding systems being called quantum wells, quantum wires, and quantum dots (QDs), respectively. Therefore, electrons are confined to move in a plane in quantum wells, are confined to move along a line in quantum wires, and are mostly localized in QDs.

This chapter describes the electronical properties of QDs and addresses general modeling issues, mostly with the density functional formalism. After giving an introduction to the different QD systems in Section 2, we review in detail the phenomena of Coulomb blockade and single-electron charging in Section 3. The distinction is made between the classical regime discovered earlier in metallic QDs [4–6] and the quantum regime recently evidenced in semiconductor QDs [7–9]. The latter QDs are often called *artificial atoms* [10] because, as in real atoms, the electron energies are totally quantized and the total charge is restricted to an integer number of electrons. Section 4 describes the first experiment, carried out by Tarucha and co-workers, that evidences a shell structure in semiconductor vertical quantum dots (VQD) [11]. Through the observation of single-electron charging effects, shell filling in VQD was interpreted as governed by Hund's first rule, thereby manifesting single spin effects. However, it is worth emphasizing the differences between artificial atoms and their real counterparts as shown in Figure 1. In atoms, the attractive potential arising from the nucleus positive charge is central symmetric and varies as the inverse of the distance from the nucleus. In QDs, the confining potential depends on the way the nanostructure is fabricated. In many cases, the electron gas is strongly confined in one direction and subject to a more or less harmonic potential in the other two directions. Furthermore, while the nuclear charge is an integer number of positive elementary charges, the net attractive charge of a QD varies continuously by means of an external gate [10]. Sections 5 and 6 describe the physics background for our numerical model. In particular, the density functional formalism and the important concept of addition energy are introduced. In Sections 7 and 8, we present the results of simulations on cylindrical QDs, rectangular QDs, and electrostatically deformable QD structures, respectively. Section 7 discusses symmetric structures while Section 8 concentrates on the consequences of breaking this symmetry, structurally and electrostatically. Finally, Section 9 studies the electronic properties of double QDs, or *artificial molecules*.

Motivation for studies on QDs arises from the many possibilities they afford for investigation of many-body interactions in an atomlike setup. Whereas it is not possible to

* This chapter first appeared in *Quantum Dots and Nanowires*, Edited by S. Bandyopadhyay and H. S. Nalwa. © 2003, American Scientific Publishers.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

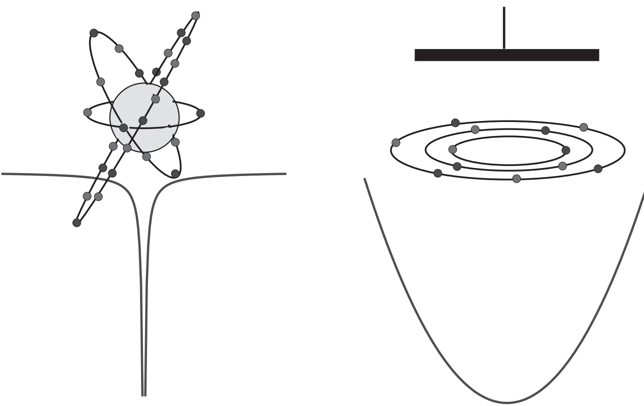


Figure 1. Schematic representation of a 3D real atom (left) with central symmetry and a 2D artificial atom (right) with a parabolic confining potential.

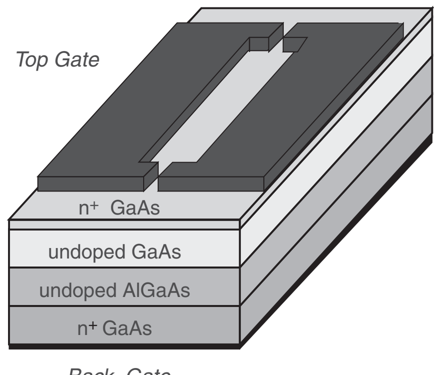


Figure 2. Schematic diagram of a PQD showing the gates and different material layers.

alter the nature and strength of the confining potential in atoms as they are fixed by the nuclear charge, it can be varied more freely in QDs. Furthermore, it is possible to populate the QD with an arbitrary number of electrons, corresponding to a specific atom in the periodic table. An additional important feature is that the length scale on which such phenomena occur is translated from the atomic dimensions (a few angstroms) to the mesoscopic dimensions (thousand angstroms), making the observations easier. All these factors and the many challenges that they pose to theorists have made the study of mesoscopic devices hugely popular in recent years. The interest from an engineering point of view comes from their many promising applications as optical memory elements, floating gate flash memory cells, lasers, photodetectors, current standards, etc. [12–16]. Moreover, with the emergence of quantum information, artificial molecules appear to be a very favorable ground to generate entangled states and therefore to be the core of quantum computers [17, 18].

2. QUANTUM DOTS

QDs are nanostructures electrostatically or structurally isolated from the outside in which electrons are confined. Typically QDs consist of a few hundreds to a few millions of atoms, but only a small number of electrons ($\leq$100) are free. Depending on electron confinement, it is possible to distinguish between planar QDs, vertical QDs, and self-assembled QDs.

2.1. Planar Quantum Dots

The first experiments with semiconductor QDs were made with planar structures by patterning several Schottky metal gates above the surface of a two-dimensional electron gas (2DEG) which isolates electrostatically a small region and form a planar quantum dot (PQD) [7, 19–25] (for a more exhaustive list see, e.g., [26]). An example of such structure is depicted on Figure 2. By applying a negative bias on the top gate, electrons are repelled and a small island is isolated at the center of the 2DEG, at the n^+GaAs/AlGaAs interface. A back gate allows one to change the dot confinement and thus the number of electrons inside the dot.

PQDs have two main characteristics. (i) The current flows in the 2DEG plane; (ii) the tunnel barrier are created electrostatically which has an advantage: their height is controllable. However, electrostatic barriers are weak (a few meV high) and thick ($\approx$100 nm). Moreover they have variable shape which is strongly influenced by the gate voltage. That makes PQDs quite sensitive to external perturbations.

2.2. Vertical Quantum Dots

In vertical quantum dots (VQDs), the zero-dimensional region is sandwiched vertically between a double resonant tunneling barrier made of heterostructure layers, and lateral confinement is achieved by a combination of deep mesa etching and electrostatic depletion [8, 9, 11, 27–37]. The dot confinement is controlled by a Schottky gate that is deposited on top of the mesa [30] or is wrapped around it [31]. VQD atomiclike properties were first evidenced by Ashoori, by capacitance spectroscopy measurements [8, 29, 30]. Later, in addition to the quantum features mentioned above, spin effects were clearly observed by Tarucha in experiments with a device similar to the one depicted in Figure 3 [9, 11, 31, 34, 35].

The main characteristics of VQD are: (i) the current flows vertically, perpendicularly to the 2DEG plane; (ii) the tunnel

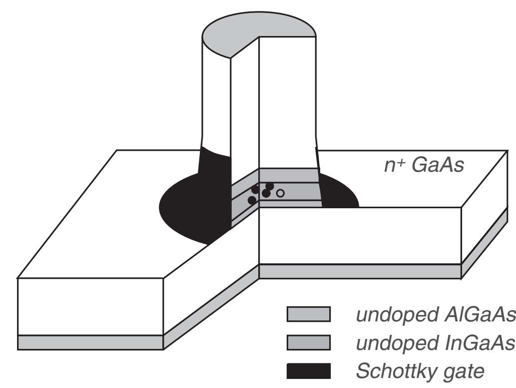


Figure 3. Schematic diagram of Tarucha's VQD. Adapted with permission from [11], S. Tarucha et al., *Phys. Rev. Lett.* 77, 3613 (1996). © 1996, American Institute of Physics.

barriers are high (>100 meV) and thin (≈10 nm) because they are due to the large bandgap material such as GaAs, but then, they are not electrically tunable. VQDs are therefore less sensitive to external fluctuations.

The most popular device made with PQDs and VQDs is the *single-electron transistor* (SET) in which the QD replaces the inversion channel of conventional MOSFETs [38]. Potential applications for SETs are metrology [39, 40], optical detectors [16], and ultradense digital logic circuits [15]. However, due to their high sensitivity to random background charges, SETs future is not clear [6].

2.3. Self-Assembled Quantum Dots

In PQDs and VQDs, electrostatic confinement leads typically to dimensions around 100 nm, and structural confinement is of the order of 10 nm. The eigenlevel separation of a rectangular box of such dimensions is about 1 meV. To prevent quantum effects from being smeared out by thermal fluctuations, experiments with PQDs and VQDs are conducted at very low temperature, typically below 1 K [7, 11].

When a few monolayers of InAs are grown on a GaAs substrate, the lattice mismatch between the two materials induces a strain which leads to the breakdown of a 2D layered structure and the emergence of randomly distributed self-crystallized islands called self-assembled quantum dots (SAQDs) on the GaAs substrate [41–44] (for a more complete list of references see, e.g., Jiang [45]). The confinement in SAQDs is determined by the conduction band offset between InAs and GaAs and is therefore much stronger than in PQDs and VQDs. This leads to pyramidal or lens-shape structures with sizes of approximately 100 Å. Pyramidal QDs are very promising structures for laser applications [46–49] because of their spectral purity, relative insensitivity to temperature, low current threshold, and high differential gain [50, 51]. However, they are relatively small to be addressed individually by electrical means, which constitutes an important drawback for electronic applications.

Even more promising candidates for room temperature devices are the so called "silicon nanocrystal s" which consist of 50 Å diameter spherical silicon particles embedded in silicon dioxide [52–55]. Here the eigenlevel separation is typically about 100 meV [56]. Several groups have fabricated flash memories based on nanocrystal s which surpass classical CMOS devices in terms of retention time and power consumption.

3. COULOMB BLOCKADE EFFECT AND SINGLE-ELECTRON TUNNELING

The ability to observe the single-electron charging effect relies essentially on the *Coulomb blockade effect* (CBE), which can be described as discrete changes in charging energy of an ultrasmall capacitor due to electrostatic interactions between electrons. We follow Likharev [6] and consider a small piece of conducting material called the *island* as being "somehow" isolated from its surrounding. Originally, the island is electrically neutral (Fig. 4a) and therefore, little work is required to bring an electron on the island. However, since the island is well isolated, after accepting the

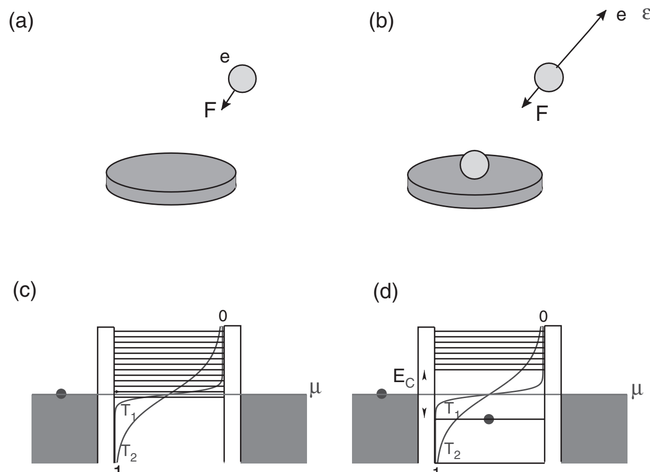


Figure 4. Principle of the CBE: spatial (a and b) and energetic (c and d), representation before (a and c) and after (b and d) the addition of a single electron as explained in detail in the text. Adapted with permission from [60], K. Likharev, *Proc. IEEE* 87, 606 (1999). © 1999, IEEE.

electron, the island charge becomes $-e$ (Fig. 4b). In order to add another electron on the island, one must then overcome the repulsive force originating from the electric field ε. The required energy is called the *charging energy* E_C. This situation is contrasted with a large conductor where charges are allowed to redistribute easily, and the individual electronic charges in the capacitor can be readily screened by polarizing charges in the system. Here, the island is a QD isolated from a reservoir by tunnel barriers and the injection of electrons is controlled by a gate capacitively coupled to the dot; that is, the electric field from the QD electrons can only be screened by adjusting the external control gate potential. If C is the effective capacitance between the dot and the reservoir, the charging energy is [6]

$$E_C = \frac{e^2}{C} \tag{1}$$

Although CBE was already observed in thin metallic films as early as 1950 by Gorter [4], single electronics gained popularity in the mid 1980s with the development of the so-called "orthodox theory" by Kulik and Shekhter [5] and Averin and Likharev [6]. This theory immediately supported by almost all experiments with metallic conductors. A theory similar to the orthodox theory, but taking into account level quantization in the dot, was later developed by Averin [57] and Beenakker [58].

3.1. The Coulomb Blockade Regime

It is worth mentioning that there is an visible contradiction between this classical picture—an integer number of electron N in the dot—and the quantum mechanical picture which associates to any observable an average value, here $\langle N \rangle$, and fluctuations with respect to this value, here the standard deviation $\langle \delta N \rangle$. In order to reconcile these two pictures, $\langle \delta N \rangle$ must be much smaller than 1. This criterion

is met if the average time $\langle\tau\rangle$ spent by the electron inside the dot is much larger than the charging time, $\Delta\tau > \hbar/SE$, where SE is the electron level broadening. In other words, the tunnel barriers must be opaque enough to prevent the electron from going back and forth between the dot and the reservoir. To quantify the opacity of the barrier by a resistance R, we use Thouless's argument [58]: let $i_e = e/\langle\tau\rangle$ be the current provided by one electron flowing across the barrier and let v_B be the bias allowing this current to flow. Then $R = v_B/i_e$. From Heisenberg uncertainty relation $\Delta\tau\,\delta E \geq \hbar$ and the condition for observation of CBE (i.e., $\langle\tau\rangle \gg \Delta\tau$ and $\delta E \ll ev_B$), we obtain

$$R \gg \frac{\hbar}{e^2} \tag{2}$$

$\hbar/e^2 \approx 4\ k\Omega$ is called the *quantum resistance*. When Eq. (2) is satisfied, it can be shown [26] that *cotunneling*—a two-electron process in which one electron from the lead tunnels into the dot while a second electron tunnels simultaneously from another state in the dot to the lead—can be neglected as well.

In the other regime (i.e., $R \leq \hbar/e^2$), when the barriers are too transparent (not high or wide enough), the electron wave function extends largely over the barriers and the electron cannot be treated as a localized particle. This regime corresponds to the *Kondo effect* [59] and is not treated here.

Moreover, CBE can only be observed if the charging energy is much larger than thermal fluctuations in the system. According to Likharev [60], $E_C \approx 100k_BT$ is a good condition for preventing CBE to be smeared out by thermal fluctuations. This situation is illustrated in Figures 4c and 4d where the Fermi–Dirac distribution has been superimposed on the energy levels inside the dot to account for thermal fluctuations at $T = T_1$ and T_2 ($T_1 < T_2$). When the dot is empty (Fig. 4c), an electron outside the dot can tunnel at both temperatures. When there is one electron inside the dot, the CBE creates a gap in the energy spectrum of the dot (Fig. 4d). At $T \leq T_1$, in the *Coulomb blockade regime* (CBR), there are no levels available to the electron outside the dot, and the conduction is suppressed. At $T = T_2$, thermal fluctuations make higher levels available to the outside electron, and the CBE is washed out.

In summary, the CBR can be characterized by these two criteria:

$$R \gg \frac{\hbar}{e^2} \qquad E_C \gg k_BT \tag{3}$$

3.2. Quantum Dot Statistics

A QD weakly coupled to a reservoir must be treated as a grand canonical system since, on one hand, the QD exchanges particles and energy with the reservoir and, on the other hand, the reservoir determines the temperature T and the electrochemical potential μ in the dot.

Under this framework, let $\{n_j\} = \{n_1, n_2, \dots, n_k, \dots\}$ be a dot configuration; that is, n_k represents the number of electrons occupying the many-body eigenlevel ε_k ($n_k = 0$ or 1 for fermions) and such that $\sum_k n_k = N$, the number of electrons in the dot. From Gibbs distribution, the probability for the dot to be in the configuration $\{n_j\}$ reads [61]

$$P(\{n_j\}) = \frac{1}{\Xi}\exp\left[-\beta\left(\sum_i n_i\varepsilon_i + U(N) - N\mu\right)\right] \tag{4}$$

where μ is the electrochemical potential of the reservoir measured with respect to the bottom of the conduction band, $\beta = (K_BT)^{-1}$, and Ξ is the grand partition function,

$$\Xi = \sum_{\{n_j\}}\exp\left[-\beta\left(\sum_i n_i\varepsilon_i + U(N) - N\mu\right)\right] \tag{5}$$

and where

$$U(N) = \int_0^{Ne}\phi(Q)\,dQ \tag{6}$$

is the electrostatic energy of the dot with ϕ being the potential difference between the dot and the reservoir.

To obtain the probability that the dot contains N electrons, we need to sum Eq. (4) over all configurations such that $\sum_k n_k = N$

$$P(N) = \sum_{\{n_j\},\,\sum_k n_k = N} P(\{n_j\}) \tag{7}$$

which can be rewritten

$$P(N) = \frac{1}{\Xi}\exp[-\beta(F(N) - \mu N)] \tag{8}$$

where $F(N)$ is the Helmholtz free energy

$$F(N) = -\frac{1}{\beta}\ln\left\{\sum_{\{n_i\},\,\sum_i n_i = N}\exp\left[-\beta\left(\sum_i n_i\varepsilon_i + U(N)\right)\right]\right\} \tag{9}$$

In order to derive a simple relation between the potential and the number of electrons in the dot at very low temperature, and following Beenakker [58], it is assumed that there is a dominant configuration $\{n_k\}$ in the summation so that $F(N)$ takes the simplified form

$$F(N) = \sum_i n_i\varepsilon_i + U(N) \tag{10}$$

The equilibrium configuration corresponds to the integer N that minimizes $F(N) - N\mu = \Omega(N)$, the grand potential. At very low temperature, either $P(N) \neq 0$ for one single value of N, which corresponds to a stable configuration in the dot, that is, there is no fluctuation allowed (Fig. 5a, b, d), or there are two nonzero values N and $N-1$ such that $P(N) = P(N-1)$ (Fig. 5c). According to Eq. (8), this leads to

$$\mu = F(N) - F(N-1) \tag{11}$$

which states that the electrochemical potential in the dot is equal to the electrochemical potential in the reservoir. Then, the number of electrons in the dot can fluctuate between N and $N-1$; the sequence $N-1 \to N \to N-1$ is allowed and corresponds to a single-electron tunneling event.

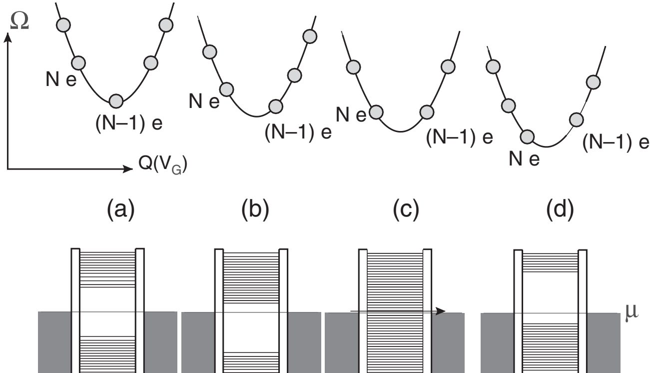


Figure 5. Illustration of single-electron tunneling and Coulomb blockade effect as the gate voltage V_G is swept. (a) The grand potential is minimized for $Q = (N-1)e$. The number of electrons in the dot is fixed, a gap is created in the energy levels, and tunneling is suppressed by CBE. (c) When the grand potential is minimized for $Q = (N-1/2)e$, the number of electrons can fluctuate between $N-1$ and N, the gap has vanished, and single-electron tunneling is allowed. (b) and (d) show intermediate situations. Adapted with permission from [10], M. Kastner, *Phys. Today*, 24, Jan (1993). © 1993, American Institute of Physics.

3.3. The Single-Electron Transistor

Single-electron transistors (SETs) are three terminal devices based on the CBE where the number of electrons inside a dot and the transfer of electrons one by one from a reservoir called the *source* to another reservoir called the *drain* through two series tunnel junctions is controlled by a *gate* [7]. They are schematically represented by Figure 6a, and the operation principle is similar to a MOSFET.

When the conditions (3) for CBR are satisfied, the tunnel junctions can be approximated by capacitors and the equivalent circuit for single electron transistor is shown in Figure 6b.

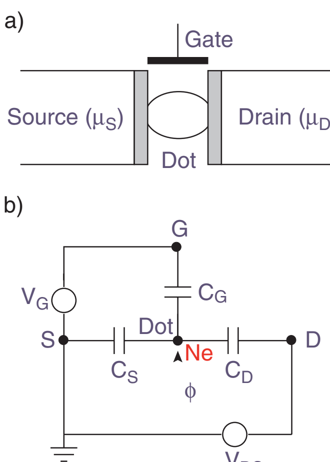


Figure 6. (a) Schematic diagram of a SET and (b) equivalent circuit in the CBR. Adapted with permission from [60], K. Likharev, *Proc. IEEE* 87, 606 (1999). © 1999, IEEE.

Solving the equivalent circuit, we find for the potential of the dot [62]

$$\phi = \frac{1}{C_\Sigma}\left(Ne - C_D V_{DS} - C_G V_G\right) \tag{12}$$

where C_G, C_S, and C_D are the gate, source, and drain capacitance, respectively, and $C_\Sigma = C_G + C_S + C_D$.

3.3.1. Linear Regime: Coulomb Oscillations

Here, we derive the condition for single-electron tunneling as a function of the gate voltage when the SET operates close to zero source-drain bias.

Using Eqs. (10) and (12) with $V_{DS} = 0$, the grand potential reads

$$\Omega(N) = \frac{(Ne)^2}{2C_\Sigma} - Ne\frac{C_G}{C_\Sigma}V_G + \sum_{i=1}^{N}\varepsilon_i - \mu N \tag{13}$$

and the condition (11) to observe single-electron tunneling becomes

$$\frac{(N-1/2)e^2}{C_\Sigma} - eV_G\frac{C_G}{C_\Sigma} + \xi(N) = \mu \tag{14}$$

where $\xi(N) = \sum_{i=1}^{N}\varepsilon_i(N) - \sum_{i=1}^{N-1}\varepsilon_i(N-1)$.

In summary, at equilibrium, as the gate voltage is swept continuously, there is alternation between Coulomb blockade and single-electron tunneling events. Typical experimental data of the current probed when a small bias ($\approx 150\,\mu$V) is applied between the source and the drain are shown in Figure 7. For each single-electron tunneling, a current peak or *Coulomb oscillation* (CO) is measured. It can be shown that the height of the CO is proportional to V_{DS} [58] and thus the small source-drain bias regime is usually called the *linear regime*.

The interval between two single-electron tunneling events reads

$$\Delta V_G = \eta^{-1}\left(\frac{e}{C_\Sigma} + \frac{1}{e}\left(\xi(N+1) - \xi(N)\right)\right) \tag{15}$$

since the electrochemical potential of the reservoir is insensitive to the fluctuations in the dot. The ratio $\eta = C_G/C_\Sigma$

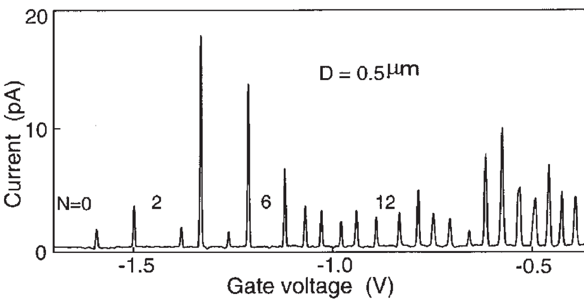


Figure 7. Coulomb oscillations as a function of the gate voltage for Tarucha's structure showing the irregular current peak spacing. Reprinted with permission from [11], S. Tarucha et al., *Phys. Rev. Lett.* 77, 3613 (1996). © 1996, American Institute of Physics.

can be seen as the gate efficiency if one notices that, all other quantities being unchanged,

$$\Delta\phi = \eta\Delta V_G = \left(\frac{e}{C_\Sigma} + \frac{1}{e}(\xi(N+1) - \xi(N))\right) \tag{16}$$

The quantity $E_a = e\Delta\phi$ is called the *addition energy* and its computation is a major issue of this work. It represents the amount of energy needed to add an electron in the dot. The addition energy is a very important parameter for single-electron devices, as crucial as the threshold voltage in MOSFETs.

If we set $C_Q(N) = e^2/(\xi(N+1) - \xi(N))$, Eq. (16) can be rewritten in terms of capacitances

$$\frac{1}{C(N)} = \frac{1}{C_\Sigma} + \frac{1}{C_Q(N)} \tag{17}$$

to introduce the dot capacitance $C(N)$, which consists of a classical contribution C_Σ depending only on the device geometry, that is, independent of N, a quantum contribution C_Q, highly dependent on N [63].

The addition energy (AE) is also defined in the literature as

$$E_a(N) \equiv \Delta_2(N) \equiv \mu(N+1) - \mu(N) \tag{18}$$

which can be understood by looking at Figure 8. The addition energy E_a can be thought of as the amount of energy provided by the gate to add an electron to the dot. This corresponds to the difference $\mu(N+1) - \mu(N)$, at constant V_G.

It is common to distinguish between the classical CBE, when the dot size is much larger than the de Broglie wavelength (Fig. 9a) and the quantum CBE, when the dot size is of the same order as the de Broglie wavelength (Fig. 9b).

3.3.2. Classical Limit

When the dot size is much larger than the de Broglie wavelength, the level spacing $\Delta\varepsilon \ll E_C$. Therefore, the second term of the right hand side of Eq. (15) can be safely neglected and the Coulomb oscillations are periodic, with period e/C_G: This is the classical limit as shown in Figure 10. The charging diagram that exhibits a staircase profile is also shown in this figure.

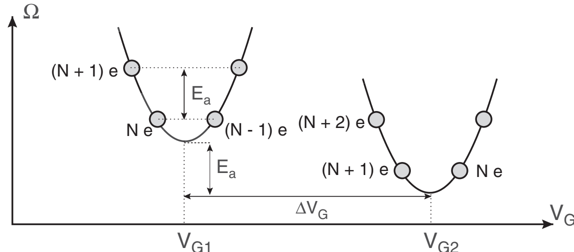


Figure 8. Grand potential Ω as a function of V_G at the transition between $N-1$ and N electrons in the dot (left curve) and the transition between N and $N+1$ electrons in the dot (right curve) that shows that the addition energy can be interpreted as the amount of energy provided by the gate in order to add an electron to the dot.

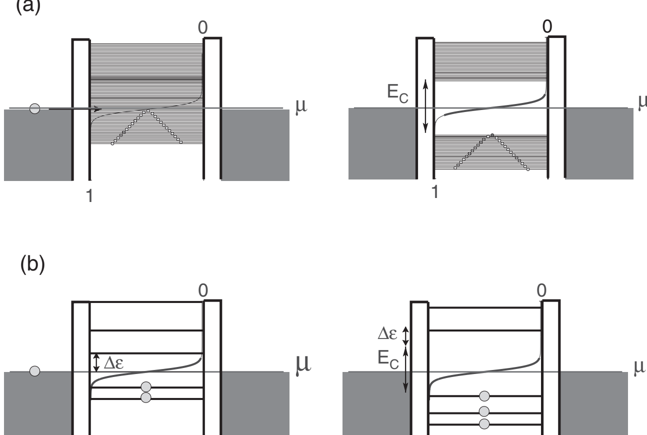


Figure 9. Illustration of the two CBRs: (a) When the dot size in much bigger than the de Broglie wavelength, the level spacing $\Delta\varepsilon$ is much smaller than the charging energy E_C which can be considered as the energy required to add an electron in the dot. (b) In smaller dots, $\Delta\varepsilon$ become of the order of E_C and can no longer be neglected in the calculations.

Based on an equation similar to (13), Averin and Likharev developed the "orthodox" theory where they used a master equation approach to derive the time evolution of $P(N)$ [6]. This theory was very successful in explaining single-electron effects in metal tunnel junctions where quantum effects such as discreteness of energy levels do not play an important role.

3.3.3. Quantum Limit

When the dot size is comparable to the de Broglie wavelength, electron energy quantization should be taken into account as shown in Figure 9b. In order to add an electron to the dot, not only does the classical charging energy have to be overcome but also $\Delta\varepsilon$, which is now comparable to E_C.

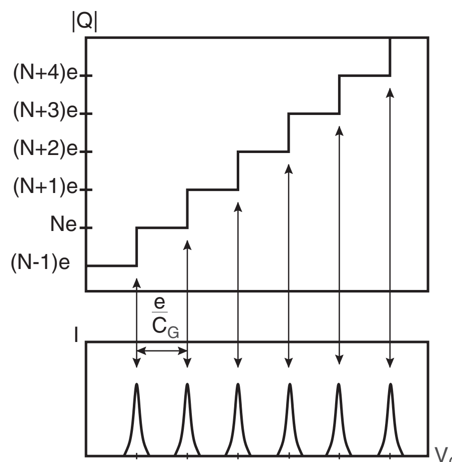


Figure 10. Coulomb staircase (top) and COs (bottom) as a function of the gate voltage V_G in the classical CBR, showing the periodicity e/C_G of the COs. Adapted with permission from [26], U. Meirav and E. Foxman, *Semicond. Sci. Technol.* 10, 255 (1996). © 1996, IOP Publishing.

In very small dots, $\Delta\varepsilon$ can even become the main contribution in ΔV_G and (15) reads $\Delta V_G \approx \xi(N+1) - \xi(N)$; that is, the COs are governed by the level spacing in the dot.

In this study, we consider QDs of intermediate sizes, for which no simplification of Eq. (15) is made. Moreover, the term $\xi(N+1) - \xi(N)$ hides a complicated many-body problem because the addition of a new electron in the dot takes into account a combination of electron–electron repulsion and exchange-correlation effects based on the occupation of quantum orbitals and electron spins. Therefore, this term can only be computed accurately by solving a many-body Schrödinger equation.

4. TARUCHA'S EXPERIMENT

The first reproducible experiments on single-electron charging effects with QDs in the quantum limit were achieved only recently by Tarucha and co-workers [9, 11, 35]. Figure 3 shows a schematic diagram of Tarucha's vertical QD [9, 11, 31, 33, 34]. The structure consists of an undoped 12 nm $In_{0.05}Ga_{0.95}As$ well and undoped $Al_{0.22}Ga_{0.78}As$ barriers of thickness 9 and 7.5 nm. This difference in the barrier thickness is required to facilitate accumulation of electrons in the dot in nonlinear transport. The inclusion of indium in the well is a major breakthrough to immune QD to fluctuations. In earlier devices, the well was made of GaAs [31]. However, to position the well bottom below the Fermi level, the AlGaAs barrier had to be doped, which increases the probability of dopant diffusion into the dot. The inclusion of indium in the well lowers the bottom of the conduction band and circumvents the drawbacks of delta doped AlGaAs barriers of earlier devices. The lead on the side of the thinner (thicker) tunnel barrier is made of n^+GaAs and is referred to as the source (drain). The doping is gradually reduced from the source (drain) to the double barrier heterostructure (DBH). The DBH is etched to a point below the DBH to form either a circular [9, 11, 31, 33–35, 64–66] or a rectangular [9, 32, 66, 67] mesa. A Schottky gate surrounding the mesa controls the number of electrons in the dot. The surrounding Schottky gate is another improvement with respect to other devices. In conventional planar technology, etching is performed to a point above the DBH and the Schottky gate is placed on top of the device. This innovation allows better control in pinching off the conducing channel [31].

Experimental measurements were made with samples cooled in a dilution refrigerator and the electron temperature was estimated to be about 0.2 K. A very small voltage is applied between the source and drain contacts and a current flowing through the dot is measured as the gate voltage is swept. The gate voltage raises or lowers the electrostatic potential of the dot relative to the source and drain. In order to observe COs, one should operate close to the zero source-drain bias for which tunnel barrier asymmetry mentioned above is irrelevant. Thus, when the source-drain bias is small, most of the time, the current is zero due to the Coulomb blockade but occasionally the Coulomb gap is lifted and current flows (i.e., the number of electrons in the dot oscillates between N and $N+1$), and current oscillations are observed as seen in Figure 7. Clearly, as N decreases, the current oscillations are more irregularly spaced. This was already observed by Dellow et al. [36] but was attributed to inhomogeneities in the dot. In contrast, in these experiments, the irregularities are systematic and depend on N, suggesting a mapping between the current peak interval and the energy level spacing in the dot. In order to confirm this idea, the addition energy, $\Delta\mu$, is measured as a function of N. From Eq. (16), $\Delta\mu = \eta\Delta V_G$, where ΔV_G is measured from the COs and η is deduced from the Coulomb diamonds of Figure 11 [11]. As shown in Figure 12a, the AE is unusually large for $N = 2, 6$, and 12, which suggests the influences of quantum effects in the AE spectrum of the QD.

5. THE MANY-BODY HAMILTONIAN OF ARTIFICIAL ATOMS

The electronic spectrum of N-electron QDs is computed by considering the many-body Hamiltonian

$$H = \sum_i H_{oi} + \sum_{i \neq j} H_{ij} \tag{19}$$

where H_{oi} is the single-particle Hamiltonian of the ith electron and

$$H_{ij} = \frac{e^2}{\epsilon |\mathbf{r}_i - \mathbf{r}_j|} \tag{20}$$

is the interaction Hamiltonian describing the Coulomb interaction amongst carriers. Here ϵ is the dielectric constant of the material. In the second term of Eq. (19), the sum is carried out on $i \neq j$ avoiding the interaction of carrier with themselves. Quite generally the Hamiltonian Eq. (19) is used for solving the Schrödinger equation for the many-particle energies and wavefunctions

$$E = E_N(1, 2, 3, \ldots, N) \tag{21}$$

$$\Psi = \Psi_N(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_N) \tag{22}$$

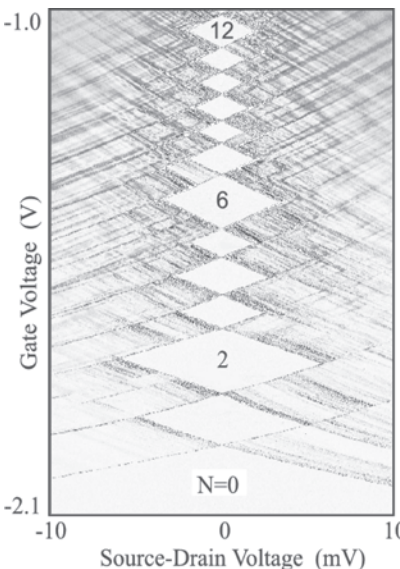


Figure 11. Differential conductance $\partial I/\partial V_{DS}$ plotted in the V_G–V_{DS} plane. In the white regions (*Coulomb diamonds*), $\partial I/\partial V_{DS} = 0$ due to Coulomb blockade. Reprinted with permission from [35], L. P. Kouwenhoven et al., *Science* 278, 1788 (1997). © 1997, American Assoc. Adv. Sci.

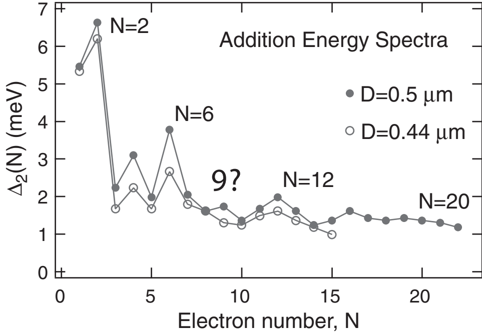


Figure 12. Typical addition energy spectra as a function of N for two different QDs with diameter $D = 0.5$ and $0.44\,\mu$m. Reprinted with permission from [11], S. Tarucha et al., *Phys. Rev. Lett.* 77, 3613 (1996). © 1996, American Institute of Physics.

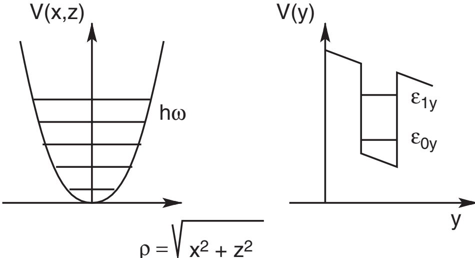


Figure 13. Schematic representation of a 2D parabolic potential (left) with cylindrical symmetry in the xy-plane showing equally spaced energy levels and the square potential (right) with the first two quantized levels in the z-direction such that $E_{1z} - E_{0z} \gg \hbar\omega$. Reprinted with permission from [112], J.-P. Leburton and S. Nagaraja, "Optical Spectroscopy of Low Dimensional Semiconductor," NATO ASI Ser. E, Vol. 344, p. 235. © 1997, Kluwer Academic.

which given the two-body interaction (20) can only be solved exactly for $N = 2$. In this section we will describe a natural approach toward the solution of this problem for a general number N of electrons by considering successive approximations.

5.1. Single-Particle Hamiltonian and Shell Structures

We start by considering a system of independent and 3D confined electrons in the conduction band. By neglecting the interaction H_{ij}, the Hamiltonian (19) is reduced to a sum of single particle Hamiltonians, each of the same form

$$H_{oi} = H_i = \frac{p_x^2 + p_y^2 + p_z^2}{2m^*} + V(\boldsymbol{r}) \tag{23}$$

Here we assume the electrons can be described with an effective mass m^*; p_x, p_x, p_x are the components of the electron momentum and $V(\boldsymbol{r})$ is the external potential which contains several contributions according to the confinement achieved in the quantum dot. We will also assume that the QD is realized by confinement of the electrons in an heterostructure quantum well along the z-direction and electrostatic confinement in the x–y plane. The latter confinement results usually from dopant atoms in neighboring semiconductor layers and from the fringing field of the metal electrodes on the semiconductor surfaces. This configuration is most commonly achieved in PQDs [7] and Tarucha's VQD structure [9], and results in a first approximation in a 2D parabolic potential in the x–y plane (Fig. 13). Confinement at the heterostructure along the z-direction is generally strong ($\approx$10 nm) with energy separation of the order of 50–100 meV, while the x–y planar confinement is much weaker with energy separation of the order of 1 meV over larger distances (>100 nm). In that case, the external potential is separable in a first approximation,

$$V(\boldsymbol{r}) = V_1(x, y) + V_2(z) \tag{24}$$

which results in the energy spectrum $\varepsilon_{\nu, n_x, n_y} = \varepsilon_\nu + \varepsilon_{n_x, n_y}$ with corresponding wavefunctions $\Psi_\nu(z)\,\Psi_{n_x, n_y}(x, y)$ where ε_ν (ε_{n_x, n_y}) is the spectrum resulting from the z-potential ($x - y$ potential). Hence, each value of the ν-quantum number gives a series of $x - y$ energy levels. At low temperature, given the large separation between the ε_ν energy states, only the first levels of the lowest series $\nu = 0$ are occupied by electrons. If one further assumes that the $V_1(x, y)$ potential is cylindrically symmetric, the $\nu = 0$ energy spectrum is written as

$$\varepsilon_{0, n_x, n_y} = \varepsilon_{0, n, l} = \varepsilon_0 + n\,\hbar\omega \tag{25}$$

where ω is frequency of the cylindrical parabolic potential. Here each n-level is $2n$-times degenerate with the factor 2 accounting for the spin degeneracy. The number n (=1, 2, 3, ...) is the radial quantum number and the number l (=0, ±1, ±2, ...) is the angular momentum quantum number. Hence the 2D cylindrical parabolic potential results in 2D orbitals supporting 2, 4, 6, 8, ... electrons which give rise to shell structures filled with 2, 6, 12, 20, ... particles, thereby creating a sequence of numbers which can be regarded as the 2D analogues of "magic numbers" in atomic physics.

In the absence of cylindrical or square symmetry, the parabolic potentials are characterized by two different frequencies, ω_x and ω_y, which lift the azimuthal degeneracy on the l-number of the 2D artificial atoms. Therefore electronic states are spin-degenerate only and determine a sequence of shell filling numbers 2, 4, 6, 8, ... of period or increment 2. Only when the ratio $K = \omega_x/\omega_y$ is commensurable does the sequence of filling numbers deviate from the period 2 and provide new sequences of numbers for particular combinations of the n_x and n_z quantum numbers in the case of accidental degeneracy.

5.2. Hartree–Fock Approximation and Hund's Rules

5.2.1. The Hartree–Fock Approximation

The natural extension of the atomic model for independent 3D confined electrons is the consideration of the Coulomb interaction between particles in the Hartree–Fock (HF) approximation. The HF scheme has the advantage of conserving the single-particle picture for the many-body state of the system by representing the total wavefunction as a

product of single-particle wavefunctions in a Slater determinant which obeys Fermi statistics. The main consequence of the HF approximation for the Coulomb interaction between particle is a correction of two terms to the single-particle energies derived from the H_o Hamiltonian;

$$\varepsilon_i = \varepsilon_{\nu,n,l} + \frac{e^2}{\epsilon}\sum_{j\neq i}\int \frac{\psi_j^*(\mathbf{r}')\psi_j(\mathbf{r}')\psi_i^*(\mathbf{r})\psi_i(\mathbf{r})\,d\mathbf{r}\,d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} - \frac{e^2}{\epsilon}\sum_{j\neq i}\int \frac{\psi_i^*(\mathbf{r})'\psi_j(\mathbf{r}')\psi_i(\mathbf{r})\psi_j^*(\mathbf{r})\,d\mathbf{r}\,d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} \tag{26}$$

where the first sum is the Hartree energy carried on all occupied j-states different from the i-state, irrespectively of their spins, and accounts for the classical repulsion between electrons. The second term is the attractive exchange interaction which occurs between carriers with parallel spins. In this scheme, the wavefunctions $\Psi_i(\mathbf{r})$ satisfy the HF integro-differential equation where the Coulomb interaction term depends upon all the other single-particle wavefunctions of the occupied states. The HF equation is therefore nonlinear and must be solved self-consistently for all wavefunctions of occupied states.

5.2.2. Hund's Rules

In atomic physics, the ground state of atoms can be approximated by first neglecting the electron–electron interaction and considering that individual electrons occupy a single-particle hydrogenic state (n, l, m). In this model, the total energy depends only on the radial quantum number n so that, for a given value of n, there are $2n^2$ degenerate eigenstates, including spin. In reality, however, electron–electron interaction lifts up these degeneracies so that the order of occupation is not arbitrary. Friedrich Hund devised a set of rules to identify the lowest energy electron configuration when such a "quasi-degenerate" set of levels has to be filled [68]:

1. *The configuration with the highest total spin S will have the lowest energy.*

   This rule relies on the fact that exchange between electrons of identical spin reduces the total energy of the configuration. Moreover, the Hartree energy of the configuration is reduced when an electron occupies a free orbital rather than an already partially occupied orbital since there is then less overlap between occupied orbitals.
2. *For a given total spin S, the term with the highest value of the total orbital angular momentum L will have the lowest energy.*

   The basis of this rule is that electrons with a high value of L (i.e., orbiting in the same direction) will meet less often than electrons with low value of L (i.e., orbiting in opposite directions), thereby reducing their Coulomb repulsion [68].
3. *For atoms with less (more) than half-filled shells, the level with the lowest (highest) value of the total angular momentum J will have the lowest energy*

   The origin of that rule is the spin–orbit coupling [68].

The orbitals of QDs studied here exhibit degeneracies due to circular rather than spherical symmetry. Therefore, the degree of degeneracy is reduced in QDs compared to real atoms so that only Hund's first rule needs to be invoked to resolve degenerate configurations.

5.2.3. Hund's First Rule in Quantum Dots

One of the important consequences of the HF approximation for interelectron interaction (20) is the prediction of spin effects in the shell filling of artificial atoms similar to Hund's rules in atomic physics. These effects are illustrated in the charging energy of a few electron QDs with cylindrical parabolic potential achieved in planar or vertical quantum structures. In Figure 14a, we show schematically the Coulomb staircase resulting from charging a QD with a few electrons as a function of the charging energy or voltage between the metal electrode or gate and the semiconductor substrate. The relative step sizes of the staircase represent the amount of energy needed to put an additional electron in the dot. The arrows on each step represent the spin of each individual electron on the successive orbitals during the charging process. The filling of the first shell (s-orbital with two electrons) consists of one electron with spin up followed by an electron with spin down. The step size of the spin-up electron measures the charging energy needed to overcome the Coulomb repulsion against the spin-down electron, which is just the Hartree energy between the two particles. The larger step size of the second (spin-down) electron is due to the fact that the charging of the third electron requires overcoming the Coulomb repulsion energy augmented by the energy to access the next quantized level, which is the first p-orbitals. The latter process starts the second shell filling with the third electron on either one of the degenerate $l = \pm 1$ orbitals of either spins (here we choose the $l = -1$ and the spin-up). At this stage, the configuration

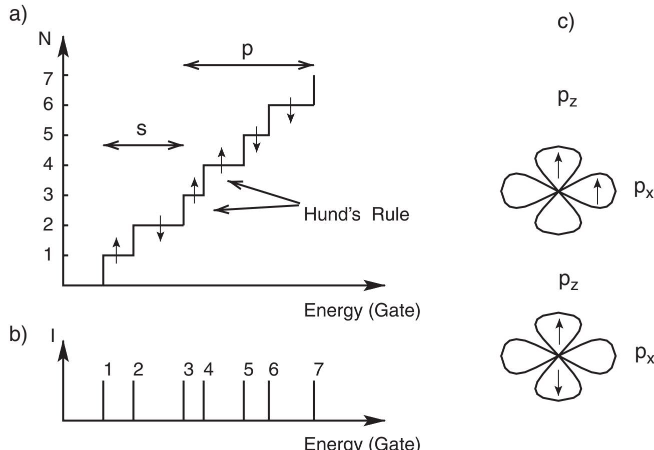


Figure 14. (a) Coulomb staircase as a function of the charging energy with the spin states of each electron. The horizontal two head arrows indicate the occupation of the s- and p-orbitals in the dot. (b) COs versus the charging energy. (c) 2D p-orbitals illustrating the two possible occupations of two electrons with parallel (top) and antiparallel spins (bottom). Reprinted with permission from [112], J.-P. Leburton and S. Nagaraja, "Optical Spectroscopy of Low Dimensional Semiconductor," NATO ASI Ser. E, Vol. 344, p. 235. © 1997, Kluwer Academic.

with the fourth electron on the $l = 1$ orbital with a parallel spin becomes more favorable because it minimizes the Hartree energy between orbitals of different quantum numbers and results in an attractive exchange energy between the two electrons (Fig. 14c). This is the reason the third step is smaller than the first and the second steps, requiring less energy and evidencing Hund's rule in the electron filling of the 2D artificial atom. The fourth step is wide because the addition of the fifth electron on either of the p-orbitals with $l = \pm 1$ must correspond to a spin-down electron which undergoes a repulsion from the two other p-electrons without benefitting from the exchange with them since its spin is antiparallel. Figure 14b shows the schematics of current peaks resulting from the single-electron charging of the QD which is obtained by differentiating the Coulomb staircase and which for the first two shells is consistent with the experimental data of Tarucha (Fig. 7).

Hence the HF approximation provides a reasonable picture of the contribution of electron–electron interaction and spin effects in the spectrum of quantum dots. However, it is well known from atomic physics and theoretical condensed matter physics that this approximation suffers from two important drawbacks which are the neglect of electron correlation and the fact that it overestimates the exchange energy. Moreover, it leads to a tedious solution of the self-consistent problem when involving a large number of electrons and makes it difficult to apply to quantum devices in accounting for the environment of heterolayers and doping as well as boundary conditions at the semiconductor interfaces with metal electrodes. Hence no single HF model to date reproduces one-to-one the absolute values of $E_a(N)$ actually observed in VQDs [66, 69–71]. Model calculations understandably but nevertheless include some arbitrary assumptions, typically a rigid or perfectly parabolic confinement, perfect circular symmetry, and the neglect of 3D effects [69, 72–74], which may not be completely realistic. In particular, it has been shown that maximal spin alignment still consistent with Hund's first rule can lead to different secondary maxima in the $0T$ third shell addition energy spectrum depending on the actual electronic configurations in realistic VQDs [71]. We argue that behavior which would widely be regarded as "anomalous" is a consequence of understandable many-body interactions and is strongly influenced by the exact profile of the 3D VQD confining potential. This issue is addressed in the following sections.

6. QUANTUM DOTS AS ELECTRONIC DEVICES

In order to investigate the effect of 3D geometry on the confining potential in the dot and describe its deviations from parabolicity, we model the CVQD as a cylindrical structure (Fig. 15) by taking into account macroscopic parameters, such as the doping of the source and the drain (N_D), the radius of the device (R), the pinning of the conduction band at the Schottky gate interface (ϕ_S), and the well width ($2z_i$), and include the effects of the third dimension.

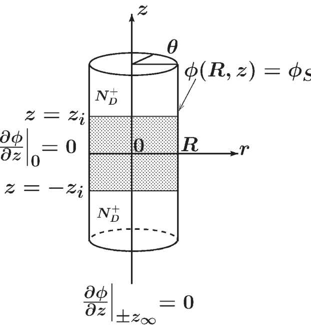


Figure 15. Cylindrical charge model for the cylindrical VQD structure with boundary conditions. N_D^+ is the ionized donor concentration, and (r, θ, z) are the cylindrical coordinates. Reprinted with permission from [71], P. Matagne et al., *Phys. Rev. B* 65 (2002). © 2002, American Institute of Physics.

6.1. Potential Model

The model consists of an intrinsic region for $-z_i \leq z \leq z_i$ surrounded by two depleted regions for $|z| > |z_i|$, which extends up to $z = \pm z_\infty$ where it becomes neutral. The structure is assumed to be symmetric around $z = 0$, and the potential $\phi = \phi(r, z)$ does not depend on θ. The dot is also assumed to be empty of electrons in order to capture only the structural effects, as opposed to the many-body effects due to electron–electron interactions, on the confining potential. We obtain the confining potential ϕ by solving the 3D Poisson equation

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r}\frac{\partial \phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \theta^2} + \frac{\partial^2 \phi}{\partial z^2} = -\frac{\rho}{\epsilon} \tag{27}$$

where ρ is the charge density such as

$$\rho = 0 \qquad (-z_i < z < z_i)$$

and

$$\rho = qN_D^+ \qquad (|z| > z_i)$$

Here, ϵ is the dielectric constant. Since $z_{\pm\infty} \approx 300$ Å is much smaller than the source (drain)–DBH distance (4000 Å), the doping variation is neglected in this analytical model. The boundary conditions at $z = \pm z_\infty$ are such as the z-field is relaxed; that is, $(\partial\phi/\partial z)|_{\pm z_\infty} = 0$, and $\phi_S = \phi(R, z)$ is the Schottky barrier potential at the gate interface ($r = R$).

6.1.1. *Intrinsic Region* ($0 \leq |z| \leq |z_i|$)

We separate variables to solve Eq. (27) and find for the potential [71]

$$\phi(r, z) = \phi_S + \sum_n A_n J_0(\gamma_n r)(e^{-\gamma_n z} + e^{\gamma_n z}) \tag{28}$$

where $\gamma_n = \mu_n/R$ and μ_j is the jth zero of the zero order Bessel function of the first kind J_0.

6.1.2. N-*Doped Region* ($|z| > |z_i|$)

By using the same technique as above, we find

$$\phi(r,z) = \phi_S - K(R^2 - r^2) + \sum_n B_n J_0(\gamma_n r) \cdot e^{\gamma_n z_\infty}(e^{-\gamma_n(z-z_\infty)} + e^{\gamma_n(z-z_\infty)}) \tag{29}$$

where $K = qN_D^+/4\epsilon$. The constants A_n and B_n are determined by matching the values of ϕ and $\partial\phi/\partial z$ with the solutions found at $z = \pm z_i$, which yields

$$A_n = \frac{2KR^2}{\mu_n^2}\frac{J_2(\mu_n)}{J_1^2(\mu_n)}\frac{sh[\gamma_n(z_\infty - z_i)]}{sh[\gamma_n z_\infty]} \tag{30}$$

and

$$B_n = \frac{-2KR^2\, e^{-\gamma_n z_\infty}}{\mu_n^2}\frac{J_2(\mu_n)}{J_1^2(\mu_n)}\frac{sh[\gamma_n z_i]}{sh[\gamma_n z_\infty]} \tag{31}$$

so that the potential inside the dot can be rewritten

$$\phi(r,z) = \phi_s - \frac{qN_D^+R^2}{\epsilon}\sum_{n=0}\frac{sh[\gamma_n(z_\infty - z_i)]}{sh(\gamma_n z_\infty)} \cdot \frac{J_2(\mu_n)}{\mu_n^2 J_1^2(\mu_n)} J_0(\gamma_n r)\, ch(\gamma_n z) \tag{32}$$

By taking the limit $z \to 0$ in Eq. (32), we obtain the potential in the center of the dot. We make use of the power expansion of $J_0(x)$ and invert the summations to get the following expression;

$$\phi(r) = \phi_0 - \alpha\frac{qN_D^+}{4\epsilon}r^2 - \Lambda\left(\frac{r}{R}\right)r^4 \tag{33}$$

If we denote

$$C_n = \frac{sh(\gamma_n(z_\infty - z_i))}{sh(\gamma_n z_\infty)}\frac{J_2(\mu_n)}{J_1^2(\mu_n)} \tag{34}$$

which is simply a structure factor depending on the problem geometry, the first term

$$\phi_0(V_G, N_D^+) = \phi_S(V_G) - \frac{qN_D^+R^2}{\epsilon}\sum_n\frac{C_n}{\mu_n^2} \tag{35}$$

represents the potential in the bottom of the dot that depends on the applied gate voltage V_G. The second term is the harmonic term with the form factor $\alpha = \sum_n C_n$. We notice that this term does not scale directly with the dot radius R, nor V_G. However, there is a hidden dependence on these factors through the α-parameter that is a function of R, z_i and z_∞, as shown in Figure 16. Our model predicts a weak dependence on the radius of the structure R (Fig. 16a) but a strong dependence on z_i and z_∞ (Fig. 16b and c). In plotting the R-dependence of the α-factor, we use $z_i = 175$ Å, according to device geometry, and $z_\infty \approx 350$ Å, which is extracted from the solution of our numerical model presented hereafter. The weak dependence on R is not surprising since the solution of the 2D harmonic oscillator (2DHO) is only a function of the doping level. The z_i-dependence can be understood by noticing that, by increasing the width of the intrinsic ($\rho = 0$) well, the field density, and thus the potential curvature in the longitudinal direction (z) decreases. In Figure 16b, we use $R = 2500$ Å, and we keep the depleted region $z_\infty - z_i$ constant at 200 Å in order to maintain the charge constant. The z_∞-dependence of α (here $R = 2500$ Å and $z_i = 175$ Å) is strong because the variation of z_∞ implies a significant change in the charge of the depleted region, which, via the Poisson equation, results in a strong variation of the electrostatic potential curvature. It is worth emphasizing that Figure 16c reflects the 3D nature of the potential variation, with correlation between its curvature along the radial and the longitudinal directions. Figure 17 shows that the 2D harmonic oscillator (dotted curve) model, with a spring constant $K = qN_D^+/4\epsilon$, overestimates the real potential curvature compared with 3D models. This is consistent with the z_i-dependence of the α factor previously obtained since the 2D harmonic oscillator picture can be considered as the limit for $z_i \to 0$ of the 3D problem (Fig. 16b). In practical problems, K is a fitting parameter

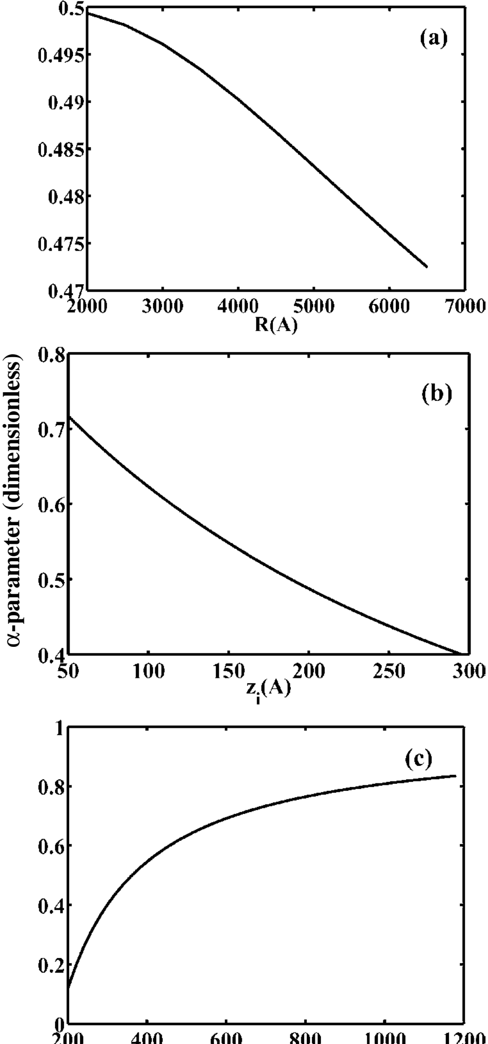


Figure 16. Variation of the form factor α as a function of (a) the dot radius R ($z_i = 175$ Å, $z_\infty \approx 350$ Å), (b) the well width $2z_i$ ($R = 2500$ Å, $z_\infty \approx 350$ Å), (c) the extension of the depletion width z_∞ ($R = 2500$ Å, $z_i = 175$ Å) in the vertical direction.

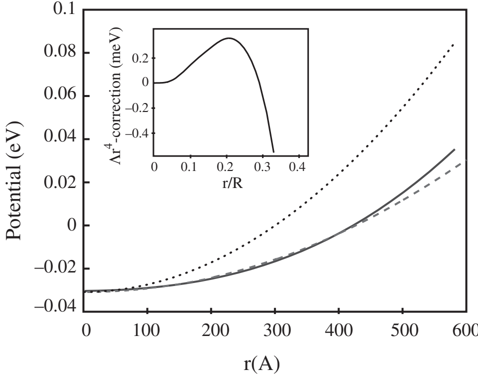


Figure 17. Comparison between the profiles of the conduction band edge ($z = 0$) computed from self-consistent simulations (solid) and from our 3D cylindrical charge model (dashed). The dotted curve shows a parabolic potential with $K = qN_D/4\epsilon$. The inset shows the Λr^4-correction to the harmonic potential as a function of r/R, where R is the dot radius.

[66, 75] fudged with a power variation of n, the electron density, but then it loses its physical meaning since it is unrelated to any of the nanostructure parameters.

The last term in Eq. (33) is the 3D correction to the "quasi-2D" parabolic potential with

$$\Lambda\left(\frac{r}{R}\right) = \frac{qN_D^+}{4\epsilon R^2}\sum_{n=0}^{\infty} C_n \sum_{l=0}^{\infty}(-1)^{l+1}\frac{\mu_n^{2l+2}}{2^{2l+2}[(l+2)!]^2}\left(\frac{r}{R}\right)^{2l} \tag{36}$$

We point out that, while retaining the cylindrical symmetry, this term is structure (N_D^+, C_n) and geometry (R) dependent; it is not necessarily negligible, growth with r/R, and could be positive or negative since J_2 oscillates with μ_n. The correction $\Lambda\, r^4$ is plotted as a function of (r/R) in the inset of Figure 17. It is seen that the correction is small but not negligible at short distance from the origin.

6.2. 2D Eigenstates Analysis

Since the effective dot radius is much smaller than R, the $\Lambda\, r^4$-correction can be treated as a perturbation of a parabolic potential in the quantum mechanical problem of the determination of the eigenvalues and the eigenfunctions in the QD [76]. Hence, we consider the actual xy-potential profile to be given by

$$\phi(r, \theta) = \phi_{\text{osc}}(r) + W(r, \theta) \tag{37}$$

where $\phi_{\text{osc}}(r) = \phi_0 - Kr^2$ is the harmonic potential of Eq. (33) and W is a perturbation which contains the Λ-correction term and, for the sake of generality, depends on r and θ (Fig. 15b). If the confining potential ϕ retains cylindrical symmetry, one can state that

1. ϕ commutes with the two-dimensional angular momentum operator $(\hbar/i)(\partial/\partial\theta)$. Therefore, H and H_0 have the same eigenfunctions $|n, l\rangle$.
2. $|n, l\rangle$ and $|n, -l\rangle$ can be generated from one another by rotation; that is, if $\mathbf{R}$ is a rotation operator with z axis, $\mathbf{R}|n, -l\rangle = |n, l\rangle$. Therefore, $|n, l\rangle$ and $|n, -l\rangle$ have the same energy $\varepsilon_{n,|l|}$.

Thus, as long as the perturbation is small with respect to $\hbar\omega$, the shell structure in VQD is preserved. However, within a shell, degeneracy is partially lifted, and subshell levels, characterized by identical absolute values in their quantum number l, emerge.

When the cylindrical symmetry is broken by, for instance, the electron occupation of non-s-orbitals, deviations from the circular shape of the dot, and/or potential fluctuations arising from the long range Coulomb interaction of ionized donors in GaAs, the eigenlevels are, in general, no longer degenerate.

7. FULL SCALE SIMULATION OF QUANTUM DOTS

Beyond the HF approximation, the density functional theory is extremely powerful in predicting with relative accuracy the electronic properties of atomic systems, and even of large ensemble of atoms. It is therefore well suited for computing the electronic spectra of QDs and artificial atoms in the nanostructure environment [77–81]. Moreover, like the HF approximation, it has the advantage of retaining the single-particle picture for the wavefunctions in the system.

In simulating quantum devices, it is often useful to partition the device into regions of different confinements. Hence, the eigenenergies and the eigenstates are computed numerically in each region based upon the degree of quantization and the 3D electrostatic potential computed from the Poisson equation. The wavefunctions are then assembled using the appropriate occupation probabilities to generate a 3D charge density throughout the device. If the quantum device contains 2DEG regions, the adiabatic approximation can be used by assuming the Schrödinger equation to be separable into vertical (z) and transverse components (x, y). This approximation assumes that the vertical component of the wavefunction responds instantaneously to the variation of the transverse potential. Since the longitudinal wavefunctions generally exhibit a propagating nature, they are cast as plane waves to enhance the performance of our code. In quasi-2D leads, therefore, the quantum mechanical problem breaks down into solving the 1D Schrödinger equation in slices down the x–y plane of the 2D region. The use of plane waves in 2D regions results in an incomplete treatment of evanescent leakage and reflection in regions where propagating states impinge on the barriers. The plane wave approximation could be justified, however, when in the calculation of charge densities, the counting of occupied propagating states is of interest rather than rigorously calculating quasi-bound energies.

7.1. The Local Spin Density Approximation Hamiltonian

In the envelope function–effective mass approximation, the Kohn–Sham equations, one for spin up (↑) and one for spin down (↓), read

$$H^{\uparrow}(\mathbf{r})\psi_m^{\uparrow}(\mathbf{r}) = \varepsilon_m^{\uparrow}\psi_m^{\uparrow}(\mathbf{r}) \qquad H^{\downarrow}(\mathbf{r})\psi_m^{\downarrow}(\mathbf{r}) = \varepsilon_m^{\downarrow}\psi_m^{\downarrow}(\mathbf{r}) \tag{38}$$

where ε_m and ψ_m are the corresponding eigenenergies and eigenstates of the Hamiltonians $H^{\uparrow}$ and $H^{\downarrow}$:

$$H^{\uparrow(\downarrow)}(\mathbf{r}) = -\frac{\hbar^2}{2}\nabla\left[\frac{1}{m_n^*(\mathbf{r})}\nabla\right] + E_c + \mu_{xc}^{\uparrow(\downarrow)}(n, \zeta) \quad (39)$$

The conduction-band edge energy $E_c(\mathbf{r})$ is given by $E_c(\mathbf{r}) = -e\,\phi(\mathbf{r}) + \Delta E_c$, where ΔE_c is the band offset between different materials. The electrostatic potential $\phi(\mathbf{r}) = \phi_H(\mathbf{r}) + \phi_{\text{ions}}(\mathbf{r})$, where $\phi_H(\mathbf{r})$ is the Hartree potential equivalent to the second term in Eq. (26) and ϕ_{ions} is the potential resulting from ionized donors and acceptors present in the nanostructure, is the solution of the 3D Poisson equation

$$\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})) = e(n(\mathbf{r}) - N_D^+(\mathbf{r}) + N_A^-(\mathbf{r})) \quad (40)$$

which is solved self-consistently with the stationary Kohn–Sham equations. In the above equation, we allow for the position dependence of the effective mass m^* and the dielectric constant ϵ; $n(\mathbf{r})$, $N_D^+(\mathbf{r})$, and $N_A^-(\mathbf{r})$ are electron, ionized donor, and ionized acceptor densities respectively. Exchange and correlation are self-consistently incorporated into the model using the Kohn–Sham density-functional method [82]. This approach accounts for high-order electron–electron interaction by adding a correction to the Hartree potential,

$$\mu_{xc}(n) = \frac{d}{dn}[n\varepsilon_{xc}(n)] \quad (41)$$

where $\varepsilon_{xc}(n)$ is the exchange-correlation energy per electron derived from the local electron density n which can be expressed in terms of the Slater exchange energy and the Perdew–Zunger parametrization of the correlation potential [83]. The effect of exchange is predominant in the localized QD region, where the conduction band is altered by as much as 1 meV. In the local spin density approximation (LSDA) formulation, the exchange correlation energy is a function of the spin polarization $\zeta = (n^{\uparrow} - n^{\downarrow})/n$, where $n^{\uparrow}$ ($n^{\downarrow}$) is the spin up (down) electron density. Here, we use Perdew and Wang's formulation [84].

7.2. Electron Equilibrium Statistics

The electron density in the dot containing N electrons reads

$$n(\mathbf{r}) = \sum_i f(\varepsilon_i|N)|\psi_i(\mathbf{r})|^2 \quad (42)$$

where $f(\varepsilon_i|N)$ is the probability that the eigenlevel ε_i is occupied, knowing that there are N electrons in the dot. In general, $f(\varepsilon_i|N)$ is not equivalent to the Fermi–Dirac factor [58] but has to be determined by filtering out from (7) those particular configurations for which ε_i is occupied,

$$f(\varepsilon_i|N) = \frac{1}{P(N)} \sum_{\{n_j\},\, \sum_k n_k = N,\, n_i = 1} P(\{n_j\}) \quad (43)$$

The computation of Eq. (43) is a combinatorial problem that becomes rapidly insurmountable, even with powerful computer resources. Fortunately, in the Coulomb blockade regime, that is, when the level spacing $\Delta\varepsilon_i$ is much larger than the thermal fluctuation k_BT, for N electrons in the dot, the first N eigenlevels are certainly occupied. The electron density in the QD thus reads

$$n(\mathbf{r}) = n^{\uparrow}(\mathbf{r}) + n^{\downarrow}(\mathbf{r}) = \sum_{i=1}^{N_{\uparrow}} |\psi_m^{\uparrow}(r)|^2 + \sum_{i=1}^{N_{\downarrow}} |\psi_m^{\downarrow}(r)|^2 \quad (44)$$

where $N_{\uparrow} + N_{\downarrow} = N$ is the number of electrons in the dot.

7.3. Bulk Region

In regions where quantum effects are negligible, all the quantities of interest are obtained under the Thomas–Fermi approximation. The electron density reads

$$n(\mathbf{r}) = N_c \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c(\mathbf{r})) \quad (45)$$

where

$$\eta_c(\mathbf{r}) = \frac{-q\phi(\mathbf{r}) + \Delta E_c(\mathbf{r})}{k_BT}$$

which sets up the Fermi level E_F as the zero energy (i.e., $-q\phi = E_C - E_F$). $F_{1/2}$ is the Fermi integral,

$$F_{1/2}(\eta) = \int_0^{\infty} \frac{\xi^{1/2}\,d\xi}{1 + e^{\xi-\eta}} \quad (46)$$

and

$$N_c = 2\left(\frac{2\pi m_n k_B T}{h^2}\right)^{3/2} \quad (47)$$

is the electron effective density of states where m_n is the electron density of states effective mass. For nondegenerate semiconductors, Eq. (45) reduces to the Boltzmann factor

$$n = N_c\, e^{\eta_c} \quad (48)$$

However, at very low temperature, a semiconductor is always degenerate and the Fermi integral (46) cannot be simplified. Here, we use the rational function approximation proposed by Aymerich-Humet and Millan [85],

$$F_{1/2}(\eta_c) \approx \left(\frac{3/2\ 2^{3/2}}{\left(2.13 + \eta_c + (|\eta_c - 2.13|^{12/5} + 9.6)^{5/12}\right)^{3/2}} + \frac{2}{\sqrt{\pi}} e^{-\eta_c}\right)^{-1} \quad (49)$$

which in the worst case shows a 0.5% relative error $(\Delta F_{1/2}/F_{1/2})$ [86].

The dopant statistics for dilute donors and acceptors read

$$N_D^+ = \frac{N_D}{1 + g_D\ e^{(E_F - E_D)/k_BT}} \qquad \text{(donors)} \quad (50)$$

$$N_A^- = \frac{N_A}{1 + g_A\ e^{(E_A - E_F)/k_BT}} \qquad \text{(acceptors)} \quad (51)$$

where N_D and N_A, N_D^+ and N_A^-, E_D and E_A are the donor and acceptor concentrations, ionized concentrations, and energy levels, respectively. g_D and g_A are coefficients

depending on the semiconductor species. This statistic relies on the assumption that the wavefunction of a dopant in a piece of semiconductor is very similar to the first ground state in an hydrogen atom and is given by

$$\psi(\mathbf{r}) = (\pi a_B^3)^{-1/2} e^{-\mathbf{r}/a_B} \tag{52}$$

where a_B is the effective Bohr radius

$$a_B = \frac{4\pi\hbar^2\epsilon_d}{m_d e^2} \tag{53}$$

and ϵ_d is the semiconductor dielectric constant. At low impurity concentrations, the impurity density of states can be approximated by delta functions (see Fig. 18a). As the doping is increased, this simple picture is no longer appropriate. The orbitals of adjacent impurities begin to overlap so that impurity energy levels smear with each other in such way to create an impurity band (Fig. 18b). At even higher doping levels, the center of the impurity band distribution moves toward the conduction band minimum and the ionization energy diminishes considerably. Finally, at a certain doping level, the conduction and impurity bands merge, leaving a zero effective ionization potential (Fig. 18c). The set of additional states below the conduction band is traditionally referred to as the *band-tail*.

The precise determination of band-tailing in semiconductors was first treated semiclassically by Kane [87] and later quantum mechanically by Halperin [88]. The complete implementation of these models would be computationally too expensive. We only retain the two main consequences of band-tailing: (i) the bandgap of the semiconductor is narrowed and (ii) all the impurities are ionized, as opposed to what would be predicted by the simple model of isolated impurities. We use the results established by Jain and Roulston to compute the bandgap narrowing in our model [89].

7.4. Boundary Conditions

Boundary conditions for the electrostatic potential ϕ are established by imposing Dirichlet conditions at the source, drain, and lateral surfaces of the device. At the source and drain, a flat band condition is assumed and ϕ is set up such that the net charge is zero in these regions. An equivalent way to express neutrality is to impose zero electric field (i.e., a Neumann boundary condition). Both approaches were successfully tried, but the latter is favored in the finite element method since homogeneous Neumann boundary conditions

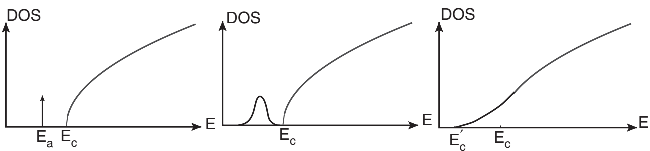


Figure 18. Variation of the density of states as a function of energy for a *n*-doped semiconductor. (a) For nondegenerate semiconductors, the donor levels are localized at E_a, in the forbidden gap; (b) for degenerate semiconductors, the orbitals of neighbor impurity overlap and their interaction lead to the formation of an impurity band; (c) for highly degenerate semiconductors, the donor band merges with the conduction band E_c.

are automatically taken into account. Along lateral surfaces, the Schottky barriers heights ϕ_s are strongly influenced by surface chemistry, so we use the experimental data of Grant et al. [90] and Best [91]. On the gated surfaces, the Schottky barriers are modified by $\phi_s - V_G$, where V_G is the gate bias.

Boundary conditions for the Kohn–Sham equation are imposed by assuming vanishing wavefunctions on any surface of the device. Since the QD is much smaller than the physical dimensions of the device, the wavefunctions vanish already before reaching those boundaries. In the direction perpendicular to the heterointerfaces, we allow wavefunctions to leak into the source and drain regions. However, this leakage is not taken into account for the computation of the charge in the source and drain regions where it is computed entirely using the Thomas–Fermi approximation in that region. This is a good approximation since it does not affect the charge within the dot and is too small compared to the bulk charge in the source and drain regions.

7.5. Conduction Band Profiles

Figure 19c shows the variation of the effective conduction band edge $E_c(\mathbf{r})$, including many-body effects, at $x = A$, B, C, and D (Fig. 19a) and at $y = 0$ (center) along the growth (z) direction for $N = 2$ electrons in a 4000 Å diameter VQD structure. Although the overall picture is not at all symmetrical with respect to a vertical line drawn at the center of the well, the central region, as shown in the inset, is perfectly symmetrical, which explains that simulations on a simpler domain were qualitatively in good agreement with experimental results [63, 66, 70]. The central InGaAs dot region (4105 Å $\leq z \leq$ 4225 Å) is defined by the 180 meV AlGaAs double barrier. The lower energy gap of InGaAs (1.448 eV) compared to the energy gap of GaAs (1.519 eV) results in the bottom of the well being, 35 meV lower than the source and drain regions. Therefore, electrons are confined in the 120 Å wide InGaAs region. The portion of curve A on the right of the barrier shows a steplike behavior which results in two different pinnings of the conduction band under the gate and at the exposed surface, respectively.

Figure 19d shows, for the same structure, the variation of the conduction band edge along the x direction, that is, perpendicular to the growth direction, along the lines E and F (in the bulk), G (underneath the gate), H (in the InGaAs well), and I (in the AlGaAs barrier). None of these curves is parabolic at all, but, as we will see, the bottom of the conduction band in the dot region is almost parabolic.

In the analytical potential model, we have already pointed out the weak potential dependence with respect to radius of the structure. This is confirmed by our simulations and is shown in Figure 20a for structures with diameter $d =$ 2000, 5000, and 8000 Å. For $0 \leq r \leq 1000$ Å, the conduction band profile of each structure is almost identical so that the effective size of the dot is practically independent of the size of the structure.

According to our experience, a variation in the gate shape or length has an impact only on the threshold voltage V_{th} of the structure as shown in Figure 20b. Indeed, in a structure with a shorter gate, the fringing fields are more important than in a structure with a larger gate so that a more negative bias is required to empty the dot. As predicted by

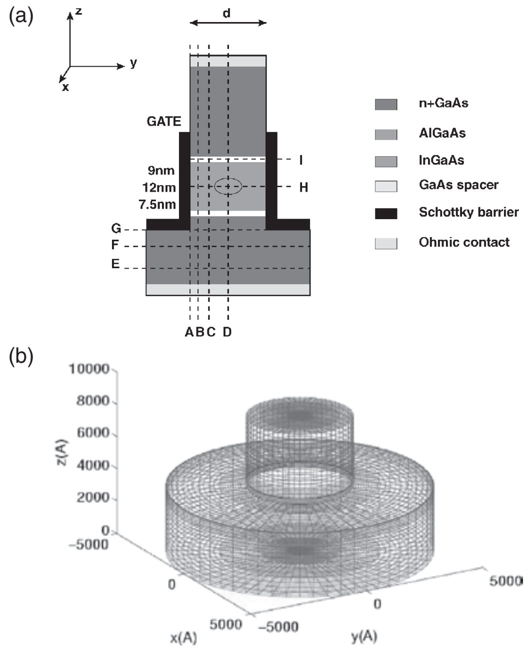


Figure 19. (a) Schematic diagram of a cylindrical vertical quantum dot tunneling heterostructure showing the different semiconductor layers. The quantum dot, represented with the oval, lies in a 12 nm wide quantum well ($In_{0.05}Ga_{0.95}As$) surrounded by two potential barriers ($Al_{0.22}Ga_{0.78}As$) whose thickness are 7.5 nm on the source side and 9 nm on the drain side. (b) Finite element mesh. (c) Conduction band edges along the z-direction at $x = 0$ along the lines A, B, C, and D in (a). (d) Conduction band edges along the x-direction at $y = 0$ along the lines E, F, G, H, and I in (a).

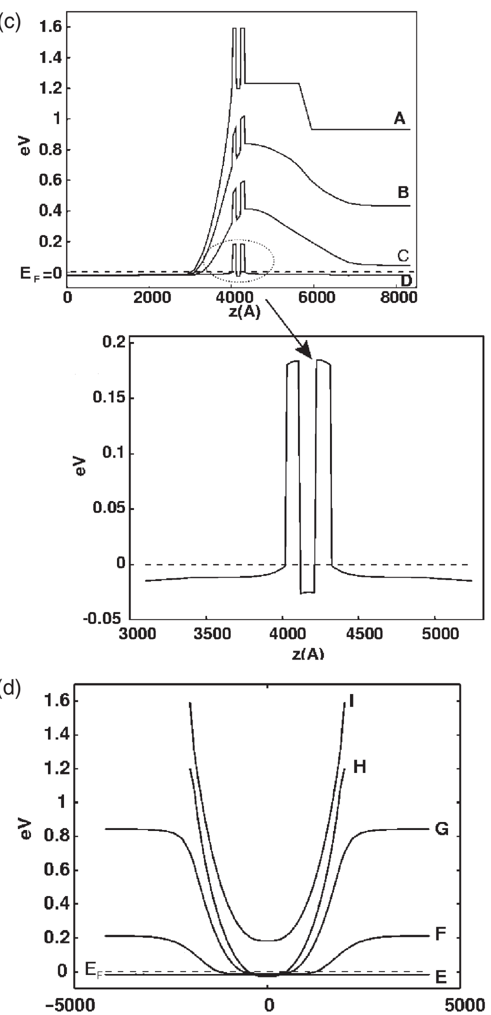


Figure 19. Continued.

the analytical model [Eq. 35], the threshold voltage varies linearly with the ionized donor concentration in the bulk N_D^+. Figure 20c shows that V_G varies almost linearly with N_D^+. It also shows a comparison between the self-consistent (SC) results, our 3D analytical model, and the 2DHO model. The slope of the curve predicted by our 3D model is in remarkable agreement with the SC results and is a clear improvement over the 2D model. There is, however, a 250 mV difference between the values predicted by the analytical model and the SC data because the analytical model does not consider the quasi-neutral region of the structure. Figure 20d shows the dependence of the confining potential on the ionized donor concentration. In practice, however, the range of doping concentration itself depends on the size of the structure. When device diameter is small, low doping concentration produces a confining potential that hardly confines anything. If device diameter is large, a too high doping concentration would make it impossible to pinch the dot off.

Figure 21 shows the variation of the conduction band edge $E_c(\mathbf{r})$ along the x direction, at $y = 0$ in the QD, for N varying from 0 to 20; that is, -1.7 V $< V_G < -0.61$ V. It is seen that when the dot is empty, the conduction band profile is almost parabolic, but as the gate voltage is varied, thus increasing the number of electrons in the dot, the bottom of the conduction band flattens, due to increasing Coulomb repulsion (Hartree potential) amongst charge carriers. The inset shows that the minimum of the conduction band edge $E_c(\mathbf{r})$ in the dot is not a monotonic function of the number of electrons N but oscillates as a combined influence of the attractive ion potential ϕ_{ION} and V_{xc}, and the repulsion between electrons ϕ_H.

7.6. Electronic Structure of Vertical Quantum Dots

The first 45 single-particle eigenlevels and the profile of the wave functions along the radial direction are shown in Figure 22, for $N = 2$. The confining potential is drawn on dashed lines. The corresponding xy (or $r\theta$) contour plots are shown in Figure 23, in agreement with the statements of Section 6.2. Although the shell structure is maintained, each shell is split into various subshells. In general, if one starts numbering shells at $n = 0$, the nth shell contains $n+1$ states. The even numbered shells ($n = 2j$) split into j twofold degenerate subshells plus one s state; the odd numbered shells ($n = 2j+1$) split into $j+1$ twofold degenerate subshells. Figure 23 shows that the levels in the shell number

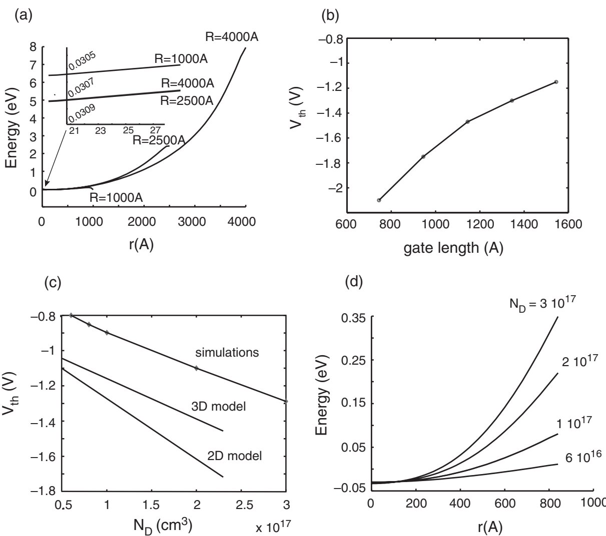


Figure 20. (a) Conduction band profiles in the dot calculated from simulations for $R = 1000$, 2500, and 4000 Å for $V_G = V_{\text{th}}$. (b) Threshold voltage V_{th} as a function of the gate height, (c) as a function of the ionized donor concentration N_D^+ in the bulk. (d) Dependence of the confining potential on N_D^+. The simulations were performed on a 2000 Å diameter cylindrical VQD. In (a), $N_D^+ = 10^{17}$ cm^{-3}. Reprinted with permission from [81], P. Matagne et al., *Comput. Modeling Eng. Sci.* 1, 1 (2000). © 2000, Tech. Science Press.

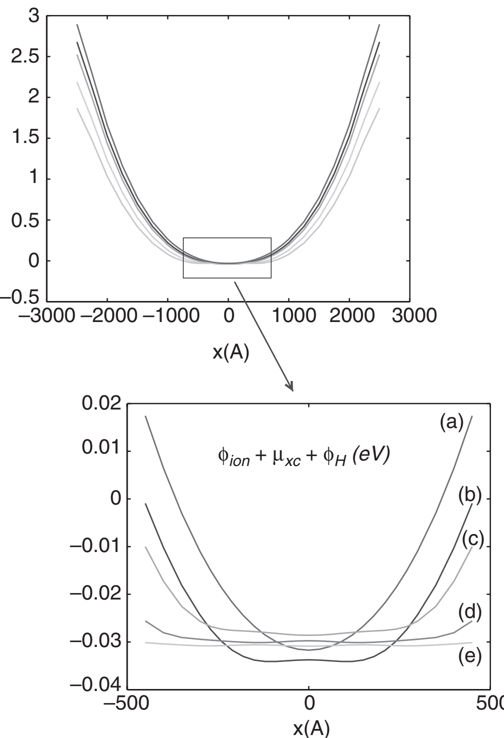


Figure 21. Conduction band profile in the x-direction (perpendicular to the growth direction) at $y = 0$ and $z = 1180$ Å, that is, in the quantum well for N equals (a) 0 ($V_G = -1.7$ V), (b) 2 ($V_G = -1.56$ V), (c) 6 ($V_G = -1.2$ V), (d) 12 ($V_G = -0.9$ V), and 20 ($V_G = -0.609$ V) electrons. The inset is a zoom of the box drawn in the general view. Reprinted with permission from [81], P. Matagne et al., *Comput. Modeling Eng. Sci.* 1, 1 (2000). © 2000, Tech. Science Press.

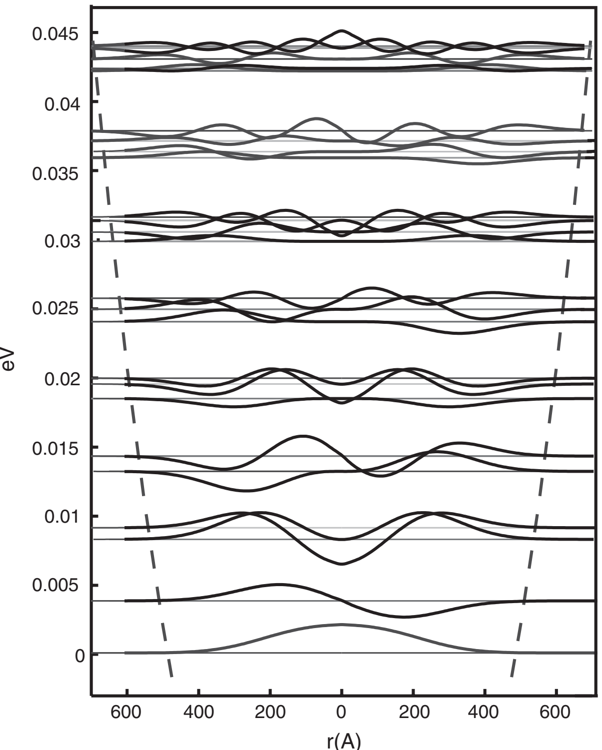


Figure 22. First 45 eigenlevels and radial projection of the eigenfunctions in the confining potential (dashed line).

j are obviously characterized by the quantum numbers (n, l)

$$(0, \pm j), (1, \pm(j-2)), \ldots, (j/2, 0) \qquad j \text{ even} \quad (54)$$

$$(0, \pm j), (1, \pm(j-2)), \ldots, ((j-1)/2, \pm 1) \qquad j \text{ odd} \quad (55)$$

It must be noticed that, inside a shell, the subshells are ordered in descending order of $|l|$.

7.7. Single-Electron Charging

Because the QD is weakly coupled to the source and drain, electrons are completely localized in the dot. At equilibrium, and for a given bias, as mentioned in Section 3.2 the integer number of electrons N minimizes the grand potential of the system Ω. In order to determine N, at equilibrium two approaches may be considered:

- The first method involves the computation of Ω for several values of N in order to find its minimum. This task is readily simplified by considering that the number of electrons in the dot lies around twice the number of eigenlevels below the Fermi level. Only a few values of N are considered. In $\Omega(N) = F(N) - N\mu$, the Helmoltz free energy $F(N)$ now reads

$$F(N) = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \iint \frac{q^2}{4\pi\epsilon_0\epsilon_r} \frac{n(\boldsymbol{r})\, n(\boldsymbol{\rho})}{|\boldsymbol{r} - \boldsymbol{\rho}|}\, d\boldsymbol{r}\, d\boldsymbol{\rho} + E_{xc}[n(\boldsymbol{r})] - \int n(\boldsymbol{r}) V_{xc}[n(\boldsymbol{r})]\, d\boldsymbol{r} \qquad (56)$$

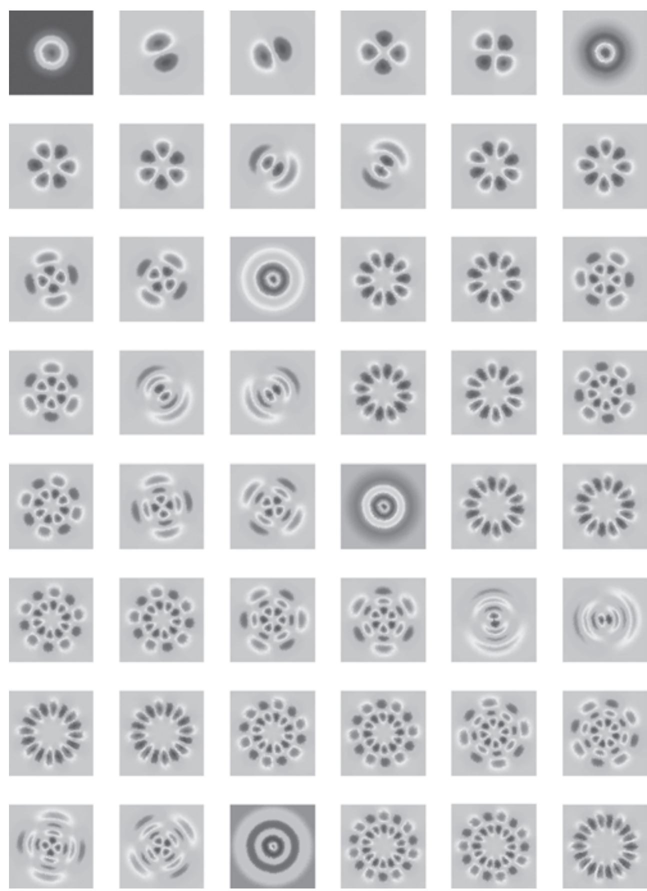


Figure 23. Eigenfunction contour plots for the first nine shells.

where E_{xc} and V_{xc} are the exchange and correlation energy and potential, respectively. In Eq. (56), the second term with the factor 1/2 on the right hand side avoids double counting in the Coulomb interaction energy [82] and the third and fourth terms on the right hand side take exchange and correlation contributions into proper account [82]. This method is computationally tedious because of the double integral in Eq. (56) and will be avoided whenever possible.

- The second method relies on the physical significance on the Kohn–Sham equation eigenvalues [92]

$$\varepsilon_i = \frac{\partial E_T}{\partial n_i} \tag{57}$$

where E_T is the total energy of the dot and n_i is the occupancy of the single-particle level ε_i. Then, if ε_i is a linear function of n_i, it is easy to show that [92] (*Slater formula*)

$$E_T(N+1) - E_T(N) = \int_0^1 \varepsilon_{\text{LUO}}(n)\, dn \approx \varepsilon_{\text{LUO}}(1/2) - E_F \tag{58}$$

where E_F is the Fermi energy and ε_{LUO} is the lowest unoccupied orbital eigenvalue. Hence, upon populating ε_{LUO} with 0.5 electron, a stable configuration of N electrons is achieved in the dot if $E_T(N+1) > E_T(N)$ [i.e., if the integral in Eq. (57) is positive]; otherwise there are $N+1$ electrons. Calculations by Fonseca et al. [93] have established the validity of this approximation in self-assembled InAs/GaAs QDs. Here, we will present a method to test the validity of this approach.

7.8. Computation of the Addition Energy

In this section, we apply Eq. (57) to the particular case where the single-particle eigenlevels ε_i are linear functions of both n_i and V_G. We then show that the AE has a very convenient geometrical interpretation.

By definition,

$$E_a = \mu(N+1) - \mu(N) \tag{59}$$

$$\mu(N) = E_T(N) - E_T(N-1) \tag{60}$$

where E_a is the AE, E_T is the total energy, and $\mu \equiv E_F$ is the chemical potential. Thus, using (58), (59), and (60), we get

$$\begin{aligned} E_a &= \mu(N+1) - \mu(N) && (61)\\ &= (E_T(N+1) - E_T(N)) + (E_T(N) - E_T(N-1)) && (62)\\ &= \varepsilon^{N}_{\text{LUO}}(1/2) - \varepsilon^{N-1}_{\text{LUO}}(1/2) && (63) \end{aligned}$$

Therefore (see Fig. 24),

$$E_a = \left(V_G^{N+1} - V_G^{N}\right) tg(B) \tag{64}$$

if we assume that for N electrons, $\varepsilon_{\text{LUO}} = \varepsilon_{i+1}$, and for $N-1$ electrons, $\varepsilon_{\text{LUO}} = \varepsilon_i$. Here, the upper script of ε gives the total number of electrons in the system and the number between brackets gives the occupancy of a single-particle level.

Let us now assume that ε_i is linear function of both n_i and V_G,

$$\varepsilon_i(n_i, V_G) = an_i + bV_G + c \tag{65}$$

Obviously from Slater's formula, we must have

$$\varepsilon_i^N(n_i, V_G) = an_i + b(V_G - V_G^{N+1}) - a/2 \tag{66}$$

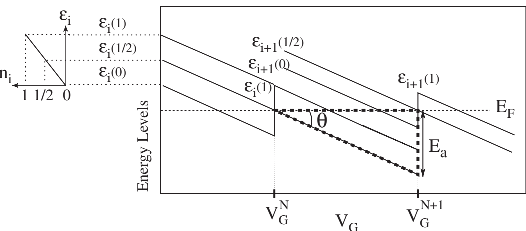


Figure 24. Schematic of the energy spectrum in the gate voltage interval corresponding to successive charging of two electrons. Single-particle energies $\varepsilon_i(n_i)$ are represented with the level index (i) and its electron occupation n_i. E_F and E_a are the Fermi level and the AE, respectively. θ is the slope of the variation of the eigenlevel as a function of the gate voltage. Side: variation of the eigenlevels $\varepsilon_i(n_i)$ as a function of their occupation number n_i.

Thus,

$$\varepsilon_i^N(n_i, V_G^{N+1}) = an_i - a/2 \tag{67}$$

and

$$\int_0^1 \varepsilon_i^N dn_i = \int_0^1 (an_i - a/2)dn_i = \left[\frac{an_i^2}{2} - \frac{an_i}{2}\right]_0^1$$
$$= 0 = \varepsilon_i^N(1/2, V_G^{i+1}) \tag{68}$$

which confirms Slater's rule. Moreover, using Eq. (67),

$$\varepsilon_i^N(1, V_G^{i+1}) = a/2 \qquad \varepsilon_i^N(0, V_G^{i+1}) = -a/2 \tag{69}$$

as illustrated on Figure 24. It is therefore possible to test the validity of Slater's formula in particular by doing a simulation at $V_G = V_G^{i+1}$ and populating ε_i with 0, 1/2, and 1 electron and checking if the shift of ε_i is symmetric with respect to E_F.

7.9. Electron Densities in Vertical Quantum Dots

Figure 25 shows the evolution of the electron density $n(\boldsymbol{r})$ in the xy plane as N increases from 2 to 30. Each subplot corresponds to a filled 2D shell. Each density profile can be understood from Figure 23 which shows the contour plots of the eigenfunctions where the electron density is equal to the summation of the square of the wavefunctions over the occupied states. Hence, for two electrons in the dot, that is, one electron of each spin on the s state (Fig. 25a), the electron density has one lobe and a maximum at the center, which is in agreement with the first wavefunction of Figure 23. When the p-orbitals are filled (i.e., $N = 6$), the maximum has moved on a ring around the center (Fig. 25b) since the two p states have a node at the center and two maxima 250 Å away from it (second and third wavefunction in Fig. 23). When the third shell is completely filled (i.e., $N = 12$), the maximum of $n(\boldsymbol{r})$ moves back to the center, but the crown is still present, although further away from the center (Fig. 25c). This is consistent with the third, fourth, and fifth wavefunctions in Figure 23: the fifth wavefunction shows a maximum at the center while the third and fourth ones have their maximum 280 Å away from it. When the fourth shell is filled (i.e. $N = 20$), the maximum of the electron density moves back again at the periphery (Fig. 25d) since all wavefunctions of this shell have their maximum at the periphery.

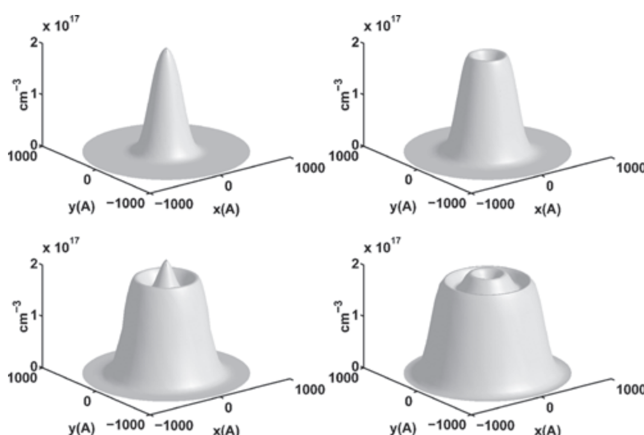


Figure 25. Electron density profile in the xy plane for (a) 2, (b) 6, (c) 12, and (d) 20 electrons in the dot. Adapted with permission from [81], P. Matagne et al., *Comput. Modeling Eng. Sci.* 1, 1 (2000). © 2000, Tech. Science Press.

Figure 26 shows the electron concentration, $n(\boldsymbol{r})$, in the xz plane at $y = 0$, for $N = 20$. The effect of the different types of confinements is evident in the figure: Along the z-direction, the electrons in the dot are squeezed into a thin layer whereas in the x-direction, the concentration profile is more spread out and exhibits four lobes, which result from the contributions of the occupied excited states The only lobe along the z-direction indicates that only the ground z-state (i.e., $n_z = 0$) is occupied, if we label the states as (n_x, n_y, n_z) where n_x, n_y, and n_z are the number of nodes in the x, y, and z directions respectively.

7.10. Shell Charging Analysis

Figure 27 shows the single-particle eigenspectrum for the first 10 orbitals in the CVQD as a function of the gate voltage V_G as obtained from the 3D computational model. Because of their spatial symmetry and spin degeneracy, the orbitals are grouped into four sets of levels which when occupied by electrons will form the first four shells. The upper horizontal scale indicates the number of electrons in the QD on a particular range of V_G. On the vertical axis, the Fermi level is taken as the reference at zero. Below $V_G = -0.62$ V, all levels are empty and decrease monotonically with V_G as the potential energy is lowered.

The variation of all eigenlevels ε_i is piecewise linear with respect to V_G. At each transition from N to $N + 1$ electrons, each curve $\varepsilon_i(V_G)$ is discontinuous with an upward shift. This is a manifestation of the Coulomb blockade effect: when a new electron enters the dot, it induces an upward shift of the whole spectrum because of the sudden increase of the total electrostatic energy in the dot due to Coulomb interaction. Here, the first electron entering the dot occurs for $V_G = -0.62$ V. It must be noticed that the higher the shell being populated, the smaller the shifts, that is, the smaller the charge increment. Unlike in orthodox Coulomb

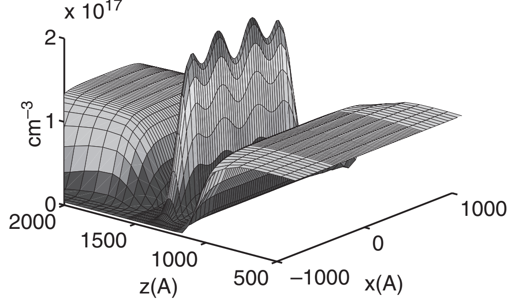


Figure 26. Electron concentration in the xz plane (growth direction) at $y = 0$ for 20 electrons in the dot. Reprinted with permission from [81], P. Matagne et al., *Comput. Modeling Eng. Sci.* 1, 1 (2000). © 2000, Tech. Science Press.

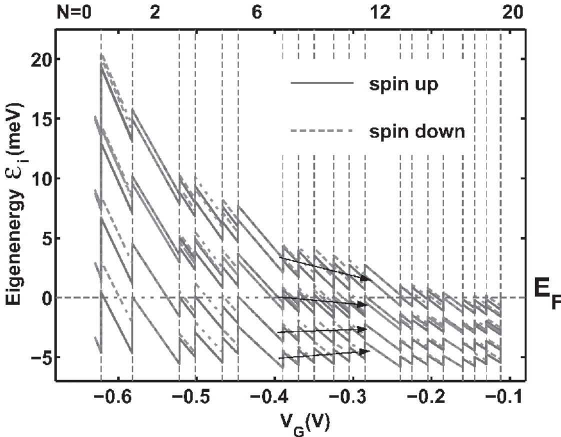


Figure 27. Energy spectrum of the first four sets of eigenlevels (solid curves: spin up, dotted curves: spin down) as a function of the gate voltage V_G (bottom horizontal axis) and number of electrons N (top horizontal axis). E_F is the Fermi level.

blockade experiments [6], ΔV_G, the voltage increment for adding an electron in the dot, takes very unequal values, irregularly distributed over the range of V_G, because of the influence of the quantization and quantum many-body effects during the charging of the dot. ΔV_G is large for $N = 2, 6$, and 12 (i.e., after the complete filling of each shell) because, in addition to the energy required to overcome the electrostatic repulsion of the electrons already present in the dot, an energy contribution is needed to lower the next orbital below the Fermi level for admitting the next electron.

We note that in the interval between the addition of two electrons, $\varepsilon_i(V_G)$ decreases monotonically with V_G because the overall potential energy of the system decreases as the gate voltage is made more positive. This behavior is different from the results of Nagaraja et al. [80] whose model did not account for individual spin states. In that case, the single-particle eigenvalues were tangential to the Fermi level during the charging of the whole orbital.

During the charging process over the whole gate voltage range (-0.62 V $< V_G < -0.082$ V), the four sets of levels (shells) remain well separated, which indicates that the shell structure is preserved. However, as predicted by the analytical model (Section 6), the degeneracy of levels belonging to the same shell, even when nonoccupied (i.e., second, third, fourth sets of levels for -0.522 V $< V_G < -0.447$ V), is partially lifted.

Another interesting feature appears by considering the overall energy variation of the different shells during a single shell charging. For instance, for -0.39 V $< V_G < -0.285$ V (i.e., during the third shell charging), the average energy of the two lower full shells is higher at the end of the charging process than at the beginning. This is easily seen by drawing a line joining the middle of each discontinuity segment. For the two lower shells, the slope is positive, which indicates an overall energy repulsion among these shells. For the third filling shell, the slope is slightly negative, which means that the attractive energy of the gate field is just sufficient to overcome the electron repulsion energy. Finally, for the upper empty shell, the slope is negative because, free of electron repulsion, the eigenstates only experience the attractive effect of the gate.

We also notice the decrease of the intershell energy separation as V_G increases, as observed experimentally [11]. This effect is generally interpreted as due to the Coulomb interaction between high energy electrons with extended wavefunctions which weakens the confining potential. However, our analytical model reveals that, in CVQDs, there is in addition a pure *3D geometrical effect* that enhances the level collapse when V_G increases *even if the dot is empty*. Indeed, from Eq. (34), the quadratic term α in the potential depends on the depletion region extension z_∞. As z_∞ decreases when V_G is made more positive, so does α (Fig. 16). Therefore, as V_G is swept, we expect the oscillator frequency to decrease and, thereby, the shell separation. This effect is at the origin of the renormalization of the harmonic oscillator frequency by the fourth root of the electron density; that is, $\omega \propto 1/n^{1/4}$, in models based on pure 2D parabolic confining potential [66, 75]. This $n^{-1/4}$ renormalization was justified to keep the electron concentration constant during charging of the dot, which is readily obtained in our simulations.

Addition energy measurements are the clearest way to evidence the shell structure in CVQDs. Figure 28 shows the variation of the addition energy with respect to the number of electrons N, as measured in Tarucha's experiment (dashed) and as computed in our simulations for three different confining potentials $\omega_1 > \omega_2 > \omega_3$ that correspond to 50, 30, and 20 meV conduction band offset ΔE_c between GaAs and InGaAs, respectively. The band offset variation is not unreasonable due to the fact that the GaAs conduction band edge with a doping density of 10^{17}/cm^3 is simply not well defined and this can result in an impurity fluctuation at the well edge ($z = z_i$). Figure 28 shows a very pronounced peak for $N = 2, 6$, and 12 reminiscent of the shell structure in the experimental curve as well as in the simulation curves for $\omega = \omega_1$ and ω_2. For $\omega = \omega_3$, the confining potential is so weak that the shell structure begins to vanish. In this case, charging the second shell ($N = 2$) requires less energy than overcoming the repulsive energy generated by the first electron ($N = 1$) so that $E_a(1) < E_a(2)$. In the experimental curve, there are also secondary peaks for $N = 4, 9$, and 16. These peaks are interpreted as the manifestation of Hund's rule with parallel spin alignment for a

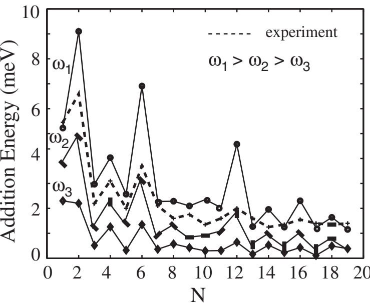


Figure 28. Comparison between computed addition energy spectra (solid) for three different confining potentials $\omega_1 > \omega_2 > \omega_3$ and the experimental spectrum [11] (dashed) as a function of the number of electrons. Notice that, by definition, $E_a(N) = E_F(N+1) - E_F(N)$, so that a value for $N = j$ in this figure refers to the energy needed to add the $(j+1)$th electron in the dot.

half-filled shell [9, 11, 34, 35]. In the simulation curve, we clearly obtain the peaks at $N = 4$ and $N = 16$, but not the peak at $N = 9$, and we have other secondary peaks at $N = 8$, 10, 14, and 18.

7.11. Spin Sequence in Vertical Quantum Dot Charging

If Hund's rule governs the filling of the second shell, then, necessarily, the addition energy spectrum will exhibit a peak at $N = 4$ surrounded by two minima at $N = 3$ and $N = 5$, as described in Section 5.2. Hence, the third and fourth electrons, with parallel spins (↑), access the empty $|0, \pm 1, \uparrow\rangle$ orbitals, maximize the exchange interaction, and induce a minimum for $N = 3$. For $N = 4$, the fifth electron, occupying the $|0, 1, \downarrow\rangle$ state, is the only ↓-electron in the second shell, with no additional exchange; that is, $E_a(4) > E_a(3)$. For $N = 5$, the sixth electron in the $|0, -1, \downarrow\rangle$ state induces exchange between the sixth and the fifth electron, which lowers $E_a(5) < E_a(4)$, leading to a peak for $N = 4$.

The idea that the peak at $N = 9$ is the manifestation of Hund's rule originates from the 2D circular oscillator model where the three states of the third shell are degenerate.

In this case, the filling sequence would be

$$L^S = 2^{1/2} \to 2^1 \to 0^{3/2} \to 2^1 \to 2^{1/2} \to 0^0 \tag{70}$$

where $L(S)$ is the total angular momentum (total spin), and leads to the configuration $1s^2\ 2p^4\ 3d^2\ 3s^1$, by adopting a terminology similar to atomic physics for representing the electronic orbitals. Here, the $|1, 0, \uparrow (\downarrow)\rangle$ (s-like) state is occupied before the $|0, \pm 2, \uparrow (\downarrow)\rangle$ (d-like) states to maximize the total orbital angular momentum. Consequently, the addition energies for $N = 7$ and 8 are low, since the exchange interaction is maximized. In addition, $|1, 0, \uparrow\rangle$ and $|0, \pm 2, \uparrow\rangle$ overlap weakly, which reduces the (repulsive) Hartree energy E_H. For $N = 9$, the same argument as for $N = 4$ applies; that is, the tenth electron is the first ↓-electron of the third shell (no exchange-correlation) that sits on an already occupied orbital (large Hartree), producing a peak for $N = 9$. However, it was demonstrated earlier (Section 6.2) that the accidental degeneracy between the $\varepsilon_2^{1,2}$ and ε_2^3 is lifted in CVQDs so that the actual configuration depends on the antagonistic action of the exchange energy and the $|1, 0\rangle$-$|0, \pm 2\rangle$ level separation E_λ.

If E_x is the exchange energy between two spin-parallel electrons, one in the $|1, 0\rangle$ state and one in a $|0, \pm 2\rangle$ state, that, for the present structure, is effectively 0.3–0.4 meV for $N = 8$ [94], three situations, shown sequentially in Figure 29, are predicted at $0T$ depending on the value of E_λ:

1. $E_\lambda \leq 0.2$ meV ($E_\lambda \ll E_x$) for which 3D effects and Coulomb repulsion interactions practically cancel each other. Then, the three levels of the third shell are practically degenerate, and $E_a(N)$ initially shows a peak at $N = 9$ and a clear dip at $N = 7$. The state occupation sequence going from $N = 7$ to 12 is $L^S = 2^{1/2} \Rightarrow 2^1 \Rightarrow 0^{3/2} \Rightarrow 2^1 \Rightarrow 2^{1/2} \Rightarrow 0^0$. The second electron in the third shell with parallel spin prefers to sit on the $|1, 0\rangle$ state rather than on the available $|0, -2\rangle$ state because of the weaker Coulomb repulsion (top inset of Fig. 29a). However, this sequence is quickly broken as soon as the accidental degeneracy between the $|1, 0\rangle$ and $|0, \pm 2\rangle$ states is slightly lifted, to be replaced by the new sequence $L^S = 2^{1/2} \Rightarrow 0^1 \Rightarrow 0^{3/2} \Rightarrow 2^1 \Rightarrow 0^{1/2} \Rightarrow 0^0$ where the weaker Coulomb (Hartree) interaction between two electrons in the $l = 0$ and $l = 2$ states cannot compensate for the level splitting (bottom inset of Fig. 29a). Nonetheless, the spectral shape, like in Figure 29a for $E_\lambda = 0.11$ meV, may change very little. Hence, negative nonharmonicity arising from 3D effects is critical to offset the influence of Coulomb interactions for achieving this configuration.
2. 0.2 meV $< E_\lambda <$ 0.6 meV ($E_\lambda \sim E_x$) for which the sequence (inset of Fig. 29b is the same as the final sequence in a) with Hund's first rule is still being fulfilled ($S = 3/2$ for $N = 9$). However, placing the third electron in the third shell on the $|1, 0\rangle$ state that is now much higher than the two $|0, \pm 2\rangle$ states requires the non-negligible energy E_λ, and this can noticeably increase $E_a(8)$ (see Fig. 29b for $E_\lambda(7) = 0.35$ meV). The fourth electron in the third shell with antiparallel spin, deprived from gain in energy due to the exchange interaction, still falls on one of the two $|0, \pm 2\rangle$ states, and so $E_a(9)$ can be reduced. Still with a clear dip at $N = 7$, $E_a(N)$ may show a peak at $N = 8$, and not $N = 9$, with a magnitude depending on the difference between E_λ and E_x.
3. $E_\lambda > 0.6$ meV ($E_\lambda > E_x$), for which the third shell splits into two distinct subshells of four and two electrons. One then expects a peak at $N = 8$ for maximum spin alignment ($S = 1$) within the lower subshell, a situation similar to that for the second shell peak at $N = 4$ (i.e., Hund's first rule). There is also a peak at $N = 10$ because of the presence of the non-negligible gap between the two subshells. $E_a(N)$ for $E_\lambda(7) = 0.65$ meV, and the new filling sequences are shown in Figure 29c.

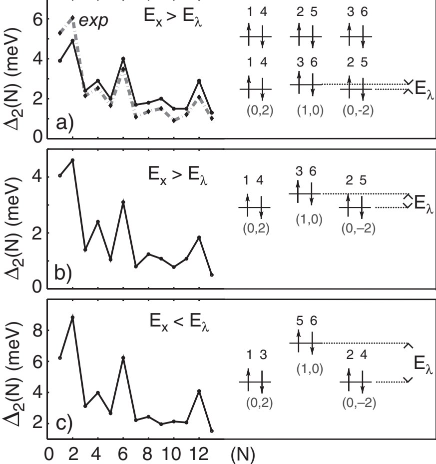


Figure 29. Addition energy spectra computed by 3D DFT versus N, for three ranges (a), (b), and (c) of nonparabolic potential energy, E_λ. Insets schematically show the filling sequences for the third shell. Bold dotted-dash trace in (a) gives $E_a(N)$ for a 0.54 μm device discussed in detail in the text. Reprinted with permission from [71], P. Matagne et al., *Phys. Rev. B* 65 (2002). © 2002, American Institute of Physics.

7.12. Magnetic Field Analysis

In order to determine the spin filling sequence within a shell, a magnetic field parallel to the current flow is applied. Level splitting is observed according to the Darwin–Fock model [95],

$$E_{n,l} = (2n + |l| + 1)\hbar\left(\frac{1}{4}\omega_c^2 + \omega_0^2\right)^{1/2} - \frac{1}{2}l\hbar\omega_c \qquad (71)$$

where $\hbar\omega_c$ is the cyclotron energy. Theoretical calculations and experimental results are shown in Figure 30a and b respectively. Theoretical calculations predict that levels with positive l shift downward and levels with negative l shift upward. Figure 30b shows the variation of the current oscillations as a function of N for B varying from 0 to $8T$. Assuming a mapping between Figure 30a and b, the electrons must therefore undergo transitions in their quantum numbers. For example, at $0.4T$, the fourth electron goes from $|0, -1\rangle$ to $|0, 1\rangle$ and the fifth electron goes from $|0, 1\rangle$ to $|0, -1\rangle$. This suggests that at $B < 0.4T$, shell filling can be predicted by Hund's rule: up to half shell filling, all electrons have parallel spins; more electrons can only be added with antiparallel spins. At $B > 0.4T$, Hund's rule breaks down because the splitting between levels originally degenerate at $B = 0T$ becomes larger than exchange between electrons of parallel spin.

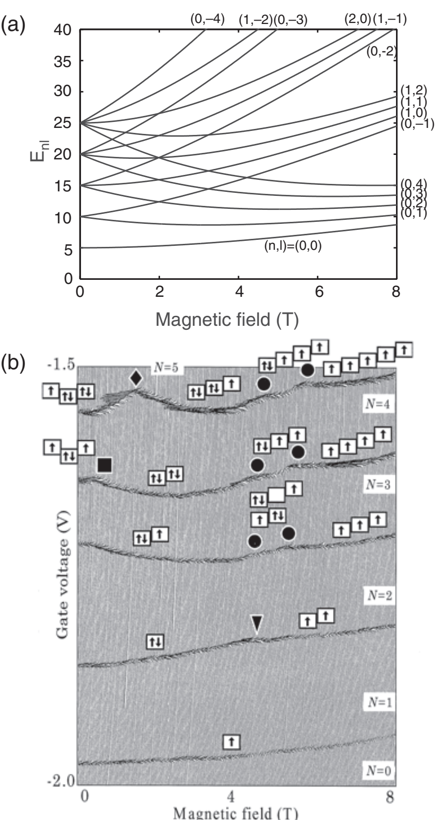


Figure 30. Single-particle energy levels as a function of the magnetic field (a) theoretical (Darwin–Fock model), (b) experimental data. Reprinted with permission from [35], L. P. Kouwenhoven et al., *Science* 278, 1788 (1997). © 1997, American Assoc. Adv. Sci.

Figure 31 shows experimental $0T$ third shell spectra for four VQDs. Applying an arbitrarily small bias ($<100\ \mu$V) between the source and drain, V_d, the spacings between the Coulomb oscillation peaks on sweeping V_g are measured to an accuracy of better than 2%, which reflect directly $E_a(N)$ [71]. For easy comparison, we actually show the peak spacings normalized by the peak spacing for $N = 6$, that is, $E_a(N)/E_a(6)$. We start our analysis with the topmost trace in Figure 31. The AE spectrum of this 0.5 μm device is shown in full in Figure 31b of ref. [11], and principal peaks at $N = 2, 6, 12$, and 20 are clear. The secondary peak at $N = 9$ has been widely interpreted as due to Hund's first rule ($S = 3/2$), apparently indicating that the three single-particle levels of the third shell are quasi-degenerate as depicted in Figure 31a. We cast doubt on whether this simple interpretation is so clear-cut. First, the weak magnetic ($B-$) field dependence close to $0T$ of the 7th to the 12th Coulomb oscillation peaks appears to show just regular pairing (Fig. 2 of ref. [11]), surprisingly implying the sequence $L^S = 2^{1/2} \Rightarrow 4^0 \Rightarrow 2^{1/2} \Rightarrow 0^0 \Rightarrow 0^{1/2} \Rightarrow 0^0$.

Second, $E_a(7)$ is larger than expected when we look at Figure 31a. Both observations bare the hallmark of the presence of small (possibly but not exclusively elliptic) deviations from the circular shape of the fabricated VQD (e.g., due to potential fluctuations arising from long range Coulomb interactions of ionized donors in the doped GaAs), which can lift the degeneracy of all third shell 3s and 3d states. Because the secondary peak in $E_a(N)$ at $N = 4$ can be demonstrated to be due to Hund's first rule ($S = 1$) [11], the deviation from the circular symmetry (κ) is of the order of 10% [66]. Nonetheless, the half-shell peak at $E_a(9)$ does not always guarantee maximum spin alignment ($S = 3/2$). Trace

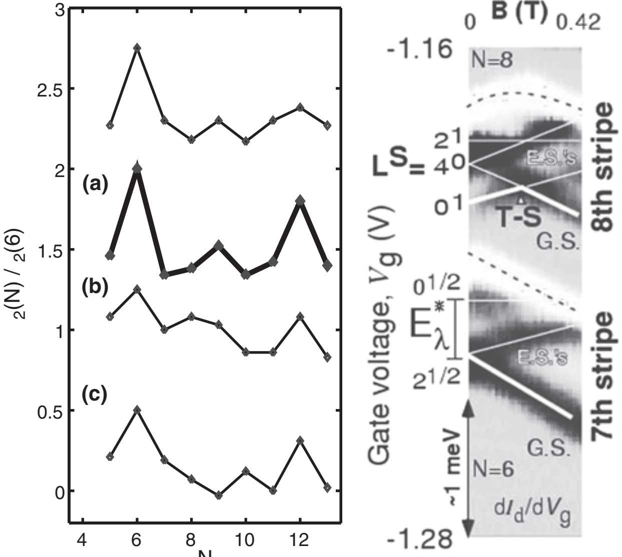


Figure 31. Left panel: Experimental $0T$ third shell $E_a(N)/E_a(6)$ spectra for four VQDs. Traces are shifted vertically for clarity. Right panel: Grey scale plot shows the weak B-field dependence of the seven- and eight-electron ground and first few excited states of the 0.54 μm device whose $E_a(N)$ spectrum is shown as the bold trace in Figure 29a. Reprinted with permission from [71], P. Matagne et al., *Phys. Rev. B* 65 (2002). © 2002, American Institute of Physics.

(a) in Figure 31 is crucial. $E_a(N)$ for this 0.54 μm device is shown in Figure 31a (bold trace). Principal peaks at $N = 2, 6, 12$, and 20 are clear, and the general shape is in good agreement with the theoretical trace. The secondary peak here at $N = 9$ can be demonstrated to be due to Hund's first rule ($S = 3/2$ at $N = 9$). We make this interpretation because: (i) the weak B-field dependence close to $0T$ of the 7th to the 12th peaks reveals clear deviations from regular pairing (Figure 2 of [94]), and (ii) there is a clear dip in $E_a(N)$ at $N = 7$. In Figure 31, we also show dI_d/dV_g in the V_g–B plane for a weak B-field applied parallel to the current (I_d) for the same mesa showing the 7th and 8th current stripes ($V_d \gg 1$ mV) so the width of each stripe is equivalent to an energy $\gg 1$ meV. Positive (negative) dI_d/dV_g is black (white). The lower (upper) edge of the each stripe is black (white). Features within these stripes provide information about the ground state (G.S.) and first few excited states (E.S.) [35] of the 7 and 8 electron systems. For each stripe, the G.S. (E.S.) is marked with a thick (thin) white line at (above) the lower edge. From the $B = 0T$ dependence, we can identify the quantum numbers, L^S, of these many-body states. For $N = 7$, the G.S. (first E.S.) corresponds to the 1st third shell electron entering a $3d$ ($3s$) state at $0T$. For a small B-field, the $L^S = 2^{1/2}$ G.S. splits because the $|n, l\rangle = |0, \pm 2\rangle$ states split. We do not claim that the dot inside this mesa is perfectly circular. Nevertheless, it is sufficiently circular that within our measurement resolution these two states appear degenerate at $0T$ ($\kappa \sim 0\%$). From the separation, E^*_λ, between the $2^{1/2}$ G.S. and $0^{1/2}$ E.S., an estimate for the effective value at $N = 7$ of the subthird shell splitting at $0T$ yields $E_\lambda(7) \sim 0.26$ meV [94]. For $N = 8$, the 2nd third shell electron initially enters the $|0, -2\rangle$ state so the $0T$ G.S. is $L^S = 0^1$, a spin-triplet consistent with Hund's first rule [11, 94]. This electron can only go into the $|1, 0\rangle$ state if E_λ is much smaller than that deduced. The 4^0 singlet E.S. splits for a small B-field, again because the $|0, \pm 2\rangle$ states split. At about $0.2T$, there is a triplet–singlet (T–S) transition in the $N = 8$ G.S. Based on Figure 29, for this near-ideal artificial atom, the filling sequence for the six electrons to enter the third shell is $L^S = 2^{1/2} \Rightarrow 0^1 \Rightarrow 0^{3/2} \Rightarrow 2^1 \Rightarrow 0^{1/2} \Rightarrow 0^0$.

With principal peaks at $N = 2, 6$, and 12, experimental trace (b) ($d = 0.5$ μm) of Figure 31 most resembles the situation of Figure 29b. This trace has a prominent peak at $N = 8$ and a clear dip at $N = 7$ ($\kappa \sim 0$–10%). The key point here is that this spectrum also bares the signature of three parallel spin filling up to $N = 9$, and this is not in violation of Hund's first rule.

Lastly, experimental trace (c) ($d = 0.54$ μm) of Figure 31 reflects the situation depicted in Figure 29c. $E_a(N)$ again reveals principal peaks for $N = 2, 6$, and 12. The peak at $N = 10$ is clear, but there must be a non-negligible splitting of the $|0, \pm 2\rangle$ states. This splitting reduces the effect of exchange between two parallel spin electrons, increasing $E_a(7)$ and reducing $E_a(8)$. The $B = 0T$ dependence of second and third shell Coulomb oscillation peaks reveals barely noticeable deviations from regular pairing. The dot is therefore probably deformed to an extent that puts κ just below that required to induce the $N = 4$ T–S transition, namely $\sim$20% [66].

The different ways electrons fill shells depend sensitively on the 3D nature of the VQD confinement, including small deviations from perfect parabolicity and circular symmetry, and the resulting complex Coulomb interactions between electrons. We have focused on the third shell, for which a clear dip at $N = 7$ followed by a peak at $N = 8$ or 9 is the signature of maximum spin alignment at half-shell, but without loss of generality we have shown that: (i) maximal spin alignment does not guarantee a peak at half-shell filling (conversely a half-shell peak does not always mean maximum spin alignment), and (ii) a particular spin filling sequence does not result in a unique addition energy spectrum. Finally, the model discussed here is very relevant for the general subject of electrostatics in any confined interacting electron system. The third shell of an artificial atom is a particularly good test for this realistic approach.

It is now possible to suggest an interpretation for the peak at $N = 16$ in the fourth shell. Since the accidental degeneracy between the $\psi_3^{1,2}$ and $\psi_3^{3,4}$ states is lifted in the CVQDs (see Section 6.2)—here, the spacing between the two subshells is about 1.5 meV—the filling sequence predicting the four spin alignment up to $N = 16$, as shown in Figure 32a, seems unlikely. As for $N = 10$, the peak at $N = 16$ results from the excess energy needed to access the second subshell in the fourth shell as shown in Figure 32b which also depicts the ordering sequence obtained in our density functional theory-based simulations and for which the addition energy spectrum shows a jagged curve with alternative peaks and minima for even and odd numbers of electrons, respectively (Fig. 28). In this sequence, the secondary peaks at $N = 14$ and $N = 18$ result from the fulfillment of Hund's rule with two parallel spins at midoccupation within each subshell. In order to explain the experimental spectrum with a single peak at $N = 16$, we again consider the complete lifting of degeneracy caused by external static perturbations as illustrated in Figure 32c. The minima at $N = 14$ and $N = 18$ are then due to the filling of two empty states $|0, -3\rangle$ and $|1, -2\rangle$. The intermediate values at $N = 13, 15$, and 19 result from the Coulomb repulsion due to the filling of the already occupied states $|0, 3\rangle$, $|0, -3\rangle$, and $|1, -2\rangle$, respectively. The maximum at $N = 16$, in this scenario, is also due to the energy increment needed to access the second subshell, made of the $|1, 2\rangle$ and $|1, -2\rangle$ orbitals.

8. ASYMMETRICAL SINGLE QUANTUM DOTS

The confining potential of a QD can be altered either structurally, by etching mesas of various shapes, or electrostatically, by patterning several independent gates on a square VQD and applying different voltages at the gates.

8.1. Rectangular Quantum Dots

Experiments performed on rectangular structures lead to the conclusion that the shell structure that would be observed from addition energy measurements in a square or cylindrical VQD is disrupted [9].

For this reason, we consider a rectangular vertical QD shown in Figure 33. The material sequence is identical to the cylindrical structure investigated in the previous chapter with

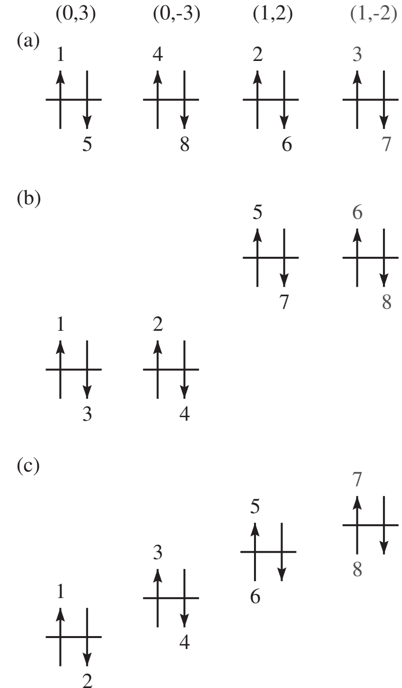


Figure 32. Schematic possible sequences for the filling of the fourth shell. (a) In a 2D circular oscillator with four degenerate levels and four spin alignment. (b) Separation between $(0, \pm 3)$ and $(1, \pm 2)$ leading to the formation of two twofold degenerate subshells. (c) Total lifting of degeneracy leading to the usual pairing sequence $\uparrow \downarrow$.

the vertical confinement realized in the InGaAs well. In-plane confinement is achieved by applying a negative bias on a Schottky gate wrapped around the rectangular pillar, as in the cylindrical structure. The aspect ratio δ of the rectangle is $S = W/D = 1.4$ ($W \cdot D = 7000$ Å $\cdot$ 5000 Å).

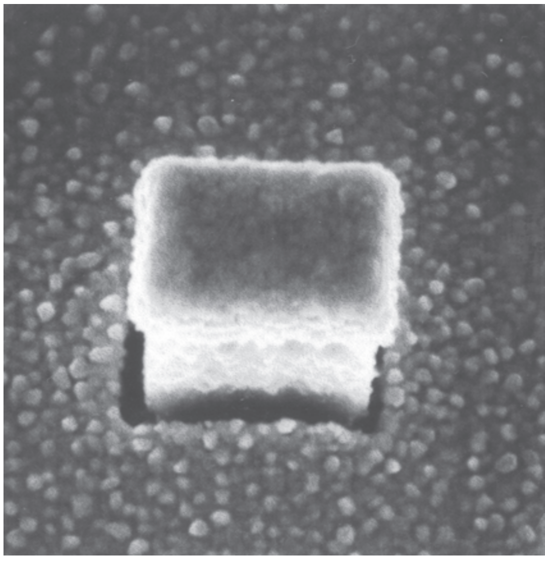


Figure 33. Scanning electron micrograph of a rectangular VQD. Reprinted with permission from [66], D. G. Austing et al., *Phys. Rev. B* 60, 11514 (1999). © 1999, American Institute of Physics.

Using a simple model of an anisotropic harmonic oscillator whose eigenstate reads

$$\varepsilon_{n_{xy}} = \hbar\omega_x\left(n_x + \frac{1}{2}\right) + \hbar\omega_y\left(n_y + \frac{1}{2}\right) \qquad n_x, n_y = 0, 1, \ldots \tag{72}$$

since $\omega_x \neq \omega_y$, there is no reason to expect systematic degeneracies due to spatial symmetry as was the case for the cylindrical structure. Depending on the aspect ratio, only accidental degeneracies may arise.

In Figure 34, we observe level crossings when charging the dot. Level crossings are pointed by arrows and labeled A, B, C, D, and E. For example, at the transition $N = 2 \longrightarrow 3$ electrons, the third and fourth eigenlevels switch, as well as the seventh and eighth eigenlevels. In order to be able to identify the eigenlevels, we need to look at the eigenfunctions. As seen in Figure 35, n_x, the quantum number associated with the larger dimension, changes faster than n_y, the quantum number associated with the smaller dimension. This implies that the aspect ratio of the dot increases with the number of electrons, as the gate is made more positive. This is confirmed by Figure 36 which shows equipotential in the xy plane for $N = 0, 5$, and 10 electrons at $V_G = -1.887$, -1.726, and -1.579 V, respectively. At the center of Figure 36b and c, we can even see how the equipotentials are influenced by the eigenfunctions. The fact that the eigenlevel sequence depends on the number of electrons can be explained easily. For $N = 0$ already, due to the aspect ratio of the QD, the quantum number n_x changes faster than n_y so that the electrons occupy orbitals that extend more in the x-direction than in the y-direction (Fig. 35a).

However, level crossings are observed even if the dot is empty, by solely varying the gate voltage, which is unexpected. Besides, one would expect that high order wavefunctions would be more sensitive to the outer equipotentials that follow more closely the rectangular shape of the device, and consequently have a smaller aspect ratio than inner equipotentials whose shapes are more elliptic. The alteration of the eigenlevel sequence is actually the combination of the many-body effect described above and a three-dimensional effect that can be understood by looking at Figure 37 that shows equipotentials in xz and yz planes, for $V_G = -1.887$ V and $V_G = -1.579$ V, $N = 0$ in both cases. The changes in

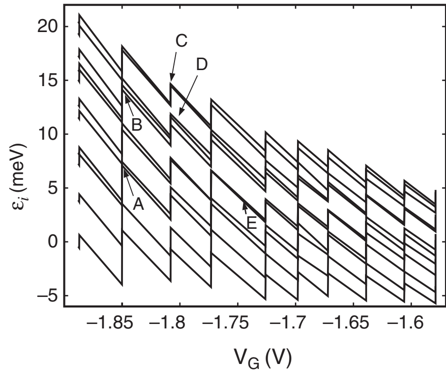


Figure 34. Eigenlevel spectrum ($\uparrow$) for the rectangular quantum dot. The arrows A, B, C, D, and E indicate level crossings.

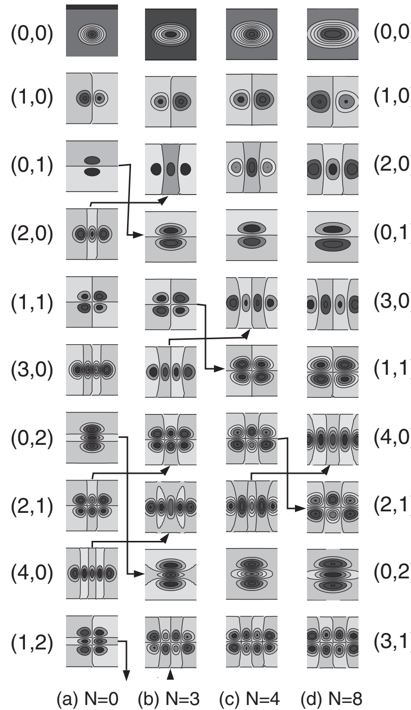


Figure 35. Eigenvector (↑) contour plots for a rectangular VQD as N varies from 0 to 8, showing the alteration of the eigenlevel sequence due to the electrostatic deformation of the confining potential.

the potential profile outside the dot are carried over inside the dot. This effect can also be explained by Eq. (34): the depletion in the bulk influences the curvature of the potential in the dot. Here, however, since the field is stronger in one direction, the rate of change is different along x than along y, which enhances the effect.

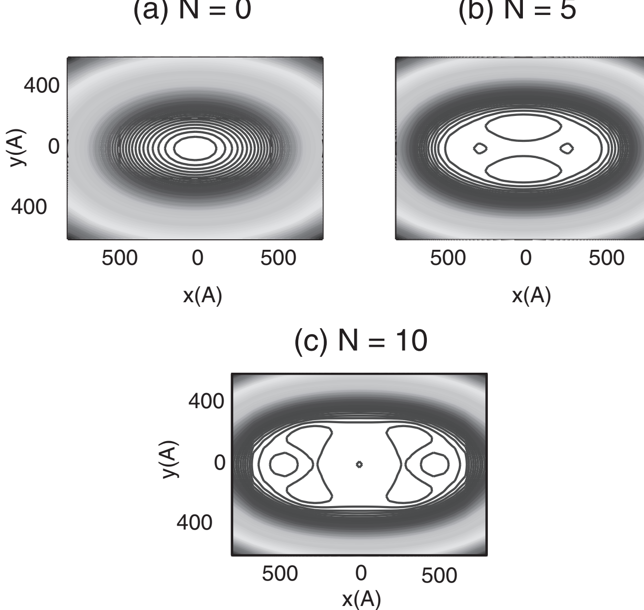


Figure 36. Potential contour plots in the xy plane for $N = 0$, 5, and 10, respectively, showing that the QD aspect ratio increases with N.

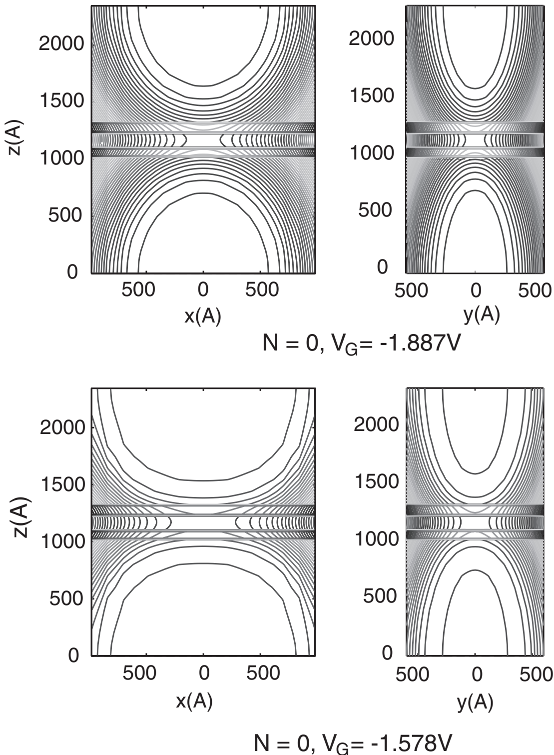


Figure 37. xz potential contour plot for $N = 0$ at $V_G = -1.887$ V and $V_G = -1.578$ V showing that, in addition to the many-body effect, there is a geometrical effect that modifies the potential curvature of the dot.

Figure 38 shows the comparison between the addition energy spectra of a cylindrical QD and a rectangular QD. By comparing Figure 38 to Figure 34, the differences between the two spectra are easily explained. The peak at $N = 2$ is considerably reduced, because, in the rectangular QD with $L = 7000$ Å, the second shell is closer to the first shell than for the 5000 Å diameter cylindrical QD. Hence, the confirmation of Hund's rule fulfillment is rather a matter of how deep the minima are rather than how high the peaks are in the addition energy spectrum as seen for $N = 3$. The rectangular QD has a second shell with only one spatially nondegenerate level whose filling involves no exchange. The

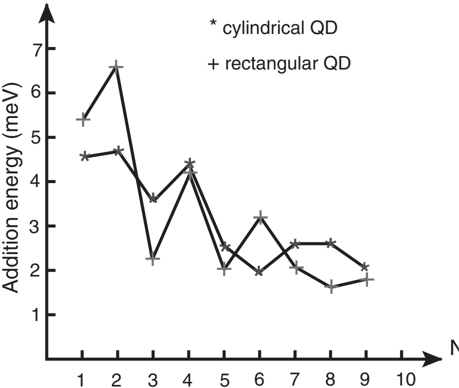


Figure 38. Addition energy spectrum comparison between the cylindrical (+) and the rectangular (∗) QD.

addition energy for $N = 3$ is therefore much larger for the rectangular QD than for the cylindrical QD. The peaks exhibited by both structures for $N = 4$ have two completely different origins. For the rectangular QD, the increase in addition energy stands for the access of a new shell while for the cylindrical QD, it is due to the maximization of the Hartree contribution and minimization of exchange when a shell is half filled plus one electron. For $N = 7$, there is a minimum in the rectangular QD addition energy spectrum because the third shell is half filled and the exchange is maximized, then an increase for $N = 8$ to reach the fourth shell, etc .... The main consequence of the modification of the eigenlevel sequence is that electrons might undergo level transition. Although it is not observed here—the crossings occur at empty level—it will certainly do at a later stage.

Austing et al. studied experimentally and theoretically the charging of rectangular QDs with aspect ratios (δ) varying from 1 to 3.2 [66]. A similar theoretical study has also been performed by Lee et al. [96], for $1 \leq \delta \leq 2$. Both works assume an a priori, lateral CP given by $V_{\text{ext}} = \frac{1}{2}\omega^2(\delta x^2 + \frac{y^2}{\delta})$ and show that, for ellipticity factor $\delta > 1.2$, the shell structure is almost completely destroyed and the total spin sequence is strongly altered. Unsurprisingly, all that was described so far is confirmed by our simulations. But, in addition, we found that the eigenlevel sequence changes when the dot is filled with electrons and the gate is made more positive.

Finally, Figure 39 shows a two-electron configuration for $\delta = 3$ and 5, respectively. In the first case, the attractive confining potential still dominates the electrostatic repulsion and the electrons are in the center of the dot. In the second case, we have the opposite situation where the electron–electron repulsion dominates and the electrons are now clearly localized on each side of the dot.

8.2. Electrostatically Deformable Quantum Dots

Since ideal features are very sensitive to the electrostatic potential confining the electrons [66, 71, 96], Tarucha and co-workers investigated a square VQD with four independent gates [32, 67, 97] that causes deformation of the lateral potential by acting differently on each gate. The interest in this study was to analyze the influence of the spatial distortion of the lateral confinement on the atomiclike properties of the VQD. Evidence of the deformation of the confining potential (CP) was deduced from measurements of the source-drain current peak spacing, which arises from single-electron tunneling between Coulomb blockade regimes, as a function of the gate voltages. In particular, it was found that the peak spacings were more separated when fewer gates were operated, evidencing different CPs for different gate voltage sequences.

Figure 40 shows a schematic view of a 4GVQD with 0.8 μ side length similar to the device investigated by Tarucha et al. [32]. The material sequence of this QD is the same as for the cylindrical QD. The top view (Fig. 40) reveals the four independent Schottky gates A, B, C, D with corresponding voltages V_A, V_B, V_C, and V_D.

Figure 41 shows the source-drain current peak spacing as a function of the gate voltage V_G and the number of electrons, N, inside the dot for two different gate voltage sequences. Each peak, schematically represented by a vertical arrow, corresponds to an additional electron in the dot. The arrow orientation indicates the spin polarity. The peak height is arbitrary since our model does not compute the transport properties. For each sequence, the total spin S as a function of N in the dot is shown on the inset. In Figure 41a, the VQD is pinched off with the four gates at $V_G = -2.98$ V. Then, the four gates are interconnected and swept continuously from -2.98 to -2.73 V to charge 13 electrons in the dot (symmetric charging). If ΔV_G^i denotes the incremental gate voltage required to add the ith electron in the dot, the large gaps ΔV_G^3, ΔV_G^7, and ΔV_G^{13} reflect the shell structure of the orbitals proper to the quasi-parabolic confining potential of a symmetrical structure [11, 70, 81]. For $N = 4$ and 9, $S = 1$ and 3/2, respectively; that is, the dot is fully polarized at half occupied shells because these configurations are more favorable with electrons of parallel spins (exchange energy is maximized) and orthogonal orbitals (Hartree energy is minimized) [71]. In Figure 41b, the two opposite gates A and B are kept at constant voltage $V_A = V_B = -2.73$ V and V_C and V_D are swept backward until the dot is empty again at $V_G = -3.24$ V (asymmetric charging). We first notice that the gate voltage swing with two acting gates is almost exactly twice the gate voltage swing for four acting gates, reflecting the fact that the strength of four gates is twice that of two gates, which confirms the experiments [67]. This effect can also be noticed by observing that the separation between individual peaks

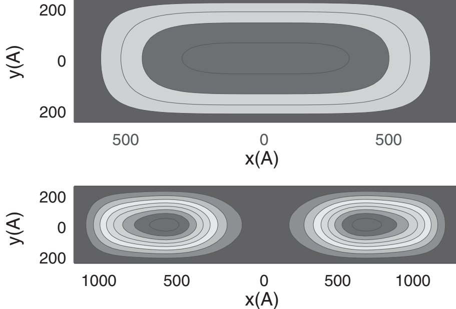


Figure 39. Electron density contour plot for two rectangular QDs with two electrons. Top: Highly confined QD (aspect ratio = 3) with electron delocalized in the QD. Bottom: Weakly confined rectangular QD (aspect ratio = 5) with electrons localized at each side of the dot.

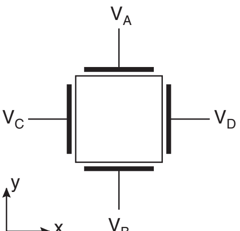


Figure 40. Top view of the VQD showing the four independant gates with applied voltages V_A, V_B, V_C, and V_D. Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002, American Institute of Physics.

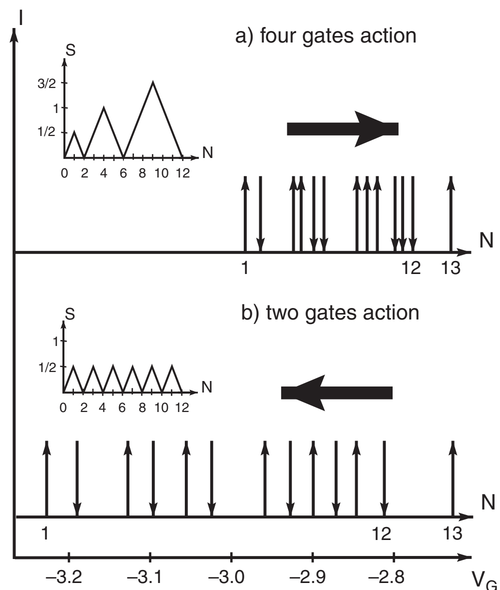


Figure 41. Source-drain current peak spacing as a function of the gate voltage V_G and the number of electrons N (a) when the four gates are swept uniformly, (b) when only V_C and V_D are swept while V_A and V_B are held at $V_G = -2.73$ V. Inset: total spin S of the QD as a function of N. Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002, American Institute of Physics.

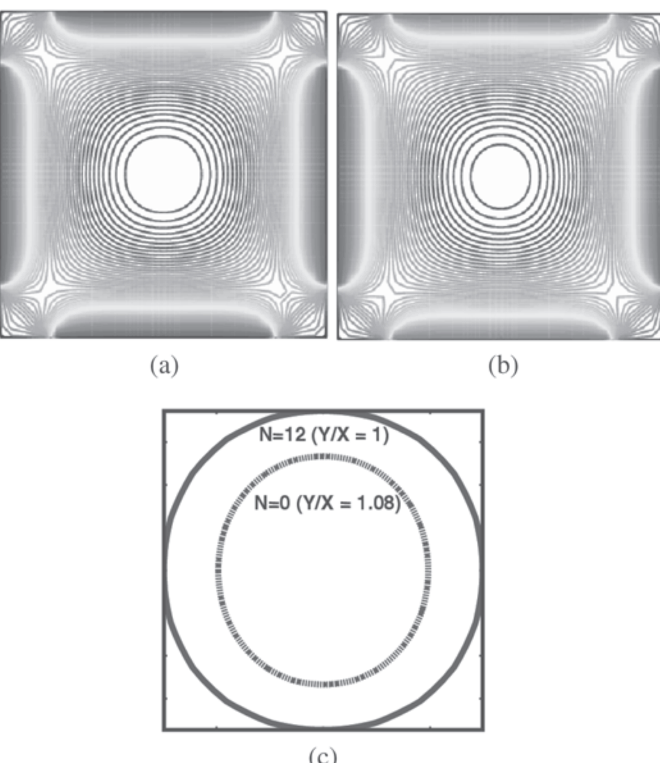


Figure 42. x–y potential contour plots (a) for $N = 12$ and $V_A = V_B = V_C = V_D = -2.8$ V, (b) for $N = 0$, $V_A = V_B = -2.73$ V, and $V_C = V_D = -3.24$ V; (c) potential contour at $\phi = E_F/q$ for $N = 12$ (solid line) and $N = 0$ (dashed line). Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002, American Institute of Physics.

is larger when only two gates are swept. In this respect, it is interesting to compare ΔV_G^j and ΔV_G^k, where the $j-1$th and jth electrons belong to the same shell, whereas the $k-1$th and kth belong to different shells. In Figure 41a and b, $\Delta V_G^k > \Delta V_G^j$ (i.e., the shell structure is preserved even in the asymmetric charging). However, the ratio $\Delta V_G^k/\Delta V_G^j$ is smaller in the asymmetric configuration which indicates that the shell structure is weaker when the CP loses symmetry. As seen in the inset of Figure 41b, the total spin of the dot reveals orbitals filling successively with electrons of antiparallel spins. Indeed, due to the CP deformation, the degeneracies inside a shell are gradually lifted and, at some point, sequential state filling with antiparallel spin electrons becomes more favorable because exchange between electrons does not compensate any more for the degeneracy lifting between occupied levels.

Figure 42 shows the equipotential contours in the x–y plane of the dot, for $N = 12$ at $V_G = -2.73$ V (Fig. 42a), when the dot is still uniformly biased, and for $N = 0$ and $V_C = V_D = -3.25$ V (Fig. 42b), after the asymmetric bias has been applied. In both cases, the very dense equipotentials on the four sides of the plots denote the proximity of the gates. In Figure 42a, the contours are circular at the center and become square at the periphery while in Figure 42b, they become slightly elliptic in the center, which reflects the fact that the electric field is now stronger in the x-direction than in the y-direction. The deformation is weak though, as shown in Figure 42c where the equipotential is drawn at the Fermi level before (solid line, $N = 12$) and after deformation (dashed line, $N = 0$). The area spanned by the Fermi contours reflects the effective size of the QD. Since the solid (dashed) contour corresponds to $N = 12$ ($N = 0$), it simply shows that the QD expands while filled with electrons. It is also worth noticing that, at $T = 0$ K, the electronic properties are mainly determined by the confining potential at the Fermi level, which explains why large square and circular QDs have almost identical electronic properties, as already reported by Kumar [98]. The aspect ratio of the dot after deformation is $\delta = 1.08$, for gate voltage swing ratio of $\Delta V_G = 0.49/0.25 = 1.96$, which shows that we cannot expect to obtain large deformations in electrostatic deformable dots as opposed to rectangular dots where δ, determined by the geometric aspect ratio, is arbitrarily fixed.

Figure 43 shows the contour plots of the first six eigenfunctions as a function of the number of electrons in the case of asymmetric voltage sweep. The upper-most row corresponds to $N = 12$, before deformation ($V_A = V_B = V_C = V_D = -2.73$ V), and the lower-most row corresponds to $N = 0$ ($V_A = V_B = -2.73$ V, $V_C = V_D = -3.24$ V), that is, the strongest deformation. Intermediate rows show the progressive transformation from circularly confined orbitals to elliptically confined orbitals as N varies from 12 to 0. Before deformation, the first orbital is a s-like state according to atomic physics and constitutes the first shell. The second and third orbitals are p-like and form the second shell. The fourth and fifth d-like orbitals and the sixth s-like orbital form the third shell. In general, all the orbitals are less spread out for $N = 0$ than for $N = 12$ because, as mentioned above, the dot has shrunk.

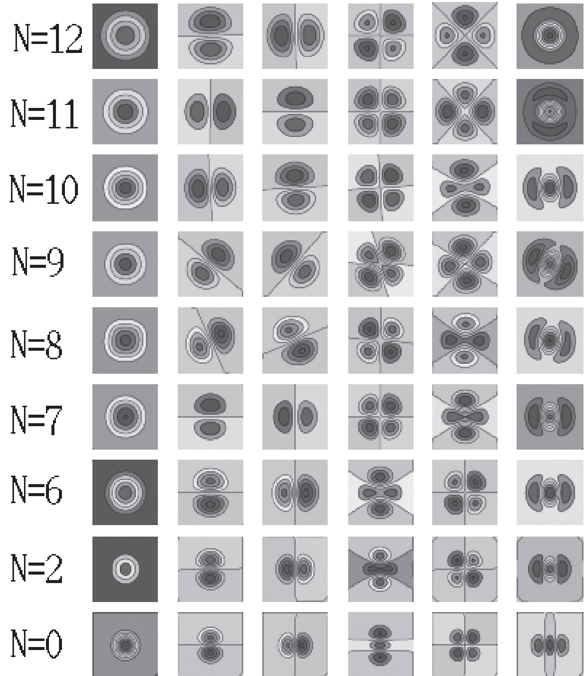


Figure 43. Wavefunction contour plots for the first three shells (first shell: first column, second shell: second and third column, third shell: fourth, fifth, and sixth column) of a 4GVQD. From top row to bottom row, N varies from 12 to 0; V_C and V_D are swept from −2.8 to −3.24 V. V_A and V_B are help at −2.73 V. Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002, American Institute of Physics.

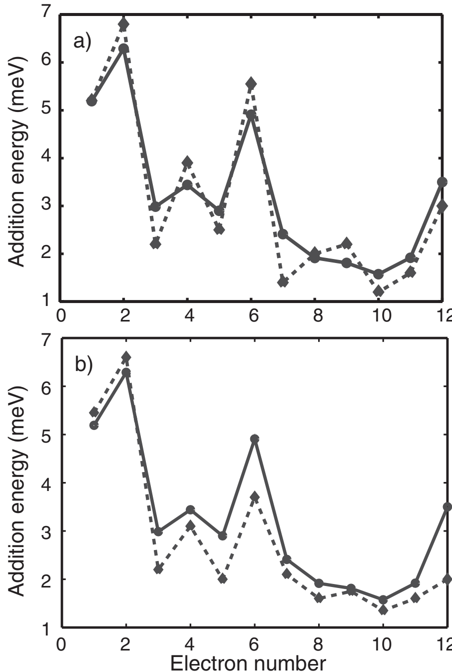


Figure 44. (a) Theoretical AE spectra of 4GVQD for the symmetric charging (dashed) and the asymmetric charging (solid) as a function of the number of electrons. (b) Comparison between computed AE spectrum (solid) of the asymmetric 4GVQD and the experimental spectrum [11] of a symmetric VQD (dashed) as a function of the number of electrons. Notice that, by definition, $E_a(N) = E_F(N+1) - E_F(N)$, so that a value for $N = j$ in this figure refers to the energy needed to add the $(j+1)$th electron in the dot. Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002, American Institute of Physics.

The first orbital is very weakly affected by the deformation. From $N = 12$ to $N = 8$, the two p-states do not show any particular favored orientation. Physically, as long as they are degenerate, all linear combinations of the original eigenfunctions are equally valid solutions. Numerically, the orientation of the two degenerate wavefunctions is mainly influenced by the mesh. On the contrary, from $N = 7$ to $N = 0$, V_A and V_B are now substantially different than V_C and V_D so that the states are no longer degenerate and align according to the symmetry axis of the dot. The third shell orbitals (last three columns) show the most spectacular transformations, that is, from circular state symmetry with principal and azimuthal quantum number $(n, l) = (0, 2)$, $(0, -2)$, and $(1, 0)$ to rectangular state symmetry with quantum numbers $(n_x, n_y) = (0, 2)$, $(1, 1)$, and $(2, 0)$ states, respectively. One also notices the crossing between the fourth and fifth eigenlevels at the transition $N = 7 \longrightarrow 6$.

Various addition energy spectra (AES) are shown in Figure 44. Figure 44a shows a comparison between the theoretical AES corresponding to the symmetric charging (dashed line) and the AES corresponding to the asymmetric charging (solid line). The symmetric and asymmetric AES have similar global features. First, there are large peaks at $N = 2, 6$, and 12 due to the shell structure, which reflects the energy contribution needed to lower the next orbital below the Fermi level for admitting the next electron, in addition to the energy required to overcome the electrostatic repulsion of the electrons already present in the dot. These peaks are approximately 0.5 meV lower in the asymmetric AES though, since the intershell spacing is reduced in this configuration. Second, there are secondary peaks at $N = 4$, which has been so far explained in term of Hund's rule (HR) [11, 70]: the third and fourth electrons, with parallel spins (↑), access empty p-like orbitals, maximize E_{xc}, and minimize E_H, inducing a minimum for $N = 3$. For $N = 4$, the fifth electron with antiparallel spin sits on an already occupied state, which increases E_H. Moreover, this electron is the only ↓-electron in the second shell, with no additional exchange interaction; thus, $E_a(4) > E_a(3)$. For $N = 5$, the sixth electron sits on an already occupied orbital sharing exchange interaction with the fifth electron, which lowers $E_a(5) < E(4)$, leading to a peak for $N = 4$. This explanation applies, of course, to the symmetric AES (solid curve, Fig. 44a), which demonstrates that HR is a necessary condition for observing a peak at $N = 4$. However, it is not a sufficient condition since the asymmetric AES (dashed curve, Fig. 44a) also exhibits a peak at $N = 4$, and, as shown before, HR is not fulfilled for asymmetric charging (inset of Fig. 41b). The peak is weaker, however, since there is no exchange between the third (fifth) and four (sixth) which increases the addition energy for $N = 3$ and $N = 5$. Moreover, the fifth electron sits now on an unoccupied orbital, which reduces the addition energy for $N = 4$. For the symmetric charging, there is also a secondary peak for $N = 9$ which, with a dip at $N = 7$, reflects three parallel spins in the third shell

[71, 99]. This peak is not present in the asymmetric AES, which shows that, due to symmetry breaking, levels are more randomly distributed inside the third shell. In summary, secondary peaks in the energy spectrum are not necessarily the signature of parallel spin alignment.

Figure 44b shows a comparison between the theoretical asymmetric charging (solid) and the experimental AES of a circular device [11]. The striking feature here is that the asymmetric AES follows very closely the experimental AES of the cylindrical structure. The asymmetric AES is actually closer to the experimental AES than the symmetric AES, in particular for the third shell charging. This agreement is not fortuitous since it was recently demonstrated that the peak structure in the experimental curve [11] reflects a sequence of alternate spins in the filling of the third shell [71], similar to the sequence achieved in the theoretical asymmetric charging. This observation leads us to conclude that real VQDs are probably never completely symmetric.

Based on the observations above, and on the fact that the total energy of the dot is spin dependent, we investigate the possibility of changing the spin polarization by deforming the dot electrostatically (Fig. 45). Here, we assume that the first shell is completely filled, and we exclusively focus on filling the second shell, which consists of two degenerate p-like states, each of them being twofold spin degenerate. Let us call them $\psi_x^\uparrow$, $\psi_x^\downarrow$, $\psi_y^\uparrow$, $\psi_y^\downarrow$ with eigenvalues $\varepsilon_x^\uparrow$, $\varepsilon_x^\downarrow$, $\varepsilon_y^\uparrow$, $\varepsilon_y^\downarrow$, respectively. Let us assume that, with $N = 3$, $\varepsilon_x^\uparrow$ is occupied. Now, let us increase the four gates positively until the fourth electron enters the dot. By using the same argument as above, as long as the orbitals are degenerate, the energetically most favorable configuration for $N = 4$ is achieved with the occupation of $\varepsilon_y^\uparrow$ with the fourth electron, as predicted by HR. As soon as the fourth electron is in the dot, $V_C = V_D \equiv V_Y$ is made more negative and $V_A = V_B \equiv V_X$ is made more positive to confine electrons stronger along the x- than along y-direction, while keeping $N = 4$ constant. Figure 45 shows the evolution of $\varepsilon_x^\uparrow$, $\varepsilon_y^\uparrow$, and $\varepsilon_y^\downarrow$ with respect to $\Delta V_G \equiv V_X - V_Y$. When $\Delta V_G = 0$, the dot is spatially symmetric; hence, $\varepsilon_x^\uparrow = \varepsilon_y^\uparrow$ and $\varepsilon_x^\downarrow = \varepsilon_y^\downarrow$. The dot is spin polarized, however, and the exchange energy between $\uparrow$-electrons lowers $\varepsilon_x^\uparrow$ and $\varepsilon_y^\uparrow$ with respect to $\varepsilon_x^\downarrow$ and $\varepsilon_y^\downarrow$. Another interpretation of the gap between $\varepsilon_x^\uparrow$, $\varepsilon_y^\uparrow$ and $\varepsilon_x^\downarrow$, $\varepsilon_y^\downarrow$ is given by the Coulomb blockade that should be overcome to populate the empty orbitals. Let us mention that $\varepsilon_x^\downarrow$, which is equal to $\varepsilon_y^\downarrow$ for $\Delta V_G = 0$ and runs parallel to $\varepsilon_x^\uparrow$, is not shown for clarity, since it is never occupied. As V_X (V_Y) is made more positive (negative), the confining potential along y (x) is weaker (stronger), ε_y (ε_x) are shifted down (up), and the energy separation $\varepsilon_x^\uparrow - \varepsilon_y^\uparrow$ grows linearly. It must be noticed that the rate of decrease of ε_y is larger than the rate of increase of ε_x because ΔV_Y should be smaller than ΔV_X in order to maintain four electrons inside the dot. Otherwise, if $\Delta V_Y = \Delta V_X$, the three-electron configuration is more favorable. For $V_G < V_G^S$, $\varepsilon_x^\uparrow$ is smaller than $\varepsilon_y^\downarrow$, and the parallel spin configuration, schematically shown in inset (a) of Figure 45, is favored with respect to the configuration shown in inset (b), because the attractive exchange interaction between electrons sitting on $\psi_x^\uparrow$ and $\psi_y^\uparrow$ is larger than the energy separation $\varepsilon_x^\uparrow - \varepsilon_y^\uparrow$. At $V_G = V_G^S$, $\varepsilon_x^\uparrow - \varepsilon_y^\uparrow = 0.15$ meV exactly offsets the exchange energy between the two parallel spin electrons with wave functions $\psi_x^\uparrow$ and $\psi_y^\uparrow$. For $V_G > V_G^S$, $\varepsilon_x^\uparrow$ is larger than $\varepsilon_y^\downarrow$; that is, the exchange interaction between parallel spin electrons is no longer capable of overcoming the energy separation $\varepsilon_x^\uparrow - \varepsilon_y^\uparrow$, and the $\psi_y^\downarrow$-state becomes energetically more favorable than the $\psi_x^\uparrow$-state. Thus, through electrostatic deformation, it is possible to control the spin polarization of a 4VGQD.

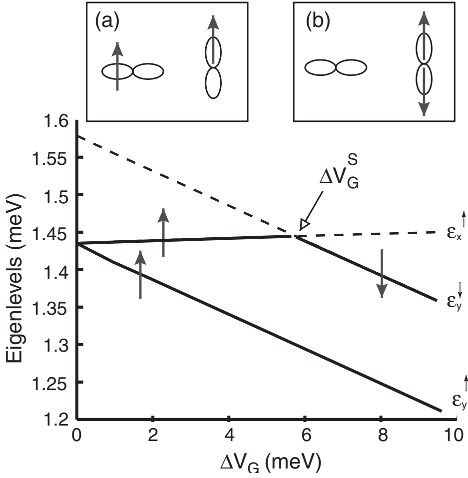


Figure 45. Variation of the single-particle eigenlevels of the second shell $\epsilon_x^\uparrow$, $\epsilon_y^\uparrow$, and $\epsilon_y^\downarrow$ as a function of $\Delta V_G \equiv V_X - V_Y$ ($X = A, B$; $Y = C, D$). A solid (dashed) eigenlevel is connected to an occupied (empty) orbital. For $\Delta V_G < \Delta V_G^S = 5.8$ meV, two parallel electron spins sit on the orbitals associated with $\epsilon_x^\uparrow$ and $\epsilon_y^\uparrow$ so that the favored configuration is shown in the top left inset (i.e., a triplet state). For $\Delta V_G > \Delta V_G^S$, two antiparallel electron spins sit on the orbitals associated with $\epsilon_y^\uparrow$ and $\epsilon_y^\downarrow$ so that the favored configuration is shown on the top right inset (i.e., a singlet state) ($S = 0$). Reprinted with permission from [113], P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002). © 2002. American Institute of Physics.

9. DOUBLE QUANTUM DOTS—ARTIFICIAL MOLECULES

While a single QD displays features characteristic of an atom, an array of QDs coupled through tunnel junctions shows many similarities to molecules and is therefore also called an *artificial molecule*. Just as in molecules, electron states (in coupled QDs) can couple to form covalent states that are delocalized over the entire array, making it possible for an occupying electron to tunnel between the various dots without being localized to any [100, 101]. These *bonding states* are lower in energy than the constituent dot states by an amount that is equivalent to the binding energy of the molecule. For instance, a two-dot system may be compared to a diatomic molecule. Such artificial molecules provide an advantage in that the number of electrons in the coupled dot, equivalently the constituent "atoms" in the periodic table, may be varied by varying a external potential. Different molecular analogs can be realized by varying the size of the dots, simultaneously or independently, and their number of electrons. Furthermore, in the case of coupled PQDs the vibrational motion of a molecule may be simulated by driving such an array between weak and strong tunneling regimes.

Most of the experimental [101–103] and theoretical studies [104–109] on coupled dots have so far been on linear arrays of PQDs, with the exception of some recent work on vertically coupled SADs [75, 93]. In particular, Ruzin et al. [106] studied the Coulomb blockade structure for two nonidentical dots using the activation-energy approach. Stafford and Sharma [107] and Klimeck et al. [108] have used a Mott–Hubbard approach with and without interdot capacitances to determine the many-body wavefunction for an array of dots. More recently Golden and Harpen [109] have studied the problem of Coulomb blockade peak splitting in the weak and strong coupling limits to explain the experimental data in [110]. Partoens and Peeters [75] studied the ground state of two vertically coupled QDs as a function of the interdot distance to explain an experiment performed by Tarucha et al. [64]. Here, we will focus on a planar double quantum dot (DQD) structure because this configuration provides an easy way to control the coupling between the dots electrostatically [111].

9.1. Double Quantum Dot Structure

The schematic of the double dot is shown in Figure 46. It consists of an inverted GaAs–$Al_{0.3}Ga_{0.7}As$ heterostructure which confines the electrons to a 2D gas at the interface. In our model, the simulated structure consists of a 22.5 nm layer of undoped $Al_{0.3}Ga_{0.7}As$ followed by a 125 nm layer of undoped GaAs, and finally an 18 nm GaAs cap layer. The cap layer is uniformly doped to 5×10^{18} cm^{-3} so that the conduction band edge is just above the Fermi level at the GaAs cap layer–undoped GaAs boundary. The inverted heterostructure is grown on a GaAs substrate and charge control is achieved by varying the voltage on the back gate V_{back}. We assume a negligible voltage drop across the substrate; hence, we apply V_{back} directly to the bottom of the $Al_{0.3}Ga_{0.7}As$ layer. The two dots are defined by energizing the 10 metallic gates shown in Figure 46b, with the coupling between them varied by means of the voltage V_t on the tuning gates. Electron charging of the two dots is possible through tunnel injection from the adjacent two-dimensional regions through the 35 nm opening between the opposite stubs. The tuning gates, besides demarcating the two dots, also control the extent to which the two dots are coupled.

Figure 47a shows the total potential for the empty double-dot in the plane of the heterointerface for $V_t = -0.67$ V obtained by LSDA-based computer simulations. The two dot regions are visible as depressions in the region 4000 Å $\leq x \leq$ 8000 Å and 2500 Å $\leq y \leq$ 4200 Å. The potentials in the dots are parabolic at low energy as seen more clearly in Figure 47b, which shows the potential along $x-$, the direction of coupling of the two dots for two values of V_t. For $V_t = -0.67$ V, hereafter referred to as the weak-coupling regime, the interdot barrier (Δ_{67} in Fig. 47) is 4 meV, while for $V_t = -0.60$ V (i.e., the strong coupling regime), the

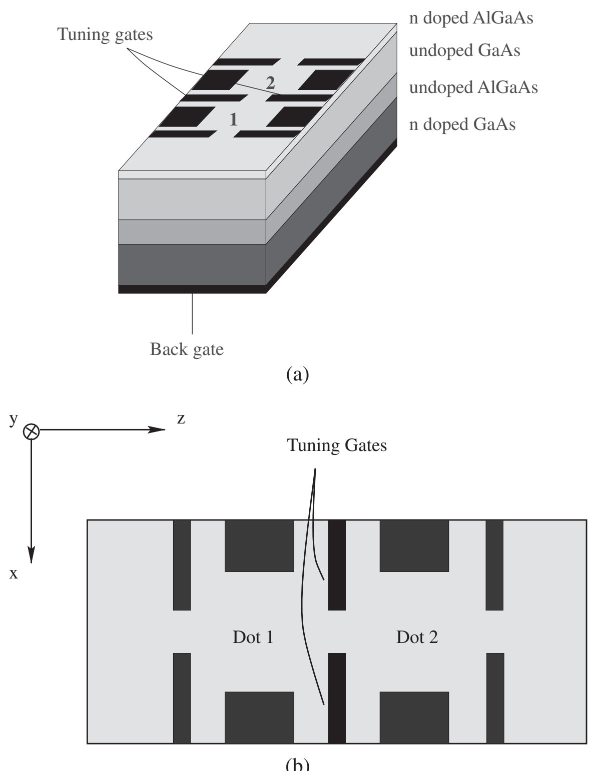


Figure 46. The double dot system: (a) layer structure of the double dot, and (b) schematic representation along the x–y plane. The gate stub voltage is −1.2 V and the gate pad voltage is −0.40 V. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

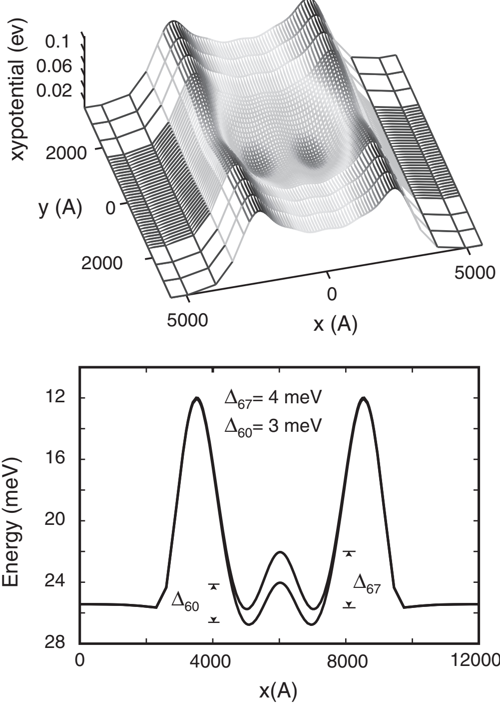


Figure 47. Potential profile for the empty double dot (a) in the plane of the heterointerface, and (b) along y in the weak (dashed line), $V_t = -0.67$ V, and strong (solid line), $V_t = -0.60$ V, coupling regimes. The corresponding interdot barriers, Δ_{67} and Δ_{60}, are seen to be 4 and 3 meV, respectively. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

interdot barrier (Δ_{60} in Fig. 47) is reduced to 3 meV. The entire double-dot is bounded by 14 meV tunnel barriers at the source and drain ends. Confinement along the vertical y-direction is provided by a combination of the band offset of 255 meV between the GaAs and $Al_{0.3}Ga_{0.7}As$ and a large vertical electrostatic field, $F_y \simeq 35$ kV/cm, in GaAs. This results in the separation between energy levels along y of 30–40 meV, much greater than that in the x–y plane, which is of the order of 1 meV. This large level separation results in only the ground state (along the z-direction) being occupied at low temperatures. All the results present correspond to $T = 0.25$ K.

9.2. Electronic Structure and Charging Effect in Double Quantum Dots

Figure 48 shows the schematic of the lowest four states with their wavefunctions in the double dot for $V_t = -0.67$ V and $V_t = -0.60$ V. For both values of V_t the ground states in the individual dots are 1s type and form a degenerate pair. However, the first excited states, which are p_x- and p_y-like, are degenerate for weak coupling (i.e., $V_t = -0.67$ V), whereas for strong coupling ($V_t = -0.60$ V), the p_x-like states couple to form symmetric (bonding) and antisymmetric (antibonding) states which are lower in energy than the p_y-like states as seen in Figure 48. This reordering of the states has an important bearing on the spin polarization of the double-dot system, as shall be fully explained later.

9.2.1. Weak Coupling

Figure 49a shows the *Coulomb staircase* indicating the variation of the number of electrons in the dot with V_{back} at $V_t = -0.67$ V. For $V_{\text{back}} = 0.9769$ V, $\epsilon_{\text{LAO}}(0.5)$ is just negative, implying that the dot can accept one electron in the lowermost 1s state. At this point no distinction is made as to whether Dot 1 or Dot 2 or both are occupied since, in our model, the two dots are identical from the electrostatic and quantum mechanical points of view. The dots are decoupled quantum mechanically, since the leakage of the 1s wavefunctions into the adjacent dot is negligible, and electrostatically, since the distance between the two electron distributions is large, and the resulting Coulomb repulsion, screened by the high dielectric constant of GaAs ($\epsilon_{\text{GaAs}} = 13.2$), is negligible, or at least weak enough to be overcome by less than 0.1 mV change in V_{back} (which is the minimum increment in V_{back} considered in this work). Thus both dots can be charged simultaneously with an electron each, resulting in N jumping from zero to two. Also, the orientation of the individual spins of the two electrons does not matter; we choose both electrons to be ↑. This simultaneous (double) charging persists as long as the two dots are isolated. However, as V_{back} is increased to 0.9804 V, when the next charge degeneracy point occurs, even though Dots 1 and 2 are identical in all respects, only one of them can be charged (with a ↓ electron), but not both. This is due to the fact that charging, say, Dot 1 first increases the total Coulomb repulsion experienced by the incoming electron to Dot 2; that is, electrostatic coupling is established between the dots prior to any appreciable quantum mechanical coupling. Overcoming this repulsion requires a 0.1 mV increment in V_{back}, resulting in the termination of double charging, and is evident as a narrow step for $N = 3$. At $V_{\text{back}} = 0.9805$ V, Dot 2 also can be charged with a ↓ electron increasing N to 4. At this point the double-dot is spin unpolarized and the 1s ground states of both the dots are completely filled with two electrons (↑↓) each. The next available states for occupation are the p_x and p_y states in each of the dots which are above the ground state by $\simeq$1 meV. These states are almost degenerate; the slight splitting of the states caused by the nonisotropicity of the two-dimensional parabolic confining potential in the x–y plane is very small ($\simeq 10^{-6}$ eV). Though

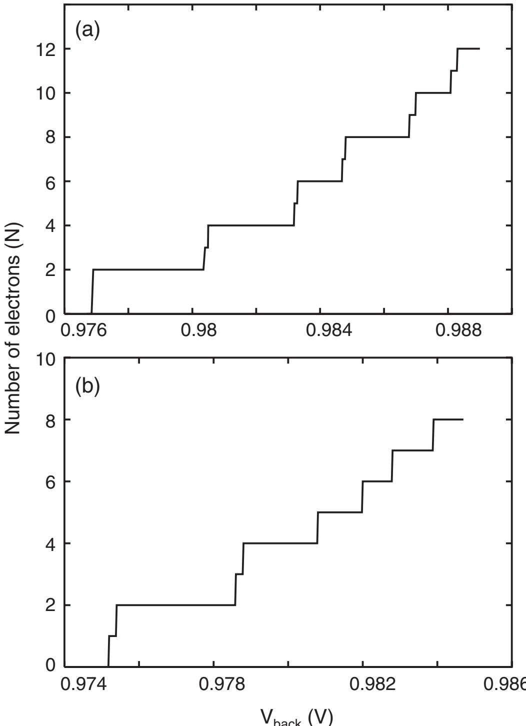


Figure 49. Coulomb staircase diagram for (a) $V_t = -0.67$ V, where the transitions that do not follow Hund's rules are shown in dashed lines, and (b) $V_t = -0.60$ V. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

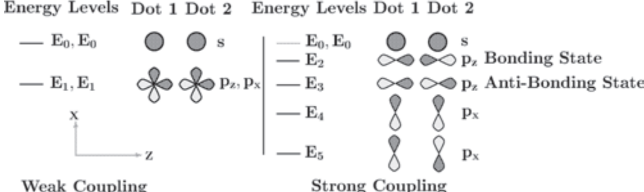


Figure 48. Schematic representation of the lowest six states in the empty double dot in the weak ($V_t = -0.67$ V) and the strong ($V_t = -0.60$ V) interdot coupling regimes. In the former the p_z- and p_x-like states are shown together since they are degenerate within each dot and are decoupled from the corresponding states in the other dot. The reordering of these single-particle states due to increased interdot coupling is also seen. The shaded areas denote the positive portion of the wavefunctions. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

either of the two states could be occupied we choose to occupy the p_x state since it is slightly lower in energy than the p_y state by $\simeq 10^{-6}$ eV. Even though the leakage of the p_x state into the interdot barrier is greater than that of the 1s state, it is still not coupled to the corresponding p_x state of the adjacent dot to form a bonding state. Consequently, charging of the two p_x states proceeds in the same manner as done for the $N = 2$ to $N = 3$ and $N = 3$ to $N = 4$ electrons. The charging of the dot with the fifth electron occurs at $V_{\text{back}} = 0.9832$ V, where we choose to occupy the p_x state of Dot 1 with an $\uparrow$ electron. Subsequent occupation of the p_x state of Dot 2 with an $\uparrow$ electron ($N = 5$ to $N = 6$ transition) occurs at $V_{\text{back}} = 0.9833$ V, the 0.1 mV increment in V_{back} being required to overcome the Coulomb repulsion due to the fifth electron. It must be noted that the sixth electron occupying p_x state in Dot 2 is also of $\uparrow$ spin, which makes the double-dot spin polarized. This configuration is energetically favorable due to the attractive nature of the exchange interaction between the two $\uparrow$ electrons in the p_x states of Dots 1 and 2. The unpolarized configuration, where the sixth electron (in p_x state of Dot 2) is of spin $\downarrow$ is less favourable and occurs at a slightly higher value V_{back} (0.98335 V) (dashed curve in Fig. 49a). Likewise, occupying the p_y state of Dot 1 with a spin $\uparrow$ electron, thereby increasing N to 7, occurs at a lower voltage than occupying the p_y state of Dot 1 with a $\downarrow$ electron or completely filling the p_x state of Dot 1. This is again a demonstration of Hund's first rule. Similarly, the eighth electron of $\uparrow$ spin occupies the p_y state of Dot 2 at $V_{\text{back}} = 0.9848$ V, requiring a 0.1 mV increment in V_{back} after the p_y state in Dot 1 has been occupied. As before, this 0.1 mV increase in V_{back} is required to overcome the Coulomb repulsion due to the p_y electron in Dot 1. The double-dot at this state is spin polarized with six $\uparrow$ and two $\downarrow$ electrons. The ninth through twelfth electrons (of $\downarrow$ spins) now complete the partially occupied p_y and p_x states in Dots 1 and 2, at $V_{\text{back}} = 0.9868$, 0.9870, 0.9881, and 0.9883 V, thereby reverting the double-dot to a zero spin state. A feature that is conspicous over the $N = 9$ to $N = 12$ range is the increasing interaction between the electrons occupying similar single-particle states in the two dots, as evident in the 0.2 mV increment in V_{back} for $N = 9 \rightarrow N = 10$ ($\downarrow$ electrons in the p_x states of Dots 1 and 2) and $N = 11 \rightarrow N = 12$ transitions, compared to the 0.1 mV increment for the $N = 5 \rightarrow N = 6$ and $N = 7 \rightarrow N = 8$ transitions. This is a consequence of increasing leakage of the wavefunctions into the adjacent dot and the center of the charge distributions in the two dots moving toward each other. Experimental investigations of Waugh et al. [110] on arrays of two and three coupled dots confirm our findings.

9.2.2. Strong Coupling

Figure 49b shows the Coulomb staircase diagram for $V_t = -0.60$ V. The interdot coupling in this case is stronger than in the previous one ($V_t = -0.67$ V), which enhances the leakage of the 1s states into the interdot barrier, thereby increasing quantum mechanical as well as electrostatic couplings between the dots. The increase in electrostatic coupling between the dots (quantum mechanical coupling, although present, is still negligible for s states), expectedly, terminates double charging and charging proceeds sequentially from $N = 1$ ($V_{\text{back}} = 0.9752$ V) to $N = 2$ ($V_{\text{back}} = 0.9754$ V). The increase in interdot electrostatic interaction is evident from the 0.2 mV increment required to charge Dot 2 with a $\downarrow$ electron after Dot 1 has been charged with an $\uparrow$ electron. This increment in V_{back} as may be observed from Figure 49a is more than the interaction between the third and fourth electrons that go to complete the 1s states in Dots 1 and 2 for $V_t = -0.67$ V. The $N = 2 \rightarrow N = 3$ transition occurs at $V_{\text{back}} = 0.7586$ V and the $N = 3 \rightarrow N = 4$ transition at $V_{\text{back}} = 0.9788$ V. The four electrons occupy the two 1s states in Dots 1 and 2. The p_x states which are the lowest available for occupation couple strongly to form symmetric and antisymmetric states akin to *bonding* and *antibonding* states in molecules. A consequence of strong interdot coupling is that for states including p_x, and for higher states, the double-dot appears as a single dot about twice as long along the x-direction as along the y-direction. These lead to a reordering of states in the eigenenergy spectrum, as the states with increasing n_x (number of nodes along x-direction) get closer and move below the p_y state. Thus, the fifth through eighth electrons occupy sequentially the p_x bonding and antibonding states. This is to be contrasted with the partial filling of the p_y and p_x states, for $V_t = -0.67$ V, by the fifth through eighth electrons, which create a spin polarized state with a net polarization $S = 2\hbar$. It is thus clear that an increase in the interdot coupling drives the double-dot from a spin polarized to an unpolarized state. Also, the Coulomb staircase steps assume a more uniform width for $N > 4$ as charging proceeds similar to a single QD.

A complete transformation of the two-dot system into a single large dot requires a de-energizing of the tuning gates (i.e., $V_t = 0$ V), a case that has not been considered in this work. However, from the cases considered so far, it can be extrapolated that the separation of the peaks would be maximum and equal to e/C_T, where C_T is the capacitance of the single large dot. This corresponds to the case in Figure 2d in the work of Waugh et al. [110].

Figure 50 shows the variation of the spin $\uparrow$ and $\downarrow$ electron densities at the GaAs/$Al_{0.27}Ga_{0.73}As$ interface for $V_t = -0.60$ V, along the x-direction for $N = 3, 5$, and 8. Figure 50a shows the variation for $N = 3$, a spin polarized state with two $\uparrow$ and one $\downarrow$ electrons. The two $\uparrow$ electrons occupy the 1s states in Dots 1 and 2, respectively, while the $\downarrow$ electron occupies the 1s state in Dot 1. Consequently, the two electron densities have a 1s type distribution in the two dots. Since the two dots are identical the electron distributions in them are identical. In reality, however, a very small voltage is applied between the drain and source to facilitate electron tunneling injection into the two QDs. This does result in a breaking of symmetry between the two dots and might lead to a small differences between the two electron distributions. The electron densities in the interdot barrier region ($z \simeq 6000$ Å) are small as the coupling of the 1s states is weak for this value of V_t. However, as N increases the stronger electron–electron interaction induces a lowering of the barrier, an increase in the leakage of the wavefunctions, and an increase in the electron densities in the barrier. This is seen in Figure 50b for $N = 5$; the dot is still spin polarized with three $\uparrow$ and two $\downarrow$ electrons. The two $\downarrow$ and two $\uparrow$ electrons occupy the 1s states of Dots 1 and 2, respectively, while the third $\downarrow$ electron occupies the p_z bonding state that gives the jagged profile to the $\downarrow$ electron concentration, as

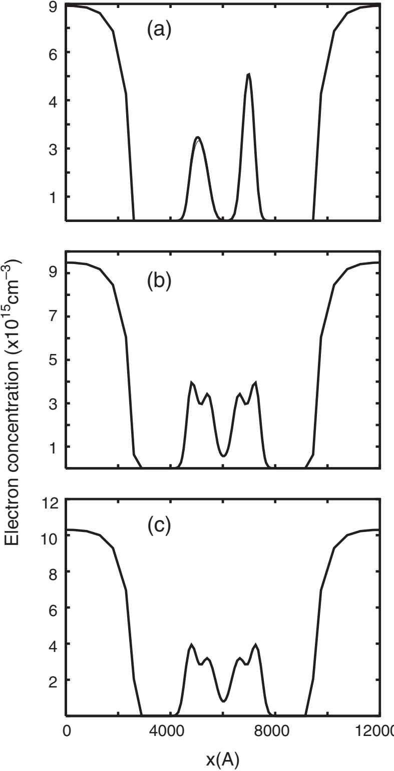


Figure 50. The variation of the electron spin ↑ (solid) and the ↓ (dashed) electron densities at the GaAs/$Al_{0.27}Ga_{0.73}$ interface along the length of the double dot for (a) $N = 3$, $V_{back} = 0.9786$ V, (b) $N = 5$, $V_{back} = 0.9808$ V, and (c) $N = 8$, $V_{back} = 0.9839$ V. Resolution of spin is confined to the dot region only (i.e., $3000 \leq z \leq 9000$ Å). The tuning voltage $V_t = -0.60$ V. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

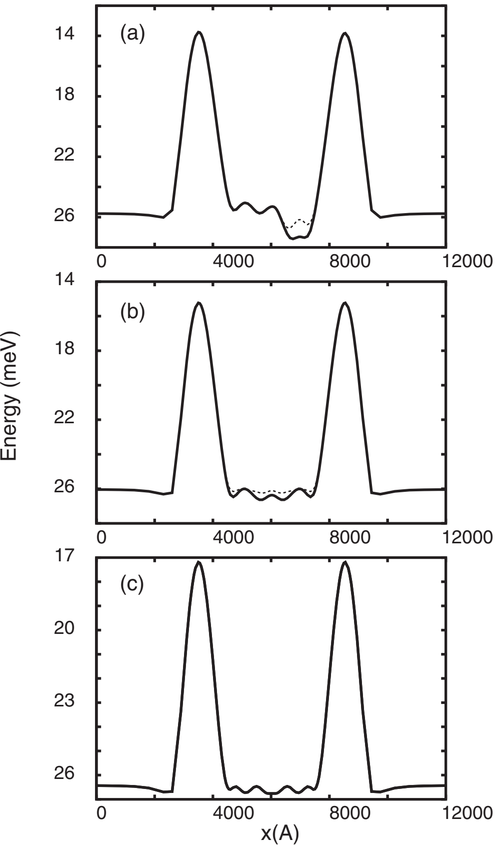


Figure 51. The variation, along $z-$, of the effective potential energy at the GaAs/$Al_{0.27}Ga_{0.73}$As interface for the spin ↑ (solid) and the ↓ (dashed) electrons for (a) $N = 3$, $V_{back} = 0.9786$ V, (b) $N = 5$, $V_{back} = 0.9808$ V, and (c) $N = 8$, $V_{back} = 0.9839$ V. The tuning voltage $V_t = -0.60$ V. Reprinted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

seen in the figure. Figure 50c shows the variation for $N = 8$. This is a spin unpolarized state with four ↑ and four ↓ electrons. These fill completely the two 1s states in Dots A and B and the p_x state. Hence, the two (↑ and ↓) distributions look identical.

It is also seen in Figure 50 that the peak electron densities decrease with increasing N. This apparent anomaly can be explained by noting that as more electrons are added to the dot, the volume of the dot increases (in the plane of the interface where the confinement is the weakest). Consequently, the electron density spreads out in the x–y plane while the height of the distribution decreases. This is manifest in the figure as an increase in the width electron distribution.

Figure 51 shows the variation of the effective potential energy for the up and down spin electrons at the GaAs/$Al_{0.27}Ga_{0.73}$As interface for $V_t = -0.60$ V, along the x-direction, for $N = 3$, 5, and 8. The profile for $N = 0$ is similar to that shown in Figure 47 but the height of the interdot barrier is lower at 3 meV. In a spin polarized state the effective potential for ↑ electrons is different from that for the ↓ electrons, since $\mu_{xc}(\mathbf{r})$ for each of them is different. In the $N = 3$ case shown in Figure 51a the higher ↑ density leads to a larger $\mu_{xc}^{\uparrow}(\mathbf{r})$, which, being attractive, lowers the net electron–electron interaction energy. This also leads to the interdot barrier seen by the ↑ electrons to be slightly higher than that seen by the ↓ ones. When N increases to 5, as in Figure 51b, the large ↑ density in the barrier region and the resulting $\mu_{xc}^{\uparrow}(\mathbf{r})$ leads to a drastic reduction in the barrier height. The potential energy profile for the ↑ electrons then resembles a single large dot, instead of a coupled dot. In contrast, for the ↓ electrons whose density in the barrier is small, the potential energy is characteristic of a coupled-dot system. With a further increase in N to 8 (four each of ↑ and ↓ electrons), as in Figure 51c, the profiles for the ↑ and ↓ electrons are identical. The interdot barrier separation is extremely small and is barely noticeable in the figure. A feature that is evident in Figure 51, typical of electrostatically defined dots, is the decrease in confinement with increasing N. The barrier at the source and drain ends decreases from $\simeq$15 meV for $N = 0$ to about 9 meV for $N = 8$.

The variation of interdot coupling by varying V_t is comparable to the vibration of a diatomic molecule. Weak coupling (a large negative V_t) is comparable to the situation wherein the distance between the atoms is large, and strong coupling when the distance is a minimum. Thus by varying V_t the level crossing as a function of separation between the artificial atoms in the molecule can be studied. From the Coulomb staircase diagrams of Figure 49a and b and the ensuing discussions it is seen that a lowering of the interdot

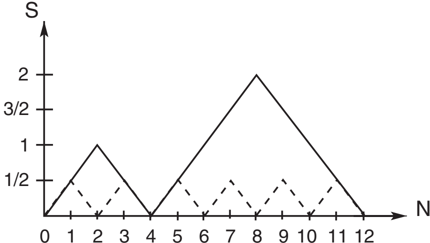


Figure 52. Variation of the total electron spin S in the double dot with N for $V_t = -0.67$ V (solid line), $V_t = -0.60$ V (dashed line). Adapted with permission from [111], S. Nagaraja et al., *Phys. Rev. B* 60, 8759 (1999). © 1999, American Institute of Physics.

barrier results in a reordering of the single-particle levels, thereby transforming the double-dot (for $N = 8$) from a spin polarized to an unpolarized state. The variation of total spin with the number of electrons in the dot in the weak ($V_t = -0.67$ V) and strong ($V_t = -0.60$ V) coupling regimes is illustrated in Figure 52. For weak coupling the total spin of the dot increases in steps of $\hbar/2$ between $N = 5$ and $N = 8$ as the p_x and p_z states in Dots 1 and 2 get charged with an electron each of $\uparrow$ spin, which is an illustration of the familiar Hund's rule in atomic shell filling. For strong coupling such a spin polarization of the dot is precluded by a lifting of degeneracy of the p_x and p_y states, and the total spin is never greater than $\hbar/2$. We would like to draw attention to the fact that for $V_t = -0.67$ V the total spin is seen to be $\hbar$ for $N = 2$ (Fig. 52). This, however, does not imply that the double dot is spin polarized, but rather that since the dots are electrostatically decoupled the spins of the two electrons are not correlated in any way. Though in reality there may be a weak interaction between the two electrons, such an interaction is not observable within the resolution of our model, and our choice of the two electrons being of parallel spin is incidental.

GLOSSARY

Addition energy Amount of energy required to overcome Coulomb blockade and add an electron in a quantum dot.

Artificial molecule Coupled quantum dot structures that exhibit molecular properties due to the quantum nature of the electronic motion.

Coulomb blockade Suppression of conductance through a QD due to electrostatic interactions between electrons.

Quantum dot (QD) Electronic structure in which electrons are completely confined in the three directions of space. Confinement can be electrostatic or structural. When the size of the confining region is comparable to the electron de Broglie wave length, QDs exhibit full quantum behavior: three dimensional energy quantization with orbital motion, shell structure, typical to atoms. For these reasons, QDs are often called artificial atoms.

Quantum well Material structure consisting of heterojunctions that creates a potential well in which electrons are trapped, so their motion is restricted onto a plane parallel to the interface of the two materials.

Quantum wire Material structure in which electrons are confined in two dimensions of space so their motion is restricted to move along a line.

Single electron charging Ability to control the charging of an ultra-small capacitor by one electron at a time due to Coulomb blockade.

Single electron devices Devices based on Coulomb blockade in conjunction with tunneling of single-electron from externally controlled reservoirs.

Single-electron transistor Three-terminal quantum dot device based on the Coulomb blockade effect, in which single electron transfer from a reservoir called the source to another reservoir called the drain is executed through two series tunnel junctions and controlled by a electric gate.

ACKNOWLEDGMENTS

This work was supported by NEDO and NSF Grant DESCARTES ECS-98-02730. P.M. gratefully acknowledges the support of the Beckman Institute. We are indebted to Professor S. Tarucha, Professor R. Martin, and Dr. G. Austing for valuable discussions. We are also indebted to Dr. S. Nagaraja for advice and technical assistance.

REFERENCES

1. K. Hess and G. Iafrate, *IEEE Spectrum* 44–49 (1992).
2. E. Corcoran, *Sci. Am.* 122–131 (1990).
3. K. K. Likharev and T. Claeson, *Sci. Am.* 80–85 (1992).
4. C. Gorter, *Physica* 15, 777 (1951).
5. I. Kulik and R. Shekhter, *Sov. Phys. JETP* 41, 308 (1975).
6. D. Averin and K. K. Likharev, *J. Low Temp. Phys.* 62, 345 (1986).
7. U. Meirav, M. A. Kastner, and S. Wind, *Phys. Rev. Lett.* 65, 771 (1990).
8. R. C. Ashoori, H. L. Störmer, J. Weiner, L. Pfeiffer, K. Baldwin, and K. West, *Surf. Sci.* 305, 558 (1994).
9. S. Tarucha, D. G. Austing, and T. Honda, *Japan. J. Appl. Phys.* 37, 3917 (1997).
10. M. A. Kastner, *Phys. Today* 24 (1993).
11. S. Tarucha, D. G. Austing, T. Honda, R. J. van der Hage, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 77, 3613 (1996).
12. K. Imamura, Y. Sugiyama, Y. Nakata, S. Muto, and N. Yokoyama, *Japan. J. Appl. Phys.* 34, 1445 (1995).
13. C. Wasshuber, H. Kosina, and S. Selberherr, *IEEE Trans. Electron Devices* 45, 2365 (1998).
14. K. K. Likharev, *FED* 6, 5 (1995).
15. J. R. Tucker, *J. Appl. Phys.* 72, 4399 (1992).
16. J. L. Jimenez, L. R. Fonseca, D. Brady, J.-P. Leburton, D. R. Wohlert, and K. Cheng, *Appl. Phys. Lett.* 71, 3558 (1997).
17. G. Burkard and D. P. DiVincenzo, *Phys. Rev. A* 57, 120 (1998).
18. G. Burkard, D. Loss, and D. P. DiVincenzo, *Phys. Rev. B* 59, 2070 (1998).
19. C. Smith, M. Pepper, H. Ahmed, J. Frost, D. Hasko, D. Peacock, D. Ritchie, and C. Jones, *J. Phys. C* 21, 893 (1988).
20. A. Sachrajda, R. Taylor, C. Dharmawardana, J. Adams, P. Zawadzki, P. Coleridge, and M. Davies, *Canad. J. Phys.* 70, 1148 (1992).
21. A. Yacoby, M. Heiblum, D. Mahalu, and H. Shtrikman, *Phys. Rev. Lett.* 74, 4047 (1995).
22. C. Marcus, A. Rimberg, R. Westervelt, P. Hopkins, and A. Gossard, *Phys. Rev. Lett.* 59, 506 (1992).
23. L. Kouwenhoven, A. Johnson, N. van der Vaart, C. Harmans, and C. Foxon, *Phys. Rev. Lett.* 67, 1626 (1991).

24. B. van Wees, L. Kouwenhoven, C. Harmans, J. Williamson, C. Timmering, M. Broekaart, C. Foxon, and J. Harris, *Phys. Rev. Lett.* 62, 2523 (1989).
25. R. Haug, J. Hong, and K. Lee, *Surf. Sci.* 263, 415 (1992).
26. U. Meirav and E. Foxman, *Semicond. Sci. Technol.* 10, 255 (1996).
27. M. Reed, *Sci. Am.* 118–123 (1993).
28. R. C. Ashoori, H. L. Störmer, J. Weiner, L. Pfeiffer, K. Baldwin, and K. West, *Phys. Rev. Lett.* 71, 613 (1993).
29. R. C. Ashoori, H. L. Störmer, J. Weiner, L. Pfeiffer, S. Pearton, K. Baldwin, and K. West, *Phys. B* 184, 378 (1993).
30. R. C. Ashoori, H. L. Störmer, J. Weiner, L. Pfeiffer, S. Pearton, K. Baldwin, and K. West, *Phys. Rev. Lett.* 68, 3088 (1992).
31. S. Tarucha, D. G. Austing, and T. Honda, *Superlatt. Microstruct.* 18, 121 (1995).
32. D. G. Austing, T. Honda, and S. Tarucha, *Japan. J. Appl. Phys.* 36, 4151 (1997).
33. D. G. Austing, T. Honda, and S. Tarucha, *Semicond. Sci. Technol.* 11, 388 (1996).
34. D. G. Austing, T. Honda, and S. Tarucha (1996).
35. L. P. Kouwenhoven, T. H. Oosterkamp, M. W. Danoesastro, M. Eto, D. G. Austing, T. Honda, and S. Tarucha, *Science* 278, 1788 (1997).
36. M. Dellow, P. Beton, C. Langerak, T. Foster, P. Main, L. Eaves, M. Henini, S. Beaumont, and C. Wilkinson, *Phys. Rev. Lett.* 68, 174 (1992).
37. T. Schmidt, R. Haug, K. von Klitzing, A. Förster, and H. Lüth, *Phys. Rev. Lett.* 78, 1544 (1997).
38. M. A. Kastner, *Rev. Mod. Phys.* 64, 849 (1992).
39. M. Yoo, T. Fulton, H. Hess, R. Willet, L. Dunkleberger, R. Chichester, L. Pfeiffer, and K. West, *Science* 276, 579 (1997).
40. P. McEuwen, E. Foxman, U. Meirav, M. Kastner, Y. Meir, N. Wingreen, and S. Wind, *Phys. Rev. Lett.* 66, 1926 (1991).
41. M. Fricke, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, *Europhys. Lett.* 36, 197 (1996).
42. M. Grundmann, J. Christen, N. Ledentsov, J. Börer, S. Ruminov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V. Ustinov, et al., *Phys. Rev. Lett.* 74, 4043 (1995).
43. F. Adler, M. Geiger, A. Bauknecht, F. Scholz, H. Schweizer, M. H. Pilkuhn, B. Ohnesorge, and A. Forchel, *J. Appl. Phys.* 80, 4019 (1996).
44. M. Grundmann, N. N. Ledentsov, O. Stier, D. Bimberg., V. M. Ustinov, P. Kopev, and Z. I. Alferov, *Appl. Phys. Lett.* 69, 979 (1996).
45. H. Jiang and J. Singh, *IEEE J. Quantum Electron.* 34, 1188 (1998).
46. S. Fafard, K. Hinzer, S. Raymond, M. Dion, J. M. Y. Feng, and S. Charbonneau, *Science* 274, 1350 (1996).
47. M. Grundntann, O. Stier, and D. Bimberg, *Phys. Rev. B* 52, 11969 (1995).
48. H. Huang and D. G. Deppe, *IEEE J. Quantum Electron* 37, 691 (2001).
49. P. Bhattacharyaa, K. Kamath, J. Singh, D. Klotzkin, J. Phillips, H. Jiang, N. Chervela, T. Norris, T. Sosnowski, J. Laskar, et al., *IEEE Trans. Electron Devices* 46, 871 (1999).
50. Y. Arakawa and H. Sakaki, *Appl. Phys. Lett.* 40, 939 (1982).
51. Y. Arakawa and A. Yariv, *IEEE J. Quantum Electron.* 21, 1666 (1985).
52. S. Tiwari, F. Rana, H. Hanafi, A. Harstein, E. Crabble, and K. Chan, *Appl. Phys. Lett.* 68, 1377 (1996).
53. K. Yano, T. Ishii, T. Kobayashi, F. Murai, and K. Seki, *IEEE Trans. Electron Devices* 41, 1628 (1994).
54. L. G. Guo, E. Leobandung, and S. Y. Chou, *Appl. Phys. Lett.* 70, 850 (1997).
55. I. Kim, S. Han, H. Kim, J. Lee, B. Choi, S. Hwang, D. Ahn, and H. Shin, *IEDM Tech. Dig.* 111 (1998).
56. A. Thean and J. Leburton, *IEEE Electron Device Lett.* 22, 148 (2001).
57. D. Averin, A. Korotkov, and K. Likharev, *Sov. Phys. JETP* 41, 308 (1975).
58. C. Beenakker, *Phys. Rev. B* 44, 1646 (1991).
59. L. Kouwenhoven and L. Glazman, *Phys. World* 14, 33 (2001).
60. K. K. Likharev, *Proc. IEEE* 87, 606 (1999).
61. B. Diu, G. Guthmann, D. Lederer, and B. Roulet, "Physique Statistique." Hermann, Paris, 1997.
62. J. Weis, R. Haug, K. v. Klitzing, and K. Ploog, *Phys. Rev. B* 46, 12837 (1992).
63. M. Macucci, K. Hess, and G. Iafrate, *Phys. Rev. B* 48, 17354 (1993).
64. D. G. Austing, T. Honda, and S. Tarucha, *Japan. J. Appl. Phys.* 36, 1667 (1997).
65. D. G. Austing, T. Honda, K. Muraki, Y. Tokura, and S. Tarucha, *Phys. B* 249–251, 206 (1998).
66. D. G. Austing, S. Sasaki, S. Tarucha, S. Reimann, M. Koskinen, and M. Manninen, *Phys. Rev. B* 60, 11514 (1999).
67. D. G. Austing, T. Honda, and S. Tarucha (1997).
68. P. Atkins and R. Friedman, "Molecular Quantum Mechanics." Oxford Univ. Press, London, 1997.
69. M. Macucci, K. Hess, and G. Iafrate, *Phys. Rev. B* 55, 4879 (1997).
70. I.-H. Lee, V. Rao, R. M. Martin, and J.-P. Leburton, *Phys. Rev. B* 57, 9035 (1998).
71. P. Matagne, J.-P. Leburton, D. Austing, and S. Tarucha, *Phys. Rev. B* 65 (2002).
72. F. Bolton, *Phys. Rev. B* 54, 4780 (1996).
73. T. Ezaki, N. Mori, and C. Hamaguchi, *Phys. Rev. B* 56, 6428 (1997).
74. C. Yannouleas and U. Landman, *Phys. Rev. Lett.* 82, 5325 (1999).
75. B. Partoens and F. Peeters, *Phys. Rev. Lett.* 84, 4433 (2000).
76. W. de Heer, *Rev. Modern Phys.* 65, 611 (1993).
77. M. Macucci, K. Hess, and G. Iafrate, *J. Appl. Phys.* 77, 3267 (1995).
78. M. Stopa, *Phys. Rev. B* 54, 13767 (1996).
79. D. Jovanovic and J.-P. Leburton, *Phys. Rev. B* 49, 7474 (1994).
80. S. Nagaraja, P. Matagne, V.-Y. Thean, J.-P. Leburton, Y.-H. Kim, and R. M. Martin, *Phys. Rev. B* 56, 15752 (1997).
81. P. Matagne, J.-P. Leburton, J. Destine, and G. Cantraine, *Comput. Modeling Eng. Sci.* 1, 1 (2000).
82. W. Kohn and L. Sham, *Phys. Rev.* 140, 1133 (1965).
83. J. Perdew and Y. Wang, *Phys. Rev. B* 45, 13244 (1992).
84. J. Perdew and Y. Wang, *Phys. Rev. B* 45, 13244 (1992).
85. F. S.-M. X. Aymerich-Humet and J. Millan, *Solid State Electron.* 24, 981 (1981).
86. J. Blakemore, *Solid State Electron.* 25, 1067 (1982).
87. E. O. Kane, *Phys. Rev.* 131, 79 (1963).
88. B. Halperin and M. Lax, *Phys. Rev.* 148, 722 (1966).
89. S. Jain and D. Roulston, *Solid State Electron.* 34, 453 (1991).
90. R. Grant, J. Waldrop, S. Kowalczyk, and E. Kraut, *J. Vac. Sci. Technol.* 19, 477 (1981).
91. J. Best, *Appl. Phys. Lett.* 34, 522 (1979).
92. J. Slater, *Adv. Quantum Chem.* 6 (1972).
93. L. R. Fonseca, J. L. Jimenez, J.-P. Leburton, and R. M. Martin, *Phys. Rev. B* (1997).
94. S. Tarucha, D. G. Austing, Y. Tokura, W. van der Wiel, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 84, 2485 (2000).
95. C. Darwin, *Proc. Cambridge Philos. Soc.* 27 (1930).
96. I.-H. Lee, Y.-H. Kim, and K.-H. Ahn, *J. Phys. Condensed Matter* 13, 1993 (2001).
97. D. G. Austing, T. Honda, and S. Tarucha, *Semicond. Sci. Technol.* 12, 631 (1997).
98. A. Kumar, S. E. Laux, and F. Stern, *Phys. Rev. B* 42, 5166 (1990).
99. P. Matagne, J.-P. Leburton, D. Austing, and S. Tarucha, *Phys. E*, in press.
100. L. P. Kouwenhoven, *Science* 268, 1440 (1995).
101. R. H. Blick, D. Pfannkuche, R. J. Haug, K. v. Klitzing, and K. Eberl, *Phys. Rev. Lett.* 80, 4032 (1998).
102. R. J. Haug, , R. H. Blick, and T. Schmidt, *Phys. B* 212, 207 (1995).

103. N. C. van der Vaart, S. F. Godijn, Y. V. Nazarov, C. J. P. M. Harmans, J. E. Mooij, L. W. Molenkamp, and C. T. Foxon, *Phys. Rev. Lett.* 74, 4702 (1995).
104. G. W. Bryant, *Phys. Rev. B* 48, 8024 (1993).
105. K. A. Matveev, L. I. Glazman, and H. U. Baranger, *Phys. Rev. B* 53, 1034 (1996).
106. I. M. Ruzin, V. Chandrasekhar, E. I. Levin, and L. I. Glazman, *Phys. Rev. B* 45, 13469 (1992).
107. C. A. Stafford and S. D. Sarma, *Phys. Rev. Lett.* 72, 3490 (1994).
108. G. Klimeck, G. Chen, and S. Datta, *Phys. Rev. B* 50, 2316 (1994).
109. J. M. Golden and B. I. Halperin, *Phys. Rev. B* 53, 3893 (1996).
110. F. R. Waugh, M. J. Berry, D. J. Mar, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Phys. Rev. Lett.* 75, 705 (1995).
111. S. Nagaraja, J.-P. Leburton, and R. Martin, *Phys. Rev. B* 60, 8759 (1999).
112. J.-P. Leburton and S. Nagaraja, Electronic properties of Quantum dots and artificial atoms, in "Optical Spectroscopy of Low Dimensional Semiconductor," NATO ASI Ser. E, Vol. 344, pp. 235–256. Kluwer Academic, Amsterdam, 1997.
113. P. Matagne and J.-P. Leburton, *Phys. Rev. B* 65, 1553 (2002).

Quantum Well Infrared Detectors

W. Lu

*Shanghai Institute of Technical Physics, Chinese Academy of Sciences,
Shanghai 200083, China*

Y. Fu

Chalmers University of Technology, Göteborg S-412 96, Sweden

CONTENTS

1. Introduction
2. Molecular Beam Epitaxy Technology for QWIP Material
3. Physical Model
4. Processing of Infrared Photodetectors
5. Device Development
6. Conclusions
Glossary
References

1. INTRODUCTION

Nanotechnology has shown its powerful ability to control electron behavior in a potential well, and the size of the potential well is of the order of the wavelength of the electron wave function [1, 2]. In this quantum mechanical regime, the electron properties depend strongly on the shape of the potential well. The most intensively studied system is the semiconductor quantum well, which consists of three semiconductor layers, one narrow bandgap material layer (called quantum well) sandwiched between two wide bandgap material layers (potential barriers, or just barriers). The potential barriers confine the electrons in the quantum well, and the energy band structure of the confined electron in the quantum well is made of discrete energy sublevels along the confinement direction. The energy positions of the sublevels are determined by the thickness of the quantum well (quantum well width) and the difference between the energy bandgaps of the quantum well and barrier materials (i.e., the height of the potential barrier). In one application of the quantum well system, the inter-sublevel energy is tuned to match the energy of the incident infrared photon [3, 4]. An electron originally occupying one sublevel can then transit to another sublevel by absorbing one infrared photon. This is the core physical nanostructure for infrared detection. Up to now, the quantum well in infrared detection has been one of most successful nanotechnology applications. With the reverse process of absorption, one designs the quantum well in such a way that the electron transits from one sublevel to another by emitting a photon, and with this we are in the field of light emitter and laser.

Infrared detector technology has been well developed based on narrow bandgap materials, like InSb and $Hg_{1-x}Cd_xTe$ [5]. Due to the high performance and flexibility in the infrared detection application, $Hg_{1-x}Cd_xTe$-based infrared detector technology has been studied extensively in the past 30 years. The infrared response of $Hg_{1-x}Cd_xTe$ is based on the interband transition, that is, from the valence band to the conduction band (Fig. 1a). The transition energy (i.e., the energy band gap, E_g) is determined by the mole fraction of Cd atoms, x, in $Hg_{1-x}Cd_xTe$. In such a detection system, the response wavelength can be easily adjusted from 1 to 20 μm. After absorbing the infrared photons, photogenerated electrons and holes transport in opposite directions either in an external electrical field when the device is biased in the photoconductive mode, or in a built-in electric field in the photovoltaic mode. The photoconductive detector is based on the resistance variation due to the increased density of conduction carriers excited from the valence band by the interband transition after absorbing infrared photons. Such a photoconductive operation mode has been widely used for the infrared detectors of $Hg_{1-x}Cd_xTe$.

Since the response time of the $Hg_{1-x}Cd_xTe$ photoconductive detector is normally in the region of 10^{-4} to 10^{-6} second, an image frame from a medium size image format of 256 × 256 pixels takes about 0.1–1 second, whereas the time interval between image frames should be less than 0.01 second when shooting moving objects. A focal plan

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

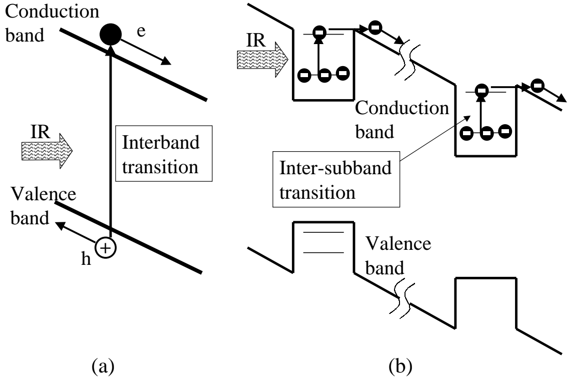


Figure 1. Optical transitions in infrared detection systems. (a) $Hg_{1-x}Cd_xTe$; (b) quantum well infrared photodetector. Interband transition is used for infrared detection in $Hg_{1-x}Cd_xTe$ material, while the QWIP system utilizes the inter-subband transition.

array (FPA) device has therefore been developed to replace the two-dimensional scanner. In such a system, the mechanical scanner is replaced by an electrical scanner fulfilled by a readout circuit, which can be well fabricated by the commercial microelectronic foundry. The integration time for one pixel in the FPA device is now extended to the time duration of the image frame. With this, the image recording speed can be high (e.g., more than 100 frames per second). A spin-off advantage of the electrical readout circuit in the FPA device over the mechanical scanner is the much reduced weight of the detection system, and the reduction can be more than 50%. At present, the infrared technology is mainly based on the FPA technology.

However, the dark current in the $Hg_{1-x}Cd_xTe$ photoconductive detector is normally rather high, which limits the FPA applications. A high dark current of one pixel in a FPA device easily saturates the capacity of the readout circuit in the photoconductive operation mode. Meanwhile, a FPA device consists of a huge number of sensors which are monolithically integrated in one semiconductor chip. A high dark current from one pixel means a huge dark current in the semiconductor chip, which can destroy easily the device. Thus, the $Hg_{1-x}Cd_xTe$ FPA system cannot work in the photoconductive mode, even if the fabrication process of a photoconductive sensor is rather simple and the sensor exhibits very high detectivity and responsibility.

To avoid the large dark current, the photovoltaic mode was developed. In this mode, the dark current is almost zero under the zero external bias condition. The photon excited carriers are swept out by the built-in electric field in the *pn* junction. With the photovoltaic operation mode, very large format FPAs (e.g., 2048 × 2048) have been demonstrated from the *pn* junctions of narrow bandgap semiconductors including InSb and $Hg_{1-x}Cd_xTe$. For long wavelength applications, the 256 × 256 format FPA device has already been commercialized. Up to now, the most widely used infrared FPA devices have been based on InSb and $Hg_{1-x}Cd_xTe$.

However, with the breakthrough of $Hg_{1-x}Cd_xTe$ based FPA device technology, one finds that the yield of the device is still too low, especially for very large format FPA devices. The infrared detectors always work at low temperature at about 77 K, but they are processed at room temperature. The temperature change degrades the device performance caused by the thermal expansion induced defects originating from the interface between the substrate and the $Hg_{1-x}Cd_xTe$ film. Moreover, due to the thermal instability and the material growth difficulties of the $Hg_{1-x}Cd_xTe$ material, the cost is rather high. This strongly limits the $Hg_{1-x}Cd_xTe$ FPA device application.

Since 1980, the infrared detector based on the intersubband transition in quantum well systems, called the quantum well infrared photoconductor (QWIP), has been studied and developed intensively.

By inter-subband transition, one can utilize wide bandgap materials like GaAs for infrared detection [3, 4, 6–20]. As shown in Figure 1b, the photoexcited electrons have a much higher mobility compared to ground-state carriers, which are basically not mobile due to the quantum confinement of the quantum well. This large difference in the mobilities forms the principal infrared photoconductive signal in the QWIP material. Quantum well systems based on wide bandgap III–V materials are much more mature in both the system fabrication and the device processing as compared to narrow bandgap $Hg_{1-x}Cd_xTe$ material. Up to now, most research and commercialization effects have been concentrated on the multiple quantum well system of GaAs/AlGaAs. This quantum well system has an almost perfect lattice match, which makes a good interface between the quantum well material and barrier material (very low density of interface states). Excluding silicon, the compound semiconductor of GaAs is the best developed material among the commonly used semiconductors. By combining the GaAs semiconductor technology and nanotechnology through optimization of the quantum well structure, the QWIP will show its advantages with its low cost, high stability, very large format FPA, easy wavelength extension (especially for the wavelengths longer than 12 μm), easy multicolor integration, and easy monolithic integration of the readout circuit with infrared sensor technology. Figure 2 shows thermal images of a man taking a glass in hand taken by a 128 × 1 linear array and a hand by a 64 × 64 QWIP FPA with a cutoff wavelength of 9 μm. A huge number of research papers have been published since 1985. The number becomes saturated in 1997; from then the commercialization project of the QWIP has focused on the optimization of the device performance.

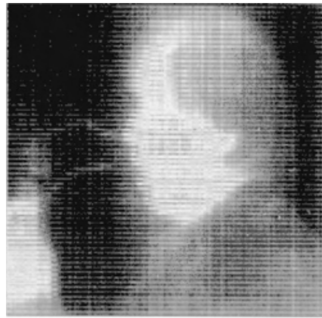


Figure 2. Thermal images of a man with a glass in his hand taken by a 128 × 1 linear array and a hand taken by a 64 × 64 QWIP FPA with a cutoff wavelength of 9 μm.

2. MOLECULAR BEAM EPITAXY TECHNOLOGY FOR QWIP MATERIAL

The quantum well is the core for the infrared detector operation. Conventional infrared detector materials of HgCdTe and InSb utilize inter-band transitions for photon excitation, whereas the QWIP utilizes the inter-subband optical transitions. In HgCdTe and InSb, the optimized material size is of micrometer order determined by the diffusion length of photoexcited carriers and surface-state induced energy band bending. The subband alignment in the QWIP is directly determined by the quantum well width and barrier height, so that both the infrared response wavelength and responsibility are determined by the energy separation between the ground state and first excited state and the energy position of the excited state with respect to the barrier height. A one monolayer change in the quantum well width can significantly affect the QWIP performance. Due to this sensitive condition, almost all QWIP materials are grown by the molecular beam epitaxy (MBE) technique. For the most widely used QWIP material of lattice-matched GaAs/AlGaAs quantum wells, the MBE process has been very well developed. Meanwhile, by using the MBE method one can control the doping profile in the nanometer scale. Therefore the MBE process can in principle grow all kinds of desired quantum well structures for QWIP applications. And normally MBE provides better control over the quantum well potential profile than is observed in the metal-organic chemical vapor deposition method [21].

As we will see in Section 2.2 the best alignment of the subband in quantum wells for the QWIP application is that the first excited state is at the quasi-bound condition. In this case, the photoexcited electrons are easily driven by an external bias. For this, the quantum well structure, namely the quantum well width and barrier height, is to be precisely controlled. During the material growth, the reflection high energy electron diffraction (RHEED) specular intensity oscillation is one of the best monitoring methods.

A typical MBE material growth chamber consists of effusion cells with shutters, a RHEED gun with an electron beam detection fluorescent screen, an ionization gauge for source flux detection, a gate valve to separate the growth chamber and the sample load chamber, a main shutter, and a view port. The effusion cells contain different sources for QWIP material growth, including Ga, Al, and As for quantum well structure growth and Si for *n*-type doping in GaAs material. The RHEED equipment is used to monitor the growth speed and material surface reconstruction. The sample holder in the center of the chamber keeps the (001) GaAs substrate rotating and heated. A typical GaAs substrate temperature for QWIP material growth is about 600 °C. All the effusion cells and substrate holder heater are well controlled by the electrical power with a temperature stability which is better than 0.2 °C, so that a homogeneous QWIP material is guaranteed for the FPA (which will be discussed later) device fabrication. For the QWIP FPA device, the homogeneity of the sample structure is much more critical than that of any other conventional infrared material.

The RHEED measurement is normally in the forward configuration and the energy of the electron beam is in the range of 5–40 keV. The de Broglie wavelength of such electrons ranges from 0.17 to 0.06 Å. The incident angle between the direction of the electron beam and the substrate surface is 1–3°. The penetration length of the electron beam into the GaAs substrate is limited within several atomic layers. In this way, the roughness of the growth surface is clearly reflected by the RHEED pattern. Basic properties of the RHEED pattern are explained by the kinetic theory of the electron diffraction [22]. For a flat growth surface, the lattice can be approximated as a two-dimensional grating which results in a RHEED diffraction pattern in the form of parallel lines. When the surface is deposited with lattice islands in the nanometer scale, the islands act as a three-dimensional grating so that diffraction points will be generated in the RHEED pattern [23–25].

For the GaAs substrate used for the QWIP material growth, the (001) orientation is used. The surface of the (001) GaAs can be either a Ga-rich surface or an As-rich surface. The two surfaces have different surface reconstructions. The Ga-rich surface, which has a 4×2 reconstruction, is to be avoided in QWIP materials. Instead, the As-rich surface with a 2×4 reconstruction is utilized. A high As-atom partial pressure can protect the MBE growth surface against the Ga-rich surface. However, a too high As-atom partial pressure will degrade the migration of Ga atoms on the growth surface, resulting in the generation of lattice islands. In order to grow a high-quality quantum well structure, a layer by layer growth mode is preferred to keep the growth surface flat. Thus the conditions of the MBE growth are very critical so that one has an As-rich surface and, at the same time, a balanced As-atom partial pressure. Obviously, the transition point in the RHEED pattern from 2×4 to 4×2 reconstruction is a good reference condition for the As-atom partial pressure.

The RHEED specular intensity shows a clear oscillation of the intensity during the MBE growth. This oscillation has been used to measure the epitaxy growth rate [26] and to control both the composition and layer thickness in semiconductor superlattice [27]. Moreover, the atom diffusion on the surface [28, 29], growth mechanism [30, 31], and doping behavior [32] have been also studied by the oscillation measurement. The RHEED intensity oscillation period corresponds to the growth time of one atomic layer, that is, one layer of Ga–As atoms along the (001) direction, and its thickness is 2.83 Å. This oscillation behavior is characteristic in the two-dimensional growth mode [33]. We can directly correlate the RHEED intensity with the roughness of the growth surface: A flat surface will result in a high reflectivity, while a rough surface will decrease the specular intensity due to the increased diffusion reflection. Starting at a flat growth surface, the electrons of the RHEED beam are largely reflected into the special specular points. With a constant flux of molecular beams, the anion and cation atoms combine randomly on the growth surface, so that the local growth rates vary. Since the atomic layer step of the (001) GaAs substrate is 2.83 Å, which is much longer than the de Broglie wavelength 0.1–0.2 Å of the electron beam used in the RHEED, the combined Ga–As atoms which are randomly distributed make the originally flat surface "rough." The RHEED electrons are reflected and the electron densities on the special RHEED specular points

decrease. When half of the growing surface becomes occupied by the islands, the degree of roughness reaches its maximum, and the RHEED specular intensities reach their minimum [34]. Further growing of the surface reduces the isolated island density and a complete flat surface gradually emerges. The degree of the surface roughness becomes small, and enhanced RHEED specular point intensities are expected. The completion of the new flat surface thus forms a period of the RHEED intensity oscillation. Normally, a new layer will start to grow before an old one is complete so that gradually the growing surface will consist of several different atomic layers, which decreases the RHEED specular intensity oscillation amplitude.

In general, it is observed from the RHEED intensity oscillation that the QWIP material growth is in the two-dimensional mode. The RHEED oscillation period provides us the absolute growth rate with a precision on the scale of one subatomic layer. Thus, with the help of RHEED intensity oscillation, we can grow the quantum well structure with a layer thickness precision of 1 Å required for the optimized QWIP device operation.

In order to have a high integrity QWIP, a high carrier mobility in the $Al_xGa_{1-x}As$ barrier region is also required, which means a very pure material growth condition. During the material growth, the wall of the MBE growth chamber is normally cooled down to 77 K to prevent foreign atoms or molecules from the chamber wall. And the background pressure should be lower than 10^{-11} Torr. In addition, adsorbed atoms or molecules on the substrate surface should be eliminated. This is done by degassing the substrate at 150, 260, and 370 °C separately for several minutes in a buffer chamber. After that, the substrate is heated to about 580 °C to evaporate the oxidation layer under the protection of As flux. At this moment, the RHEED pattern will change from the amorphous-structure-induced ringlike pattern to the crystal-structure-induced linelike pattern. Finally, the substrate will be heated at 620 °C for about 20 minutes to evaporate other remaining impurity atoms.

2.1. Growth of $Al_xGa_{1-x}As$/GaAs Heterostructure

Unlike all other conventional materials of infrared detector applications, the QWIP system contains a lot of heterointerfaces between GaAs and $Al_xGa_{1-x}As$ layers. Up to now, among all heterostructure semiconductor materials, the GaAs/$Al_xGa_{1-x}As$ system is the best developed in terms of the MBE growth technology. Such a mutual structure has established an excellent basis to industrialize the QWIP materials. Moreover, the system is also used to test and evaluate new epitaxy technologies [35].

The first step to grow GaAs/$Al_xGa_{1-x}As$ heterostructure is to obtain a correct Al composition x. This is done by adjusting the fluxes of Ga and Al sources (i.e., P_{Ga} and P_{Al}). Since the adsorption coefficients of Ga and Al on the GaAs substrate are almost identical, we can approximate the ratio between Al and Ga molecular beam fluxes as their composition ratio of $x/(1-x)$. In the MBE growth process, the flux ratio is determined by the partial pressures measured by the ionization gauge using the equation

$$\frac{J_{Al}}{J_{Ga}} = \frac{P_{Al}\eta_{Ga}}{P_{Ga}\eta_{Al}}\sqrt{\frac{T_{Al}M_{Ga}}{T_{Ga}M_{Al}}} \tag{1}$$

where J_{Ga} and J_{Al} are the Ga and Al fluxes, η_{Ga} and η_{Al} represent the relative sensitivities of the ionization gauge to Ga and Al, T_{Ga} and T_{Al} are the temperatures of the effusion cells, and M_{Ga} and M_{Al} are the Ga and Al atomic weights.

However, this is only the first approximate adjustment of the Al composition. A fine tuning of the composition x is to be achieved by the RHEED specular intensity oscillation measurement. Even for a small Al composition change of $x = 0.02$, the difference in the oscillation periods is unambiguous after 10 oscillation periods. From the oscillation periods of GaAs and $Al_xGa_{1-x}As$ materials (e.g., $\tau_{GaAs} =$ 1548 ms and $\tau_{Al_xGa_{1-x}As} = 1517$ ms), we can determine the composition x:

$$x = 1 - \frac{\tau_{Al_xGa_{1-x}As}}{\tau_{GaAs}} = 0.02 \tag{2}$$

The quantum well width of quantum well is normally equal to 4–5 nm for long wavelength QWIP applications. One atomic layer change means a fluctuation of 6% in the quantum well width. A large quantum well width fluctuation degrades the device performance. There have been extended works on the study of the quantum well interface quality using photoluminescence (PL) spectra and transmission electron microscopy [36]. Since the migration mobility of Ga on the growing surface is larger than that of Al, GaAs-stopped and AlGaAs-stopped surfaces shall have different roughness sizes. Two type of interfaces are expected; one is called the normal interface formed by $Al_xGa_{1-x}As$ grown on GaAs, and the other is the inverse interface which is formed by GaAs grown on $Al_xGa_{1-x}As$. Normally, interface growth pause is used to improve the interface flatness. There are four possible ways: (A) growth pause at the two interfaces of quantum well; (B) growth pause only at the normal interface; (C) growth pause only at the inverse interface; and (D) without any growth pause. At 77 K, the PL linewidths from types A and B are about 4 meV, while they are about 12 meV when types C and D are used. In addition, the quantum wells in the QWIP have to be heavily doped by Si at a level of about 1×10^{18} cm^{-3}. The PL lines of interband transitions from these doped quantum wells are about 22 meV wide when the samples are grown by type A; growth pause type D results in a 39 meV wide PL line.

2.2. Si Doping in Quantum Wells

Shallow impurity doping is another key point in obtaining high quality QWIP materials. In QWIP materials, impurities are normally doped in quantum wells (i.e., the GaAs layers), and Si and Be are used as n- and p-type doping impurity atoms, respectively. The Si atoms are the best shallow donors among IV group elements like Ge and Sn [37]; they can result in an electron concentration as high as 1.8×10^{19} cm^{-3} in GaAs-based materials [38]. As an element in group IV, Si can be either a donor having an ionization

energy of 5.8 meV or an acceptor with an ionization energy of 35 meV in the GaAs material. During the Si doping process in GaAs, there are two possible reactions,

$$\begin{aligned} \mathrm{Si}_{(g)} + V_{\mathrm{Ga}} &\longleftrightarrow \mathrm{Si}_{\mathrm{Ga}}^{+} + e \\ \mathrm{Si}_{(g)} + V_{\mathrm{As}} &\longleftrightarrow \mathrm{Si}_{\mathrm{As}}^{-} + h \end{aligned} \tag{3}$$

where the label (g) represents the gas phase, V_{Ga} and V_{As} are the vacancies of Ga and As, $\mathrm{Si}_{\mathrm{Ga}}^{+}$ and $\mathrm{Si}_{\mathrm{As}}^{-}$ are the cation and anion of Si occupying Ga and As lattice positions, respectively, and e and h are electron and hole.

In the MBE material growth process, the flux ratio of As and Ga sources, $J_{\mathrm{As}}/J_{\mathrm{Ga}}$, determines the behavior of Si in GaAs to be either a donor or an acceptor [39, 40]. If the flux ratio is small, the GaAs surface will be Ga-rich with a (4×2) reconstruction. More Si atoms will occupy the As vacancies. The epilayer becomes p-type. With the increase of the flux ratio, the GaAs surface will change its reconstruction from (4×2) to (3×1) and finally to As-rich (2×4). The Si atoms then mainly occupy the Ga vacancies, resulting in an n-type epilayer. As described previously, the QWIP material should be grown in the As-rich (2×4) reconstruction surface condition; thus the Si doping will mainly result in n-type epilayers.

Increasing the doping concentration to 1×10^{18} cm^{-3}, the impurity scattering will dominate the free electron mobility [40]. Thus, the Si atoms are normally doped in the central parts of quantum wells; the barriers are undoped. In this way, the Si impurities are constrained inside the quantum wells to avoid possible diffusion of Si into the AlGaAs barriers where Si can induce deep levels. Meanwhile, this will also strongly suppress the scattering of photoexcited carriers, since the photoexcited carriers transport mainly in the barriers.

3. PHYSICAL MODEL

3.1. General Consideration of Optical Transition

Figure 3 shows the typical heterostructure of the QWIP system. Denoting the electronic state by its quantum number k, the photon state by photon energy $\hbar\omega$, and photon density by n_{ph}, the absorption rate between two electronic states k_i and k_j is [41]

$$W_{ji}(\hbar\omega) = \frac{\pi\hbar^2 e^2 n_{\mathrm{ph}}}{m_0^2 \omega\epsilon} |\langle k_j | \boldsymbol{a} \cdot \nabla | k_i \rangle|^2 \delta[E(k_i) - E(k_j) + \hbar\omega] \tag{4}$$

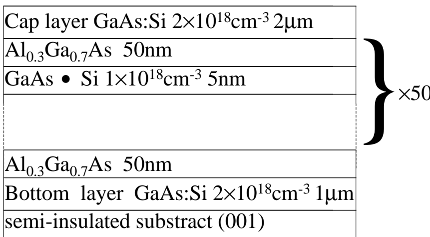


Figure 3. The layer structure of a typical QWIP with a response wavelength of 8 μm.

where $|k_i\rangle$ and $E(k_i)$ are the wave function and eigenvalue of state k_i, respectively. The net absorption rate (per unit time) of the system in a photon field is calculated by

$$W(\hbar\omega) = \sum_{ji} W_{ji}(\hbar\omega)\{f[E(k_i)] - f[E(k_j)]\} \tag{5}$$

where $f[E(k_j)]$ and $f[E(k_i)]$ are occupations of the final and initial electronic states, introduced here due to the Pauli exclusion principle (the initial state occupied and the final state empty).

Considering a beam of photons traveling along the z-axis, we can write the continuity equation for the photon density as

$$\frac{dn_{\mathrm{ph}}(\hbar\omega, z)}{dt} = \frac{\partial n_{\mathrm{ph}}(\hbar\omega, z)}{\partial t} - \frac{\partial[v n_{\mathrm{ph}}(\hbar\omega, z)]}{\partial z} \tag{6}$$

where the first term on the right side of the above equation represents the absorption rate of photons and the second term represents the photons leaving due to the photon current. At the steady state of $dn_{\mathrm{ph}}(\hbar\omega, z)/dt = 0$,

$$n_{\mathrm{ph}}(\hbar\omega, z) = n_0(\hbar\omega) e^{-\alpha(\hbar\omega)z} \tag{7}$$

which defines the absorption coefficient $\alpha(\hbar\omega)$. Since

$$\frac{\partial n_{\mathrm{ph}}(\hbar\omega, z)}{\partial t} = -W(\hbar\omega) \tag{8}$$

we have

$$\begin{aligned} \alpha(\hbar\omega) = \sum_{ij} \frac{\pi\hbar^2 e^2}{m_0^2 v\omega\epsilon} |\langle k_j | \boldsymbol{a} \cdot \nabla | k_i \rangle|^2 \delta(E_j - E_i - \hbar\omega) \\ \times \{f[E(k_i)] - f[E(k_j)]\} \end{aligned} \tag{9}$$

We now begin to calculate the optical absorption when a semiconductor quantum well sample is illuminated. By defining the growth direction of the sample as the z-axis and the plane perpendicular to this direction as the xy-plane, the Hamiltonian is

$$H = -\frac{\hbar^2 \nabla^2}{2m^*} + V(z) \tag{10}$$

in the effective-mass approximation, where m^* is the effective mass and $V(z)$ is the quantum well potential energy. By solving Eq. (10), one can easily get the energy states of ground state and the first excited state. However, the measured absorption line positions can differ from this simple prediction due to different physical effects, including nonparabolic band structures in the z-direction [42–47] and Coulomb interactions [48, 49]. For typical QWIP materials, all of these effects will not shift the absorption peak position by more than 10%. In general they are neglected in the first-order approximation, whereas they should be included in practical device simulations.

The interaction between the incident radiation and the electron is $e\hbar\boldsymbol{A} \cdot \nabla/m^*$ for an incident light of $\boldsymbol{A} = \boldsymbol{A}_0 e^{-i\omega t}$ [41]. Since H is translation symmetric in the xy-plane, the wave function can be expressed as

$$\Phi_{j\boldsymbol{k}}(\boldsymbol{r}) = \phi_j(z) u(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{b}} \tag{11}$$

where $u(\boldsymbol{r})$ is the Bloch function and $\boldsymbol{r} = (\boldsymbol{b}, z)$. For all values of j, $u(\boldsymbol{r})$ is identical, since only intra-band transitions among sublevels are involved in the photodetector. $u(\boldsymbol{r})$ can be well approximated as $\boldsymbol{k}$-independent. $\boldsymbol{k}$ and $\boldsymbol{b}$ are the wave vector and coordinate in the xy-plane, respectively. $\phi_j(z)$ is the envelope function and is normalized in the z-direction. Between ground sublevel $|\Phi_{0\boldsymbol{q}}\rangle$ and excited ones, it is easy to show that

$$\left\langle \Phi_{jk} \left| \frac{e\hbar}{m^*} \boldsymbol{A} \cdot \nabla \right| \Phi_{0\boldsymbol{q}} \right\rangle = \delta_{\boldsymbol{k},\boldsymbol{q}} \frac{e\hbar A_z}{m^*} \left\langle \phi_j \left| \frac{\partial}{\partial z} \right| \phi_0 \right\rangle \tag{12}$$

Equation (12) indicates the momentum conservation in the xy-plane, and the transition is between the envelope functions in the z-direction.

Moreover, the most important result indicated by Eq. (12) is the condition of the optical coupling: The coupling of a radiation with the carriers inside a quantum well structure is not possible at a normal incident condition (the light propagates along the sample growth direction so that the component of its vector potential along this direction, A_z, is zero). This has been clearly observed experimentally [50–52]. Optical coupling is therefore a necessity in many practical quantum-well-based photodetector devices. The Bruster angle incidence, 45 degree facet waveguide coupling, and grating coupling are the most frequently used optical coupling methods.

The probability of an electron optically excited from the ground sublevel to sublevel j becomes

$$W_j = \int \frac{2d\boldsymbol{q}}{(2\pi)^2} \left| \frac{e\hbar A_z}{m^*} \left\langle \phi_j \left| \frac{\partial}{\partial z} \right| \phi_0 \right\rangle \right|^2 \frac{2e^2\Gamma}{\hbar(\hbar^2\Omega_j^2 + \Gamma^2)} \times \left[f(E_{0\boldsymbol{q}}) - f(E_{j\boldsymbol{q}}) \right] \tag{13}$$

where we have replaced the energy conservation of $\delta(E_j - E_i - \hbar\omega)$ by the lifetime Γ of electrons in excited states.

3.2. Optical Coupling

As a good example, we consider a typical QWIP material structure for long wavelength detection. The QWIP is n-type and has an optical active layer consisting of 50 periods of $Al_{0.3}Ga_{0.7}As$–GaAs multiple quantum wells (MQWs) along the z-direction. Each $Al_{0.3}Ga_{0.7}As$ barrier is 500 Å thick and the GaAs well is 50 Å thick. The GaAs layer is Si-doped at a level of 5×10^{17} cm^{-3}. Doping levels in the cap and bottom layers are 2×10^{18} cm^{-3}. From the I–V characterization of the QWIP device, the current density is about 0.25 mA/cm^2 at the saturation level.

The active Γ electrons are described by an effective mass of $m^* = 0.067m_0$ in GaAs layers and $0.091m_0$ in $Al_{0.3}Ga_{0.7}As$ barriers, where m_0 is the free electron mass. The conduction band offset is $0.65 \times 1.247x$ eV between $Al_xGa_{1-x}As$ and GaAs.

Due to the isotropic effective mass of the Γ electrons, normal incident photon absorption is not possible, as discussed in the previous section. A component of the optical electric field along the QWIP growth direction, A_z, is required. To utilize Γ electrons, diffraction gratings are used as one of the optical coupling methods. Electron–photon coupling, by Eq. (12), is achieved normally by a two-dimensional optical grating. A typical diffraction grating structure consists of $2a \times 2a$ rectangular apertures, or circular apertures with radius a periodically arranged on the optical pixel, at distance b from each other. The depth of the apertures is denoted as h. Figure 4 shows the scanning electron microscope (SEM) pictures of such an optical grating.

The QWIP structure is not exactly nonconductive. However, the current density at normal device working conditions is very low so that the approximation of nonconduction with respect to the electromagnetic field is acceptable. When analyzing unpolarized infrared radiation, the vector nature of the wave amplitude will not be important, so that we can assume that we are dealing with a scalar amplitude function $\psi(\boldsymbol{r})$ which is a solution to the wave equation of [53]

$$\nabla^2\psi_{\text{ph}} + 4\pi^2k^2\psi_{\text{ph}} = 0 \tag{14}$$

for purely elastic diffraction, where $2\pi k = \sqrt{\epsilon\mu}\omega$ is the wave number. By Huygen's principle and considering a light source far away from the diffraction plane XY, the optical wave function for the refractive diffraction grating defined by its reflection coefficient $q(X, Y)$, which is perpendicular to the direction of radiation incidence, becomes [54, 55]

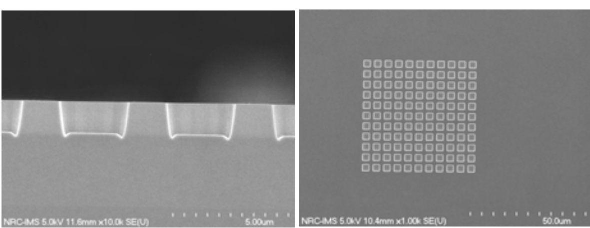
$$\psi_{\text{ph}}(x, y, R) = \iint q(X, Y) \frac{e^{-i2\pi kr}}{r} cr \times \left[\left(\frac{1}{r} + i2\pi k \right) \cos\,\theta - i2\pi k \right] dX\,dY \tag{15}$$

$r^2 = (x - X)^2 + (y - Y)^2 + R^2$, R is the distance from the observation plane xy to the diffraction grating plane XY, and θ is the angle between the z axis and vector $\boldsymbol{r}$.

From the calculated electric field as shown in Figure 5 for a quarter of a QWIP sensor with typical square apertures, one can observe a periodic behavior of optical field ψ in the xy-plane which is determined by the wavelength of the incident radiation. Such periodicity is expected when considering a one-dimensional grating structure so that the oscillation factor becomes $e^{-i2\pi kr} = e^{-i2\pi kx}e^{i2\pi kX}$. This is very much like the so-called near-field effect extensively applied to the surface properties [56].

Knowing the distribution of the optical field, we can calculate the amplitudes of different optical plane waves propagating in different directions. The optical field can be expressed as

$$\psi_{\text{ph}}(x, y, z) = e^{i2\pi kz} + \sum_{\boldsymbol{k}} t_{\boldsymbol{k}} e^{i2\pi(\boldsymbol{k}cdot\boldsymbol{b} - k_z z)} \tag{16}$$



SEM figure of grating, side view SEM figure of grating, top view

Figure 4. SEM pictures of a typical optical grating.

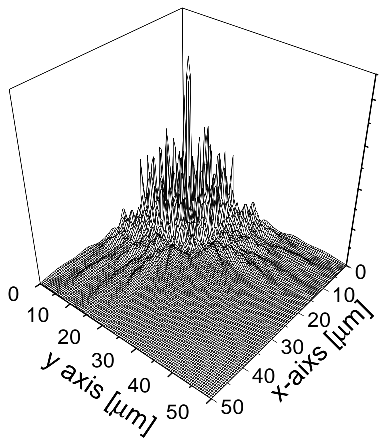


Figure 5. Intensity distribution of the optical field of 10 μm from square diffraction apertures. Reprinted with permission from [276], Y. Fu et al., *J. Appl. Phys.* 84, 5750 (1998). © 1998, American Institute of Physics.

where $\boldsymbol{r} = (\boldsymbol{b}, z)$, $e^{i2\pi kz}$ is the original incident radiation propagating along the z-direction, and $t_{\boldsymbol{k}}$ is the amplitude of the plane wave propagating along the $-z$-direction with wave vector $\boldsymbol{k} = (k_x, k_y)$ in the xy-plane after diffraction. Since we consider only elastic scattering, $k_x^2 + k_y^2 + k_z^2 = k^2$.

The diffraction grating is to be designed for large values of $t_{\boldsymbol{k}}$ corresponding to small values of k_z. The larger these amplitudes, the larger the effective optical strength to excite optical transitions, and thus a larger optical coupling between the incident photons and active electrons in the QWIP. Defining the effective strength of the diffracted radiation along the z-direction which can absorbed by the Γ-electrons in the GaAs quantum well as $\langle |A_z|^2 \rangle = A^2 B/2$, where A is the amplitude of the incident infrared radiation and the factor of $1/2$ is due to the unpolarization of the incident radiation, the value of the B parameter for the grating structure described above is 0.66. Similar results were reported earlier based on the EM mode approach [57].

Based on these design rules for diffraction grating, many different grating structures have been studied, including linear gratings [58, 59], two-dimensional (2D) periodic gratings [57, 60, 61], random reflectors gratings [62, 63], and quantum grid structure [64–67]. By the random reflectors, more passes of infrared light are achieved by the randomly roughened reflecting surface. So the absorption is efficiently enhanced. The minimum feature sizes of random reflectors for the FPA with cutoff wavelengths at 15 and 9 μm are about 1.25 and 0.6 μm, respectively [68, 69]. Combined with a AlGaAs waveguide structure or a thinned substrate, the two-dimensional periodic grating can enhance the detector responsivity by a factor of 2–3 with respect to the standard 45° polished facet illumination device [57, 60]. The so-called corrugated structure also shows its efficient optical coupling. In these structures, linear V-grooves are chemically etched to create a collection of angled facets within a single detector pixel in the FPA device [70, 71]. A general concept of utilizing microlenses was also achieved by monolithic integration. The responsivity is increased by a factor of 2.7 at a response wavelength of 15 μm [72].

3.3. Dark Current and Photocurrent

Dark current is one of key sources of the detection noise [73, 74], which determines the detectivity of the detector. A simple physical model was proposed for an order-of-magnitude estimation [75]. Another model was called the 3D carrier drift model where the three-dimensional electron density on the top of the barriers is estimated by taking into account the drift contribution while the diffusion process is neglected [76]. There was also the so-called emission-capture model which give adequate results for a quite extended range of applied fields [77–79]. A self-consistent drift-diffusion model was proposed in 1995 [80, 81]. This model was used to calculate the QWIP dark current by self-consistently solving the Poisson equation, the continuity equation for electrons in barriers, and the rate equations for electrons in the quantum wells. The self-consistent emission-capture model was also proposed, in which a nonuniform electric field is self-consistently determined by the Gauss law [82]. For the ultrafast electron transport properties, the semiclassical Monte Carlo model was advantageous [83]. Even with such intensive study on the modeling of dark current, a fundamental model is still missing for proper simulation and optimization of a realistic QWIP device, in which a thick barrier and a thin quantum well put the electron transport mechanism in the regime between a complete ballistic transport and a complete drift diffusion [4].

Distributed effects of external bias across the system recently began attracting attention [84–91]. Since the barrier regions are normally undoped and, in general, the carrier concentration in the barrier region can be neglected, the electric field there can be well approximated as constant, as assumed in many theoretical analyses [92, 93]. The electric field in each quantum well is negligible as compared with the ones in the barriers because of the high density of free carriers and the coincidental spatial distributions of electrons in the quantum well with the doped impurities [18].

Figure 6 shows a very small occupation probability $f(E_k)$ of excited states E_k above the barriers (the AlGaAs conduction bandedge is zero). On the other hand, the electron transport from one quantum well to the next quantum well

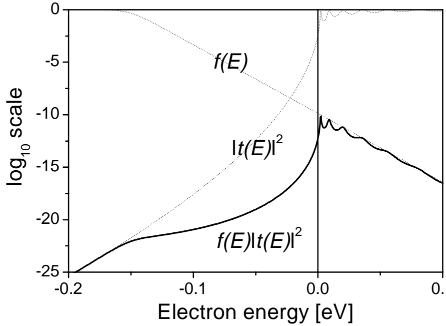


Figure 6. Occupation probability and transmission from one quantum well to the next as functions of the electron energy in the z-direction (the wave vector in the xy-plane is zero). $L = 2.9$ Å, $E_0 = -0.158$ eV, $E_f = -0.151$ eV at 77 K. Reprinted with permission from [277], Y. Fu et al., *J. Appl. Phys.* 87, 511 (2000). © 2000, American Institute of Physics.

is in the context of coherent wave transmission [94, 95],

$$\phi_k(z) = e^{ikz} + re^{-ikz} \tag{17}$$

in the quantum well that emits the electron wave, and

$$\phi_k(z) = t_k e^{ikz} \tag{18}$$

in the quantum well that collects the transmission wave. Here the QWIP is not biased. $\hbar^2 k^2/2m^* = E_k$. The transmission probability $|t_k|^2$ as a function of E_k is plotted in Figure 6, showing that due to the thick barriers the escape of carriers occupying the ground state E_0 in one quantum well to the next is almost impossible. By the complex eigenvalue approach it was concluded [20] that the lifetime of the electrons in the localized ground state E_0 is rather long so that these carriers have a rather small rate to escape to the adjacent wells through barriers.

By Figure 6 it is observed that in the QWIP system, thermal excitation dominates. In other words, the dark current consists of mainly thermally excited carriers over the barriers. It is also reconfirmed here that the current density is rather low so that the approximation of the quasi-equilibrium state is well established (we have used such an approximation by introducing the Fermi level E_f). The approximation of the nonconductive QWIP system when discussing the propagation of the electromagnetic field is also established due to this low conduction current.

It is concluded that electrons in one quantum well transmit to other spatial regions (adjacent quantum wells and contacts) via thermal activation above the barrier (forming the dark current); they become photoexcited to excited states above the barrier by absorbing the incident infrared radiation (forming the photocurrent). The current density has been modeled as $J = ev_d n'$ [92, 93, 96, 97], where v_d is the carrier drift velocity:

$$v_d = \mu F_z \left[1 + \left(\frac{\mu F_z}{v_s} \right)^2 \right]^{-1/2} \tag{19}$$

v_s is the saturation drift velocity ranging from 0.1×10^6 to 5×10^6 cm/s, μ is the low-field carrier mobility (2000 $\text{cm}^2/\text{V} \cdot \text{s}$ for n-type AlGaAs QWIPs), and F_z is the electric field intensity in the barrier region induced by the external bias. The density of mobile carriers n' in the quantum well consists of two parts, $n' = n'_{\text{th}} + n'_{\text{ph}}$, where n'_{th} is the carrier density due to the thermal excitation

$$n'_{\text{th}} = \int_{E_z \geq 0} \frac{2\, dk\, d\boldsymbol{k}}{(2\pi)^3} f(E_{\boldsymbol{k},k}) \tag{20}$$

It is easy to see that, in general, the optical transition probability is very small so that the carrier distribution of the ground state is almost unchanged. The photon excitation is

$$n'_{\text{ph}} = n_0 \int \langle |A_z(\hbar\omega)|^2 \rangle d(\hbar\omega) \int \frac{2dk}{2\pi} \left| \left\langle \phi_k \left| \frac{e\hbar}{m^*} \frac{\partial}{\partial z} \right| \phi_0 \right\rangle \right|^2 \times \frac{\Gamma}{(E_k - E_0 - \hbar\omega)^2 + \Gamma^2} \tag{21}$$

where n_0 is the carrier density occupying the ground state (E_0, ϕ_0).

The device operation of the QWIP is in photoconductive mode. It is similar to the extrinsic semiconductor [98]. Unlike conventional photoconductors, the incident photon is only absorbed in the quantum well region by the intersubband transition, so the absorption region is much narrower than the inactive barrier region. This difference causes the dependence of photoconductive gain to depend on the number of quantum wells [99]. The photoconductive gain is defined as the number of electrons flowing through the external circuit for each photon absorbed. The magnitude of optical responsivity is determined by both quantum efficiency and photoconductive gain [100, 101].

3.4. Boundary Conditions of Continuum States

We consider the multiple quantum well structure in the form of $V[z + n(d_B + d_W)] = V(z)$, where

$$V(z) = \begin{cases} \Delta E_c & -d_W/2 - d_B \leq z < -d_W/2 \\ 0 & -d_W/2 \leq z < d_W/2 \\ \Delta E_c & d_W/2 \leq z < d_W/2 + d_B \end{cases} \tag{22}$$

n is an integer, d_B is the AlGaAs barrier width, and ΔE_c is the barrier height. There is one localized state confined in the GaAs quantum well and its boundary conditions are unambiguous so that the corresponding wave function is zero deep inside the AlGaAs barriers. It is consistent numerically with other theoretical considerations, for example, the complex eigenvalue approach [20, 102–104]. Carriers occupying the localized state are thermal-excited as well as photoexcited to the extended continuum states above the barrier. Theoretically there exist three types of boundary conditions for these extended states. In early works [105–108], the GaAs quantum well and its adjacent two AlGaAs barriers were approximately embedded in an infinitely high barrier media, which is denoted as a box approximation. Because of the infinitely high barrier at the two ends, extended states above the AlGaAs barrier become discrete and are denoted as E_i, $i = 1, 2, \ldots$. Knowing the wave functions it is then easy to calculate the optical matrix element of $\langle \phi_i | \partial/\partial z | \phi_0 \rangle$.

However, due to the consideration of the carrier transport in the QWIP under normal device working conditions, we must envisage the conduction of both the thermal-excited and photoexcited electrons from one quantum to the next one. We first consider the running wave

$$\psi_{\text{running}}(z) = \begin{cases} e^{ikz} + re^{-ikz} & z < -d_W/2 \\ Ae^{iqz} + Be^{-iqz}: & -d_W/2 \leq z < d_W/2 \\ te^{ikz} & z \geq d_W/2 \end{cases} \tag{23}$$

for $E > \Delta E_c$. Here

$$\frac{\hbar^2 q^2}{2m^*} = E - \Delta E_c \qquad \frac{\hbar^2 k^2}{2m^*} = E \tag{24}$$

The third alternative is the transport of the excited continuum states in the Bloch wave form

$$\psi_{\text{Bloch}}(z) = u(z) e^{iKz} \tag{25}$$

where $u(z) = u[z + n(d_W + d_B)]$ accounts for the periodic boundary conditions of the GaAs/AlGaAs multiple quantum wells.

Knowing the optical matrix elements, the optical absorption spectra are easily calculated and numerical results are presented in Figure 7. Fine structures were observed experimentally in QWIPs consisting of multiple quantum wells, whereas single-quantum-well QWIPs display simple spectra [11]. Figure 7 suggests the Bloch-state boundary conditions as the most proper ones for continuum states in characterizing QWIPs consisting of multiple quantum wells (the discrepancy between the theoretical and experimental spectra at low wavelength range can be the result of the complicated conduction band structure at higher electron energies, where the simple effective mass approximation is not valid). The boundary conditions of running waves are definitely proper for QWIPs consisting of a single quantum well or only a few quantum wells, as demonstrated by the tunneling effect experimentally [109].

3.5. Alloy Scattering and Carrier Mobility

Recently, researchers have been interested in the application of quantum well intermixing techniques [110, 111] to modify energy levels of heterosystems. Ion implantation using materials such as Si [112] and As [113] and then further encapsulation by a dielectric layer such as SiO_2 and Si_3N_4 have been proven to be rather effective.

Figure 8 shows dark current spectra from annealed and ion-implanted QWIP samples. Note that the dark current increases by nearly a factor of 2 after proton implantation while it is almost unchanged after annealing. Theoretically it is expected that due to the ion implantation and annealing, the Al atoms diffuse from barrier regions into well regions. The bottom of the quantum well is lifted up and the width of the quantum well is effectively reduced so that subband energies of carriers are increased.

We continue considering the QWIP system consisting of 50 periods of 50-nm-$Al_{0.3}Ga_{0.7}As$/5-nm-GaAs. Generally speaking, due to the surface kinetic processes, a certain degree of GaAs and AlGaAs intermixing is inevitable at the heterointerface [114]. The intermixing is enhanced by postgrowth treatments [110–113, 115–118]. The GaAs/AlGaAs heterointerfaces are described by the degree of the Al interdiffusion [119–121]

$$x(z) = \frac{x_0}{2}\left[2 + \mathrm{erf}\left(\frac{z - d_W/2}{2L}\right) - \mathrm{erf}\left(\frac{z + d_W/2}{2L}\right)\right] \quad (26)$$

where x_0 is the initial Al concentration in the barrier, d_W is the quantum well width, L is the diffusion length, and erf is the error function. The center of the well is located at $z = 0$.

The Al diffusion length is determined by matching the peak positions of the theoretical absorption and experimental photocurrent spectra. Numerical results of the Al diffusion length are listed in Table 1. It is concluded by fitting the peak positions that the Al diffusion length increases from 2.9 to 6.5 Å when the dose of ion implantation increases up to 2.5×10^{15} cm^{-3}. In addition to the redshift of the absorption peak we have observed an increased absorption intensity following the increase of the Al diffusion length. It is due to the increased coupling between the ground state and excited states when the energy separation between them decreases. However, the photocurrent decreases experimentally following the increased H^+ dose.

On the other hand, the density of thermally excited carriers increases by a factor of 4.6 when L increases from 2.9 to 6.5 Å, while experimentally the dark current increases by a factor of only 2. It must thus be concluded that we have to take into account the microscopic mechanism (which has reduced the carrier lifetime) when studying the carrier transport property. Postgrowth processes enhance various scatterings so that the lifetime of carriers at excited states is reduced, and the absorption coefficient is therefore decreased. The concept of carrier lifetime is introduced for steady states. In studying the carrier transport we consider

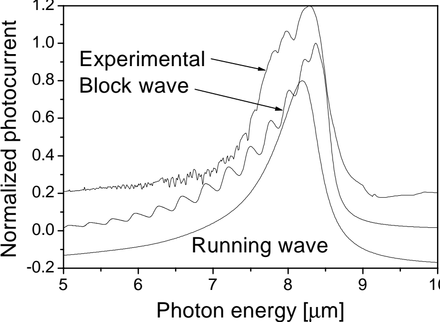


Figure 7. Calculated optical absorption rates and a typical photocurrent spectrum of a QWIP measured at 77 K. Reprinted with permission from [278], Y. Fu, *Superlattices Microstruct.* 30, 69, (2001). © 2001, Elsevier Science.

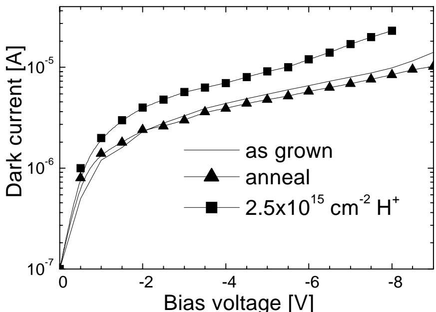


Figure 8. Dark current spectra of an as-grown QWIP sample, annealed and ion-implanted with a H^+ dose of 2.5×10^{15} cm^{-3}. Reprinted with permission from [279], Y. Fu and M. Willander, *J. Appl. Phys.* 88, 288 (2000). © 2000, American Institute of Physics.

Table 1. Numerical values of physical quantities.

H^+ dose [10^{15} cm^{-3}]	0	2.5	Exp.
Diffusion length L [Å]	2.9	6.5	
Peak absorption coefficient	24.75	34.61	
n'_{th} [10^7 cm^{-3}]	6.875	31.978	
Carrier mobility	0.82139	0.32895	
Dark current	1	1.8	2.0
Peak photocurrent	1	0.56	0.65

unperturbed electronic states and the transport properties of carriers occupying these states are perturbed by various scattering centers. At the normal QWIP device working temperature of 77 K, alloy scattering is expected to be dominant in GaAs/AlGaAs-based materials. Phonon and impurity scatterings are limited due to the low device working temperature and spatially localized doping profile.

We consider a ternary compound $A_xB_{1-x}C$, where the mole fraction x of atom A varies along the sample growth direction (z-direction) so that $x = x(z)$. Now the potential energy $V(\boldsymbol{r})$ for the one-electron Schrödinger equation is divided into a virtual crystal part

$$V_0(\boldsymbol{r}) = \sum_{\boldsymbol{a}} [xV_A(\boldsymbol{r}-\boldsymbol{a}) + (1-x)V_B(\boldsymbol{r}-\boldsymbol{a})] \tag{27}$$

and a random potential part

$$V'(\boldsymbol{r}) = \sum_{\boldsymbol{a}} c_a [V_A(\boldsymbol{r}-\boldsymbol{a}) - V_B(\boldsymbol{r}-\boldsymbol{a})] \tag{28}$$

where $V_A(\boldsymbol{r}-\boldsymbol{a})$ and $V_B(\boldsymbol{r}-\boldsymbol{a})$ are periodic atomic potentials of compound AC and BC. $c_{\boldsymbol{a}}$ is a random function which is defined only at lattice site $\boldsymbol{a}$,

$$c_{\boldsymbol{a}} = \begin{cases} (1-x) & \text{for an A atom at } \boldsymbol{a} \\ -x & \text{for a B atom at } \boldsymbol{a} \end{cases} \tag{29}$$

The total scattering rate of state $\boldsymbol{ik}$ [122] is

$$M(\boldsymbol{ik}) = \Delta^2[1 - f(E_{ik})] \int [1 - x(z)]x(z)|\phi_i(z)|^4\, dz \tag{30}$$

The velocity relaxation time associated with alloy scattering is

$$\frac{1}{\tau(\boldsymbol{ik})} = \frac{2\pi\Delta^2}{\hbar}[1-f(E_{ik})] \int [1-x(z)]x(z)|\phi_i(z)|^4 dz \tag{31}$$

The above expression reduces to the ones of Ando [123] and Bastard [124] when the alloy composition is constant in the barrier region. It further reduces to the one of bulk material when the alloy composition is constant in the whole sample (e.g., see [125]).

The mobility μ_z along the z-direction determined by the alloy scattering then becomes

$$\mu_z = \frac{e}{n'} \int \frac{2d\boldsymbol{K}}{(2\pi)^3} \tau(\boldsymbol{K}) v_z^2(\boldsymbol{K}) \frac{\partial f(E_{\boldsymbol{K}})}{\partial E_{\boldsymbol{K}}} \tag{32}$$

where $\boldsymbol{K} = (\boldsymbol{k}, k)$, and $\boldsymbol{k}$ is the wave vector in the xy-plane. $\hbar v_z = \partial E_{\boldsymbol{K}}/\partial k$. Increasing L from 2.9 to 6.5 Å (corresponding to the H^+ dose of 2.5×10^{15} cm^{-3}), the calculated mobility decreases from 0.82 to 0.33, a factor of 2.5. We then expect an increasing factor of $31.978 \times 0.33/(6.875 \times 0.82) = 1.8$ in the dark current, which agrees very well with Figure 1 in Section 1, where a factor of about 2.0 has been observed. For the photocurrent, it is 0.56–0.65. Detailed data are listed in Table 1.

4. PROCESSING OF INFRARED PHOTODETECTORS

An infrared FPA device consists of two parts. The first one is the sensor array which is used to convert the infrared radiation into the electrically readable signal. In a QWIP, this signal is the photocurrent. The second part of the QWIP is the electrical circuit array, also called the readout circuit (ROC), which is used to transfer the electrically readable signals generated in very pixels of the array in a series manner. Such a signal transferring mode is very convenient for further data acquisition technology. Normally, the infrared sensor array is hybridized with the readout circuit by indium bumps. This hybridization method is very efficient to improve the total FPA performance, with the combination of the best infrared sensor array and the very well developed silicon based integrated circuit technology. The disadvantage of the hybridization method is its complexity during device processing. A brief description of the QWIP FPA device is schematically shown in Figure 9. The up contact layer of the QWIP material is also used for fabricating the optical grating to couple the infrared (IR) radiation with the inter-subband transition, while the bottom contact layer is a common electrode of the FPA device. By the hybridization method, the QWIP sensor array is connected with the Si readout circuit to form a QWIP FPA. The IR radiation is introduced to the system from the back side of the QWIP material.

A typical QWIP material for long wavelength infrared detection has 50 periods of quantum well and barrier sandwiched between two n^+-GaAs contact layers. Each period consists of one 4.5-nm-thick quantum well and one 50-nm-thick $Al_{0.3}Ga_{0.7}As$ barrier. The quantum wells are doped with Si at a level of about 5×10^{17} cm^{-3}. The contact layers are doped at a level of 1×10^{18} cm^{-3} for ohmic contact. Such a material structure can be further optimized for specific device application requirements. There are two basic considerations in designing a QWIP FPA device: (1) matching the quantum well width and barrier height with the detecting wavelength and (2) increasing the doping concentration to enhance the absorption coefficient within the limit of

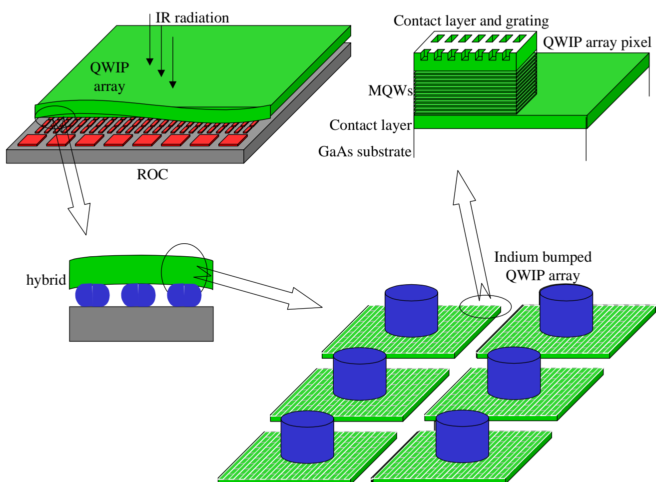


Figure 9. Schematic presentation of the QWIP FPA device.

the dark current that should be acceptable by the readout circuit.

A detailed device process is described schematically by Figure 10. Unlike the single infrared detector fabrication, special attention concerning the uniformity of all pixels in the FPA is to be emphasized in the FPA fabrication. Almost every step will introduce some nonhomogeneous factors. For example, the first step in the process is to fabricate the optical grating on all pixels in Figure 10 to couple the infrared radiation with the inter-subband transition in the quantum wells. The nonhomogeneity in the grating fabrication will result in different optical coupling efficiencies in different pixels. The second step is to make the mesa for all pixels. Both the wet etching and dry etching can be used to create a square mesa with a typical size from 30×30 to 50×50 μm^2. Any fluctuation in the mesa etching process among different pixels will induce different mesa surface states. These surface states will in turn induce different dark currents. In order to make a good ohmic contact between the In bump and the GaAs contact layer, the alloying between the GaAs contact layer and an AuGeNi/Au alloy metal is to be made, on which the In bump is evaporated. The In bump is typically a cylinder with a diameter of about 10–20 μm and a height of 10 μm. Using these In bumps the QWIP array is hybridized with the readout circuit. In a very large FPA format, fluctuations in the In bump diameter and height will affect the electric connection property between the QWIP array and the readout circuit. If an optical pixel is not properly connected with the readout circuit, a blind point appears in the FPA. Finally the QWIP array is mechanically thinned to release the thermal expansion mismatch between the GaAs based material and Si material. A thin QWIP array is also preferred so as to avoid optical cross-talking and to enhance the multi-reflection-induced infrared radiation absorption. At this moment, the QWIP FPA device process is completed.

For the destination application, the device is packaged into a cooling Dewar at a temperature of about 77 K. Normally

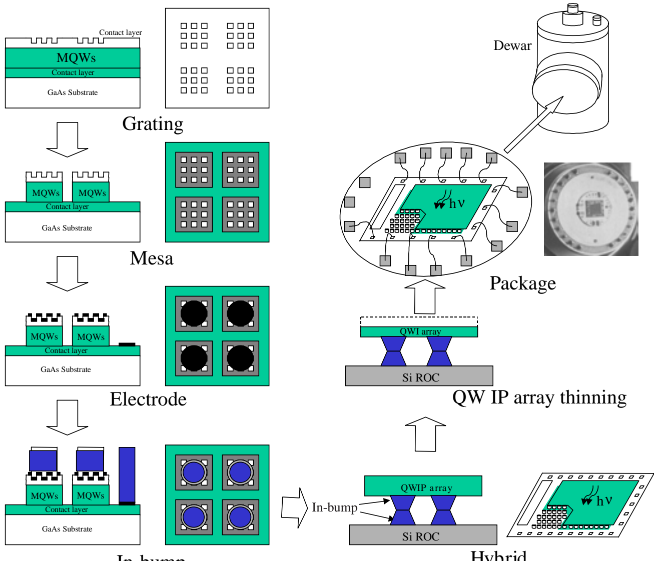


Figure 10. Schematic description of QWIP FPA device process steps.

for the same response wavelength, the dark current of the QWIP detector increases much faster than that of a HgCdTe detector with temperature because of the thermal contact. The packaging process should be much more careful to ensure perfect thermal contact between the QWIP chip and the cooling head.

One important aspect in the FPA fabrication is the optical coupling efficiency. A detailed electromagnetic wave modeling about the light coupling mechanism in the QWIP has been developed [126]. Optical gratings with different spacings were used to match the two-color QWIP device at 4.8 and 9.4 μm [127]. A corrugated QWIP structure was also proposed by chemically etching linear V-grooves through the detector active region, where normal incident light is modified through internal reflections for optical absorption so that the coupling efficiency of the structure does not depend on the wavelength or the grating size [128]. High coupling efficiency was obtained by random reflectors in a very long wavelength (15 μm) infrared QWIP 128×128 FPA device [129], and the prism–film coupler concept was also proposed, which can turn the incident electric field from parallel polarization to perpendicular polarization with respect to the quantum well layer surface, thereby increasing dramatically the absorption coefficient [130]. Direct optical coupling in inter-subband transitions utilizes the AlGaSb multiple-valley quantum-well material, where strong normal-incidence electron absorption can be achieved using asymmetrically stepped quantum wells in contrast to rectangular wells [131].

5. DEVICE DEVELOPMENT

In 1982, a GaAs/GaAlAs quantum well structure was proposed as a potentially new type of infrared photodetector for 3–5 and 8–10 μm bands [132]. Up to now, there has been very extended progress on both the common QWIP structures and some new structures based on the inter-subband transitions.

5.1. *n*-Type QWIPs

Many *n*-type QWIP device structures have been investigated, including bound-to-bound, bound-to-continuum, bound-to-quasibound, bound-to-bound miniband, bound-to-continuum miniband, and bound-to-miniband structures [133–141]. The first bound-to-bound state QWIP was reported at an absorption peak at 10.9 μm and a net absorption of 5% [133]. An increased barrier thickness structure was used to reduce the dark current [134]. Similar to the bound-to-bound QWIP structure, a bound-to-continuum QWIP was realized by decreasing the quantum well width so that the energy position of the first excited state is higher than the barrier height. This makes the photoexcited electrons free to move away from the quantum well. A detectivity of 3×10^{10} cm$\sqrt{\text{Hz}}$/W at 68 K with a cutoff wavelength of 10 μm was reported [135]. In order to improve the QWIP performance, a bound-to-quasibound structure was designed to further reduce the dark current [136]. In this structure

the first excited state is placed exactly at the well top. The bound-to-bound miniband QWIP was experimentally demonstrated in the response wavelength range from 3.6 to 6.3 μm [137], using the superlattice miniband to replace the discrete sublevel in the single quantum well. A square step barrier at the end of superlattice was used to reduce the ground state miniband tunneling dark current [138]. By placing the excited state miniband in the continuum levels, one can improve the transportation ability of photoexcited electrons. This kind of QWIP has been demonstrated in the 5–9 μm spectral range [139]. Finally the bound-to-miniband structure has shown its excellent long wavelength infrared images in a 256×256 FPA camera [141]. The structure consisted of two bound states; the higher energy level is resonant with the ground state miniband in the superlattice barriers. For the thermal infrared image and thermal infrared imaging spectrometer applications, a broadband QWIP in the 6–20 μm spectral range has been developed [142]. Such a detector has been mated to a 640×480 multiplexer.

5.2. *p*-Type QWIPs

The valence band structure is normally much more complicated compared to the conduction band structure due to the nonparabolic and anisotropic effects and the valence subband mixing effects. However, the complications result in the direct absorption of the normal incident IR radiation by the *p*-type material-based QWIPs. These *p*-type QWIPs have thus been introduced and studied [143–146]. The electronic structure and wave functions of the quantum wells have been calculated by the 8×8 envelope-function model to study the photoresponse as a function of the quantum well width, alloy concentration, polarization, and temperature [147]. The dark current behavior has also been studied [148], and the noise was attributed to a number of fluctuation noises associated with generation and recombination of holes from and to the quantum-well bound states and the extended valence-band states [149].

InGaAs/GaAs, InGaAs/InP, InGaAs/AlGaAs, and GaAs/GaInAsP MQWs have very good detectivities in the middle IR wavelength region [150–159]. AlGaAs/GaAs and AlGaAs/InGaAs MQWs grown on (311) GaAs have also been investigated for *p*-type QWIP device applications [160]. The GaAs/AlGaAs *p*-type QWIP device has been systematically studied as far as operation performance as a function of the barrier height and the quantum well width [161, 162]. The photoresponse of the detector ranges from 1.4 to 15 μm [163]. By aligning the first heavy hole subband as the ground state and the second light hole subband in resonance with the top of the barrier, the detector shows a background-limited operation temperature of 100 K for the 3–5 μm response [164]. The optimum two-dimensional doping density obtained for this structure is in the range of $(1\text{–}2) \times 10^{12}$ cm^{-2} [165]. A resonant cavity can increase the absorption efficiency of the *p*-type GaAs/AlGaAs quantum-well peak absorption by 25% [166]. It was also found that a QWIP with three wells of $In_{0.25}Ga_{0.75}As$ of different thicknesses (4.5, 5.5, and 6.2 nm) and an undoped GaAs barrier (40 nm thick) shows a broadband IR detection in the range of 7–14 μm [167]. A very long wavelength detection with a peak response at 19.2 μm has also been demonstrated by the *p*-type compressively strained $In_{0.27}Ga_{0.73}As/Al_{0.15}Ga_{0.85}As$ QWIP [168]. The detectivity of 5.9×10^{10} cm$\sqrt{\text{Hz}/\text{W}}$ at the peak wavelength of 8.1 μm and operation temperature 77 K was obtained [169]. More recently, the *p*-type GaAs/GaInP QWIPs show background-limited performance up to 120 K, while the *p*-type GaInAsP/GaInAsP QWIPs exhibit a rather broadband response window of 2.5–10 μm [170].

5.3. QWIP Light Emitting Device

An innovative concept of frequency up-conversion based on QWIP devices has been proposed [171, 172]. The conversion of the middle infrared radiation into near infrared or visible radiation utilizing inter-subband transitions in the QWIP and interband transitions in the light emitting device (LED) has been demonstrated. A large area QWIP is covered by a near infrared LED to fulfill the frequency up-conversion. The up-converted photons generated by the detection of longer wavelength infrared radiations were collected by a Si charge coupled device, resulting in the up-converted image from the infrared source image [172–174]. Recently, a large pixelless thermal imager based on the up-conversion technique has shown its temperature resolution of 1 K by converting a mid-infrared (9 μm) scene to a near-infrared (0.82 μm) emission image [175]. A multicolor infrared photodetector based on this up-conversion concept has been achieved. A remarkable selectivity of the device has been achieved between wavelengths of 9.1 and 4.85 μm by bias voltage adjustment [176].

5.4. Multicolor QWIP

With the development of the IR technology, the integrated multicolor detector is very advantageous in increasing the target recognition capability. The multicolor, especially having response wavelengths in two atmosphere windows, 3–5 and 8–12 μm, is of great interest. Very extended works on multicolor QWIP devices have been published. There are two typical methods for multicolor detection. One is the voltage tuning method based on the Stark effect of the confined electron states; the other is based on multistack MQWs in the QWIP structure [177–186]. Theoretically, it was observed that a strong Stark effect can tune the response wavelength from 2.9 to 4.2 μm by an electric field variation of 90 to -90 kV/cm [187, 188], and from 7.4 to 14 μm when the electric field changes from 60 to -60 kV/cm [189, 190]. Experimentally a response wavelength tuning from 8.2 to 9.1 μm and 10.8 to 11.5 μm was demonstrated [183].

In the simplest structure, one can electrically tune the response wavelength of a GaAs/AlGaAs asymmetric step quantum well from 8.5 to 13.5 μm [191]. In the multistack MQW structure, the wavelength tuning behavior is determined by both the dc resistance and the dynamic resistance of individual detectors [180, 181, 184, 192, 193]. A structure with four stacks of MQWs for four different detection wavelengths at 4.7, 8.5, 9.0, and 12.3 μm has been fabricated from InGaAs/AlGaAs and GaAs/AlGaAs systems [194].

Recently, a two-color QWIP device has also been demonstrated based on electron transfer between two asymmetric

coupled quantum wells under an applied bias [195, 196]. The intermixing technology in the nanostructure of quantum wells has also shown its potential application of the multicolor detection [197–201].

5.5. Photovoltaic Quantum Well Detector

Similar to the HgCdTe materials, a quantum well photoconductive device has a much larger dark current than the photovoltaic (PV) device, so that the PV device is more suitable in FPA device applications to match with the external readout circuit. Unlike the conventional PV device with the *p–n* junction, an asymmetric nanostructure is used in the quantum well device to generate the necessary built-in field. Many types of structures have been used, including AlGaAs/GaAs and InAs/GaInSb superlattice structures [202–204], modulation doping [205], nanosize AlAs layers induced asymmetric structures [206–209], asymmetric quantum well structures [210], and double-barrier quantum wells [211–216]. A structure consisting of graded barrier between three quantum wells also shows the PV effect [217]. A MQW structure operating in a photovoltaic detection mode with an intrinsic event discrimination capability has also been developed, and the device shows a peak responsivity of 0.4 A/W and a detectivity of 6.0×10^{11} Jones [218]. The dynamic process of the photoexcited carrier was also studied, both theoretically [219] and experimentally, by infrared pulse excitation [220–222]. In order to enhance the PV effect, a four-zone structure was proposed, where every period of the active region consists of an excitation zone, a drift zone, a capture zone, and a tunneling zone [223]. The PV structure was also combined with a PV structure for a two-color device application [224–226].

5.6. High Frequency Response Detectors

Infrared communication needs a very fast response infrared detector. Like optical communication at the wavelength of 1.3–1.5 μm, operation frequencies in the range of GHz are strongly required. For the CO_2 laser based communication, a GHz frequency response detector at the wavelength of 10.6 μm becomes the key device. For conventional infrared detectors, such a high frequency requirement does not impose any problems since the lifetime of photoexcited carriers is rather long. The QWIP device can also operate at such a high GHz response frequency, since the lifetime of photoexcited carriers is about 5 ps [227]. By the optical heterodyne detection technique, an experiment with an intermediate operation frequency up to 26.5 GHz was reported [228]. Furthermore, by combining the conventional heterodyne experimental method with an additional microwave excitation, a frequency separation of two infrared beams up to 82.16 GHz has been achieved [229]. Moreover, theoretical and experimental works focused on the high frequency capability of QWIP are useful for specific applications [230–234].

5.7. FPA Devices

For FPA device applications, the QWIP device should be matched with the external readout circuit electrically. To achieve this, optimization and trade-off of the QWIP operation parameters, including responsivity, dark current, dynamic resistance, noise current, optical gain, and detectivity, are very important and, at the same time, rather complicated [235]. The dark current is one of the key parameters, which should be reduced as much as possible from the viewpoint of perfectly matching the QWIP arrays with the readout circuit. An infrared hot-electron transistor was proposed

Table 2. Optical transition structures of quantum well infrared photodetectors.

Optical transition	Material	Advantage	Disadvantage	Ref.
Bound–bound	GaAs/AlGaAs	high absorption, low dark current	low responsibility, normal incident absorption forbidden	[133, 134]
Bound–continuum	GaAs/AlGaAs InGaAs/InAlAs InGaAs/InP GaInAsP/InP GaAs/GaInP GaAs/AlInP InGaAs/GaAs	high responsibility	low absorption, high dark current, normal incident absorption forbidden	[135, 261–271]
Bound–quasibound	GaAs/AlGaAs	high responsibility, high absorption, low dark current	normal incident absorption forbidden	[136]
Bound–bound miniband	GaAs/AlGaAs	low dark current, low noise, good for very long wavelength response	low responsibility, normal incident absorption forbidden	[137, 138, 272, 273]
Bound–continuum miniband	GaAs/AlGaAs	high responsibility	high dark current, normal incident absorption forbidden	[139]
Bound–miniband	GaAs/AlGaAs, GaAs/AlGaAs/InGaAs	high responsibility, low dark current	normal incident absorption forbidden	[140, 141, 274, 275]
p-Type	GaAs/AlGaAs InGaAs/InP	normal incident absorption	low responsibility, high dark current	[143–158] [159, 161, 162] [166–170]

Table 3. Device structures of QWIPs.

Device structure	Material	Advantage	Disadvantage	Ref.
QWIP-LED	GaAs/AlGaAs, GaAs/InGaAs	optical readout, pixelless for FPA	fabrication complex	[171–176]
Multicolor QWIP	GaAs/AlGaAs, GaAs/InGaAs/AlGaAs	two-color or multicolor detection	fabrication complex	[177–190, 192–201]
Photovoltaic QWIP	GaAs/AlGaAs	low dark current	low resposibility	[202–226]
High frequency response QWIP	GaAs/AlGaAs	very fast response	low responsibility	[227, 228, 230–234]
QWIP FPA	GaAs/AlGaAs	good for infrared technology application	fabrication complex	[235–244, 246, 247]

to reduce the dark current [236]. For other conventional QWIP structures, a low-bias current should be used [237, 238]. Recently, the feasibility of monolithic integration of QWIP with the readout circuit on the same GaAs/AlGaAs material was demonstrated, where a charge storage capability of 2×10^7 electrons in a 50×50 μm^2 area was obtained [239].

As a single device for high temperature operation (>70 K), the HgCdTe material is better than the QWIP material due to the fundamental limitations associated with inter-subband transitions. However, the QWIP is advantageous in the fabrication of large format FPAs due to its high impedance, fast response time, long integration time, and low power consumption [240]. The 128×128 QWIP FPAs at 7.7 μm have been shown to have a high performance at 78 K [241]. For long wavelengths, the FPAs with 256×256 and 640×512 pixels have been demonstrated to have a noise equivalent temperature difference $NE\Delta T$ of only 10 and 20 mK, respectively [242]. A similar result was also reported from a 640×480-pixel single-color FPA, which has a $NE\Delta T$ of 30 mK at 60 K operation temperature [243]. A very useful two-color array for both mid- and long wavelength was reported having a FPA format of 256×256 [243]. Moreover, a 640×480 format monolithically integrated 8–9 and 14–15 μm two-color QWIP FPA was reported [244].

In the practical applications of QWIP FPAs, the 256×256 handheld QWIP long wavelength infrared camera has shown its various possible applications in science, industry, and medicine fields. The hot spots, which are not normally visible, have been clearly revealed in QWIP cameras. The QWIP cameras are perfect fire surveillance devices [245], and the volcanic observation feature (with a temperature variation from 300 to 1000 °C) demonstrates the wide dynamic range of the QWIP [246]. In medical applications, the sensitivity of the QWIP revealed arterial plaque inside a heart [247]. The excellent $1/f$ noise performance of the QWIP FPA operating at 35 K has been successfully used in the slow scan strategy, which is often required in infrared observation from space [245].

6. CONCLUSIONS

Quantum well technology is a typical and a major one-dimensional nanotechnology. With this, one is able to control the electron behavior in semiconductors. For quantum well infrared photodetector applications, detector key parameters can be adjusted from the nanoscale of atomic monolayers to optimize the infrared detection. Tables 2 and 3 list important aspects of the QWIP materials and devices. This novel QWIP will be an important member in the infrared detector family in the very near future. Meanwhile, the QWIP detector performance can be further optimized and new device concepts have been reported based on deeper theoretical investigation and the advancing MBE material growth technology. It has been clearly shown that the commercialization of the QWIP has been very successful due to its maturation of the GaAs based material and related device process technology. Moreover, the intensive research on the QWIP in the past 20 years has greatly improved its IR detecting ability, which is now almost identical to the typical IR materials of HgCdTe and InSb. Current and future research activities concentrate on the normal incident response, dark current reduction, high temperature operation, very long wavelength detection, very large format FPA devices, and monolithic integration. It has been suggested that by further increasing the electron confinement (e.g., to work with quantum dots), higher detector operation temperature and normal incidence absorption coefficients can be achieved [248–258]. The QWIP material grown on Si substrate has also shown its high application potential [259], and special structures have been proposed to reduce the dark current [260].

GLOSSARY

Focal plan array (FPA) An FPA consists of either a one-dimensional or a two-dimensional array of QWIPs sensor. The optical response is fulfilled by the QWIP sensor. Then the electrical signal from the QWIP sensor is picked up by the readout circuit in one by one mode.

Infrared photodetector A device converts that the infrared photon into an electric signal which is ready for use.

Multicolor QWIP One type of QWIP device. With this device, more than one wavelengths can be selectively detected. The electric field can tune the detection wavelength due to the Stark effect in asymmetric quantum well structures. Or one can simply stack QWIP structures with different response wavelengths together during material growth and then pick out the detection signal separately.

Photovoltaic quantum well detector Similar to a normal QWIP device, but a built-in electrical field is created in

the quantum well region by the asymmetric quantum well structure design. Then the photoexcited carrier will be driven by the built-in field and result in the electrical signal without external bias.

Quantum well infrared photodetector (QWIP) Energy sublevels form in the multiple GaAs/AlGaAs quantum wells. Electrons at low-energy sublevels (normally confined in the GaAs quantum wells) absorb infrared photons and then jump to high-energy sublevels (in the AlGaAs barriers), forming a photocurrent in an external bias induced electric field.

REFERENCES

1. Y. Fu and M. Willander, in "Handbook of Advanced Electronic and Photonic Materials and Devices" (H. S. Nalwa, Ed.), Ch. 5. Academic Press, San Diego, 2001.
2. S. Kim and M. Razeghi, in "Handbook of Advanced Electronic and Photonic Materials and Devices" (H. S. Nalwa, Ed.), Vol. 2, Ch. 3. Academic Press, San Diego, 2001.
3. K. K. Choi, "The Physics of Quantum Well Infrared Detector" (I. Dzyaloshinski, S. Lundqvist, and Y. Lu, Eds.), Series in Modern Condensed Matter Physics, Vol. 17. World Scientific, Singapore, 1997.
4. H. C. Liu and F. Capasso, "Intersubband Transitions in Quantum Wells, Physics and Device Applications I" (R. K. Willardson and E. R. Weber, Eds.), Semiconductors and Semimetals, Vol. 62. Academic Press, San Diego, 2000.
5. A. Rogalski, *Infrared Phys. Technol.* 43, 187 (2002).
6. M. A. Gadir, P. Harrison, and R. A. Soref, *J. Appl. Phys.* 91, 5820 (2002).
7. K. K. Choi, S. V. Bandara, S. D. Gunapala, W. K. Liu, and J. M. Fastenau, *J. Appl. Phys.* 91, 551 (2002).
8. T. N. Casselman, *Proc. SPIE* 2999, 2 (1997).
9. A. Rogalski, *Infrared Phys. Technol.* 38, 295 (1997).
10. S. D. Gunapala, S. V. Bandara, J. K. Liu, W. Hong, M. Sundaram, P. D. Maker, R. E. Muller, C. A. Shott, and R. Carralejo, *IEEE Trans. Electron Devices* 45, 1890 (1998).
11. B. F. Levine, *J. Appl. Phys.* 74, R1 (1993).
12. M. F. Wan, H. J. Ou, W. Lu, X. Q. Liu, X. S. Shen, N. Li, N. Li, S. C. Shen, W. X. Wang, Q. Huang, and J. M. Zhou, *J. Infrared Millimeter Waves* 17, 76 (1998).
13. S. C. Shen, *Microelectron. J.* 25, 713 (1994).
14. W. Lu, H. J. Ou, M. H. Chen, Z. H. Ma, X. Q. Liu, X. L. Huang, H. B. Mao, W. Jiang, S. C. Shen, Y. H. Gu, and L. B. Ye, *J. Infrared Millimeter Waves* 13, 9 (1994); W. Lu, H. J. Ou, M. H. Chen, X. L. Huang, S. C. Shen, R. H. Gu, and L. B. Ye, *J. Infrared Millimeter Waves* 15, 137 (1994).
15. P. R. Bratt and T. N. Casselman, *J. Vac. Sci. Technol. A* 3, 238 (1985).
16. K. Kosai, *J. Electron. Mater.* 24, 635 (1995).
17. C. R. Helms, J. L. Melendez, H. G. Robinson, S. Holander, J. Hasan, and S. Halepete, *J. Electron. Mater.* 24, 1137 (1995).
18. Y. Fu, M. Willander, W. Lu, and W. L. Xu, *J. Appl. Phys.* 84, 5750 (1998).
19. Y. Fu, M. Willander, W. Lu, W. L. Xu, Ning Li, Na Li, X. Q. Liu, Y. D. Chen, and S. C. Shen, *J. Appl. Phys.* 85, 1237 (1999).
20. Y. Fu, N. Li, M. Karlsteen, M. Willander, Na Li, W. L. Xu, W. Lu, and S. C. Shen, *J. Appl. Phys.* 87, 511 (2000).
21. Na Li, L. Fu, Ning Li, Y. C. Chan, W. Lu, S. C. Shen, H. H. Tan, and C. Jagadish, *J. Cryst. Growth* 222, 786 (2001).
22. E. Bauer, Reflection electron diffraction, in "Techniques of Metals Research" (R. F. Bunshah, Ed.), Vol. 2, Ch. 15. Wiley-Interscience, New York, 1969.
23. P. K. Larsen, P. J. Dobson, J. H. Neave, B. A. Joyce, and B. Bolger, *Surf. Sci.* 169, 176 (1986).
24. B. A. Joyce, P. J. Dobson, J. H. Neave, K. Woodbridge, J. Zhang, and P. K. Larsen, *Surf. Sci.* 168, 423 (1986).
25. B. A. Joyce, P. J. Dobson, J. H. Neave, and J. Zhang, *Surf. Sci.* 178, 110 (1986).
26. J. H. Neave, B. A. Joyce, P. J. Dobson, and N. Norton, *Appl. Phys. A* 31, 1 (1983).
27. P. J. Dobson, B. A. Joyce, J. H. Neave, and J. Zhang, *J. Cryst. Growth* 81, 1 (1987).
28. J. H. Neave, P. J. Dobson, B. A. Joyce, and J. Zhang, *Appl. Phys. Lett.* 47, 400 (1985).
29. J. M. Van Hove and P. I. Cohen, *J. Cryst. Growth* 81, 13 (1987).
30. B. F. Lewis, J. C. Lee, F. J. Gruthaner, A. Madhukar, B. Fernandez, and J. Maserjian, *J. Vac. Sci. Technol. B* 2, 419 (1984).
31. P. Chen, A. Madhuka, J. Y. Kim, and J. C. Lee, *Appl. Phys. Lett.* 48, 650 (1986).
32. J. M. Van Hove, P. R. Pukite, and P. I. Cohen, *J. Vac. Sci. Technol. B* 3, 563 (1985).
33. J. D. Weeks and G. H. Gilmer, *Adv. Chem. Phys.* 40, 157 (1979).
34. B. A. Joyce, P. J. Dobson, J. H. Neave, K. Woodbridge, J. Zhang, P. K. Larsen, and B. Boelger, *Surf. Sci.* 168, 423 (1986).
35. M. A. Herman, H. Sitter, "Molecular Beam Epitaxy: Fundamentals and Current Status" (M. B. Panish, Ed.), Springer Series in Materials Science, Vol. 7, p. 134. Springer-Velag, Berlin/Heidelberg, 1989.
36. M. Tanaka and H. Sakaki, *J. Cryst. Growth* 81, 153 (1987).
37. C. E. C. Wood, D. Desinione, and S. Judaprawira, *J. Appl Phys.* 51, 2074 (1980).
38. M. Ogawa and T. Baba, *Jpn. J. Appl. Phys.* 24, L572 (1985).
39. M. Heiblum, W. I. Wang, L. E. Osterling, and V. Deline, *J. Appl. Phys.* 54, 6751 (1983).
40. D. H. Zhang, K. Radhakrishnan, and S. F. Yoon, *J. Cryst. Growth* 135, 441 (1994).
41. Y. Fu and M. Willander, "Physical Models of Semiconductor Quantum Devices," pp. 75–92. Kluwer Academic, Boston, 1999.
42. D. F. Nelson, R. C. Miller, and D. A. Kleinman, *Phys. Rev. B* 35, 7770 (1987).
43. K. H. Yoo, L. R. Ram-Mohan, and D. F. Nelson, *Phys. Rev. B* 39, 12808 (1989).
44. U. Ekenberg, *Phys. Rev. B* 36, 6152 (1987).
45. U. Ekenberg, *Phys. Rev. B* 40, 7714 (1989).
46. D. J. Newson and A. Kurobe, *Semicond. Sci. Technol.* 3, 786 (1988).
47. P. von Allmen, M. Berz, G. Petrocelli, F.-K. Reinhart, and G. Harbeke, *Semicond. Sci. Technol.* 3, 1211 (1988).
48. T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.* 54, 437 (1982).
49. E. Batke, G. Weimann, and W. Schlapp, *Phys. Rev. B* 43, 6812 (1991).
50. C. W. Cheah, G. Karunasiri, L. S. Tan, and L. F. Zhou, *Appl. Phys. Lett.* 80, 145 (2002).
51. J. Hernando, J. L. Sanchez-Rojas, A. Guzman, E. Munoz, J. M. G. Tijero, D. Gonzalez, G. Aragon, and R. Garcia, *Appl. Phys. Lett.* 78, 2390 (2001).
52. E. Dupont, H. C. Liu, S. R. Schmidt, and A. Seilmeier, *Appl. Phys. Lett.* 79, 4295 (2001).
53. K. Kishino and S. Arai, "Handbook of Semiconductor Lasers and Photonic Integrated Circuits," Ch. 11, p. 350. Chapman & Hall, London, 1994.
54. J. C. Stover, "Optical Scattering: Measurement and Analysis," p. 51. McGraw–Hill, New York, 1990.
55. J. Cowley, "Diffraction Physics," p. 11. Elsevier, Amsterdam, 1995.
56. J. Barenz, A. Eska, O. Hollricher, O. Marti, M. Wachter, U. Schoffel, and H. Heinecke, *Appl. Opt.* 37, 106 (1998).
57. J. Y. Andersson, L. Lundqvist, and Z. F. Paska, *Appl. Phys. Lett.* 58, 2264 (1991); J. Y. Andersson and L. Lundqvist, *Appl. Phys.*

Lett. 59, 857 (1991); J. Y. Andersson and L. Lundqvist, *J. Appl. Phys.* 71, 3600 (1992).
58. K. W. Goosen and S. A. Lyon, *Appl. Phys. Lett.* 47, 1257 (1985).
59. G. Hasnain, B. F. Levine, C. G. Bethea, R. A. Logan, J. Walker, and R. J. Malik, *Appl. Phys. Lett.* 54, 2515 (1989).
60. G. Sarusi, B. F. Levine, S. J. Pearton, K. M. S. V. Bandara, and R. E. Leibenguth, *J. Appl. Phys.* 76, 4989 (1994).
61. S. Bandara, S. Gunapala, J. Liu, W. Hong, and J. Park, *Proc. SPIE* 2999, 103 (1997).
62. G. Sarusi, B. F. Levine, S. J. Pearton, K. M. S. V. Bandara, and R. E. Leibenguth, *Appl. Phys. Lett.* 64, 960 (1994).
63. B. Xing and H. C. Liu, *J. Appl. Phys.* 80, 1214 (1996).
64. L. P. Rokhinson, C. J. Chen, D. C. Tsui, G. A. Vawter, and K. K. Choi, *Appl. Phys. Lett.* 74, 759 (1999).
65. J. Mao, Amlan Majumdar, K. K. Choi, D. C. Tsui, K. M. Leung, C. H. Lin, T. Tamir, and G. A. Vawter, *Appl. Phys. Lett.* 80, 868 (2002).
66. L. P. Rokhinson, C. J. Chen, K. K. Choi, D. C. Tsui, G. A. Vawter, L. Yan, M. Jiang, and T. Tamir, *Appl. Phys. Lett.* 75, 3701 (1999).
67. K. K. Choi, *Proc. SPIE* 4646, 79 (2002).
68. S. D. Gunapala, J. S. Park, G. Sarusi, T. L. Lin, J. K. Liu, P. D. Maker, R. E. Muller, C. A. Shott, T. Hoelter, and B. F. Levine, *IEEE Trans. Electron Devices* 44, 45 (1997).
69. S. D. Gunapala, J. K. Liu, J. S. Park, M. Sundaram, C. A. Shott, T. Hoelter, T. L. Lin, S. T. Masie, P. D. Maker, R. E. Muller, and G. Sarusi, *IEEE Trans. Electron Devices* 44, 51 (1997).
70. K. K. Choi, A. C. Goldberg, N. C. Das, M. D. Jhabvala, R. B. Bailey, and K. Vural, *Proc. SPIE* 3287, 118 (1998).
71. K. K. Choi, C. J. Chen, A. C. Goldberg, W. H. Chang, and D. C. Tsui, *Proc. SPIE* 3379, 441 (1998).
72. F. S. Pool, D. W. Wilson, P. D. Maker, R. E. Muller, J. J. Gill, D. K. Sengupta, J. K. Liu, S. V. Bandara, and S. D. Gunapala, *Proc. SPIE* 3379, 402 (1998).
73. M. Ershov, *J. Appl. Phys.* 89, 6253 (2001).
74. A. Carbone and P. Mazzetti, *Infrared Phys. Technol.* 42, 185 (2001).
75. H. C. Liu, L. Li, M. Buchanan, and Z. R. Wasilewski, *J. Appl. Phys.* 82, 889 (1997).
76. M. J. Kane, S. Millidge, M. T. Emeny, D. Lee, D. R. P. Guy, and C. R. Whitehouse, in "Intersubband Transitions in Qautum Wells" (E. Rosencher, B. Viter, and B. F. Levine, Eds.), pp. 31–42. Plenum, New York.
77. H. C. Liu, A. G. Steele, M. Buchanan, and Z. R. Wasilewski, *J. Appl. Phys.* 73, 2029 (1993).
78. B. F. Levine, C. G. Bethea, G. Hasnain, V. O. Shen, E. Pelve, R. R. Abbott, and S. J. Hsieh, *Appl. Phys. Lett.* 56, 851 (1990).
79. S. R. Andrews and B. A. Miller, *J. Appl. Phys.* 70, 993 (1991).
80. M. Ershov, V. Ryzhii, and C. Hamaguchi, *Appl. Phys. Lett.* 67, 3147 (1995).
81. M. Ershov, C. Hamaguchi, and V. Ryzhii, *Jpn. J. Appl. Phys. Lett.* 35, 1395 (1996).
82. L. Thibaudeau, P. Bois, and J. Y. Duboz, *J. Appl. Phys.* 79, 446 (1996).
83. M. Ryzhii and V. Ryzhii, *Appl. Phys. Lett.* 72, 842 (1998).
84. V. Ryzhii, *J. Appl. Phys.* 81, 6442 (1997).
85. L. Thibaudeau, P. Bois, and J. Y. Duboz, *J. Appl. Phys.* 79, 446 (1996).
86. N. E. I. Etteh and P. Harrison, *J. Appl. Phys.* 92, 248 (2002).
87. V. Ryzhii, M. Ryzhii, and H. C. Liu, *J. Appl. Phys.* 92, 207 (2002).
88. R. Rehm, H. Schneider, M. Walther, and P. Koidl, *Appl. Phys. Lett.* 80, 862 (2002).
89. M. Ryzhii, V. Ryzhii, R. Suris, and C. Hamaguchi, *Semicond. Sci. Technol.* 16, 202 (2001).
90. D. A. Cardimona, A. Singh, D. Huang, C. Morath, and P. Varangis, *Infrared Phys. Technol.* 42, 211 (2001).
91. V. Letov, M. Ershov, S. G. Matsik, A. G. U. Perera, H. C. Liu, Z. R. Wasilewski, and M. Buchanan, *Appl. Phys. Lett.* 79, 2094 (2001).
92. S. R. Andrews and B. A. Miller, *J. Appl. Phys.* 70, 993 (1991).
93. R. L. Whitney, K. F. Cuff, and F. W. Adams, in "Semiconductor Quantum Wells and Superlattices for Long-Wavelength Infrared Detectors" (M. O. Manasreh, Ed.), Ch. 3, pp. 55–108. Artech House, Boston, 1993.
94. J. O. Sofo and C. A. Balseiro, *Phys. Rev. B* 42, 7292 (1990).
95. Y. Fu and M. Willander, *J. Appl. Phys.* 71, 3877 (1992).
96. B. F. Levine, C. G. Bethea, G. Hasnain, V. O. Shen, E. Pelve, R. R. Abbott, and S. J. Hsieh, *Appl. Phys. Lett.* 56, 851 (1990).
97. G. M. Williams, R. E. DeWames, C. W. Farley, and R. J. Anderson, *Appl. Phys. Lett.* 60, 1324 (1992).
98. N. Sclar, *Prof. Quantum Electron.* 9, 149 (1984).
99. H. C. Liu, *Appl. Phys. Lett.* 60, 1507 (1992).
100. A. G. Steele, H. C. Liu, M. Buchanan, and Z. R. Wasilewski, *Appl. Phys. Lett.* 59, 3625 (1991).
101. H. C. Liu, *J. Appl. Phys.* 73, 3062 (1993).
102. T. B. Bahder, C. A. Morrison, and H. D. Bruno, *Appl. Phys. Lett.* 51, 1089 (1987).
103. J. D. Bruno, T. B. Bahder, and C. A. Morrison, *Phys. Rev. B* 37, 7098 (1988).
104. N. Zou, J. Rammer, and K. A. Chao, *Phys. Rev. B* 46, 15912 (1992).
105. Z. Ikonic, V. Milanovic, and D. Tjapkin, *Appl. Phys. Lett.* 54, 247 (1989).
106. H. C. Liu, *J. Appl. Phys.* 73, 3062 (1994).
107. Rusli, T. C. Chong, and S. K. Chua, *Jpn. J. Appl. Phys.* 32, 1998 (1993).
108. W. Xu, Y. Fu, and M. Willander, *J. Infrared Millimeter Waves* 16, 86 (1997).
109. K. M. S. V. Bandara, B. F. Levine, and M. T. Asom, *J. Appl. Phys.* 74, 346 (1993).
110. B. Elman, E. S. Koteles, P. Melman, and C. A. Armiento, *J. Appl. Phys.* 66, 2104 (1989).
111. H. H. Tan, J. S. Williams, C. Jagadish, P. T. Burke, and M. Gal, *Appl. Phys. Lett.* 68, 2401 (1996).
112. P. Gavrilovic, D. G. Deppe, K. Meehan, N. Holonyak, Jr., J. J. Coleman, and R. D. Burnham, *Appl. Phys. Lett.* 47, 130 (1985).
113. A. G. Steele, M. Buchanan, H. C. Liu, and Z. R. Wasilewski, *J. Appl. Phys.* 75, 8234 (1994).
114. A. Madhukar, in "Physics of Quantum Electron Devices" (F. Capasso, Ed.), p. 50. Springer-Verlag, Berlin, 1990.
115. J. Y. Chi, X. Wen, E. S. Koteles, and B. Elman, *Appl. Phys. Lett.* 55, 855 (1989).
116. S. Yuan, Y. Kim, H. H. Tan, C. Jagadish, P. T. Burke, L. V. Dao, M. Gal, M. C. V. Chan, E. H. Li, J. Zou, D. Q. Cai, D. J. H. Cockayne, and R. M. Cohen, *J. Appl. Phys.* 83, 1305 (1998).
117. N. Li, Y. Fu, M. Karlsteen, M. Willander, Na Li, X. Q. Liu, X. Z. Yuan, W. Lu, and S. C. Shen, *Appl. Phys. Lett.* 75, 2238 (1999).
118. N. Li, N. Li, W. Lu, X. Q. Liu, X. Z. Yuan, Z. F. Li, H. F. Dou, S. C. Shen, Y. Fu, M. Willander, L. Fu, H. H. Tan, C. Jagadish, M. B. Johnston, and M. Gal, *Superlattices Microstruct.* 26, 317 (1999).
119. J. Crank, "The Mathematics of Diffusion." Clarendon, Oxford, 1956.
120. G. F. Redinbo, H. G. Craighead, and J. M. Hong, *J. Appl. Phys.* 74, 3099 (1993).
121. W. Feng, F. Chen, W. Q. Cheng, Q. Huang, and J. M. Zhou, *Appl. Phys. Lett.* 71, 1676 (1997).
122. Y. Fu and M. Willander, *J. Appl. Phys.* 88, 288 (2000).
123. T. Ando, *J. Phys. Soc. Japan* 51, 3900 (1982).
124. G. Bastard, *Appl. Phys. Lett.* 43, 591 (1983); G. Bastard, *Surface Sci.* 142, 284 (1984).
125. B. K. Ridley, "Quantum Processes in Semiconductors," p. 181. Clarendon, Oxford, 1988.
126. J. Mao, A. Majumdar, K. K. Choi, D. C. Tsui, K. M. Leung, C. H. Lin, T. Tamir, and G. A. Vawter, *Appl. Phys. Lett.* 80, 868 (2002).
127. M. Z. Tidrow, K. K. Choi, A. J. DeAnni, W. H. Chang, and S. P. Svensson, *Appl. Phys. Lett.* 67, 1800 (1995).

128. C.-J. Chen, K. K. Choi, W. H. Chang, and D. C. Tsui, *IEEE Trans. Electron Devices* 45, 1431 (1998); C.-J. Chen, K. K. Choi, M. Z. Tidrow, and D. C. Tsui, *Appl. Phys. Lett.* 68, 1446 (1996).
129. S. D. Gunapala, J. S. Park, G. Sarusi, True-Lon Lin, J. K. Liu, P. D. Maker, R. E. Muller, C. A. Shott, and T. Hoelter, *IEEE Trans. Electron Devices* 44, 45 (1997).
130. T. Cwik and C. Yeh, *J. Appl. Phys.* 86, 2779 (1999).
131. J. M. Jancu, F. Bassani, and P. Voisin, *J. Appl. Phys.* 92, 641 (2002).
132. J. S. Smith, L. C. Chiu, S. Margalit, A. Yariv, and A. Y. Cho, *J. Vacuum Sci. Technol. B* 1, 376 (1983).
133. B. F. Levine, K. K. Choi, C. G. Bethea, J. Walker, and R. J. Malik, *Appl. Phys. Lett.* 50, 1092 (1987).
134. K. K. Choi, B. F. Levine, N. Jarosik, J. Walker, and R. J. Malik, *Appl. Phys. Lett.* 50, 1814 (1987).
135. B. F. Levine, R. J. Malik, J. Walker, K. K. Choi, C. G. Bethea, D. A. Kleinman, and J. M. Vandenberg, *Appl. Phys. Lett.* 50, 273 (1987).
136. S. D. Gunapala, J. K. Liu, M. Sundaram, J. S. Park, C. A. Shott, T. Hoelter, T. L. Lin, S. T. Massie, P. D. Maker, R. E. Muller, and G. Sarusi, *Elec. Chem. Soc.* 95-28, 55 (1995).
137. A. Kastalsky, T. Duffield, S. J. Allen, and J. Harbison, *Appl. Phys. Lett.* 52, 1320 (1988).
138. K. M. S. V. Bandara, J.-W. Choe, M. H. Francombe, A. G. U. Perera, and Y. F. Lin, *Appl. Phys. Lett.* 60, 3022 (1992).
139. S. D. Gunapala, B. F. Levine, and N. Chand, *J. Appl. Phys.* 70, 305 (1991).
140. L. S. Yu and S. S. Li, *Appl. Phys. Lett.* 59, 1332 (1991).
141. T. S. Faska, J. W. Little, W. A. Beck, K. J. Ritter, A. C. Goldberg, and R. LeBlanc, in "Innovative Long Wavelength Infrared Detector Workshop," Pasadena, CA, 1992.
142. S. V. Bandara, S. D. Gunapala, J. K. Liu, E. M. Luong, J. M. Mumolo, W. Hong, D. K. Sengupta, and M. J. McKelvey, *Appl. Phys. Lett.* 72, 2427 (1998).
143. R. Melliti, P. Tronc, E. Mao, A. Majerfeld, and J. Depeyrot, *Superlattices Microstruct.* 23, 1037 (1998).
144. P. Man and D. S. Pan, *Appl. Phys. Lett.* 64, 321 (1994).
145. P. Man and D. S. Pan, *Appl. Phys. Lett.* 67, 3653 (1995).
146. Y. Fu, M. Willander, and W. Xu, *J. Appl. Phys.* 79, 3103 (1996).
147. F. Szmulowicz, A. Shen, H. C. Liu, G. J. Brown, Z. R. Wasilewski, and M. Buchanan, *Phys. Rev. B* 61, 13798 (2000).
148. P. Man and D. S. Pan, *Appl. Phys. Lett.* 66, 192 (1995).
149. D. C. Wang, G. Bosman, and S. S. Li, *J. Appl. Phys.* 79, 1486 (1996).
150. Y. H. Wang, J. Chu, S. S. Li, and P. Ho, *J. Appl. Phys.* 76, 6009 (1994).
151. J. Hoff, S. Kim, M. Erdtmann, R. Williams, J. Piotrowski, E. Bigan, M. Razeghi, and G. J. Brown, *Appl. Phys. Lett.* 67, 22 (1995).
152. J. R. Hoff, M. Razeghi, and G. J. Brown, *Phys. Rev. B* 54, 10773 (1996).
153. S. S. Li, J. Chu, and Y. H. Wang, *Superlattices Microstruct.* 19, 229 (1996).
154. J. Chu, S. S. Li, and P. Ho, *Appl. Phys. Lett.* 69, 1258 (1996).
155. D. K. Sengupta, S. L. Jackson, D. Ahmari, H. C. Kuo, J. I. Malin, S. Thomas, M. Feng, G. E. Stillman, Y. C. Chang, L. Li, and H. C. Liu, *Appl. Phys. Lett.* 69, 3209 (1996).
156. J. Chu and S. S. Li, *IEEE J. Quantum Electron.* 33, 1104 (1997).
157. J. Chu, S. S. Li, and P. Ho, *J. Cryst. Growth* 175, 964 (1997).
158. D. K. Sengupta, S. L. Jackson, A. P. Curtis, W. Fang, J. I. Malin, T. U. Horton, H. C. Kuo, A. Moy, J. Miller, K. C. Hsieh, K. Y. Cheng, H. Chen, I. Adesida, S. L. Chuang, M. Feng, G. E. Stillman, M. Wu, J. Tucker, Y. C. Chang, L. Li, and H. C. Liu, *J. Electron. Mater.* 26, 1382 (1997).
159. D. H. Zhang, W. Shi, P. H. Zhang, S. F. Yoon, and X. A. Shi, *Appl. Phys. Lett.* 74, 1570 (1999); D. H. Zhang, W. Shi, P. H. Zhang, and S. F. Yoon, *Jpn. J. Appl. Phys. 2 Lett.* 38, L360 (1999).
160. A. Chin, C. C. Liao, J. Chu, and S. S. Li, *J. Cryst. Growth* 175, 999 (1997).
161. H. C. Liu, L. Li, M. Buchanan, Z. R. Wasilewski, G. J. Brown, F. Szmulowicz, and S. M. Hegde, *J. Appl. Phys.* 83, 585 (1998).
162. G. J. Brown, F. Szmulowicz, and S. M. Hegde, *J. Electron. Mater.* 24, 559 (1995).
163. H. C. Liu, T. Orgarah, E. Dupont, Z. R. Wasilewski, M. Byloos, M. Buchanan, E. Szmulowicz, J. Ehret, and G. J. Brown, *Electron. Lett.* 38, 909 (2002).
164. H. C. Liu, F. Szmulowicz, Z. R. Wasilewski, M. Buchanan, and G. J. Brown, *J. Appl. Phys.* 85, 2972 (1999).
165. A. Shen, H. C. Liu, F. Szmulowicz, M. Buchanan, M. Gao, G. J. Brown, and J. J. Ehret, *Appl. Phys.* 86, 5232 (1999).
166. A. Shen, H. C. Liu, M. Gao, E. Dupont, M. Buchanan, J. Ehret, G. J. Brown, and F. A. Szmulowioz, *Appl. Phys. Lett.* 77, 2400 (2000).
167. J. Chu, S. S. Li, and A. Singh, *Appl. Phys. Lett.* 73, 3414 (1998); J. Chu, S. S. Li, and A. Singh, *IEEE J. Quantum Electron.* 35, 312 (1999).
168. J. Chu, C. S. Li, A. Singh, and P. Ho, *Appl. Phys. Lett.* 73, 1664 (1998).
169. Y. H. Wang, S. S. Li, J. Chu, and P. Ho, *Appl. Phys. Lett.* 64, 727 (1994).
170. M. Razeghi, M. Erdtmann, C. Jelen, F. Guastavinos, G. J. Brown, and Y. S. Park, *Infrared Phys. Technol.* 42, 135 (2001).
171. H. C. Liu, *Proc. SPIE* 2999, 19 (1997).
172. L. B. Allard, H. C. Liu, M. Buchanan, and Z. R. Wasilewski, *Appl. Phys. Lett.* 70, 2784 (1997).
173. V. Ryzhii and H. C. Liu, *J. Appl. Phys.* 92, 2354 (2002).
174. V. Ryzhii, M. Ryzhii, and H. C. Liu, *J. Appl. Phys.* 91, 5887 (2002).
175. E. Dupont, M. Byloos, M. Gao, M. Buchanan, C.-Y. Song, Z. R. Wasilewski, and H. C. Liu, *IEEE Photon. Technol. Lett.* 14, 182 (2002).
176. E. Dupont, M. Gao, Z. Wasilewski, and H. C. Liu, *Appl. Phys. Lett.* 78, 2067 (2001).
177. B. F. Levine, C. G. Bethea, V. O. Shen, and R. J. Malik, *Appl. Phys. Lett.* 57, 383 (1990).
178. E. Martinet, E. Rosencher, F. Luc, Ph. Bois, E. Costard, and S. Delaitre, *Appl. Phys. Lett.* 61, 246 (1992).
179. K. Kheng, M. Ramsteiner, H. Schneider, J. D. Ralston, F. Fuchs, and P. Koidl, *Appl. Phys. Lett.* 61, 666 (1992).
180. I. Grave, A. Shakouri, N. Kuze, and A. Yariv, *Appl. Phys. Lett.* 60, 2362 (1992).
181. H. C. Liu, J. Li, J. R. Thompson, Z. R. Wasilewski, M. Buchanan, and J. G. Simmons, *IEEE Electron Device Lett.* 14, 566 (1993).
182. M. Z. Tidrow, K. K. Choi, C. Y. Lee, W. H. Chang, F. J. Towner, and J. S. Ahearn, *Appl. Phys. Lett.* 64, 1268 (1994).
183. J. C. Chiang, S. S. Li, M. Z. Tidrow, P. Ho, M. Tsai, and C. P. Lee, *Appl. Phys. Lett.* 69, 2412 (1996).
184. J. C. Chiang, S. S. Li, and A. Singh, *Appl. Phys. Lett.* 71, 3546 (1997).
185. A. Guzman, J. L. Sanchez-Rojas, J. M. G. Tijero, J. Hernando, E. Calleja, E. Munoz, G. Vergara, R. Almazan, F. J. Sanchez, M. Verdu, and M. T. Montojo, *Semicond. Sci. Technol.* 16, 285 (2001).
186. J. Lin, S. S. Li, M. Z. Tidrow, W. R. Dyer, W. K. Liu, J. M. Fastenau, and T. R. Yurasits, *Appl. Phys. Lett.* 79, 2982 (2001).
187. Y. Huang and C. Lien, *J. Appl. Phys.* 76, 3181 (1994).
188. Y. Huang and C. Lien, *J. Appl. Phys.* 78, 2700 (1995).
189. Y. Huang and C. Lien, *J. Appl. Phys.* 77, 3433 (1995).
190. Y. Huang, C. Lien, and T. Lei, *J. Appl. Phys.* 74, 2598 (1993).
191. E. Martinet, F. Luc, E. Rosencher, Ph. Bois, and S. Delaitre, *Appl. Phys. Lett.* 60, 895 (1992).
192. L. C. Lenchyshyn, H. C. Liu, A. N. Buchanan, and Z. R. Wasilewski, *J. Appl. Phys.* 79, 8091 (1996).
193. L. C. Lenchyshyn, H. C. Liu, M. Buchanan, and Z. R. Wasilewski, *J. Appl. Phys.* 79, 3307 (1996).
194. M. Z. Tidrow, X. D. Jiang, S. S. Li, and K. Bacher, *Appl. Phys. Lett.* 74, 1335 (1999).

195. A. Majumdar, K. K. Choi, J. L. Reno, L. P. Rokhinson, and D. C. Tsui, *Appl. Phys. Lett.* 80, 707 (2002); A. Majumdar, K. K. Choi, L. P. Rokhinson, and D. C. Tsui, *Appl. Phys. Lett.* 80, 538 (2002).
196. A. Majumdar, K. K. Choi, L. P. Rokhinson, J. L. Reno, and D. C. Tsui, *J. Appl. Phys.* 91, 4623 (2002).
197. N. Li, N. Li, W. Lu, X. Q. Liu, X. Z. Yuan, Z. F. Li, H. F. Dou, S. C. Shen, Y. Fu, M. Willander, H. H. Tan, C. Jagadish, M. B. Johnston, and M. Gal, *Superlattices Microstruct.* 26, 317 (1999).
198. L. Fu, H. H. Tan, C. Jagadish, Na Li, Ning Li, X. Q. Liu, W. Lu, and S. C. Shen, *Infrared Phys. Technol.* 42, 171 (2000).
199. A. S. W. Lee and E. H. Li, *Appl. Phys. Lett.* 69, 3581 (1996).
200. J. Radovanovic, V. Milanovic, Z. Ikonic, and D. Indjin, *J. Appl. Phys.* 91, 525 (2002).
201. L. Fu, H. H. Tan, C. Jagadish, Na Li, Ning Li, X. Q. Liu, W. Lu, and S. C. Shen, *Appl. Phys. Lett.* 78, 10 (2001).
202. O. Byungsung, J.-W. Choe, M. H. Francombe, K. M. S. V. Bandara, D. D. Coon, Y. F. Lin, and W. J. Takei, *Appl. Phys. Lett.* 57, 503 (1990).
203. J. L. Johnson, L. A. Samoska, A. C. Gossard, J. L. Merz, M. D. Jack, G. R. Chapman, B. A. Baumgratz, K. Kosai, and S. M. Johnson, *J. Appl. Phys.* 80, 1116 (1996).
204. B. O. J.-W. Choe, M. H. Francombe, K. M. S. V. Bandara, E. Sorar, D. D. Coon, Y. F. Lin, and W. J. Takei, *J. Vac. Sci. Technol. B* 9, 1789 (1991).
205. K. W. Goossen, S. A. Lyon, and K. Alavi, *Appl. Phys. Lett.* 53, 1701 (1988).
206. H. Schneider, K. Kheng, M. Ramsteiner, J. D. Ralston, F. Fuchs, and P. Koidl, *Appl. Phys. Lett.* 60, 1471 (1992).
207. J. D. Ralston, H. Schneider, D. F. G. Gallagher, K. Kheng, F. Fuchs, P. Bittner, B. Dischler, and P. Koidl, *J. Vac. Sci. Technol. B* 10, 998 (1992).
208. H. Schneider, P. Koidl, F. Fuchs, B. Dischler, K. Schwarz, and J. D. Ralston, *Semicond. Sci. Technol.* 6, C120 (1991).
209. H. Schneider, *J. Appl. Phys.* 74, 4789 (1993).
210. Z. Y. Yuan, Z. H. Chen, D. F. Cui, J. W. Ma, Q. Hu, J. M. Zhou, and Y. L. Zhou, *Appl. Phys. Lett.* 67, 930 (1995).
211. S. Ehret and H. Schneider, *J. Appl. Phys.* 79, 9369 (1996).
212. H. Schneider, P. Koidl, C. Schonbein, S. Ehret, E. C. Larkins, and G. Bihlmann, *Superlattices Microstruct.* 19, 347 (1996).
213. H. Schneider, C. Schoenbein, K. Schwarz, and M. Walther, *Physica E* 2, 28 (1998).
214. W. G. Wu, D. S. Jiang, L. Q. Cui, C. Y. Song, and Y. Zhuang, *Solid-State Electron.* 43, 723 (1999).
215. J. H. Lee, J. C. Chiang, S. S. Li, and P. J. Kannam, *Appl. Phys. Lett.* 74, 765 (1999).
216. H. Schneider, P. Koidl, M. Walther, J. Fleissner, R. Rehm, E. Diwo, K. Schwarz, and G. Weimann, *Infrared Phys. Technol.* 42, 283 (2001).
217. M. Z. Tidrow and K. Bacher, *Appl. Phys. Lett.* 69, 3396 (1996).
218. C. S. Wu, C. P. Wen, P. Reiner, C. W. Tu, and H. Q. Hou, *Solid-State Electron.* 39, 1253 (1996).
219. K. Pozela, S. B. Naran, J. P. Leburton, and J. N. Heyman, *Appl. Phys. Lett.* 77, 265 (2000).
220. H. Schneider, C. Schonbein, G. Bihlmann, P. van Son, and H. Sigg, *Appl. Phys. Lett.* 70, 1602 (1997).
221. S. Ehret, H. Schneider, J. Fleissner, P. Koidl, and G. Bohm, *Appl. Phys. Lett.* 71, 641 (1997).
222. S. R. Schmidt, A. Seilmeier, and H. C. Liu, *J. Appl. Phys.* 91, 5545 (2002).
223. H. Schneider, C. Schonbein, M. Walther, K. Schwarz, J. Fleissner, and P. Koidl, *Appl. Phys. Lett.* 71, 246 (1997).
224. Y. H. Wang, J. C. Chiang, S. S. Li, and P. Ho, *J. Appl. Phys.* 76, 2538 (1994).
225. Y. H. Wang, S. S. Li, and P. Ho, *Appl. Phys. Lett.* 62, 621 (1993).
226. Y. H. Wang, S. S. Li, and P. Ho, *Appl. Phys. Lett.* 62, 93 (1993).
227. H. C. Liu, in "Long Wavelength Infrared Detectors" (M. Razeghi, Ed.), pp. 1–59. Gordon and Breach, Amsterdam, 1996; H. C. Liu, J. Li, M. Buchanan, and Z. R. Wasilewski, *IEEE J. Quantum Electron.* 32, 1024 (1996).
228. H. C. Liu, G. E. Henkins, E. R. Brown, K. A. McIntosh, K. B. Nichols, and M. J. Manfra, *IEEE Electron Device Lett.* 16, 253 (1995).
229. H. C. Liu, J. Li, E. R. Brown, K. A. McIntosh, K. B. Nichols, and M. J. Manfra, *Appl. Phys. Lett.* 67, 1594 (1995).
230. M. Ershov, *Appl. Phys. Lett.* 69, 3480 (1996).
231. V. Ryzhii, *J. Appl. Phys.* 81, 6442 (1997).
232. M. Ryzhii and V. Ryzhii, *Appl. Phys. Lett.* 72, 842 (1998).
233. S. Ehret, H. Schneider, J. Fleissner, P. Koidl, and G. Bohm, *Appl. Phys. Lett.* 71, 641 (1997).
234. R. K. Richards, D. P. Hutchinson, C. A. Bennett, M. L. Simpson, H. C. Liu, and M. Buchanan, *ECS Proc.* 98-21, 95 (1998).
235. G. Sarusi, S. D. Gunapala, J. S. Park, and B. F. Levine, *J. Appl. Phys.* 76, 6001 (1994).
236. C. Y. Lee, M. Z. Tidrow, K. K. Choi, W. H. Chang, L. F. Eastman, F. J. Towner, and J. S. Ahearn, *J. Appl. Phys.* 75, 4731 (1994).
237. T. R. Schimert, S. L. Barnes, A. J. Brouns, F. C. Case, P. Mitra, and L. T. Claiborne, *Appl. Phys. Lett.* 68, 2846 (1996).
238. C. J. Chen, K. K. Choi, W. H. Chang, and D. C. Tsui, *Appl. Phys. Lett.* 72, 7 (1998).
239. D. Mandelik, M. Schniederman, V. Umansky, and I. Bar-Joseph, *Appl. Phys. Lett.* 78, 472 (2001).
240. A. Rogalski, *Infrared Phys. Technol.* 40, 279 (1999).
241. L. J. Kozlowski, G. M. Williams, G. J. Sullivan, C. W. Farley, R. J. Anderson, J. Chen, D. T. Cheung, W. E. Tennant, and R. E. DeWames, *IEEE Trans. Electron Devices* 38, 1124 (1991).
242. H. Schneider, M. Walther, C. Schonbein, R. Rehm, J. Fleissner, W. Pletschen, J. Braunsteina, P. Koidl, G. Weimann, J. Ziegler, and W. Cabanski, *Physica E* 7, 101 (2000).
243. A. Goldberg, T. Fischer, S. Kennerly, W. Beck, V. Ramirez, and K. Garner, *Infrared Phys. Technol.* 42, 309 (2001).
244. S. D. Gunapala, S. V. Bandara, A. Singh, J. K. Liu, B. Rafol, E. M. Luong, J. M. Mumolo, N. Q. Tran, D. Z.-Y. Ting, J. D. Vincent, C. A. Shott, J. Long, and P. D. LeVan, *IEEE Trans. Electron Devices* 47, 963 (2000).
245. S. D. Gunapala, S. V. Bandara, J. K. Liu, W. Hong, E. M. Luong, J. M. Mumolo, M. J. McKelvey, D. K. Sengupta, A. Singh, C. A. Shott, R. Carralejo, P. D. Maker, J. J. Bock, M. E. Ressler, M. W. Werner, and T. N. Krabach, *Proc. SPIE* 3379, 382 (1998).
246. V. J. Realmuto, A. J. Sutton, and T. Elias, *J. Geo. Phys. Res.* 102, 15057 (1997).
247. S. D. Gunapala, S. V. Bandara, J. K. Liu, W. Hong, M. Sundaram, P. D. Maker, R. E. Muller, C. A. Shott, and R. Carralejo, *IEEE Trans. Electron Devices* 45, 1890 (1998).
248. A. Sergeev, V. Mitin, and M. Stroscio, *Physica B* 316, 369 (2002).
249. S. Raghavan, P. Rotella, A. Stintz, B. Fuchs, S. Krishna, C. Morath, D. A. Cardimona, and S. W. Kennerly, *Appl. Phys. Lett.* 81, 1369 (2002).
250. J. Phillips, *J. Appl. Phys.* 91, 4590 (2002).
251. E. Finkman, S. Maimon, V. Immer, G. Bahir, S. E. Schacham, F. Fossard, F. H. Julien, J. Brault, and M. Gendry, *Phys. Rev. B* 63, 045323/1 (2001).
252. S. Y. Lin, Y. J. Tsai, and S. C. Lee, *Jpn. J. Appl. Phys. 2 Lett.* 40, L1290 (2001).
253. V. Ryzhii, I. Khmyrova, V. Pipa, V. Mitin, and M. Willander, *Semicond. Sci. Technol.* 16, 331 (2001).
254. M. D. Lim, A. G. Choo, T. I. Kim, S. S. Ko, D. H. Baek, and S. C. Hong, *J. Cryst. Growth* 227, 1162 (2001).
255. V. Ryzhii, I. Khmyrova, V. Mitin, M. Stroscio, and M. Willander, *Appl. Phys. Lett.* 78, 3523 (2001).
256. V. Ryzhii, *Appl. Phys. Lett.* 78, 3346 (2001).
257. Z. H. Chen, O. Baklenov, E. T. Kim, I. Mukhametzhanov, J. Tie, A. Madhukar, Z. Ye, and J. C. Campbell, *J. Appl. Phys.* 89, 4558 (2001).

258. E. Kim, Z. H. Chen, and A. Madhukar, *Appl. Phys. Lett.* 79, 3341 (2001).
259. J. Jiang, C. Jelen, M. Razeghi, and G. J. Brown, *IEEE Photon. Technol. Lett.* 14, 372 (2002).
260. J. Park, S. Jo, S. Hong, and J. Song, *Solid-State Electron.* 46, 651 (2002).
261. G. Hasnain, B. F. Levine, C. D. L. Sivco, and A. Y. Cho, *Appl. Phys. Lett.* 56, 770 (1990).
262. S. D. Gunapala, B. F. Levine, D. Ritter, R. A. Hamm, and M. B. Panish, *Appl. Phys. Lett.* 58, 2024 (1991).
263. D. Ritter, R. A. Hamm, M. B. Panish, J. M. Vandenberg, D. Gershoni, S. D. Gunapala, and B. F. Levine, *Appl. Phys. Lett.* 59, 552 (1991).
264. C. Jelen, S. Slivken, T. David, M. Razeghi, and G. Brown, *IEEE J. Quantum Electron.* 34, 1124 (1998).
265. C. Jelen, S. Slivken, T. David, G. Brown, and M. Razeghi, *Proc. SPIE* 3287, 96 (1998).
266. J. Hoff, S. Kim, M. Erdtmann, R. Williams, J. Piotrowski, E. Bigan, and M. Razeghi, *Appl. Phys. Lett.* 67, 22 (1995).
267. C. Jelen, S. Slivken, G. Brown, and M. Razeghi, *Mater. Res. Soc. Symp. Proc.* 450, 195 (1997).
268. M. O. Watanabe and Y. Ohba, *Appl. Phys. Lett.* 50, 906 (1987).
269. B. Elman, E. S. Koteles, P. Melman, C. Jagannath, J. Lee, and D. Dugger, *Appl. Phys. Lett.* 55, 1659 (1989).
270. X. Zhou, P. K. Bhattacharya, G. Hugo, S. C. Hong, and E. Gulari, *Appl. Phys. Lett.* 54, 855 (1989).
271. S. D. Gunapala, K. S. M. V. Bandara, B. F. Levine, G. Sarusi, J. S. Park, T. L. Lin, W. T. Pike, and J. K. Liu, *Appl. Phys. Lett.* 64, 3431 (1994).
272. F. Capasso, K. Mohammed, and A. Y. Cho, *IEEE J. Quantum Electron.* 22, 1853 (1986).
273. J. W. O. B. Choe, M. H. Francombe, K. M. S. V. Bandara, D. D. Coon, Y. F. Lin, and W. J. Takei, *Appl. Phys. Lett.* 57, 503 (1990).
274. L. S. Yu, Y. H. Wang, S. S. Li, and P. Ho, *Appl. Phys. Lett.* 60, 992 (1992).
275. S. S. Lim, M. Y. Chuang, and L. S. Yu, *Semicond. Sci. Technol.* 8, 406 (1993).
276. Y. Fu, M. Willander, W. Lu, and W. Xu, *J. Appl. Phys.* 84, 5750 (1998).
277. Y. Fu, N. Li, M. Larlsteen, M. Willander, N. Li, W. L. Xu, W. Lu, and S. C. Shen, *J. Appl. Phys.* 87, 511 (2000).
278. Y. Fu, *Superlattices Microstruct.* 30, 69 (2001).
279. Y. Fu and M. Willander, *J. Appl. Phys.* 88, 288 (2000).

www.aspbs.com/enn

Quantum Well Infrared Photodetectors: Theoretical Aspects

A. F. M. Anwar
University of Connecticut, Storrs, Connecticut, USA

Kevin R. Lefebvre
State University of New York Institute of Technology, Utica, New York, USA

CONTENTS

1. Introduction
2. Influence of a Perpendicular Electric Field Upon a Quantum Well
3. Electron Escape Time from a Quantum Well
4. Electron–Polar Optical Phonon Interaction within a Biased Quantum Well
5. Density of States Induced Absorption Line Broadening of a Biased Quantum Well
6. Dark Currents in Multiple Quantum Well Infrared Photodetectors
Glossary
References

1. INTRODUCTION

With the advancements in modern growth techniques such as molecular beam epitaxy and MOVPE, the addition of quantum wells within semiconductor devices has expanded considerably over the past decade. For instance, quantum well infrared photodetectors (QWIPs) [1, 2] rely upon a series of uncoupled quantum wells to detect light by exciting a bound electron into the continuum of states above the quantum well. Self-electro-optic effect devices [3, 4] use a quantum well to enhance carrier confinement and an electric field to produce the quantum confined Stark effect for optical switching.

Upon the application of an electric field, the quantum well is no longer a closed system but becomes an open system [5–7]. This changeover from a closed system to an open system results in a continuum of states arising below the band offset where, at zero electric field, there would exist no states [7]. Moreover, the states above the conduction band offset are also modified with applied electric field. It is this modification of the states that is known as the redistribution of the density of states [5–7] and occurs in both position and energy within a biased quantum well. Furthermore, as there is a redistribution in the density of states below the band offset, there exists a finite probability of the particle lying at those states. Therefore, the wavefunctions for these states are different than the zero field wavefunctions.

The present treatment also questions the validity of the use of the free particle group velocity for a quasi confined system. Traditionally, the free particle group velocity, $\hbar k_z/m^*$, has been used in calculating the current through a quantum structure. However, it has already been shown that for double barrier resonant tunneling structures [5, 6], the local group velocity is a function of position and energy and in only very special cases is the quantity equal to $\hbar k_z/m^*$. The field dependence of the local group velocity when averaged over position, by taking into account the redistribution of the density of states [7], approaches $\hbar k_z/m^*$ only for conduction well above the band offset energy.

The importance of these effects resulting from a perpendicular electric field upon a quantum well will be demonstrated through the calculation of the escape rate of an electron from a biased quantum well [8, 9], the absorption coefficient of a two quantum well QWIP as well as the dark currents for a QWIP. Furthermore, the electron–polar optical phonon interaction is also addressed in terms of the escape rate [10], capture rate, and the general scattering rates.

The determination of the dark current [1, 11] and the escape rate [8, 9] requires the computation of the current density that in turn requires parameters such as carrier concentration, carrier velocity, and the knowledge of the density of states. Furthermore, the calculation of the absorption coefficient [1, 12–21] and the electron–phonon scattering rates [22] requires the knowledge of the matrix elements

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (199–224)

between the initial and final state wavefunctions. However, for open systems, the normalization of the wavefunction forces one to use complicated numerical techniques such as the density matrix approach [23] or the Wigner function [24]. A simple calculation of the wavefunctions can be easily achieved by the transfer matrix approach [25] or through Airy functions [25]. However, these methods are (a) not suitable for self-consistent calculations, (b) require the wavefunctions to be artificially boxed normalized, and (c) are error prone.

Over the last several years, quantum wells have been incorporated into semiconductor devices to enhance carrier confinement. A large number of these semiconductor devices [1–4, 26–28] also utilize an electric field applied perpendicular to the quantum well which modifies the electron's dynamics within the quantum well. The characteristics of these devices depend to a large extent on the rate at which carriers escape from the quantum well, the absorption of light, and the interactions between electrons and phonons. For instance, the dark current in quantum well infrared detectors and multiple quantum well avalanche photodiodes has been attributed to the electrons escaping the quantum well [29, 30] while exciton saturation [31, 32] has been linked to the rate at which both the electrons and holes escape the quantum well. However, efforts in modeling such devices as SEEDs and QWIPs have lagged behind the advancements in the growth and fabrication techniques. This has especially been demonstrated through the electron–phonon interaction within a quantum well which has traditionally been linked to the broadening in the absorption spectrum [33] and excitonic peaks [34] as well as the dark currents [35, 36] in QWIPs and SEEDs.

The calculation of the carrier escape time was first reported by Schneider and von Klitzing [32] using thermionic emission to explain their experimental study of carriers escaping a biased quantum well. This model addressed only over the barrier conduction with a transmission coefficient of unity. However, the use of this simple thermionic emission model proved to be inadequate in explaining the experimental results of Cavailles et al. [37] for the escape of both the electrons and holes from a biased quantum well. Also pointed out by Cavailles et al., lack of agreement between the experimental and theoretical results at low electric fields may be due to the sudden change from the two-dimensional (2D) density of states to the 3D density of states. Moss et al. [38] introduced the concept of tunneling through a full width at half maximum (FWHM) calculation and the uncertainty principle in order to theoretically explain the results obtained by Cavailles et al.; however, the Moss et al. treatment provided only a qualitative understanding of the experimental data. Nelson et al. [39] also reported the measurement of photocurrent from a single quantum well. However, in order to explain their experimental results, an adjustable recombination time parameter was introduced to the escape time, which included both tunneling and thermionic emission components of current, to relate the experimental work to the theoretical work.

As electron transport from a biased quantum well can either be through a thermionic current, transport over the barrier, or through a tunneling current, transport through the barrier. The calculation of the current proceeds by integrating over all energies the product of the particle's group velocity, the density of states, the occupation probability, and the transmission coefficient. However, the previously discussed models apply the infinite quantum well density of states [38] to calculate the escape time even though the energy levels are calculated for a finite quantum well. Moreover, as stated previously, in the presence of an electric field the density of states is redistributed in energy as well as position; therefore, the infinite well density of states is an approximation that cannot be justified upon the application of an electric field. In addition, the calculation of the carrier velocity must be based upon the proper density of states where the simplified group velocity of $\hbar k_z/m^*$ cannot be justified [40]. Also, proper partitioning between the thermionic emission and the tunneling current needs to be accounted for which results from the redistribution of the density of states.

The calculation of the electron–phonon scattering rate within a quantum well is critical to understanding the broadening in the absorption spectrum, the capture rate, and the escape rate. Lam and Singh [35] calculated the thermalization of carriers within a biased quantum well through the absorption of a phonons. Ferreira and Bastard [41] and Zhu et al. [42] calculated the scattering rates of electron–phonons in biased and unbiased quantum wells. Zhu et al. included the effects due to confined phonons. Hernandez-Cabrera et al. [43] considered the escape of carriers through a time dependent investigation of the electron–phonon interaction, whereas Feldmann et al. [44] addressed the escape of carriers via a phonon interaction from shallow quantum wells and Shorthose et al. [45] investigated the escape of electrons via phonon assisted tunneling. However, none of the previous studies of the electron–phonon interaction within a quantum well accounted for the redistribution of the density of states when an electric field is applied. Therefore, Section 4 will address the effects of the redistribution of the density of states in the calculation of the general electron–polar optical phonon scattering rate for both emission and absorption of phonons as a function of applied electric field as well as the capture rate and the escape rate due to an electron–phonon interaction.

Determining the performance of quantum well devices requires knowledge of the transport of carriers in and around the quantum well. For instance, the gain [22, 47, 48] and dark currents [1, 29] of a quantum well infrared photodector as well as the modulation bandwidth of quantum well laser [49] are strong functions of the rate at which carriers are captured into the quantum well. Extensive investigation [50–70] over the last decade has theoretically predicted and experimentally verified capture times on the order of femto- [57, 58] to picoseconds [58–60] with an oscillatory behavior [62–69] as a function of quantum well widths. However, these studies [50–68] have been for a flat conduction band which is essential for the operation of quantum well lasers but not for quantum well infrared photodetectors, the quantum well solar cell, or the self-electro-optic device. Furthermore, Vinter et al. [70] studied a specially designed quantum structure under bias by impedance spectroscopy; the theoretical explanation yielded a weaker dependence on the field than the experimental results.

Theoretical calculation of the absorption coefficient for a QWIP [1, 12–21] that explains experimental results has yet to be achieved. Previous calculations of the absorption coefficient [14–20] introduced a dephasing time into the transition rate to account for the broadening in the absorption spectrum. The dephasing time is determined experimentally by measuring the FWHM of the absorption spectrum. This results in a theoretical model which is dependent upon experimental results, making the design of QWIPs difficult. Furthermore, a comparison between the theoretical and experimental magnitudes of the absorption coefficient is usually done on a normalized scale, eliminating any error that may occur between theoretical calculations and experiment results

The dephasing time introduced in the transition rate has been attributed to the scattering [37] of electrons by phonons [14, 21], interface roughness [21], and impurities [21] within the quantum well. Furthermore, band nonparabolicity [19], carrier lifetimes [14, 15, 17], and k_z dependence on k_t [21] have also been attributed to the broadening where k_z and k_t are the wavevectors in the perpendicular (z) and transverse directions, respectively. However, as stated by Sizov and Rogalski [12] and Manasreh et al. [16], the broadening in the experimental absorption coefficient can be different than the theoretical broadening associated with excited state lifetimes, nonparabolicity, inhomogenity of the layers, and phonon interactions. However, none of the previous works have investigated the redistribution of the density of states as a possible source of broadening in the absorption coefficient of a biased quantum well. Since the transition rate of an electron from a bound state to a continuum state is a function of the number of final states through the matrix elements as well as a summation over final states [46], the absorption coefficient is broadened as a result of the redistribution of the density of states. An important implication of the redistribution of the density of states is that the states above the quantum well in a multiple quantum well system are correlated even in the presence of wide barriers. This correlation further affects the redistribution of the density of states, thereby affecting the absorption calculation of single wells especially in the presence of other quantum wells. This effect is enhanced in the presence of an applied electric field.

Quantum well infrared photodetectors with a high detectivity over a varying range of wavelengths [1, 11, 13, 71] have been achieved over the last several years. The range of the wavelength at which the sensitivity peak occurs varies depending upon the size of the conduction band offset as well as the quantum well width. However, variations in the quantum well width are limited to a small range since a higher detectivity occurs when there exists one bound state within the quantum well and a virtual bound state just above the conduction band offset. Therefore, any changes in the location of the sensitivity peak mainly occurs by varying the conduction band offset. However, as previously demonstrated by Levine et al. [11], decreasing the conduction band offset increases the dark currents.

Theoretical calculations of the dark currents have been performed by Levine et al. [1, 11] where the dark current was attributed to the current tunneling out of the quantum well as well as the thermionic emission current above the quantum well. The agreement between theory and experiment was achieved by varying sample parameters to match the experimental results. Liu [13] has also addressed the dark currents by calculating the current exiting a quantum well. His calculation incorporated the capture of electrons into the quantum well since the current exiting a quantum well must equal the current entering the quantum well under steady state conditions.

Considering that a QWIP is fabricated in an n^+–i–n^+ structure [1, 11, 13, 71] and assuming that the current exiting the quantum well is equal to the current captured by the quantum well under steady state [13], dark current must result from the contact or n^+ regions. Furthermore, as the transmission coefficient for an electron exiting a single quantum well is not a strong function of the barrier width and the number of quantum wells, the results of Levine et al. [1, 11] will not change for increasing barrier widths or varying number of quantum wells. Liu et al. have incorporated the barrier width into the dark currents through a transit time calculation, resulting in the dark current being inversely proportional to the barrier width. However, if the dark currents decreases with increasing barrier width, the experimental result of Liu [13] is self-contradictory. Therefore the variation in the dark currents as a function of barrier width and number of quantum well must arise from some other mechanisms.

2. INFLUENCE OF A PERPENDICULAR ELECTRIC FIELD UPON A QUANTUM WELL

Upon the application of an electric field perpendicular to a quantum well, the quantum well becomes tilted (Fig. 1) changing the density of state the group velocity and the wavefunction of the electron. The effects of an applied electric field on the energy dependence of the density of states and the wavefunction for a symmetric quantum wells are investigated within this section.

The redistribution of the density of states is calculated by using the method of the logarithmic derivative of the

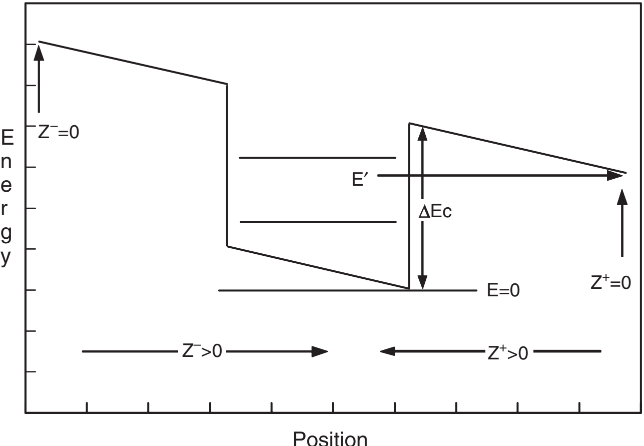


Figure 1. Symmetric quantum well under the influence of an applied electric field.

wavefunctions [5–7]. It is demonstrated that there exist discrete states as well as a continuum of states between the bottom of the quantum well to the conduction band offset which depends upon the applied field and the quantum well itself [7].

After the redistribution of the density of states is obtained, the average local velocity of electrons as well as the electron's wavefunction is calculated within a biased quantum well. Considering that the distribution of electrons within a quantum well is a function of position as well as energy, the average local velocity (group velocity) deviates from the free particle group velocity of $\hbar k_z/m^*$. A comparison between $\hbar k_z/m^*$ and the average local group velocity is made for energies both above and below the band offset as a function of applied electric field, band offset, and quantum well width.

The single electron Schroedinger equation in the effective mass approximation can be written as [46]

$$-\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m^*(z)}\frac{d\varphi(E_z, z)}{dz}\right) + (V(z) - E_z)\varphi(E_z, z) = 0 \tag{1}$$

where $V(z)$ is the potential energy which includes all externally applied fields, E_z is the z-directed carrier energy, $m^*(z)$ is the effective mass at position z, $\hbar$ is the reduced Planck's constant, and $\varphi(E_z, z)$ is the envelope function at energy E_z and position z. Using the substitution

$$G(E_z, z) = (2\hbar/jm^*(z))[\varphi'(E_z, z)/\varphi(E_z, z)] \tag{2}$$

Eq. (1) can be rewritten as [5–7]

$$\frac{d}{dz}G(E_z, z) = -j\left[\frac{m^*(z)}{2\hbar}G(E_z, z)^2 + \frac{4}{\hbar}[V(z) - E_z]\right] \tag{3}$$

where $j = \sqrt{-1}$ is the complex number and $G(E_z, z)$ is known as the logarithmic derivative of the wavefunction.

By discretizing both in energy (ΔE) and in position (Δz) (i.e., for a piecewise continuous potential profile), Eq. (3) can be solved to yield

$$\begin{aligned} G^{\pm}(E_z, z_i) &= G_{oi}\frac{G(z_{i-1})\cosh(\xi_i\Delta z) \mp G_{oi}\sinh(\xi_i\Delta z)}{G_{oi}\cosh(\xi_i\Delta z) \mp G(z_{i-1})\sinh(\xi_i\Delta z)} \\ &= G_R^{\pm}(E_z, z_i) + jG_I^{\pm}(E_z, z_i) \end{aligned} \tag{4}$$

where $G_{oi} = 2\xi_i\hbar/jm^*(z_i)$, $\xi_i = j\sqrt{(2m^*(z_i)/\hbar)/(E - V(z_i))}$, and $\Delta z = z_i - z_{i-1}$. The positive and negative signs indicate the direction of the calculation of $G^{\pm}$; for G^+ the calculation proceeds from negative to positive values of z and vice versa for G^-.

Once the logarithmic derivative of the wavefunction in both the positive and negative directions is known for all positions and energies [Eq. (4)], the one-dimensional (1D) density of states can be calculated as a function of position and energy by [5–7]

$$g_{1D}(E_z, z) = \text{Im}\left[\frac{j8/\pi\hbar}{G^+(E_z, z) - G^-(E_z, z)}\right] \tag{5}$$

The energy dependence [5–7] of $g_{1D}(E_z, z)$ can be obtained by integrating over distance, eliminating the position dependence, as

$$\langle g_{1D}(E_z)\rangle = \frac{1}{L_{\text{well}}}\int_0^{L_{\text{well}}} g_{1d}(E_z, z)dz \tag{6}$$

where L_{well} is the length of the quantum well. The total three dimensional density of states is calculated by convolving the 1D density of states with the 2D density of states. As shown in Eq. (7), this integration is trivial due to the delta function in energy:

$$\begin{aligned} g_{3D}(E'') &= \frac{m^*}{\pi\hbar^2}\int_0^{\infty} dE_t \int_0^{E''} \langle g_{1D}(E_z)\rangle dE_z\, \delta(E - E_t - E_z) \\ &= \frac{m^*}{\pi\hbar^2}\int_0^{E''} \langle g_{1D}(E_z)\rangle dE_z \end{aligned} \tag{7}$$

As shown in Figure 1, there exists an energy, the tunneling energy, below which the transmission of an electron out of the quantum well is practically zero. Since the energies below E', the tunneling energy, are still bounded, the density of states below E' is the usual quantum well density of states for a finite quantum well. This results in the total three-dimensional density of states within a biased quantum well as

$$g_{3D}(E'') = \sum_{\substack{n \\ E_n < E'}} \frac{m^*}{\pi^2\hbar^2}(k_{zn} - k_{z(n-1)}) + \frac{m^*}{\pi\hbar^2}\int_{E'}^{E''} \langle g_{1D}(E_z)\rangle dE_z \tag{8}$$

where

$$k_{zn} = \sqrt{\frac{2m^*E_n}{\hbar^2}} \qquad k_{z0} = 0$$

As stated previously, once the logarithmic derivative is known for all positions and all energies [Eg. (4)], the local group velocity can be calculated. Following the method in [5–7, 46], the local group velocity, $v_g^{\pm}(E_z, z)$, is one-half the real part of $G^{\pm}(E_z, z)$, where the positive (negative) sign indicates a positive (negative) velocity. The average local velocity is obtained by averaging the number of electrons at positive E_z with the local velocity over position as

$$\langle v_g^{\pm}(E_z)\rangle = \frac{\int_0^{L_w} n_{1d}(E_z, z)v_g^{\pm}(E_z, z)\, dz}{\int_0^{L_w} n_{1d}(E_z, z)\, dz} \tag{9}$$

where $n_{1d}(E_z, z)$ $\{=f(E_z)g_{1D}(E_z, z)\}$ is the number of carriers at position z and energy E_z and obtained by multiplying the Fermi distribution function by the one-dimensional density of states.

For the calculation of the scattering rates later addressed in this chapter, the wavefunction of the electron is required. The wavefunctions can be calculated by Eqs. (3) and (4) as

$$\varphi(z, E_z) = \exp\left[\int_0^z \zeta(z', E_z)\, dz' + \int_0^z \beta(z', E_z)\, dz'\right] \tag{10}$$

where

$$\begin{aligned} \zeta(E_z, z) &= -[m^*(z)G_R^+(E_z, z)]/2\hbar \quad \text{and} \\ \beta(E_z, z) &= i[m^*(z)G_I^+(E_z, z)]/2\hbar \end{aligned} \tag{11}$$

The structure considered is a symmetric quantum well consisting of 200 Å AlGaAs barriers and a 100 Å GaAs

quantum well. The concentration of Al is 20%, yielding a conduction band offset of 0.1621 eV and effective masses of 0.084 for AlGaAs and 0.067 for GaAs.

The energy levels and tunneling energy are plotted as a function of the applied electric field in Figure 2. As seen from the graph, the tunneling energy decreases and the energy levels increase with increasing electric field. The energy reference for the energy levels is the lowest possible energy within the quantum well.

The 1D density of states is plotted in Figure 3 as a function of energy with applied electric field as a parameter. Since both of the energy levels are below the tunneling energy when the electric field is below 20 kV/cm, the 1D densities of states are delta function spikes, as shown in Figure 3a. Also shown in Figure 3a is the existence of states between the upper most energy level and E_c. As was stated before, this is a result of the system being unbounded for energies greater than the tunneling energy when an electric field is applied. Figure 3a also shows the virtual bound states which occur above the conduction band offset (dotted vertical line), E_c. As the electric field is increased to 30 kV/cm, only the upper energy level is above the tunneling energy. This creates a spreading in the 1D density of states around the upper energy level as shown in Figure 3b. However, the spreading of the density of states is not limited to a small area around the energy level but is continuous to E_c and is more pronounced compared to the 20 kV/cm case. Also shown in Figure 3b are the virtual bound states which are becoming less defined and shifting to a higher energy. Increasing the electric field further results in the spreading of the density of states for the virtual bound states. The uppermost energy level eventually shifts above E_c, and there is an increase in the number of states between the energy levels as well as above the upper most energy level. This is demonstrated in Figure 3c–e for electric fields of 50, 100, and 140 kV/cm. The spreading out of the 1D density of states and the shifting of energy peaks suggests that as the electric field is increased, the quantum well loses the steplike characteristics of the zero field density of states and approaches the 3D density of states for bulk. This will be discussed further in the next paragraph.

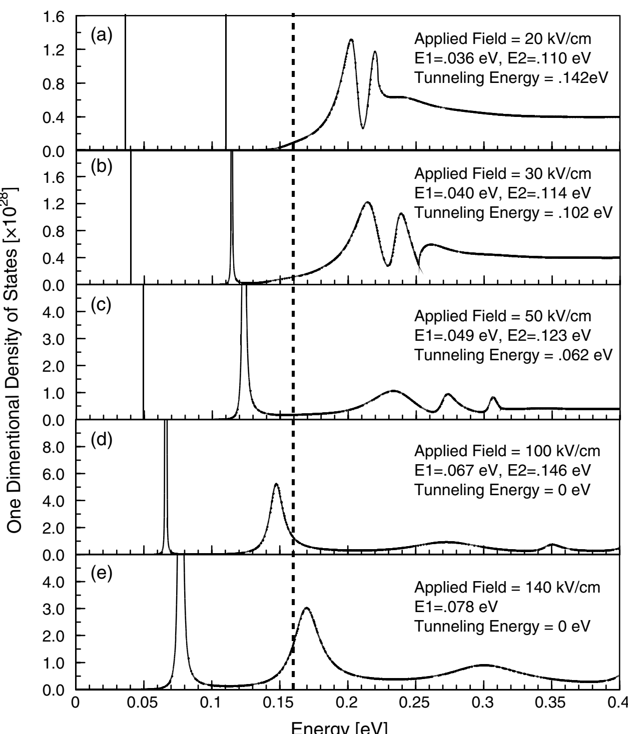


Figure 3. One-dimensional density of states for a symmetric quantum well as a function of energy and applied electric field. Reprinted with permission from [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

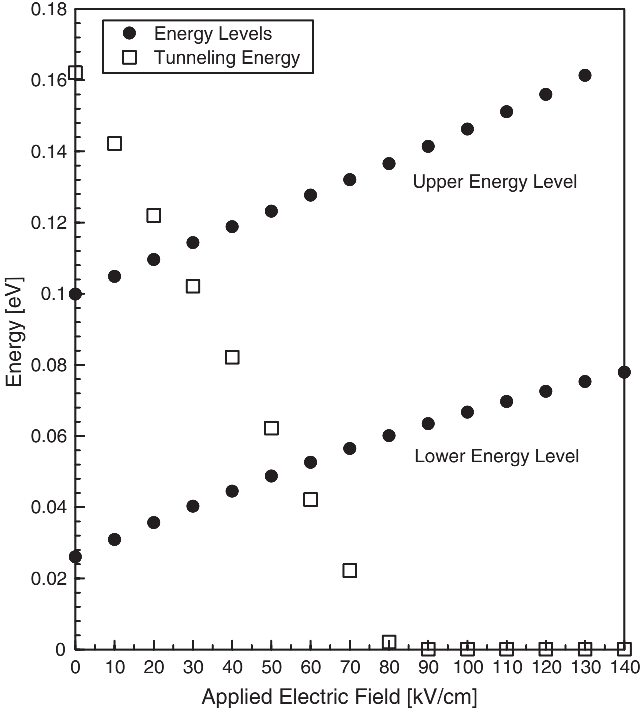


Figure 2. Energy levels and tunneling energy for a symmetric quantum well as a function of applied electric field. Reprinted with permission from [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

In Figure 4, the 3D densities of states are plotted as a function of energy for electric fields of 0, 20, 50, 90, 110, and 140 kV/cm. As shown in Figure 4a, at low fields where the tunneling energy is greater than the upper most energy level, the 3D density of states for 20 kV/cm resembles the zero electric field case except for the energy shift and a slight curvature occurring between the upper most energy level and E_c. At energies above E_c, the 3D density of states follows the bulk density of states except for slight oscillations due to the virtual bound states. However, as the electric field is increased and the tunneling energy falls below the upper most energy level, the steplike function of the density of states becomes more bulklike as shown in Figure 4b and c. Eventually the density of states will lose the steplike characteristics of the zero field density of states as the applied electric field is progressively increased. This is demonstrated in Figure 5 where the 3D density of states is plotted for electric fields of 100 and 200 kV/cm verses energy on a semilog scale.

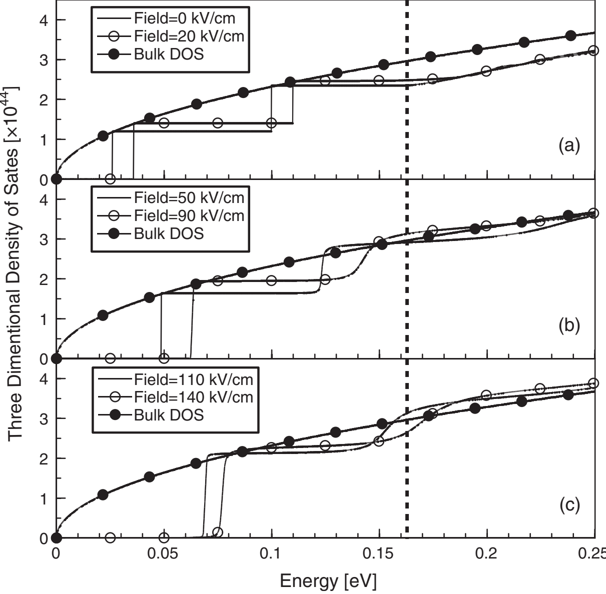


Figure 4. Three-dimensional density of states for a symmetric quantum well as a function of energy and applied electric field [46].

The calculation of the wavefunction [Eq. (10)] is shown in Figure 6 for an AlGaAs/GaAs system with an electric field of 0 kV/cm, 20% Al, a conduction band offset of 0.1621 eV, a quantum well width of 100 Å, and a barrier width of 200 Å. As shown in Figure 6, there exist two energy levels.

As previously demonstrated, there exist states from the tunneling to the band offset. If states exist then these states must have a wavefunction. This is demonstrated in Figure 7 where an electric field of 100 kV/cm is applied.

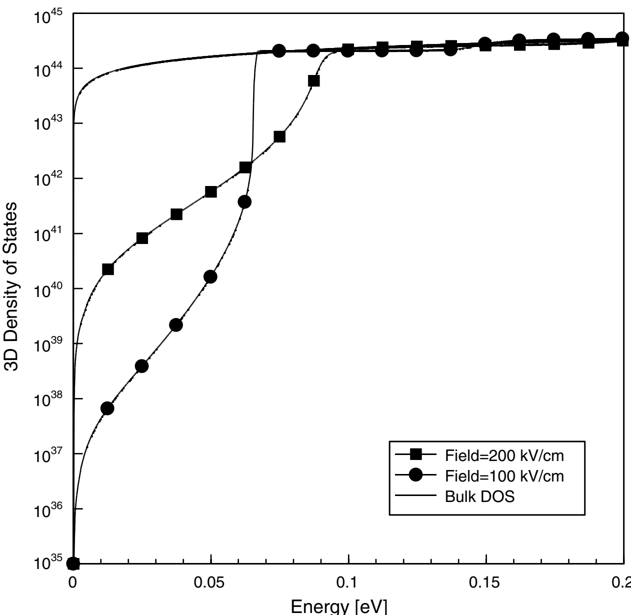


Figure 5. Three-dimensional density of states for a symmetric quantum well as a function of energy and applied electric fields of 100 and 200 kV/cm. Reprinted with permission from [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

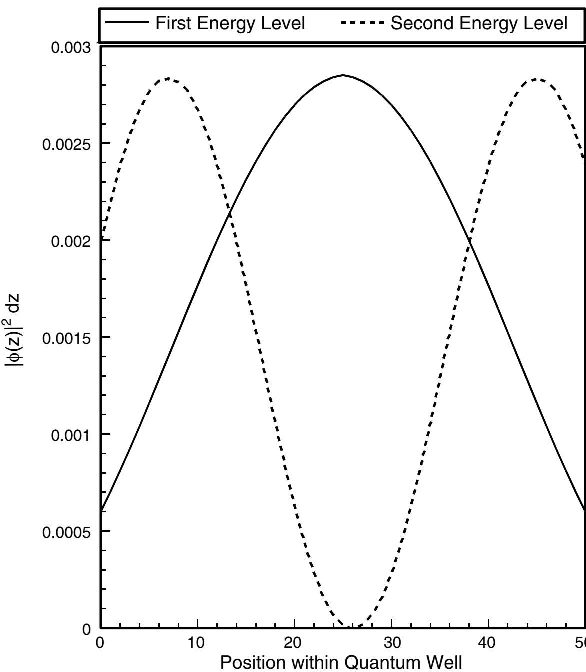


Figure 6. The wavefunctions for the first and second energy level at zero bias.

As demonstrated, the wavefunctions of the first and second energy levels shift toward the lower portion of the quantum well (inset of Fig. 7). Also, as the energy increases from the first energy level (66 meV) toward the second (161 meV), the first energy level must transform into the second. This is demonstrated for selected energies between the first and second energy level.

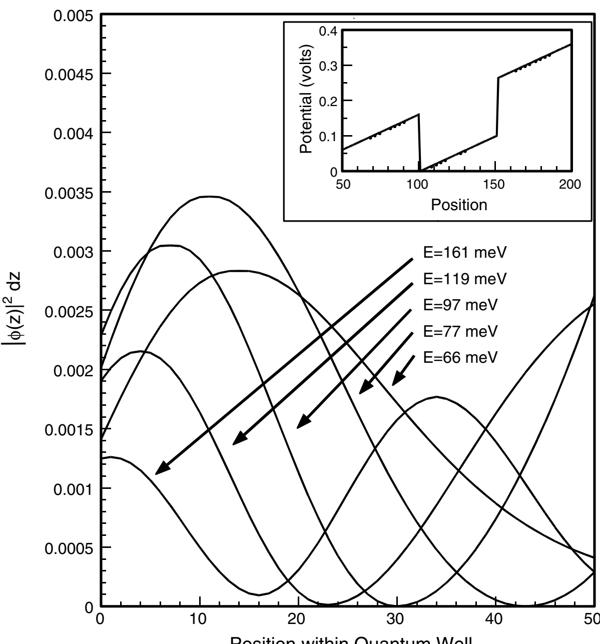


Figure 7. The wavefunctions for the first (66 meV) and second (161 meV) energy level along with the wavefunctions for selected energies between the first and second energy levels at 100 kV/cm.

3. ELECTRON ESCAPE TIME FROM A QUANTUM WELL

The current exiting a quantum well as a result of a z-directed electric field is the sum of two components (Fig. 8): the thermionic emission component (J_{th}) and the tunneling component (J_{tu}). The thermionic emission component is the current above the conduction band offset (E_c) while the tunneling component is the current below the conduction band offset. The total current is a combination of two oppositely propagating current components: J^+ along the positive z-direction and J^- along the negative z-direction [46],

$$J_{\text{tot}} = [J_{\text{tu}} + J_{\text{th}}]^+ + [J_{\text{tu}} + J_{\text{th}}]^-$$

$$[J_{\text{tu}} + J_{\text{th}}] = \frac{2e}{(2\pi)^2}\left[\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y \\ \times\left[\frac{1}{2\pi}\int_{\text{tu}} f(E)\langle v_g^{\pm}(E_z)\rangle D^{\pm}(E_z) dk_z^{\text{tu}}\right] \\ +\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y \\ \times\left[\frac{1}{2\pi}\int_{\text{th}} f(E)\langle v_g^{\pm}(E_z)\rangle D^{\pm}(E_z) dk_z^{\text{th}}\right] \end{array}\right] \tag{12}$$

where $f(E)$ is the Fermi distribution function, E_z is the energy corresponding to the wavevector in the z-direction, $\langle v_g^{\pm}(E_z)\rangle$ are the effective carrier velocities in the positive and negative z-directions, $D^{\pm}(E_z)$ is the transmission coefficient, and e is the elemental electron charge.

The integration over k_z for the tunneling current is between $\pm k_{\text{tu}}$ to $\pm k_c$ where $\pm k_{\text{tu}}$ is the z-directed wavevector associated with E' (the energy at which carriers can first tunnel out of the quantum well [$E' = (\hbar k_{\text{tu}})^2/2\text{m}^*$]: Fig. 8) and $\pm k_c$ is the wave vector associated with the energy of the conduction band offset [$E_c = (\hbar k_c)^2/2m^*$]. The integration over k_z for the thermionic emission current is from $\pm k_c$to ± 8, where the upper sign corresponds to the positive k_z integration and the lower sign is for the negative k_z integration. Assuming parabolic energy bands and converting the integrals over k into integrals over energy, we obtain the total current,

$$J_{\text{Tot}} = \frac{em^*}{2\pi\hbar^2}\left[\begin{array}{l} \int_0^{\infty} dE_t \int_{kz>0}^{E} f(E)\langle v_g^{+}(E_z)\rangle D^{+}(E_z)\langle g_{1\text{D}}(E_z)\rangle dE_z \\ +\int_0^{\infty} dE_t \int_{kz<0}^{E} f(E)\langle v_g^{-}(E_z)\rangle D^{-}(E_z)\langle g_{1D}(E_z)\rangle dE_z \end{array}\right] \tag{13}$$

where $\langle g_{1\text{D}}(E_z)\rangle$ is the effective 1D density of states [Eq. (6)], E_f is the Fermi level, $\hbar$ is the reduced Planck's constant, m^* is the effective mass, and E_t is the energy in the transverse direction.

For a given carrier concentration, the Fermi level is obtained by solving

$$n_{\text{tot}} = \sum_i \int_{E_i}^{E'} g_i(E_z) f(E) dE + \int_{E'}^{\infty} g_{3\text{D}}(E_z) f(E) dE_z \tag{14}$$

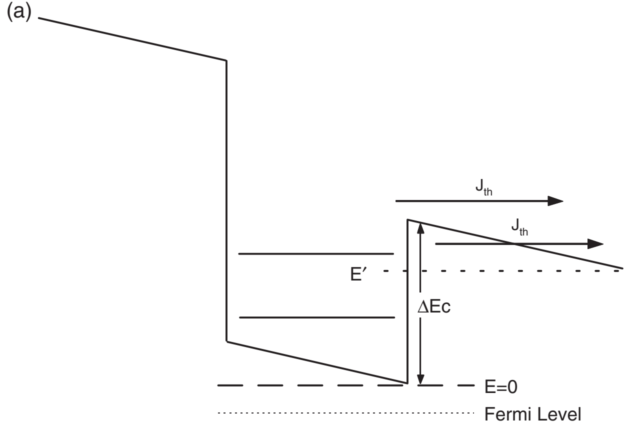
where $g_i(E_z)$ is the finite well density of states, E_i is the ith energy level below E' with a corresponding wavevector k_{zi} and $k_{z0} = 0$ and $g_{3\text{D}}(E_z)$ is the 3D density of states which is calculated by Eq. (7).

Calculation of the carrier escape time that includes the proper partitioning of the thermionic emission and tunneling components of the current is obtained by considering the total number of carriers that are able to traverse the barrier. Rearranging the usual current equation ($J = qnv = qnL/\tau$), the escape rate can be calculated as

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{un}}^{2\text{D}}}\beta + \frac{1}{\tau_{\text{un}}^{3\text{D}}}(1-\beta) \tag{15}$$

where $\tau_{\text{un}}^{2\text{D}}(=en_{2\text{D}}L_w/J_{\text{tu}})$ is the unweighted tunneling time, $\tau_{\text{un}}^{3\text{D}}(=en_{3\text{D}}L_w/J_{\text{th}})$ is the unweighted thermionic emission escape time, $\beta(=n_{2\text{D}}/n_{\text{tot}})$ is the partitioning factor, and $n_{3\text{D}}$, $n_{2\text{D}}$ are the carrier concentrations above and below E_c. According to this equation the thermionic emission and the tunneling mechanisms provide two parallel paths for carrier escape. Moreover, the partitioning of carriers in the 2D and 3D parallel paths depends upon the applied bias and the structure of the quantum well as well as the redistribution of the density of states.

The theoretical calculations are compared to the experimental work of Cavailles et al. [37] for a quantum well consisting of 200 Å AlGaAs barriers and a 96 Å GaAs well. The composition of Al is 40% and 20% yielding an asymmetric quantum well with conduction band offsets of 0.3242 and 0.1621 eV respectively (Fig. 8). The measurement of



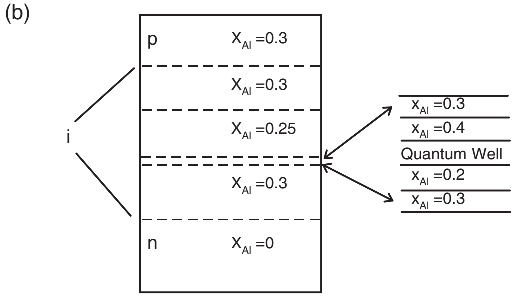


Figure 8. (a) Energy diagram of a biased quantum well and (b) *p–i–n* device structure. Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

the escape time was performed by a pump–probe method where the pumping laser produces approximately 10^{22} m^{-3} electrons within the quantum well [37] and the probing laser detected any changes in the sample at a later time with respect to the pumping laser. The material parameters used are from [8, 9, 37] and are shown in Table 1 for convenience [46].

The calculation of the electron escape time as a function of applied electric field is shown in Figure 9 (open squares). Experimental results (solid circles) [37] are also shown on the same plot and show excellent agreement with the present model. Also shown in Figure 9 is theoretical results of Moss et al. [38] (solid diamonds) which demonstrate the general trend of decreasing escape time as the electric field increases.

Three distinct regions can be identified by comparing the individual escape times (tunneling and thermionic emission) to the total escape time as shown in the inset of Figure 9: (1) a thermionic emission dominated region, (2) a tunneling dominated region, and (3) a combination tunneling and thermionic emission region. At low electric fields, the electron escape time is dominated by the thermionic emission component, whereas at high electric fields the electron escape time is tunneling dominated. At intermediate electric fields the determination of the escape time is decided by the partitioning of current/charge between the thermionic emission and tunneling components.

To better understand the role of the two current components, the ratios of the individual current components to the total current are plotted as a function of electric field in Figure 10. It is apparent that the thermionic emission component is dominant at lower bias as is evident from the ratio of J_{th}/J_{tot} whereas at high fields, J_{tu}/J_{tot} dominates. At low electric fields, the tunneling particles experience a rectangular barrier attributing to an extremely small transmission coefficient and a large tunneling energy (E' is close to E_c) resulting in a negligible tunneling current. However, for the same range of electric fields, over the barrier transport with a transmission coefficient of approximately unity contributes largely toward the total current. As the electric field increases, the barriers and the quantum well become more triangular corresponding to an increase in the transmission coefficient as well as a reduction in the tunneling energy. By reducing the tunneling energy as well as increasing the transmission coefficient, the number of carriers within the quantum well that are able to participate in J_{tu} increases, resulting in a rapid rise in J_{tu} as the electric field increases as shown in the inset of Figure 10. However, since the transmission coefficient above E_c is approximately unity and the effective velocity is approximately constant as the field increases, the steady increase in J_{th} as the electric field increases arises from an increase in the number of

Table 1. Material parameters for escape time calculations.

	$Al_{0.4}Ga_{0.6}As$	$Al_{0.2}Ga_{0.8}As$	GaAs
Band offset [eV]	0.3242	0.1621	—
m^*/m_o	0.1002	0.0836	0.067
E_g[eV]	1.92	1.67	1.424
Structure width [Å]	200	200	96

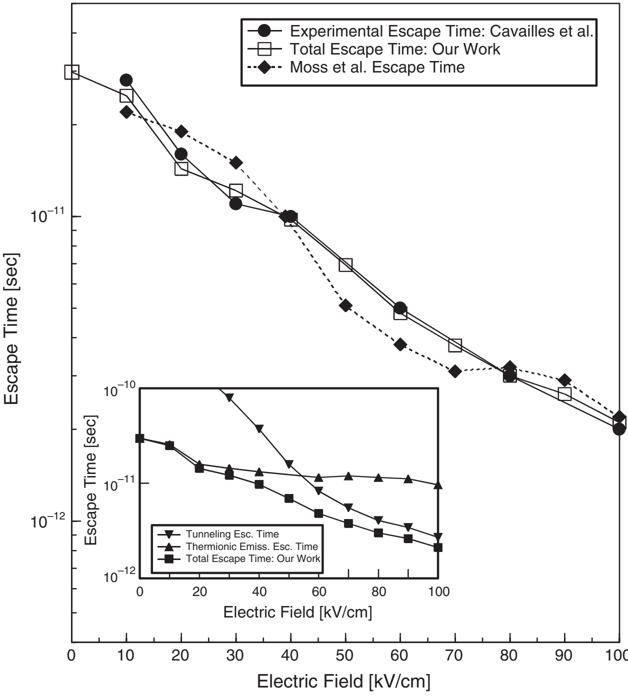


Figure 9. Comparison between the theoretical calculated escape time (□), experimental escape time (●), and Moss et al. calculated escape time (♦). The inset is the tunneling (▼) escape time, the thermionic emission escape time (▲), and the total escape time (■). Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

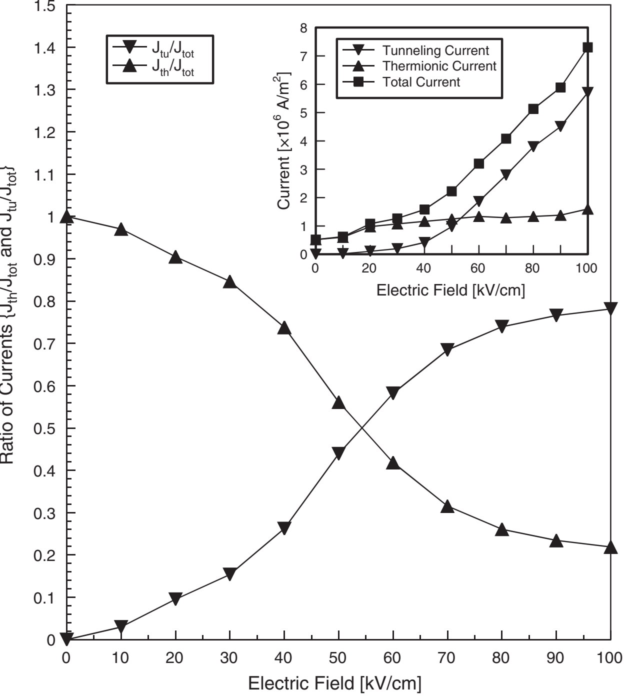


Figure 10. Comparison of the ratios J_{tu}/J_{tot} (▼) and J_{th}/J_{tot} (▲) while the inset compares the tunneling (▼), thermionic emission (▲), and the total current (■) as a function of applied electric field. Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

carriers above E_c. Therefore, as the electric field increases, J_{tu} overcomes J_{th} and eventually becomes the dominant current at high electric fields.

The increasing function of J_{th} in Figure 10 can also be explained by considering the product of $\langle\langle v_{3D}\rangle\rangle$ and n_{3D} ($\langle\langle v_{3D}\rangle\rangle$ is the local group velocity averaged over position as well as energy) which yields J_{th} as shown in Figure 11. Since the rate of increase in n_{3D} is greater than the rate of decrease in $\langle\langle v_{3D}\rangle\rangle$, J_{th} shows a steady increase with applied bias. n_{2D}, on the other hand, is much larger than n_{3D} and is a very weak function of the applied electric field. This attribute of n_{2D} along with the rapid increase in $\langle\langle v_{2D}\rangle\rangle$ results in a sharply increasing tunneling current, J_{tu}, with applied bias. Though J_{th} and J_{tu} both increase with increasing electric field, J_{tu} increases faster, becoming the major component of current at higher bias. This switch over from thermionic emission dominance to tunneling dominance for this structure occurs at around 50 kV/cm.

The calculation of the escape time based upon $\tau_{un}^{3D}(=qn_{3D}L_w/J_{th})$ and $\tau_{un}^{2D}(=qn_{2D}L_w/J_{tu})$ is not complete as carrier partitioning is not accounted for. The carrier partitioning, which is also implicit in the Moss et al. treatment, yields the correct escape time by writing it as $\tau^{-1} = \beta/\tau_{un}^{2D} + (1-\beta)/\tau_{un}^{3D}$. In Figure 12, the unweighted escape times are shown. As observed, the unweighted thermionic emission escape time always dominates, as $\langle\langle v_{3D}\rangle\rangle$ is greater than $\langle\langle v_{2D}\rangle\rangle$. Also shown in Figure 12 is the partitioning factor, β, which is close to unity for the structure considered. The weighted tunneling time ($\tau_{wt}^{3D} = \tau_{un}^{3D}/(1-\beta)$) is much greater than $\tau_{un}^{3D}$and can become comparable or even greater than the weighted tunneling time ($\tau_{wt}^{2D} = \tau_{un}^{3D}/\beta$).

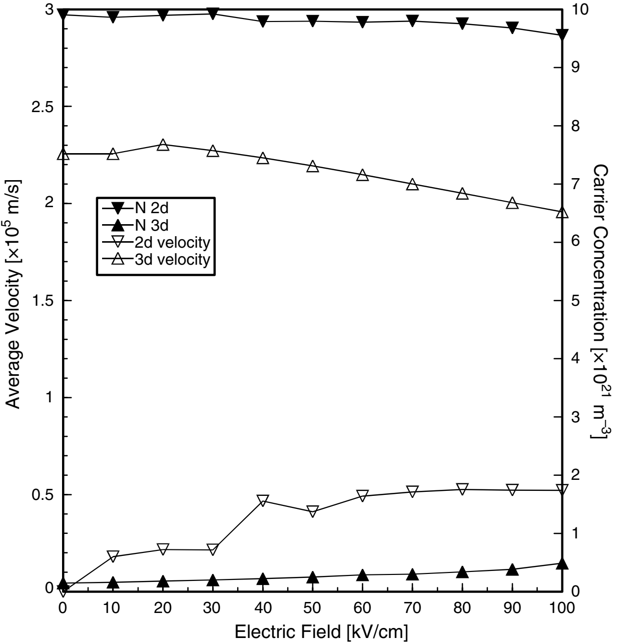


Figure 11. The carrier concentration (solid triangles) and average effective velocity (open triangles) for electrons above (▲) and below (▼) the conduction band offset as a function of applied electric field. Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

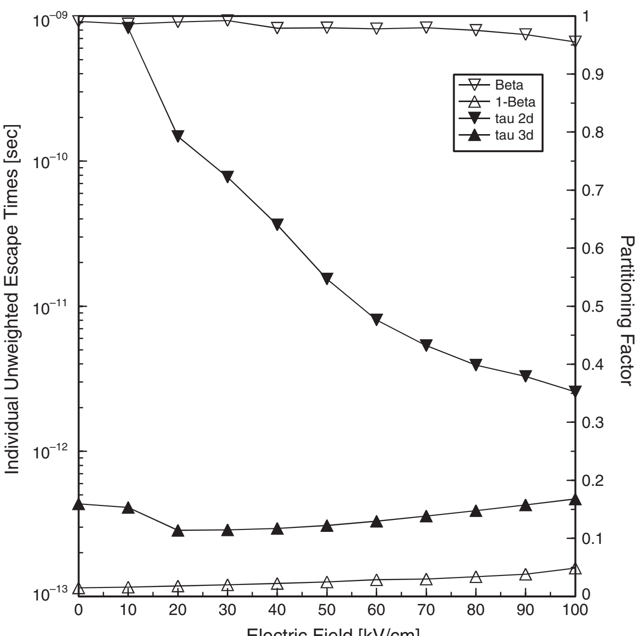


Figure 12. The partitioning factor and the unweighted tunneling escape time (▼) and the unweighted thermionic emission escape time (▲) as a function of applied electric field. Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

The importance of using the proper quantum mechanical group velocity as well as the continuous density of states is shown in Figure 13. In Figure 13 the escape time has been calculated using $\hbar k_z/m^*$ for the group velocity, including the redistribution of the density of states (solid triangles) and excluding the redistribution of the density of states (solid circles). As observed, the electron escape time is overestimated when the redistribution of the density of states is not accounted for. Also to be noticed is that the curvature of the electron escape time is opposite the experimental results when the redistribution in the density of states has been excluded; a similar situation is observed in the Moss et al. [38] calculations.

For completeness, the hole escape time is shown in Figure 14 is included. Included in the hole escape time figure are the experimental results of Cavailles et al. as well as the previously reported theoretical results of Moss et al. The theoretical escape time for holes also shows excellent agreement with the experimental results with similar plateaus in regions. In the present calculation for holes, the heavy hole masses (m_{hh}/m_o) of 0.734, 0.284, 0.454, and 0.415 and the light hole masses (m_{lh}/m_o) of 0.596 and 0.222 are used. These hole masses are obtained by fitting parabolas to the calculated band structures in [75]. The use of the data that were reported for a slightly different structure may explain the discrepancy between the experimental and the theoretical results.

Comparison between the Moss et al. work and the experimental results for the holes demonstrates large discrepancies. The large discrepancy can be attributed to the use of the conventional hole effective mass, discrete density of states, FWHM calculation for the tunneling

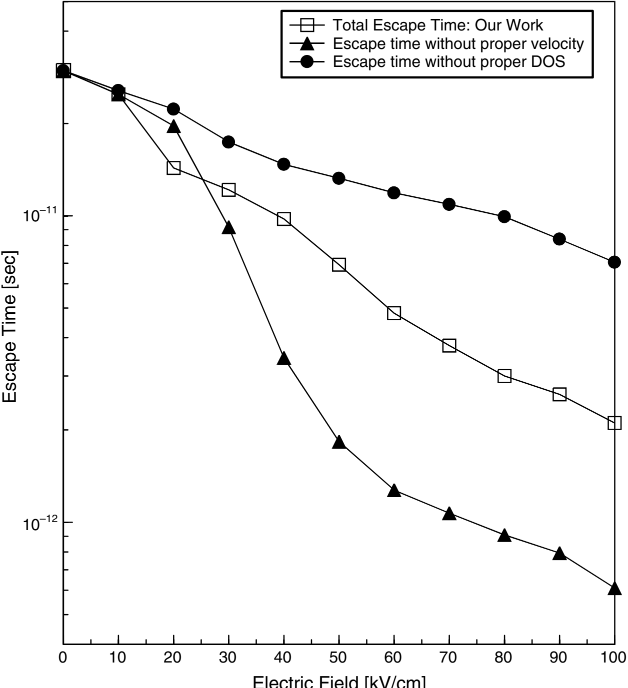


Figure 13. Comparison between of the escape times when the continuous density of states and the proper quantum mechanical group velocity are included and excluded. Reprinted with permission from [8], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997). © 1997, IEEE.

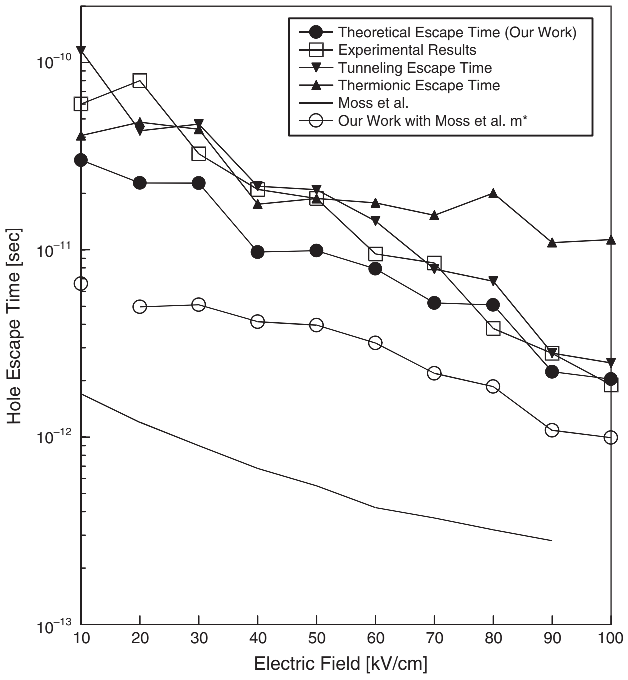


Figure 14. Comparison between the theoretical calculated hole escape time, experimental hole escape time and Moss et al. calculated hole escape time. Reprinted with permission from [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut.

times, and $\hbar k_z/m$ for the group velocity. This is demonstrated in Figure 14 where the escape time for holes has been calculated using this theory with the effective masses given by Moss et al. (open circles). As demonstrated, an improvement in the hole escape time results from using the present formulation.

4. ELECTRON–POLAR OPTICAL PHONON INTERACTION WITHIN A BIASED QUANTUM WELL

4.1. Electron–Polar Optical Phonon Interaction within an Unbiased and Biased Quantum Well

The calculation of the absorption (abs) and emission (ems) rates of phonons begins with Fermi's golden rule [46],

$$\frac{1}{\tau_{\substack{\text{abs}\\ \text{ems}}}(E_i)} = \frac{2\pi}{\hbar}\sum_{\bar{k}_f}\sum_{\bar{q}} |\langle\Psi_f|H_{ep}|\Psi_i\rangle|^2(1-f(E_f)) \times \delta(E_f - E_i \pm \hbar\omega) \tag{16}$$

where Ψ is the wavefunction in three dimensions for the initial (i) and final (f) states, E_i and E_f are the energies of the states Ψ_i and Ψ_f, $f(E_f)$ is the Fermi distribution, k_f is the final state wavevector, q is the phonon wavevector, and $\hbar\omega$ is the energy of the polar optical phonon. The positive sign in the conservation of energy ($\delta(E)$) is for the emission of a phonon while the negative sign is for the absorption of a phonon.

Assuming bulk phonons within the quantum well and neglecting screening, the interaction Hamiltonian, H_{ep}, is the Frölich Hamiltonian [46],

$$H_{ep} = \left[\frac{e^2\hbar\omega}{2V}\left[\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_o}\right]\right]^{\frac{1}{2}} \left(N_q + \frac{1}{2} \pm \frac{1}{2}\right)^{\frac{1}{2}} \left(\frac{1}{q}\right) \times \left[a_q^+ e^{-i\bar{q}\cdot\bar{r}} + a_q e^{i\bar{q}\cdot\bar{r}}\right] \tag{17}$$

where a_q is the annihilation operator, a_q^+ is the creation operator, q is the phonon wavevector, r is the position vector, e is the elemental charge, N_q is the phonon occupation number, and ε_∞ and ε_o are the high frequency and static dielectric constants, respectively.

The wavefunctions can be decomposed into transverse and perpendicular components as shown in

$$\Psi_{i,f}(r) = \frac{1}{\sqrt{V}} e^{i\bar{k}_{ti,tf}\cdot\bar{r}_t}\, \varphi_{zi,zf} \tag{18}$$

The perpendicular component ($\varphi_{zi,zf}$) is the bound (or quasi bound when an electric field is applied) particle wavefunction in the z-direction and the transverse component ($e^{i\bar{k}_{ti,tf}\bullet\bar{r}_t}$) is the free particle wavefunction parallel to the quantum well. Also, $k_{ti,tf}$ is the initial (final) wavevector in the transverse direction and r_t is the position vector. The

decomposition of the initial and final wavevectors into transverse and perpendicular components simplifies the calculation of the matrix elements, $\langle\Psi_f|H_{ep}|\Psi_i\rangle$ which yield

$$\langle\Psi_f|H_{ep}|\Psi_i\rangle = \frac{1}{V}\left[\frac{e^2\hbar\omega}{2V}\left[\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_o}\right]\right]^{\frac{1}{2}}\left(N_q+\frac{1}{2}\pm\frac{1}{2}\right)^{\frac{1}{2}} \times\left(\frac{1}{q}\right)\delta_{\bar{k}_{ti}\mp\bar{q}_t,\bar{k}_{tf}}\int_0^{L_w}\varphi_{zf}^* e^{\mp iq_z}\varphi_{zi}\,dz. \quad (19)$$

The upper sign corresponds to emission while the lower sign corresponds to absorption. The z component of the wavefunctions will be addressed later.

Using the properties $\delta(aE)=a^{-1}\delta(E)$ and $\delta((E-E_1)(E-E_2))=|E_1-E_2|^{-1}(\delta(E-E_1)+\delta(E-E_2))$, the conservation of energy and momentum in the transverse direction, as in Eqs. (16) and (19), can be combined [10] to solve for the momentum in the transverse direction as

$$\delta(E_f-E_i\pm\hbar\omega)\delta_{\bar{k}_{ti}\mp\bar{q}_t,\bar{k}_{tf}} = \frac{2m^*}{\hbar^2}\frac{1}{|Q_t^+-Q_t^-|}[\delta(q_t-Q_t^+)+\delta(q_t-Q_t^-)] \quad (20)$$

where for the absorption of phonons

$$Q_t^\pm = -k_{ti}\cos\phi\pm\sqrt{(k_{ti}\cos\phi)^2-\left(k_{zf}^2-k_{zi}^2-\hbar\omega\frac{2m^*}{\hbar^2}\right)} \quad (21)$$

and for the emission of phonons

$$Q_t^\pm = k_{ti}\cos\phi\pm\sqrt{(k_{ti}\cos\phi)^2-\left(k_{zf}^2-k_{zi}^2+\hbar\omega\frac{2m^*}{\hbar^2}\right)} \quad (22)$$

where the angle ϕ is between the initial wavevector and the phonon wavevector in the transverse direction.

As stated in Section 2, upon the application of an electric field perpendicular to the quantum well (along the z-direction), there exists a finite probability for the electrons within the quantum well to tunnel through the triangular barrier for energies greater than E'. However, for energies below E', the system is bounded and the density of states is the usual steplike function. Therefore, for the final states greater than E', the summation over the final states can be converted into an integral over energy, whereas for energies below E', the summation over final states remains. Using Eqs. (17)–(22), Eq. (16) can be rewritten (we have dropped the subscripts abs and ems for convenience) as

$$\frac{1}{\tau(E_i)} = \beta\iiint|M_{k_z q_z}|^2\frac{\langle g_{1\mathrm{d}}(E_{zf})\rangle}{|Q_t^+-Q_t^-|} \times\left[\frac{Q_t^+}{(Q_t^+)^2+q_z^2}+\frac{Q_t^-}{(Q_t^-)^2+q_z^2}\right]dq_z\,d\phi\,dE_{zf} \quad (23)$$

where

$$\beta = \frac{2m^*}{\hbar^3}\frac{L_w e^2\hbar\omega}{(2\pi)^4}\left[\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_o}\right]\left(N_q+\frac{1}{2}\pm\frac{1}{2}\right) \times(1-f(E_i\mp\hbar\omega))$$

for final states where the energy is greater than E' and

$$\frac{1}{\tau(E_i)} = \beta\sum_{k_{zf}}\iint|M_{k_z q_z}|^2\frac{1}{|Q_t^+-Q_t^-|} \times\left[\frac{Q_t^+}{(Q_t^+)^2+q_z^2}+\frac{Q_t^-}{(Q_t^-)^2+q_z^2}\right]dq_z\,d\phi \quad (24)$$

where

$$\beta = \frac{2m^*}{\hbar^3}\frac{e^2\hbar\omega}{(2\pi)^3}\left[\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_o}\right]\left(N_q+\frac{1}{2}\pm\frac{1}{2}\right) \times(1-f(E_i\mp\hbar\omega))$$

for final states in the region is below E'. Here $M_{k_z q_z} = \int_0^{L_w}\varphi_{zf}^* e^{\mp iq_z}\varphi_{zi}dz$ while the limits on the q_z and ϕ are determined by requiring the quantity beneath the radical in Eqs. (21) and (22) be positive.

The conversion of the k_{tf} summation into an integral over the final energy, E_{tf}, introduces the effective 1D density of states, $\langle g_{1\mathrm{D}}(E_z)\rangle$, into the scattering rate. Calculating the effective (1D) density of states is performed as in Section 2, Eqs. (1)–(6).

The calculation of the scattering rate, which is a function of the initial energy, can be extended to calculate the average scattering rate of phonons. The average total scattering rate is calculated by averaging $\tau_{\mathrm{tot}}^{-1}(E_{zi})(=\tau_{\mathrm{abs}}^{-1}(E_{zi})+\tau_{\mathrm{ems}}^{-1}(E_{zi}))$ over the number of carriers at that energy as shown in Eq. (24),

$$\left\langle\frac{1}{\tau}\right\rangle = \frac{\int\tau_{\mathrm{tot}}^{-1}(E_i)g_{3\mathrm{d}}(E_i)f(E_i)\,dE_i}{\int g_{3\mathrm{d}}(E_i)f(E_i)dE} \quad (25)$$

where $f(E)$ is the Fermi function and $g_{3\mathrm{d}}(E_i)$ is the 3D density of states calculated from the convolution of the effective 1D density of states $\langle g_{1\mathrm{D}}(E_z)\rangle$ and the 2D density of states [Eq. (23)].

The system under consideration consists of a single quantum well with AlGaAs barriers and a GaAs quantum well. The composition of Al is 20% producing a band offset of 0.1621 eV and a barrier effective mass of $0.0836\mathrm{m}_0$. Also, the applied electric field ranged between 0 and 70 kV/cm.

The phonon absorption, emission, and total scattering rate are plotted as a function of initial energy and applied electric field in Figure15a–d. Consider first the absorption rate (solid curves). At zero electric field (Fig. 15a) the absorption rate is zero for initial energies below the first energy level, due to a lack of initial states. With increasing initial energy, a step occurs in the absorption rate arising from the availability of initial states corresponding to the first energy level (E_1). For this structure, the separation between the two energy levels ($E_2-E_1=77$ meV) is greater than $\hbar\omega$ (36 meV); therefore, the absorption rate is determined purely by intrasubband absorption. As the initial energy is increased, there is a steady rise in the absorption rate until the initial energy equals $E_2-\hbar\omega$. Once the initial energy equals $E_2-\hbar\omega$, another step in the absorption rate results signifies that both intrasubband and intersubband scattering can occur. Further increase in the initial energy again produces a steady rise in the absorption rate

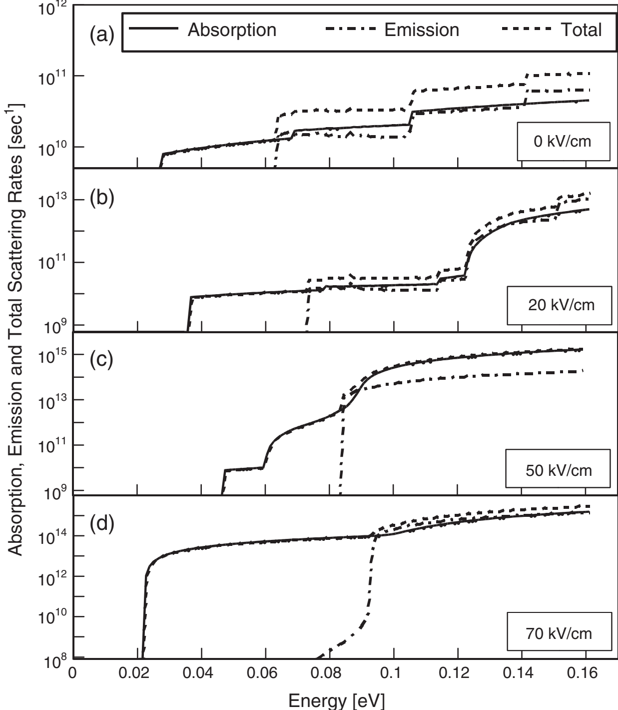


Figure 15. The electron–phonon scattering rates as a function of initial energy and the electric field as a parameter. Reprinted with permission from [10], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron*. 35, 216 (1999). © 1999, IEEE.

until the initial energy equals the second energy level. At the second energy level, a step occurs in the absorption rate as a result of an increase in the initial states. It is worth noting that if the assumption that the first energy level is the only initial state in the k_z direction and not including the second energy level, the step resulting from an increase in the initial states once the second energy level is reached (from the initial energy equaling the second energy level) would not exist in the absorption rate.

For the emission rate (dashed dotted curve) at zero electric field, the first step in the emission rate occurs when the initial energy is above the first energy by an amount equaling $\hbar\omega$. As the initial energy is further increased, there exists a slight decrease in the emission rate due to the transverse phonon wavevector increasing with an increase in initial energy and the scattering rate being inversely proportional to the transverse phonon wavevector. As the initial energy increases further another step results when the initial energy equals E_2, increasing the number of initial states which the electron can scatter from. Another step in the emission rate occurs when the initial energy equals $E_2 + \hbar\omega$ as a result in an increase in the final states of the scattered electron. The slight downward slope in the emission rate disappears since once the second energy level is reached due to the increase in the number of initial states. The total scattering rate (dashed curve) is the combination of the absorption and emission rates and demonstrates characteristics of both curves as expected.

In Figure 15b, the electric field is increased from 0 to 20 kV/cm for the phonon absorption and emission rates. As in the zero field case both the emission and absorption rates exhibit the steplike function. However, as mentioned before, the electric field redistributes the density of states above the point where the transmission coefficient becomes appreciable ($E' \sim 122$ meV for 20 kV/cm). For completeness, this is demonstrated in Figure 16a–d where the 1D density of states is plotted as a function of energy and applied electric field. With increasing electric field, the 1D density of states expands below the band offset and a spreading around the energy levels occur. Applying this knowledge to the phonon absorption and emission rates, the extra step which occurs around 122 meV corresponds to the point where the transmission is appreciable (E'). The gradual increase in the absorption and emission rates is a result of the continuous nature of the 1D density of states. Another difference between the zero field scattering rates and the 20 kV/cm scattering rates is the shift in the steps toward higher energies.

Increasing the electric field further (Fig. 15c) to 50 kV/cm demonstrates evidence of a second energy level in the absorption rate; however, the smoothing effect is more pronounced than in the previous cases. As before, the first step in the absorption curve is due to the first energy level while the second step is a result of E'. However, there exists a flattening in the absorption curve resulting from the large energy separation (>36 meV) between E' and the peak in the 1D density of states (Fig. 16c). As the initial energy is increased further to around 90 meV, the absorption rate turns up rapidly which is 36 meV from the peak in the 1D density of states. Any further increase in energy results in

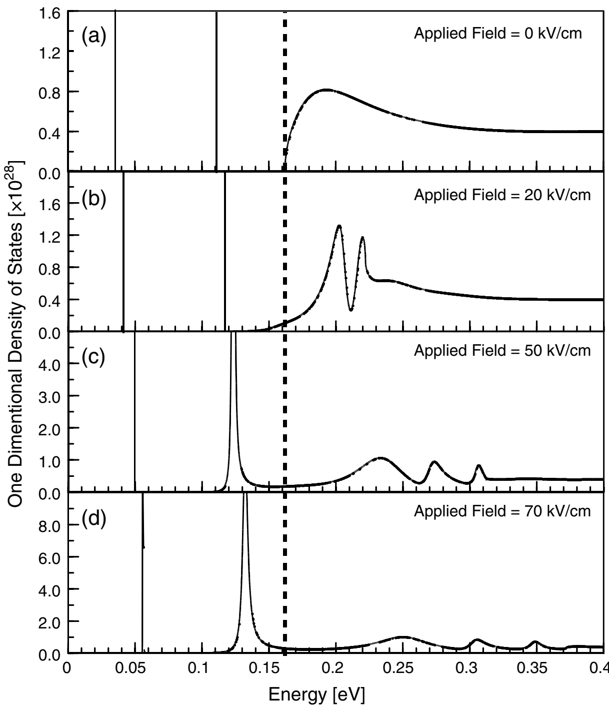


Figure 16. The effective one-dimensional density of states as a function of energy and applied electric field where the vertical dotted line is the band offset. Reprinted with permission from [10], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron*. 35, 216 (1999). © 1999, IEEE.

the gradual increase in the absorption rate as demonstrated previously. The rate at which electrons emit phonons for the 50 kV/cm case begins at 36 meV above the first energy level, as has been demonstrated before, and increases gradually with increasing energy. A further increase in the electric field to 70 kV/cm (Fig. 15d) shows a switch in the emission rates from turning on at higher energies to turning on at a lower energy. This is a result of E' falling well below the first energy level ($E' \sim 20$ meV), implying that the quantum well is almost completely unbounded. The spreading in the 1D density of states is more dramatic for the second energy level (Fig. 16d), resulting in a smoother transition in the absorption rate from the first step to the second step.

The average total scattering rate is shown in Figure 17a function of applied electric field along with the energy difference between the energy level, E', and $\hbar\omega$. which demonstrates a direct correlation between the average scattering rate and the energy separation. At electric fields of 10 and 40 kV/cm, there is a peak in the average scattering rate which occurs at a point where there is a minimum in the energy separation. As shown in Figure 17 at 40 kV/cm both energy levels are within $\hbar\omega$ from E', resulting in a sharp rise in both the emission and absorption rates. This is also demonstrated for the 20 kV/cm case (Fig. 15b), except that only the second energy level is within $\hbar\omega$ from E'.

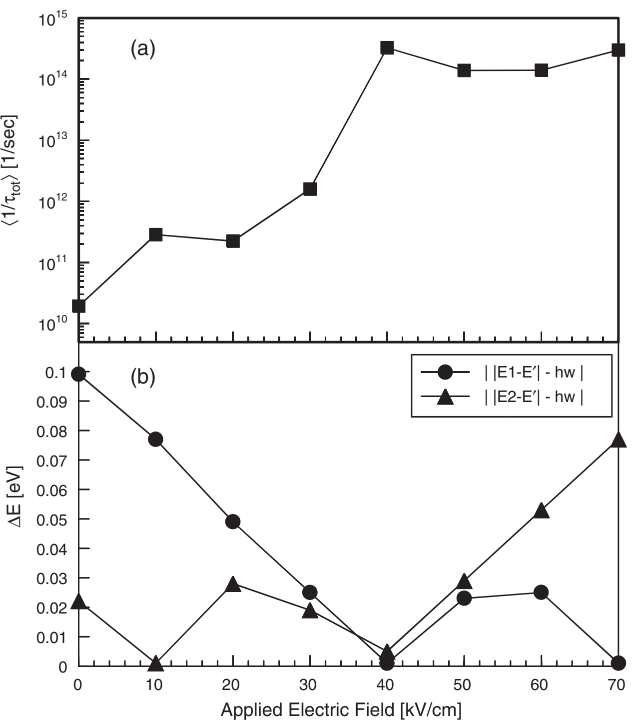


Figure 17. (a) The average electron–phonon scattering rate as a function of applied electric field. (b) The comparison between the energy level separation E_T and $\hbar\omega$. Reprinted with permission from [10], K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 35, 216 (1999). © 1999, IEEE.

4.2. Electron Escape via a Polar Optical Phonon Interaction and Tunneling from a Biased Quantum Wells

The escape rate of electrons due to a phonon interaction as a function of initial energy is calculated by first order perturbation as [46]

$$\frac{1}{\tau_{\text{abs}}(E_i)} = \frac{2\pi}{\hbar}\sum_{\bar{k}_f}\sum_{\bar{q}} |\langle \Psi_f|H_{ep}|\Psi_i\rangle|^2 \times (1-f(E_f))\,\delta(E_f - E_i - \hbar\omega) \quad (26)$$

the parameters for Eq. (26) are defined in Section 4.2.1 and only the upper sign is used in Eqs. (16) and (19).

Combining the conservation of energy and momentum in the transverse direction and solving for q_t, as in Eqs. (16) and (17) [10], yields

$$\delta(E_f - E_i - \hbar\omega)\delta_{\bar{k}_{ti}-\bar{q}_t,\bar{k}_{tf}} = \frac{2m^*}{\hbar^2}\frac{1}{|Q_t^+ - Q_t^-|} \times [\delta(q_t - Q_t^+) + \delta(q_t - Q_t^-)] \quad (27)$$

where

$$Q_t^{\pm} = -k_{ti}\cos\phi \pm \sqrt{(k_{ti}\cos\phi)^2 + \left(k_{zf}^2 - k_{zi}^2 - \hbar\omega\frac{2m^*}{\hbar^2}\right)} \quad (28)$$

and ϕ is the angle between the initial wavevector (k_{ti}) of the particle and the phonon wavevector (q_t) in the transverse direction.

The redistribution of the density of states can by included into the scattering rate calculation by converting the k_{zf} summation into an integral as the final energy, E_{zf}, introduces the effective 1D density of states, $\langle g_{1D}(E_z)\rangle$. The 1D density of states can be calculated as in Section 2. Equation (26) now may be rewritten as

$$\frac{1}{\tau(E_i)} = \beta \iiint |M_{k_z q_z}|^2 \frac{\langle g_{1d}(E_{zf})\rangle}{|Q_t^+ - Q_t^-|} \times \left[\frac{Q_t^+}{(Q_t^+)^2 + q_z^2} + \frac{Q_t^-}{(Q_t^-)^2 + q_z^2}\right] dq_z\, d\phi\, dE_{zf} \quad (29)$$

where

$$\beta = \frac{2m^*}{\hbar^3}\frac{L_w e^2\hbar\omega}{(2\pi)^4}\left[\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_o}\right](N_q)(1 - f(E_i + \hbar\omega))$$

Here $M_{k_z q_z} = \int_0^{L_w} \varphi_{zf}^* e^{-iq_z z}\varphi_{zi}\, dz$ while the limits on ϕ are determined by requiring the quantity beneath the radical in Eq. (28) be positive. The limits on the E_{zf} integration are from the top of the conduction band offset to infinity since we are considering the transition of an electron from a state below the conduction band offset to a state above the conduction band offset.

Since the tunneling rate is a function of the z-directed energy, the transition rate due to phonons needs to be converted to only a function of E_z. The calculation of the scattering rate, which is a function of the initial energy, into a

function of only E_z can obtained by the averaging scattering rate of phonons over the transverse initial states as

$$\left\langle \frac{1}{\tau(E_{z_{ti}})} \right\rangle = \frac{\int \tau^{-1}(E_i) f(E_{ti}) dE_{ti}}{\int f(E_{ti}) dE_{ti}} \tag{30}$$

where $f(E_{ti})$ is the Fermi function. Just as in Eq. (30), the average scattering rate is calculated by averaging $\langle \tau^{-1}(E_{z_{ti}}) \rangle$ over the number of electrons at E_{zi} as

$$\left\langle \frac{1}{\tau} \right\rangle = \frac{\int \langle \tau^{-1}(E_{zi}) \rangle \langle g_{1d}(E_{zi}) \rangle f(E_{zi}) dE_{zi}}{\int \langle g_{1d}(E_{zi}) \rangle f(E_{zi}) dE_{zi}} \tag{31}$$

The structure simulated is a symmetric $Al_xGa_{1-x}As/GaAs$ system with a 100 Å quantum well and 240 Å barriers. An Al mole fraction of $x = 20\%$ is used, resulting in a band offset of 0.1621 eV and two energy levels with a separation greater than the polar optical phonon energy, $\hbar\omega$ (0.036 eV). The electron's effective mass in the quantum well is $0.067m_0$ while the electron's effective mass in the barrier material is $0.084m_0$ with m_0 being the free electron mass. The two competing processes for carrier escape out of a quantum well are tunneling and phonon interaction. As demonstrated in Figure 18, tunneling is related to a spatial dislocation of the carrier whereas phonon interaction modifies the carrier's energy. The carriers that are assisted by phonons to move from within the quantum well to continuum are eventually swept away by thermionic emission. It should also be pointed out that the energy at which the transmission coefficient becomes appreciable for an electron tunneling out of the quantum well is denoted as E' as shown in Figure 18, implying that any electrons with a z-directed energy above E' and below E_c (conduction band offset) can participate in the tunneling current.

In Figure 19, the electron transition rate via a phonon interaction is plotted as a function of initial energy in the z-direction for an applied bias of 10 kV/cm. The damped oscillatory behavior of $1/\tau$ shows a peak for energies within the quantum well followed by a dip. This behavior is explained in the context of the density of states. For completeness, in Figure 20, the 1D density of states is plotted as a function of E_z. As demonstrated in Figure 20, there exist states (initial states) between the upper most energy level and the band offset as well as a virtual bound state above the band offset (final states). The initial states below the band offset allow phonon induced transitions to the final states above the band offset, creating the peak in the phonon transition rate. It should be mentioned that such a transition is only possible due to the redistribution of the states in the presence of an applied electric field. The existence of states below the band offset but above the upper most energy level make possible a phonon assisted electron transition with k_t equal to zero even though the difference between the band offset and the upper most energy level is greater than $\hbar\omega$. This is demonstrated in the inset of Figure 19 where the phonon induced transition rate has been plotted as a function of E_z for $k_t = 0$. This is unlike previous calculations where if the difference between the band offset and the energy level is greater than $\hbar\omega$, the transition of an electron from the below the band offset to above the band offset via a phonon is only possible with finite k_t.

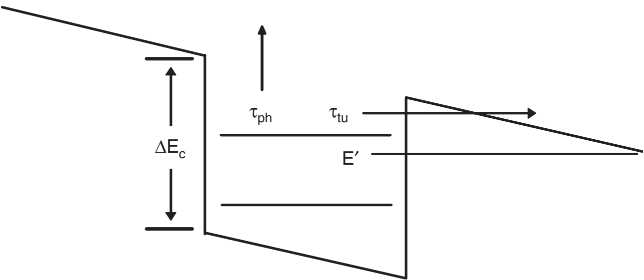


Figure 18. A quantum well under the influence of a perpendicular electric field. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

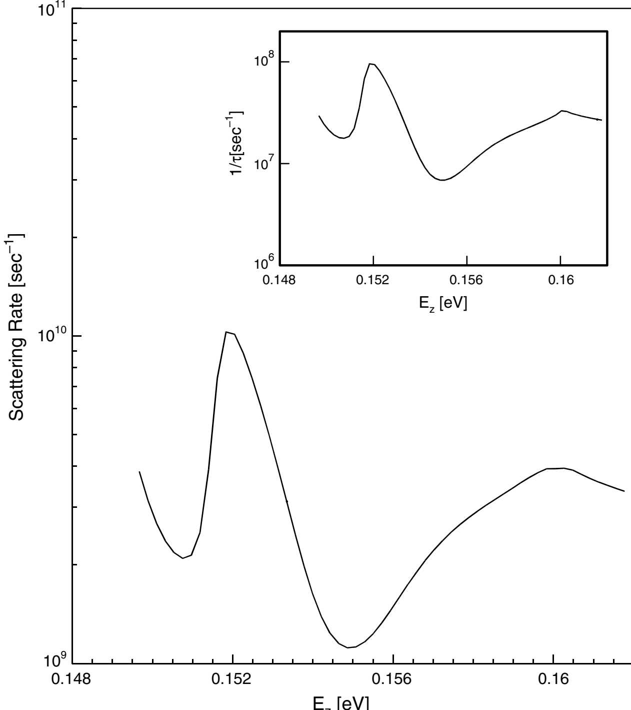


Figure 19. The electron–phonon transition rates as a function of initial z-directed energy for an electric field of 10 kV/cm. The inset is the electron–phonon transition rates for only the z-component. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

The damped oscillatory behavior of $1/\tau$ is not only due to the redistribution of the density of states, the matrix elements also play a role as is evident from Figures 21 and 22. In Figures 21 and 22, the transition rate and the density of states are plotted as a function of E_z for an applied field of 20 kV/cm. As shown from Figure 22, at an electric field of 20 kV/cm, the first virtual bound state lies beyond $E_c + \hbar\omega$. Therefore, the dip followed by the hump in $1/\tau$ (Fig. 21) is not due to the redistribution of the density of states but is a result of the matrix elements between the initial and final states. This observation is justified in the inset of Figure 21 where $1/\tau$ follows the 1D density of states and is obtained by forcing the matrix elements to a constant.

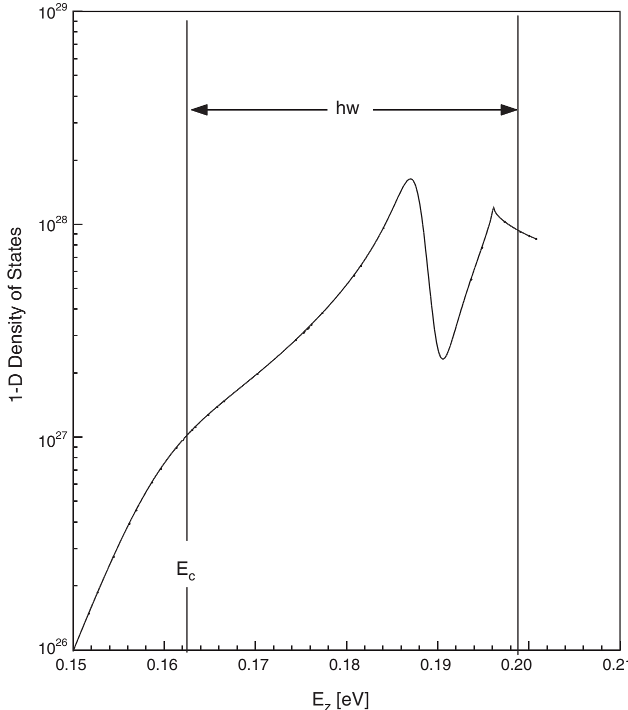


Figure 20. The effective one-dimensional density of states as a function of energy for an applied electric field of 10 kV/cm where the dotted vertical lines are for the band offset and $\hbar\omega$ above the band offset. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

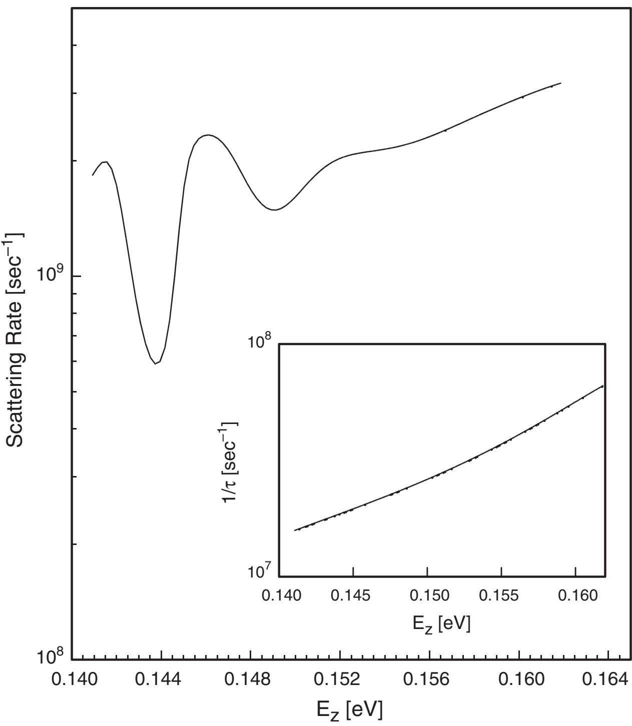


Figure 21. The electron–phonon transition rates as a function of initial z-directed energy for an electric field of 20 kV/cm. The inset is the electron–phonon transition rate for a constant value for the matrix elements. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

Further increasing the electric field to 50 kV/cm (Fig. 23) modifies the $1/\tau$ and density of states vs E_z curves considerably. As evident from the density of states plot, a bound state exists at 0.122 eV and there are no virtual bound states within $E_c + \hbar\omega$. The small hump at $E_z = 0.122$ eV results from a phonon assisted transition from the bound state to the continuum states below the band offset. In the absence of any states between the bound states and the band offset such a transition would be impossible and is attributed to the field induced redistribution of the density of states. Increasing E_z to 0.126 eV, a phonon assisted transition from within the well to the continuum is possible. Though $1/\tau$ increases sharply at 0.126 eV, the magnitude, when compared to Figures 19 and 21, is less and is due to the lower number of final states.

An electron in a biased quantum well experiences two competing processes: a change in energy due to a phonon assisted transition or a change in position via tunneling out of the quantum well. In Figure 24, the two competing processes, the phonon assisted electron transition time, $\langle\langle\tau_{\text{ph}}\rangle\rangle$, and the electron tunneling time, $\langle\langle\tau_{\text{tun}}\rangle\rangle$, are plotted as a function of applied electric field. The calculation of the tunneling times has been performed by the method of Anwar et al. in [6, 7]. As is evident, at low electric fields, phonon assisted transitions dominate over tunneling; however, with increasing electric field, tunneling dominates. To clearly understand the process, the scattering times as a function of energy, E_z, are plotted for an applied field of 50 kV/cm in Figure 25. For energies around 0.12 eV, the density of states reaches a maximum value where, from Figure 25, tunneling dominates. At higher energies, though phonon scattering time is large, the number of carriers at these energies is negligible, making its contribution insignificant toward the escape time calculation.

The dominant transport mechanism responsible for carrier transport, that mainly contributes toward escape, is a strong function of conduction band offset and width of the quantum well. In Figure 26, the phonon assisted electron transition time and the tunneling time are compared by plotting the probability of tunneling $[(1/\tau_{\text{tun}})/(1/\tau_{\text{tun}} + 1/\tau_{\text{phon}})]$ as a function of applied field for quantum well widths of 40, 70, and 100 Å and band offsets of 0.1621 eV (dashed line—20% Al) and 0.2 eV (solid line—30% Al). Starting with the 20% Al cases, for electric fields greater than 20 kV/cm the probability of tunneling is larger than the probability of a transition due a phonon interaction. For small electric fields (<20 kV/cm), there exists a virtual bound state within $\hbar\omega$ of the band offset, increasing the number of final states for phonon absorption and the probability of phonon absorption. However, with increasing electric field, the virtual bound state increases in energy beyond $\hbar\omega$ above E_c, reducing the number of final states and the probability of phonon absorption. For electric fields not exceeding 40 kV/cm the energy difference between E' and E_c is less than $E_c - \hbar\omega$ (see Fig. 18). This implies that the number of electrons available, with energy E_z, to absorb a phonon

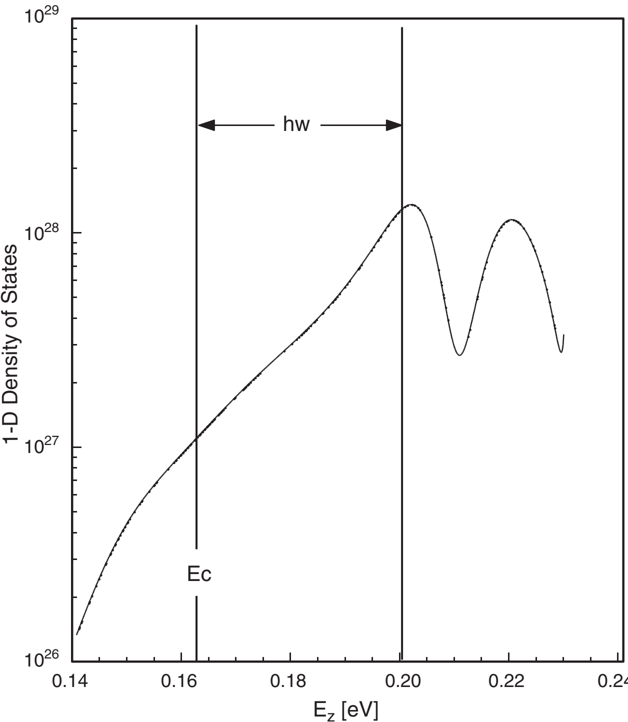


Figure 22. The effective one-dimensional density of states as a function of energy for an applied electric field of 20 kV/cm where the dotted vertical lines are for the band offset and $\hbar\omega$ above the band offset. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

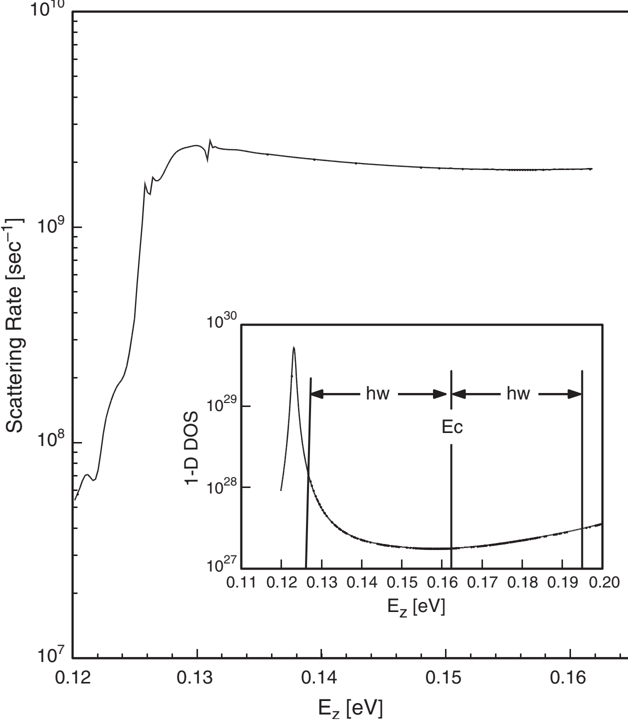


Figure 23. The electron–phonon transition rates as a function of initial z-directed energy for an electric field of 50 kV/cm. The inset is the one-dimensional density of states as a function of energy where the dotted vertical lines are for the band offset and $\hbar\omega$ above and below the band offset. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

is equal to the number of electrons available for tunneling. However, for electric fields greater than 40 kV/cm, $E_c - E'$ is greater than $E_c - \hbar\omega$, implying that the number of electron available, with energy E_z, for tunneling is greater than the number for phonon absorption. Furthermore, not only is the number of electrons available important in the removal of an electron from a quantum well; the transmission coefficient is an important parameter for tunneling while the number of final states is important for phonon absorption, both of which change with applied electric field. It is worth noting that the dip which occurs in the 70 Å quantum well between 60 to 80 kV/cm arises from an increase in the number of final states above E_c, increasing the probability of phonon absorption. However, as $E_c - E'$ is much greater than $E_c - \hbar\omega$, the effects of phonon absorption on the probability of tunneling is small. It should also be pointed out that for electric fields below 40 kV/cm the probability of tunneling decreases with increasing well width for the same electric fields. This results from the virtual bound states residing closer to E_c with increasing well width as well as being broader for larger quantum well widths, resulting in a higher probability of phonon absorption.

As the well width increases, the energy levels are located deeper in the well while the virtual bound states are located closer to the band offset at low electric fields. Therefore, with increasing well width, a larger electric field is required to move the virtual bound state above $E_c + \hbar\omega$, making the transition due to phonons more probable over a larger range of electric fields for larger well widths as shown in Figure 26. However, by increasing electric field even further, tunneling overcomes phonon absorption since there are more electrons available for tunneling than phonon absorption. This is the same situation for the case of the 30% Al composition except that the virtual bound states are broader, yielding an even larger range of electric fields where phonon transition is more probable for the 40 and 70 Å quantum wells. However, as see by Figure 26, the 100 Å quantum well, 30% Al composition demonstrates unique characteristics. This results from the 100 Å quantum well beginning with three energy levels in the well, one close to the top and two much deeper within the well, for electric fields below 20 kV/cm and the existence of a virtual bound state within $\hbar\omega$ of the band offset, producing a high probability of phonon absorption for small electric fields. However, as the electric field increases above 20 kV/cm, the probability of phonon absorption increases rapidly creating the dip in the probability of tunneling. This results from the upper most energy level being within the quantum well for small electric fields but residing above E_c within increasing electric field. This increases the probability of phonon absorption rapidly, decreasing the probability of tunneling. As the electric field is increased further, the virtual bound states exists beyond $E_c + \hbar\omega$ which reduces the probability of phonon absorption. This occurs around 80–90 kV/cm resulting in the sharp upswing in the probability of tunneling.

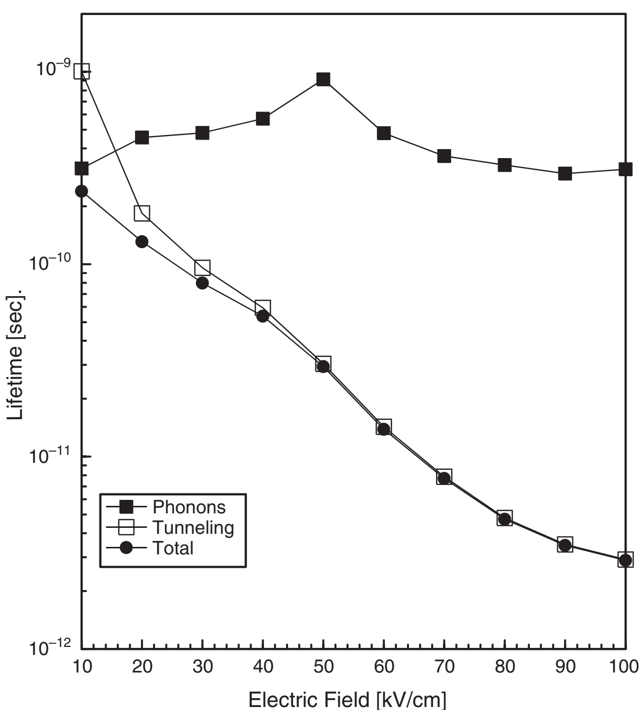


Figure 24. A comparison between the tunneling escape time and the electron–phonon transition time as a function of applied electric field. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

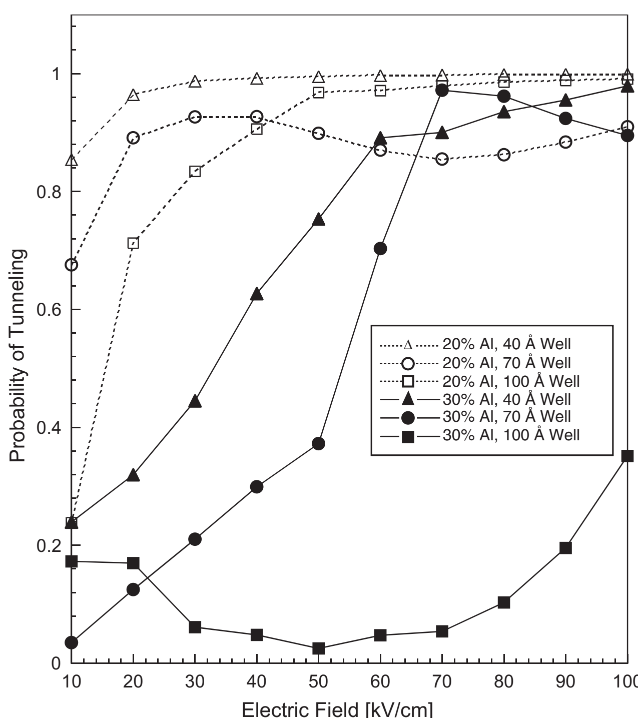


Figure 26. A comparison of the probability of tunneling as a function of applied electric field, Al concentration, and well width. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

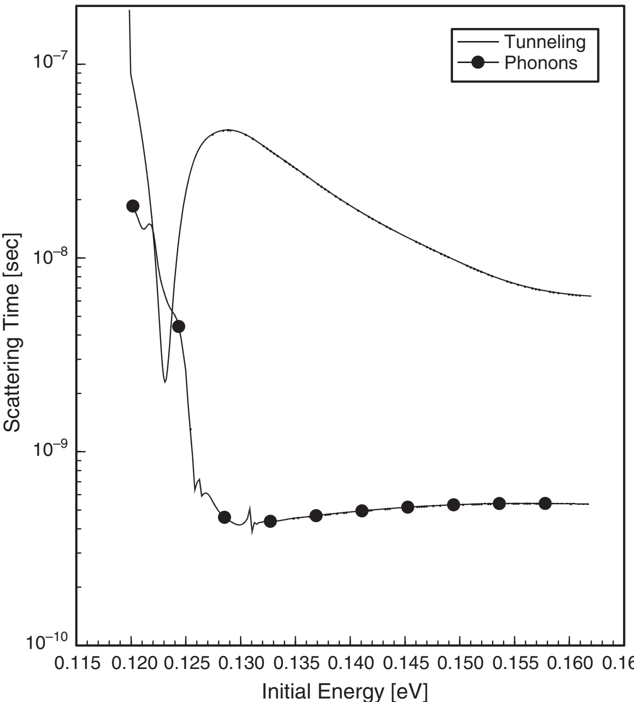


Figure 25. A comparison between the tunneling escape time and the electron–phonon transition time as a function of z-directed energy for 50 kV/cm. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

The temperature dependence on the competing escape times is demonstrated in Figure 27 where of the probability of tunneling is plotted for the 20% Al case as a function of electric field, quantum well width, and temperatures of 300 and 77 K. As one would expect, with decreasing temperature both the tunneling rate and the phonon scattering rate would increase. This is indeed the case and is demonstrated in Figure 27, where the individual scattering rates for the 40 Å quantum well have been plotted as a function of electric field and temperature. As is demonstrated in Figure 27, the tunneling rate has a stronger temperature dependence than the phonon scattering rate, making the probability for tunneling at the 77 K less than 300 K even though both interactions increase with decreasing temperature.

4.3. Dynamics of Electron Capture into a Biased Quantum Well

An electron residing above the conduction band offset of a quantum well can be captured within the quantum well by emitting a polar optical phonon. The rate at which this transition can occur is calculated by first order perturbation theory as [10]

$$\frac{1}{\tau(E_{zi})} = \frac{2\pi}{\hbar} \sum_{\bar{k}_f} \sum_{\bar{q}} |\langle \Psi_f | H_{ep} | \Psi_i \rangle|^2 (1 - f(E_f)) \times \delta(E_f - E_i + \hbar\omega) \tag{32}$$

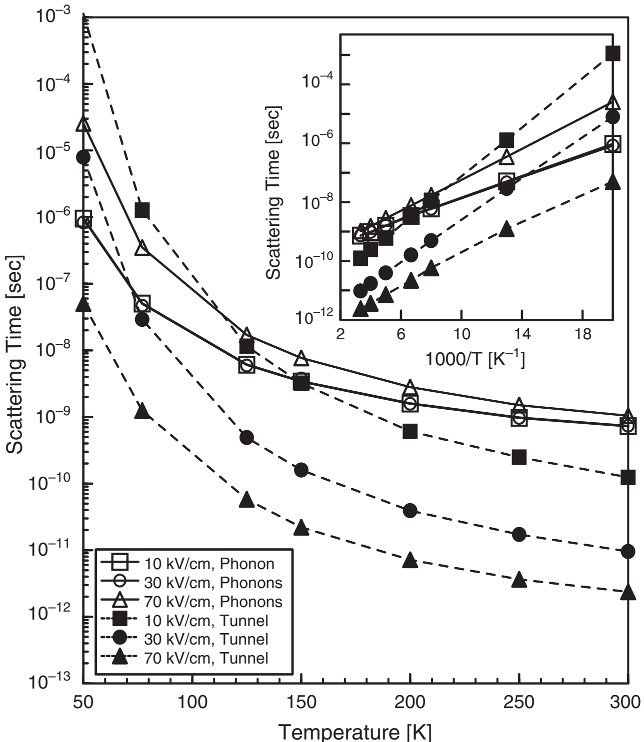


Figure 27. Scattering times as a function of temperature and electric field. Reprinted with permission form [46], K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.

where the parameters for Eq. (32) are defined in Section 4.2.1.

As stated previously, the application of an electric field perpendicular to a quantum well creates an open system, producing a continuum of states below the conduction band offset. The range of energies below the conduction band offset that becomes continuous varies as a function of electric field, a larger region of continuum with increasing electric field. If the final k_z states are within the continuum region, the summation over final k_z states can be converted over to an integral over final energy, E_z, states as

$$\frac{1}{\tau(E_{zi})}=\frac{2\pi}{\hbar}\frac{V}{(2\pi)^3}\int_{E'}^{E_c}\int_0^{\infty}\sum_{\bar{q}}|\langle\Psi_f|H_{ep}|\Psi_i\rangle|^2(1-f(E_f))$$
$$\times\,\delta(E_f-E_i+\hbar\omega)\langle g_{1\mathrm{d}}(E_{zf})\rangle dk_t\,dE_{zf} \qquad (33)$$

where the parameters for Eq. (33) are defined in Section 4.2.1. However, if the final energy, E_z, states are below E', the summation over final k_z states remains. The effective one-dimensional density of states is calculated by the method in Section 2.

Separating the wavefunctions into a transverse component ($e^{i\bar{k}_{ti,tf}\bullet\bar{r}_t}$) and a z-component ($\varphi_{zi,zf}$), the matrix elements, $\langle\Psi_f|H_{ep}|\Psi_i\rangle$, become

$$\langle\Psi_f|H_{ep}|\Psi_i\rangle=\frac{1}{V}\left[\frac{e^2\hbar\omega}{2V}\left[\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_o}\right]\right]^{\frac{1}{2}}(N_q+1)^{\frac{1}{2}}\left(\frac{1}{q}\right)$$
$$\times\,\delta_{\bar{k}_{tf}-\bar{q}_t,\bar{k}_{ti}}\int_0^{L_{\mathrm{well}}}\varphi_{zf}^{*}e^{+iq_z}\varphi_{zi}\,dz \qquad (34)$$

where e is the elemental charge, N_q is the phonon occupation number, ε_∞ and ε_o are the high frequency and static dielectric constants, φ_z is the initial (i) and final wavefunction (f), and $\delta_{\bar{k}_{tf}-\bar{q}_t,\bar{k}_{ti}}$ is the conservation of momentum in the transverse direction.

Combining the conservation of energy and momentum in the transverse direction, as in Eqs. (33) and (34) [10], yields

$$\delta(E_f-E_i\pm\hbar\omega)\delta_{\bar{k}_{ti}\mp\bar{q}_t,\bar{k}_{tf}}$$
$$=\frac{2m^*}{\hbar^2}\frac{1}{|Q_t^+-Q_t^-|}\Big[\delta(q_t-Q_t^+)+\delta(q_t-Q_t^-)\Big] \qquad (35)$$

where

$$Q_t^{\pm}=-k_{tf}\cos\phi\pm\sqrt{(k_{tf}\cos\phi)^2+\left(k_{zf}^2-k_{zi}^2+\hbar\omega\frac{2m^*}{\hbar^2}\right)} \qquad (36)$$

and the angle ϕ is between the final wavevector and the phonon wavevector in the transverse direction.

Combining Eqs. (33), (35), and (36) yields the capture rate as a function of initial energy,

$$\frac{1}{\tau(E_{zi})}=\beta\iiint|M_{k_zq_z}|^2\frac{\langle g_{1\mathrm{d}}(E_{zf})\rangle}{|Q_t^+-Q_t^-|}$$
$$\times\left[\frac{Q_t^+}{(Q_t^+)^2+q_z^2}+\frac{Q_t^-}{(Q_t^-)^2+q_z^2}\right]dq_z\,d\phi\,dE_{zf} \qquad (37)$$

where

$$\beta=\frac{2m^*}{\hbar^3}\frac{L_{\mathrm{well}}e^2\hbar\omega}{(2\pi)^4}\left[\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_o}\right](N_q+1)(1-f(E_i-\hbar\omega))$$

and $M_{k_zq_z}$ is the matrix elements.

The average capture rate can be calculated by averaging the capture rate over the number of electrons at E_{zi} as

$$\left\langle\frac{1}{\tau}\right\rangle=\frac{\int\langle\tau^{-1}(E_{zi})\rangle\langle g_{1\mathrm{d}}(E_{zi})\rangle f(E_{zi})dE_{zi}}{\int\langle g_{1\mathrm{d}}(E_{zi})\rangle f(E_{zi})dE_{zi}} \qquad (38)$$

where the integral over energy is from E_c to infinity.

An AlGaAs/GaAs single quantum well is utilized to demonstrate the electron capture rate for a biased quantum well as a function of quantum well width and conduction band offset. The parameters used in this calculation are shown in Table 2. Figure 28 demonstrates the capture of an electron by the emission of a polar optical phonon into a biased quantum well where, as stated previously, E' is the energy at which the transmission coefficient becomes appreciable.

In Figure 29, the capture rate is plotted as a function of z-directed energy above the conduction band offset (initial energy) for electric fields between 0 and 50 kV/cm. The conduction band offset (E_c) and the quantum well width

Table 2. System parameters for capture calculation.

Composition of Al	20% Al	30% Al
Band offset	0.1621 eV	0.2432 eV
Quantum well m^*	$0.067m_0$	$0.067m_0$
Barrier m^*	$0.0836m_0$	$0.092m_0$
Phonon energy	0.036 eV	0.036 eV

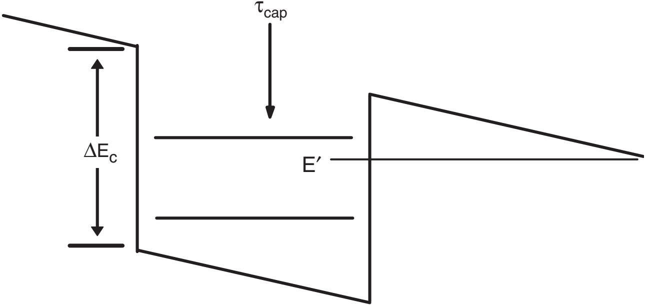


Figure 28. Energy band diagram for a quantum well with an applied electric field.

are 0.1621 eV and 100 Å, respectively. The vertical line in Figure 29 represents the energy $\hbar\omega$ above band offset. Since the bound energy levels are at least 36 meV below the band offset, the capture rate for the zero field case (▼) decreases with increasing initial z-directed energy as demonstrated in Figure 29. For a bound state removed from the band offset by $\hbar\omega$, an electron capture requires that the initial and final transverse wavevectors are nonzero for zero electric field. However, through an application of an electric field, the requirement of a tranverse momentum is relaxed for the capture of an electron and is demonstrated in Figure 29 for electric fields greater than zero.

Increasing the electric field to 10 kV/cm (◯), the capture rate for energies below 0.185 and above 0.198 eV is similar to the zero electric field case. However, for energies between 0.185 and 0.198 eV, the capture rate deviates away from the zero electric field case. This is a result of the applied electric field redistributing the final states below the band offset, creating a continuum of states below E_c. The point at which the continuum of energies below the band offset begins will be defined as E'. This is demonstrated in Figure 30 where the redistribution of the density of states is plotted, for completeness, as a function of energy and applied electric field. Through an increase in the number and location of final states, the capture rate increases in an energy range where, for zero electric field, the capture rate would be less. Furthermore, for an electric field greater than zero there exists a capture rate when the final transverse momentum equals zero and the difference between the eigenenergies and E_c is greater than $\hbar\omega$. This is demonstrated in the inset of Figure 29 where the capture rate is plotted as a function of E_z and electric fields when the momentum in the transverse direction is zero for the final states (k_{tf}). For an electric field of 10 kV/cm, the capture rate for $k_{tf} = 0$ resides in the energy range between 0.185 and 0.198 eV since electrons in this energy range can emit a phonon and be captured into the states between 0.15 eV and E_c (0.1621 eV) as shown in Figure 29 and in the inset of Figure 29.

Increasing the electric field further redistributes the states below E_c (Fig. 30) and widens the range of initial energies where a transverse final momentum is not required for a transition. This is demonstrated in Figure 29 (and inset) where the energy range $\hbar\omega + E'$ to $\hbar\omega + E_c$ expands, broadening the plateau in the capture rate which occurred for the 10 kV/cm case. It is also worth mentioning that $E_c + \hbar\omega$ occurs at 0.198 eV, implying that if there is any transition due to phonon emission, it must occur with a final transverse momentum and is demonstrated by a sharp drop in

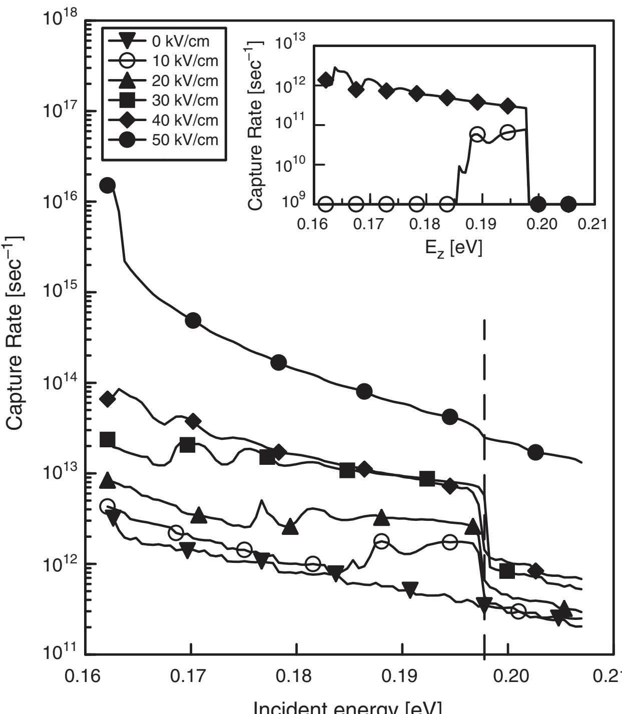


Figure 29. Capture rate as a function of incident energy and the capture rate as a function of z-directed incident energy in the inset.

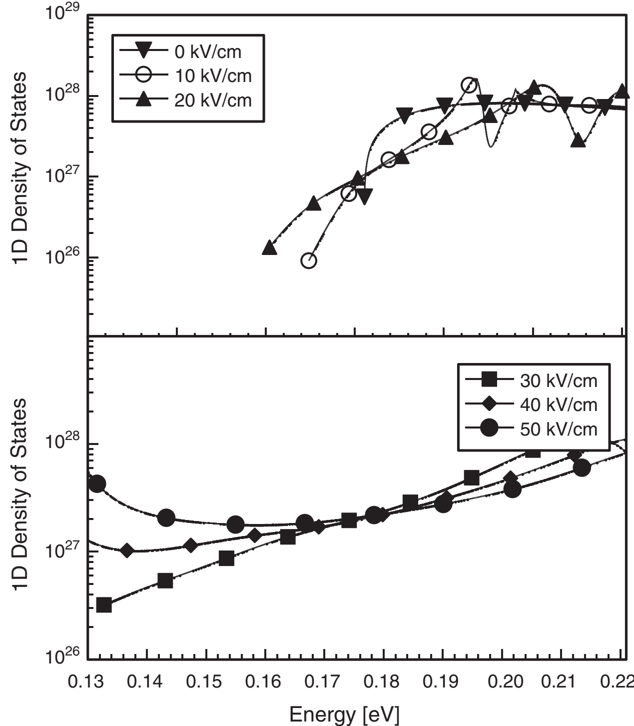


Figure 30. One-dimensional density of states as a function of energy and electric field.

the capture rate at 0.198 eV in the inset of Figure 29. In addition, the gradual rise in the capture rate for increasing electric field arises from an increase in the total number of final states (transverse and z-directed) as the electric field increases.

Increasing the band offset to 0.2434 eV (30% Al) produces an eigenenergy within $\hbar\omega$ from E_c. This allows transitions to occur when the transverse momentum is equal to zero for any electric fields. This is demonstrated in Figure 31 where the capture rate is plotted as a function of initial energy with the applied electric field as a parameter for $E_c = 0.2434$ eV. The sharp peak in the capture rate (~0.265 eV for zero electric field) is a result of the final state eigenenergy being within $\hbar\omega$ from E_c. The inset in Figure 31 is the capture rate when $k_{tf} = 0$ for various electric fields and as a function of energy. It is worth mentioning that unlike the 20% Al ($E_c = 0.1621$ eV) case, the capture rate for the 30% Al case does not generally increase with increasing electric field. This can also be explained in the context of the redistribution of the density of states which is plotted in Figure 32 as a function of energy for $E_c = 0.2434$ eV and for various applied electric fields. As the electric field increases, the density of states of the upper most energy level spreads and shifts in energy as well as reduces in magnitude. All of these combined result in a shift in the peak of the capture rate to a higher energy, a broadening in energy, as well as a reduction in magnitude for the energies around the peak.

As previously mentioned, it has been demonstrated that the capture rate oscillates with increasing quantum well width when the electric field equals zero. This is still the case for nonzero electric fields. This is demonstrated in Figure 33 where the average capture time is plotted as a function

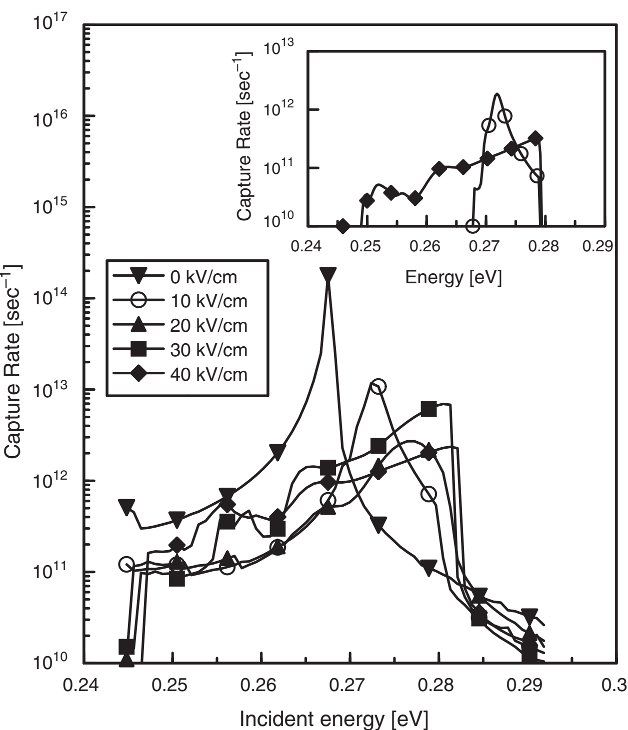


Figure 31. Capture rate as a function of incident energy for a band offset of 0.2434 eV.

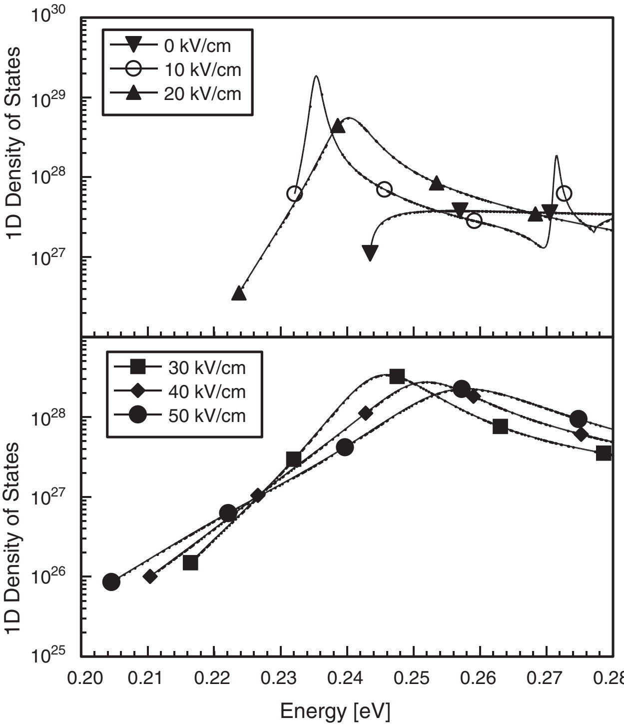


Figure 32. One-dimensional density of states as a function of electric field and energy for a band offset of 0.2434 eV.

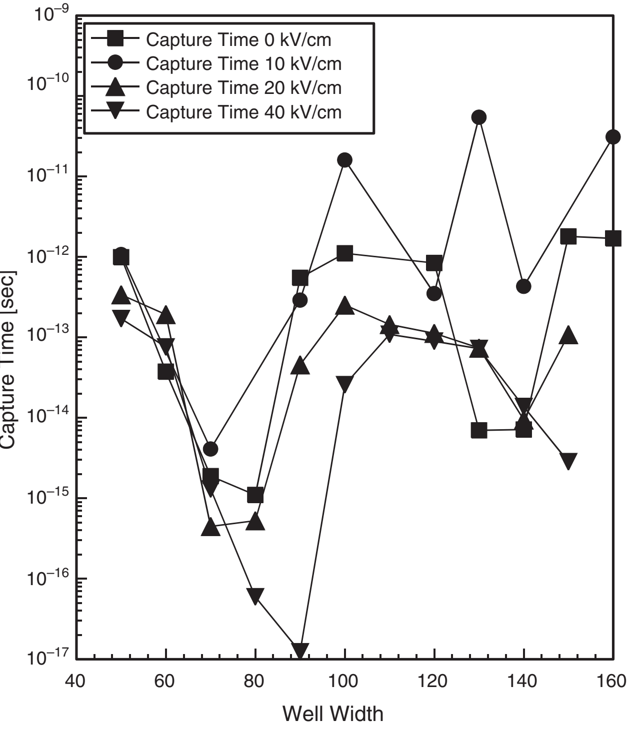


Figure 33. Average capture time as a function of quantum well width and electric field.

of quantum well width with the applied electric field as a parameter. As demonstrated by Figure 33, with increasing quantum well width and increasing electric field, the oscillatory nature dampens out, resulting from the redistribution of the density of states.

5. DENSITY OF STATES INDUCED ABSORPTION LINE BROADENING OF A BIASED QUANTUM WELL

The transition rate for an electron at initial states Ψ_i and final state Ψ_f is [46]

$$W_{if} = \frac{2\pi}{\hbar} |\langle \Psi_f | H_{\text{int}} | \Psi_i \rangle|^2 (1 - f(E_f)) f(E_i) \times \delta(E_f - E_i - \hbar\omega) \tag{39}$$

where E_i (E_f) is the initial (final) energy of the electron, H_{int} is the interaction Hamiltonian, $f(E)$ is the Fermi function, and $\hbar\omega$ is the energy of the photon. The interaction Hamiltonian is calculated by applying perturbation theory to the interaction between electrons and electromagnetic radiation [46]:

$$H_{\text{int}} = \frac{i\hbar e}{mc} \bar{A} \cdot \bar{\nabla} \tag{40}$$

Assuming z-polarized light, Eq. (40) can be written as

$$H_{\text{int}} = \frac{i\hbar e A_0}{2mc} \frac{d}{dz} \tag{41}$$

where e is the elemental charge, m is the free electron mass, c is the speed of light, $\hbar$ is the reduced Plank's constant, $i = \sqrt{-1}$, and A_0 is the vector potential. This reduces the matrix elements to

$$\langle \Psi_f | H_{\text{int}} | \Psi_i \rangle = \frac{1}{V} \frac{i\hbar A_0}{2mc} \delta_{\bar{k}_{ti}, \bar{k}_{tf}} \int_0^{L_w} \varphi_{zf}^* \frac{d}{dz} \varphi_{zi} \, dz \tag{42}$$

where the initial (i) and final (f) state wavefunctions have been separated into a z-component (φ_z) and a transverse component (φ_t) which results in a conservation of momentum in the transverse direction, $\delta_{\bar{k}_{ti}, \bar{k}_{tf}}$ V is the crystal volume and L_w is the length of the quantum well.

Absorption coefficient, α_{abs}, is defined as $\hbar\omega$ times the number of transitions per unit volume per unit time and divided by the incident power,

$$\alpha_{\text{abs}} = \frac{1}{V} \sum_{\bar{k}_f} \sum_{\bar{k}_i} \frac{2\mu c}{n_r \omega^2 A_0^2} \hbar\omega W_{if} \tag{43}$$

where n_r is the refractive index and μ is the permeability.

Combining Eqs. (39)–(42), using the conservation of momentum and energy and converting the final state summation into an integral, Eq. (42) can be written as

$$\alpha_{\text{abs}}(\hbar\omega) = \frac{e^2 c\mu\hbar^2 |\sin\,\theta|^2}{\omega\, m^2 n_r} \sum_{k_{zi}} \int k_{ti} \langle g_{1\text{d}}(E_{zi} + \hbar\omega)\rangle |M_z|^2 \times (1 - f(E_i + \hbar\omega)) f(E_i)\, dE_{ti} \tag{44}$$

where θ accounts for the incident angle of the light, M_z are the matrix elements, and $\langle g_{1\text{d}}(E_{zi} + \hbar\omega)\rangle$ is the effective one-dimensional density of states.

Converting the k_{zf} summations into an integral over the final energy, E_{zf}, introduces the effective 1D density of states, $\langle g_{1\text{d}}(E_{zf})\rangle$, into the absorption rate. The effective (1D) density of states is obtained by the method of Section 2 where $V(z)$ not only includes the conduction band potential energy and any externally applied fields but also incorporates the effects of the transverse momentum upon the z-directed wavefunction [70].

The structure investigated is the one used by Levine [2] and consists of two 40 Å GaAs quantum wells with 500 Å AlGaAs barriers. The composition of Al is 30% yielding a band offset of 0.2431 eV. The effective masses for electrons in the barriers and quantum wells are $0.092m_0$ and $0.067m_0$, respectively, where m_0 is the free electron mass. One bound energy level exists within the quantum well at 82 meV for a 25 kV/cm electric field while a virtual bound state exists at 260 meV.

In Figure 34, the calculation of the responsivity [$R \sim (1 - \exp(-2\alpha l))$] [1] of the two quantum well system is compared to the experimental two quantum well [1] system and shows excellent agreement. The responsivity of the individual wells is also shown in Figure 34 where the peaks in the responsivity of the individual wells occur at different wavelengths and are discussed later. The availability of states due to the redistribution of the one-dimensional density of states (see Fig. 35) around the band offset but above the bound energy level accounts for the excellent fit above 6 μm. For wavelengths below 6 μm, the theoretical calculations show

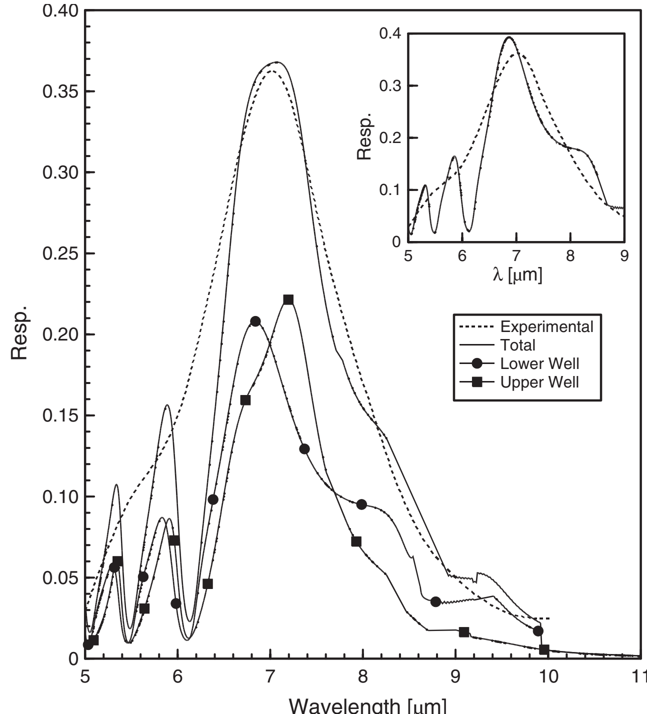


Figure 34. The calculated (solid line) and the experimental (dashed line) responsivity for the two quantum well QWIP. The inset is the responsivity of only the lower quantum well as a function of wavelength.

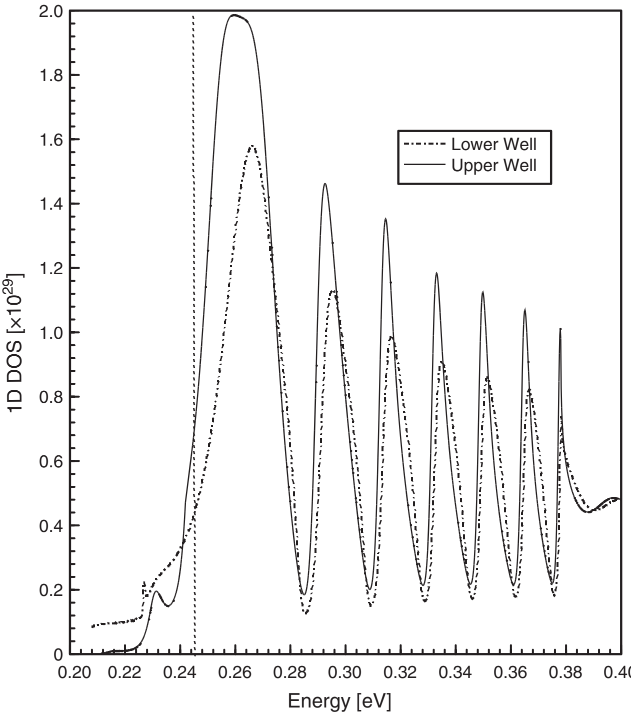


Figure 35. The one-dimensional density of states for the upper (solid) and lower (dashed–dotted) quantum wells. The vertical line is the conduction band offset.

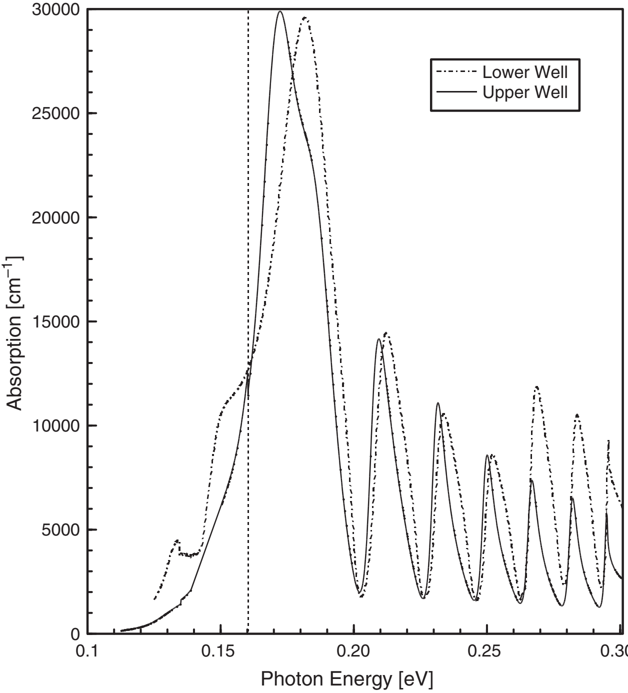


Figure 36. The absorption coefficient of the upper (solid) and lower (dashed–dotted) quantum well as a function of photon energy.

oscillatory behavior and is attributed to the nature of the one-dimensional density of states due to the virtual bound states above the band offset. This oscillatory behavior of the one-dimensional density of states, as shown in Figure 35, results from simulating an abrupt heterojunction. However, upon fabrication, the heterojunction loses its abruptness resulting in a slightly graded junction which would suppress the oscillatory nature of the 1D density of states. Furthermore, the incorporation of other scattering mechanisms (i.e., phonons, interface roughness) will also smooth out the oscillation wavelengths below 6 μm.

Figure 35 also shows the density of states of the two wells separately. As is evident, the presence of one quantum well, though separated by a large barrier, redistributes the other's density of states. This behavior affects the absorption characteristics and the responsivity as shown in Figures 35 and 36. In Figure 36, the absorption coefficient of the combined two quantum well system is shown along with the absorption coefficient of the individual quantum wells. In comparing Figures 35 and 36, it is seen that the absorption coefficient is a strong function of the density of states. Moreover, the existence of states below the band offset (vertical line in Fig. 35) but above the bound energy level results in the transition of an electron to these states. This is demonstrated in Figure 36, where there exists an absorption coefficient for states which lie below the band offset (the vertical line in Fig. 36 represents the energy difference between the energy level and the band offset). It is worth noting that if one claims that the densities of states are identical for both wells, the peak in the responsivity curve would be shifted and the broadening would be reduced as shown in the inset of Figure 34 where the responsivity is plotted using only the absorption coefficient of the lower well.

Furthermore, if the effects upon the wavefunction in the z-direction from changes in k_t are neglected, the broadening would be less and the peak in the responsivity would shift. This is demonstrated in Figure 37 where the responsivity is plotted as a function of wavelength when the effect of k_t upon the wavefunction is not accounted for. However, as shown in Figure 37, the broadening has not been drastically reduced, implying that the broadening in the absorption spectrum and the responsivity mainly results from the redistribution of the density of states and not due to scattering mechanisms or the effects of k_t upon the wavefunction.

6. DARK CURRENTS IN MULTIPLE QUANTUM WELL INFRARED PHOTODETECTORS

The dark current arising from the n^+ regions consists of a positive z-directed and a negative z-directed current where z is perpendicular to the quantum wells. The total current is the summation of the two components where each component is calculated by integrating over all k-space the local velocity ($v_g(E, z)$), the transmission coefficient ($D(E_z)$), and the Fermi function ($f(E)$) at the edge of the n^+ regions as [46]

$$J_{\text{tot}} = J^+ + J^- = \frac{2q}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dk_x \, dk_y \times \left[\frac{1}{2\pi} \int f(E) v_g^{\pm}(E_z) D^{\pm}(E_z) \, dk_z \right] \tag{45}$$

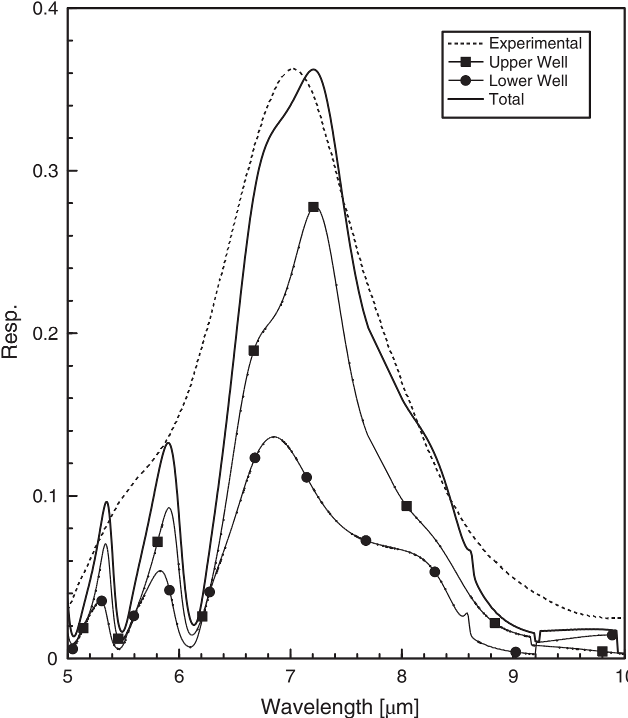


Figure 37. The responsivity as a function of wavelength for the two quantum well QWIP when the k_z dependence on k_t is not accounted for.

where q is the elemental charge and k is the wavevector in the x, y, and z direction.

Since the local velocity and the transmission coefficient are a function of the z-directed energy, the k_z integral is converted to an integral over energy (as well as the k_x and k_y integrals), E_z, as

$$J_{\text{tot}} = \frac{qm^*}{2\pi\hbar^2} \times \left[\int_0^\infty dE_t \int_{\substack{0 \\ kz>0}}^\infty f(E) v_g^+(E_z) D^+(E_z) g_{1\text{D}}(E_z) dE_z + \int_0^\infty dE_t \int_{\substack{0 \\ kz<0}}^\infty f(E) v_g^-(E_z) D^-(E_z) g_{1D}(E_z) dE_z \right] \tag{46}$$

where $g_{1\text{D}}(E_z)$ is the 1D density of states at the edge of the n^+ regions and E_t is the energy in the transverse direction (x and y).

The calculation of the 1D density of states proceeds as in Section 2. The calculation of the transmission coefficient [5–8] is obtained from the logarithmic derivative of the wavefunction by subtracting the square of the reflection coefficient at the edge of the n^+ region as

$$D^\pm(z_i, E_z) = 1 - \left| \frac{G^\pm(z_i, E_z) \mp G_{oi}}{G_{oi} \mp G^\pm(z_i, E_z)} \right|^2 \tag{47}$$

The QWIPs investigated are the same as those reported by Liu [13] with the system parameters summarized in Table 3 and an energy diagram shown in Figure 38.

Table 3. System parameters for QWIP.

Number of wells	Aluminum composition	Length of well	Length of barrier
4	26%	60 Å	254 Å
8	28%	60 Å	254 Å

In Figure 39, experimental dark current [13] is plotted as a function of applied bias for QWIPs with four and eight quantum wells. On the same graph, theoretical plots are also shown. It is observed that dark current is independent of the number of quantum wells contrary to the common belief. Transport in any structure is controlled by (a) the transmission coefficient, (b) the local velocity, and (c) density of states. In Figure 40, where the transmission coefficient, local velocity, and 1D density of states are plotted as a function of electron energy, for both four and eight quantum well systems, for a band offset of 0.216 eV. The plot suggests that the current is independent of the number of quantum wells as the parameters themselves do not change appreciably. As shown in the figure, there exist minibands of high transmission above the band offset separated by minibands of low transmission regions. The only difference in the transmission coefficient between the four well system and the eight well system is that the transmission coefficient between the minibands is smaller for the eight well system than the four well system. However, since the transmission coefficient of valleys for both systems is several orders of magnitude less than unity, there is a negligible difference between the four and eight well system. Moreover, a comparison of the local velocity and the density of states between the two systems also does not demonstrate any significant change.

However, a slight change in the growth parameters, namely, the band offset or the doping, makes a noticeable difference in the magnitude of dark current as is shown in Figure 39. It was pointed out in the table provided by Liu [13] that the Al concentration in the barrier region of the four and eight quantum well systems are 26% and 27%, respectively. The dark current for the eight well system (dashed–dotted line) is smaller than the four well system (dashed line). This results from the conduction band offset being larger for the eight well system, resulting in a smaller number of electrons participating in the current even though the peaks of the velocity and transmission coefficient are broader than the four well system, and is shown in Figure 41.

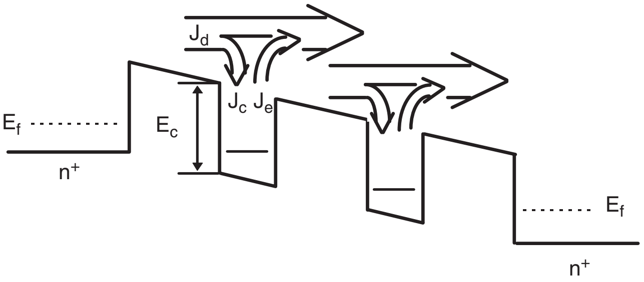


Figure 38. Energy diagram of a biased multiple quantum well n^+–i–n^+ systems demonstrating the origin of the dark currents.

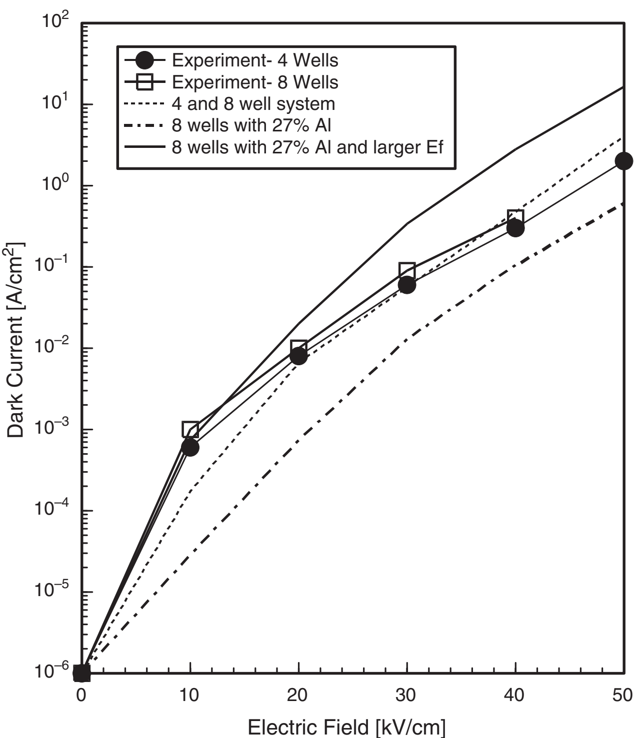


Figure 39. Comparison between the calculated and experimental dark current.

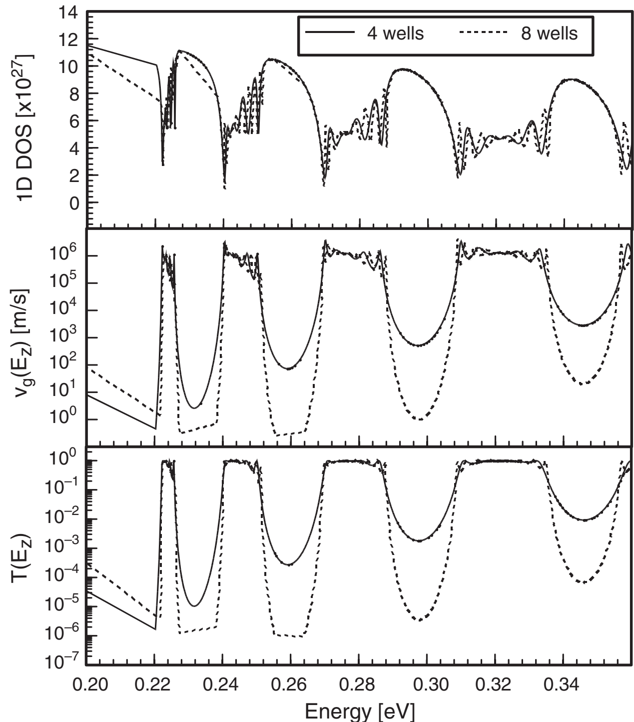


Figure 40. The velocity, density of states, and transmission coefficient for a four and eight quantum well system for Al compositions of 26%.

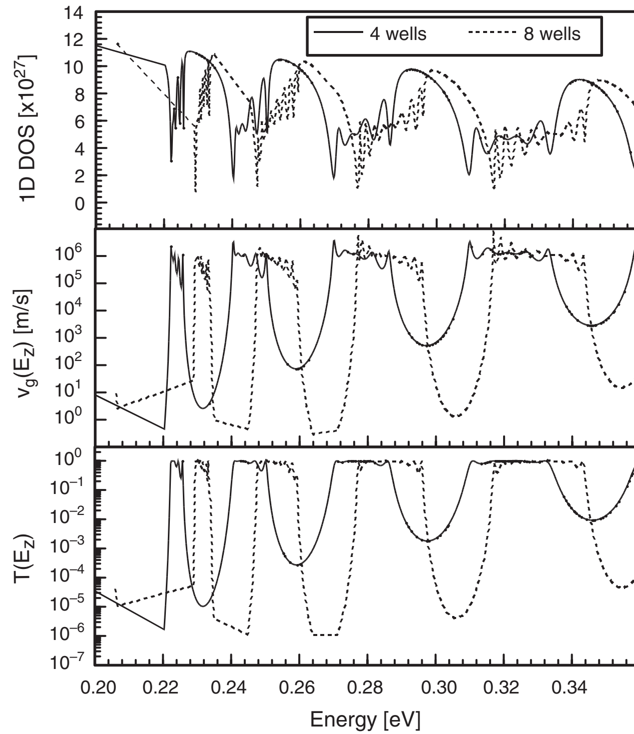


Figure 41. The velocity, density of states, and transmission coefficient for a four and eight quantum well system for Al compositions of 26% and 27%, respectively.

The experimental determination of the dark currents for the four and eight well system shows that the eight well system has a higher dark current. This can be attributed to the 10% uncertainty in the n^+ doping level as pointed out by Liu [13]. Incorporating the variation in n^+ doping level, the dark currents for the eight well system can become larger than the four well system. This is also demonstrated in Figure 39, where an increase in the dark current for the eight quantum well system (solid line) system is obtained by adjusting the Fermi level to 0.03 eV above the four well Fermi level. The increase in the dark current of the eight well system reported by Liu [13] may therefore be attributed to the difference in the donor doping level and not to the number of quantum wells.

GLOSSARY

Metalorganic vapor phase epitaxy (MOVPE) The reactor where the layered structures using different materials are grown. MOVPE growth is preferred by industry over MBE (molecular beam epitaxy) or LPE (liquid phase epitaxy) as it provides high throughput and lower cost per device while maintaining the quality especially for large wafers.

Quantum well (QW) A quantum well is like a well where a particle can be trapped. Layers made by using different semiconductor materials may result in the trapping of carriers (electrons, holes) in a particular layer. The trapped carriers and confined and are restricted to make certain movements or move in certain directions. The presence of QWs has resulted in modern electronic devices such as high

electron mobility transistors (HEMTs), heterojunction bipolar transistors (HBTs) and photonic devices such as QWIPs.

Quantum well infrared photodetector (QWIP) QWIPs operate based upon the bound to quasi-bound transition in a quantum well. It has been demonstrated in focal panel arrays in high resolution infrared cameras.

Self electro-optic device (SEED) An optically bistable device used in optical switching. An external voltage is applied to shift the wavelength of the onset of absorption, thereby, controlling the intensity of transmitted light.

REFERENCES

1. E. Rosencher, B. Vinter, F. Luc, L. Thibaudeau, P. Bois, and J. Nagle, *IEEE J. Quantum Elec.* 30, 2875 (1994).
2. B. F. Levine, *J. Appl. Phys.* 74, R1 (1993).
3. A. M. Fox, D. A. B. Miller, G. Livescu, J. E. Cunningham, and W. Y. Jan, *IEEE J. Quantum Electron.* 27, 2281 (1991).
4. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, *Phys. Rev. B* 32, 1043 (1985).
5. A. F. M. Anwar and M. M. Jahan, *Phys. Rev. B* 50, 10864 (1994).
6. A. F. M. Anwar and M. M. Jahan, *Phys. Rev. B* 49, 17440 (1994).
7. K. R. Lefebvre and A. F. M. Anwar, *Semi. Sci. Tech.* 12, 1226 (1997).
8. K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 33, 187 (1997).
9. K. R. Lefebvre and A. F. M. Anwar, *J. Appl. Phys.* 80, 3595 (1996).
10. K. R. Lefebvre and A. F. M. Anwar, *IEEE J. Quantum Electron.* 35, 216 (1999).
11. B. F. Levine, C. G. Bethea, G. Hasnain, V. O. Shen, R. R. Abbot, and S. J. Hsieh, *Appl. Phys. Lett.* 56, 851 (1990).
12. F. F. Sizov and A. Rogalski, *Progr. Quantum Electr.* 17, 94 (1993).
13. H. C. Liu, The basic physics of photoconductive quantum well infrared detectors, "Long Wavelength Infrared Detectors," Ch. 1. Gordon and Breach, New York, 1996.
14. P. J. Stevens, M. Whitehead, G. Parry, and K. Woodbridge, *IEEE J. Quantum Electron.* 24, 2007 (1988).
15. Y. Fu and M. Willander, *J. Appl. Phys.* 77, 4648 (1995).
16. M. O. Manasreh, F. Szmulowicz, T. Vaughan, K. R. Evans, C. E. Stutz, and D. W. Fischer, Intersubband infrared absorption in a GaAs/AlGaAs multiple quantum well, in "Intersubband Transitions in Quantum Wells," p. 287. Plenum, New York, 1992.
17. H. C. Liu, *J. Appl. Phys.* 73, 3062 (1993).
18. E. Rosencher, B. Vinter, F. Luc, L. Thibaudeau, P. Bois, and J. Nagle, *IEEE J. Quantum Electron.* 30, 2875 (1994).
19. D. J. Newson and A. Kurobe, *Semi. Sci. Tech.* 3, 786 (1988).
20. H. S. Li and K. L. Wang, *J. Appl. Phys.* 71, 1383 (1992).
21. Z. Ikonic, V. Milanovic, D. Tjapkin, and S. Pajevic, *Phys. Rev. B.* 37, 3097 (1988).
22. W. A. Beck, *Appl. Phys. Lett.* 63, 3589 (1993).
23. K. Blum, "Density Matrix Theory and Applications" Plenum, New York, 1937.
24. W. R. Frensley, *Rev. Modern Phys.* 62 (1990).
25. C. Weisbuch and B. Vinter, "Quantum Semiconductor Structures" Academic Press, San Diego, 1991.
26. G. W. Taylor, *Progr. Quantum Electron*. 16, 73 (1992).
27. K. W. Barnham, J. M. Barnes, B. Braun, J. P. Connolly, G. Haarpainter, J. A. Nelson, M. Paxman, C. Button, J. S. Roberts, and C. T. Foxon, A novel approach to higher efficiency—The quantum well solar cell, in "11th E.C. Photovoltaic Solar Cell Conference," 1992, p. 146
28. A. L. Lentine and D. A. B. Miller, *IEEE J. Quantum Electron.* 29, 655 (1993).
29. H. C. Liu, A. G. Steele, and M. Buchanan, *J. Appl. Phys.* 73, 2029 (1993).
30. M. Toivonen, M. Jalonen, M. Pessa, K. R. Lefebvre, and N. Anderson, *Mater. Sci. Eng. B* 21, 237 (1993).
31. T. Sizer, T. K. Woodward, U. Keller, T.-H. Chiu, D. L. Sivco, and A. Y. Cho, *IEEE J. Quantum Electron.* 30, 399 (1994).
32. H. Schneider and K. von Klitzing, *Phys. Rev. B* 38, 6160 (1988).
33. M. P. Houng, Y. H. Wang, and C. H. Chu, *J. Appl. Phys.* 77, 6338 (1995).
34. N. Susa and T. Nakahara, *Solid-State Electron.* 36, 1277 (1993).
35. Y. L. Lam and J. Singh, *IEEE J. Quantum Electron.* 31, 923 (1995).
36. L. Liang and C. S. Lent, *J. Appl. Phys.* 68, 1741 (1990).
37. Cavailles, D. A. B. Miller, J. E. Cunningham, P. Li Kam Wa, and A. Miller, *IEEE J. Quantum Electron.* 28, 2486 (1992).
38. D. J. Moss, T. Ido, and H. Sano, *IEEE J. Quantum Electron.* 30, 1015 (1994).
39. J. Nelson, M. Paxman, K. W. J. Barnham, J. S. Roberts, and C. Button, *IEEE J. Quantum Electron.* 29, 1460 (1993).
40. J. M. Ziman, "Principles of the Theory of Solids." Cambridge Univ. Press, Cambridge, UK, 1964.
41. R. Ferreira and G. Bastard, *Physi. Rev. B* 40 1074 (1989).
42. J. Zhu, W. Duan, and B. Gu, *Phys. Rev. B* 50, 5473 (1994).
43. A. Hernandez-Cabrera, P. Aceituno, and H. Cruz, *J. Appl. Phys.* 78, 6147 (1995).
44. Feldmann, K. Goossen, D. Miller, A. Fox, J. Cunningham, and W. Jan, *Appl. Phys. Lett.* 59, 66 (1991).
45. Shorthose, J. Ryan, and A. Moseley, *Solid State Elec.* 32, 1449 (1989).
46. K. R. Lefebvre, Electron Dynamics in a Biased Quantum Well, Ph.D. Dissertation, University of Connecticut, Storrs, CT, 1997.
47. H. C. Liu, *Appl. Phys. Lett.* 60, 1507 (1992).
48. K. K. Choi, *Appl. Phys. Lett.* 65, 1266 (1994).
49. S. C. Kan, D. Vassilovski, T. C. Wu, and K. Y. Lau, *IEEE Photonic Tech. Lett.* 4, 428 (1992).
50. M. Abou-Khalil, M. Goano, A. Champagne, and R. Maciejko, *IEEE Photon. Tech. Lett.* 8, 19 (1996).
51. B. Deveaud, J. Shah, T. C. Damen, and W. T. Tsang, *Appl. Phys. Lett.* 52, 1886 (1988).
52. D. Y. Oberli, J. Shah, J. L. Jewell, and T. C. Damen, *Appl. Phys. Lett.* 54, 1028 (1989).
53. L. Davis, Y. L. Lam, Y. C. Chen, J. Singh, and P. K. Bhatttacharya, *IEEE J. Quantum Electron.* 30, 2560 (1994).
54. H. J. Polland, K. Leo, K. Rother, K. Ploog, J. Feldmann, G. Pters, E. O. Gobel, K. Fujiwara, T. Nakayama, and Y. Ohta, *Phys. Rev. B* 38, 7635 (1988).
55. K. Kalna, M. Mosko, and F. M. Peeters, *Appl. Phys. Lett.* 68, 117 (1996).
56. C. Y. Tsai, L. F. Eastman, Y. H. Lo, and C. Y. Tsai, *IEEE Phot. Tech. Lett.* 6, 1088 (1994).
57. B. Deveaud, A. Chomette, D. Morris, and A. Regreny, *Solid State Comm.* 85, 367 (1993).
58. M. R. X. Barros, P. C. Becker, D. Morris, B. Deveaud, A. Regreny, and F. Beisser, *Phys. Rev. B* 47, 10951 (1993).
59. B. Deveaud, D. Morris, A. Regreny, R. Planel, J. M. Gerard, M. R. X. Barros, and P. Becker, *Semi. Sci. Tech.* 722 (1994).
60. D. Morris, B. Deveaud, A. Regreny, and P. Auvary, *Phys. Rev. B* 47, 6819 (1993).
61. C. Y. Tsai, C. Y. Tsai, Y. H. Lo, R. M. Spencer, and L. F. Eastman, *IEEE J. Select. Topics Quantum Electron.* 1 (1995).
62. K. Kalna and M. Mosko, *Phys. Rev. B* 54, 17730 (1996).
63. J. A. Brum, T. Weil, J. Nagle, and B. Vinter, *Phys. Rev. B* 34, 2381 (1986).
64. M. Preisel and J. Mork, *J. Appl. Phys.* 76, 1691 (1994).
65. D. Bradt, Y. Sirenko, and V. Mitin, *Semicond. Sci. Technol.* 10, 260 (1995).
66. J. A. Brum and G. Bastard, *Phys. Rev. B* 33, 1420 (1986).

67. B. K. Ridley, *SPIE Quantum Well Superlattice Phys. II* 943, 84 (1988).
68. P. W. M. Blom, C. Smit, J. E. M. Haverkort, and J. H. Wolter, *Phys. Rev. B* 47, 2072 (1993).
69. P. Sotirelis and K. Hess, *Phys. Rev. B* 49, 7543 (1994).
70. B. Vinter, F. Luc, L. Thibaudeau, and E. Rosencher, *Solid State Electron.* 37, 773 (1994).
71. F. F. Sizov and A. Rogalski, *Progr. Quantum Electr.* 17, 94 (1993).
72. W. Trzecialowski and M. Gurioli, *Phys. Rev. B* 44, 3880 (1991).
73. T. Bahder, J. Bruno, R. G. Hay, and C. Morrison, *Phys. Rev. B* 37, 6256 (1988).
74. Iannaccone and B. Pellegrini, *Phys. Rev. B* 53, 2020 (1996).
75. G. D. Sanders and K. K. Bajaj, *Phys. Rev. B* 35, 2308 (1987).

www.aspbs.com/enn

Raman Scattering in Nanostructures

C. E. Bottani, C. Castiglioni, G. Zerbi

Center of Excellence for the Engineering of Nanostructured Materials and Surfaces, Politecnico di Milano, Milano, Italy

CONTENTS

1. Introduction
2. Classes of Nanostructures
3. General Considerations on Vibrational Dynamics and Raman Scattering of Nanostructures
4. Confinement Effects
5. Specific Nanostructures
6. Brillouin Scattering
References

1. INTRODUCTION

1.1. Semiclassical Theory of the Raman Scattering from a Molecule

In this section a simplified description of the Raman effect is presented, with the aim of providing readers who have no experience in the field with the basic physical concepts and equations, which may allow a discussion of some peculiar aspects of the Raman response of nanostructured materials or (more generally) of chemical compounds on the order of nanometers. For a comprehensive treatment of the Raman effect the reader can refer to books and reviews listed in the References [1–3].

Several concepts needed to understand the phenomenon of inelastic scattering of light observed when a material is irradiated with a monochromatic beam can be rather easily introduced starting from a classical approach. According to this scheme, we describe scattered light in terms of the fluctuating polarization field originating in the material and radiating the Raman light. In this section the material is described as the simplest case of an isolated molecule.

The intensity I (or time-averaged power per unit solid angle) radiated by an oscillating electric dipole $\mathbf{p}$ induced in a molecule by the electric field $\mathbf{E}^I = E^I\mathbf{e}^I$ of the incident radiation beam of frequency ω^I, along a direction making an angle θ with the axis of the dipole is given by classical electrodynamics [4] as

$$I = \frac{(\omega^S)^4 p^2 \sin^2(\theta)}{32\pi^2\epsilon_0^2 c^3} = \frac{(\omega^S)^4 |\mathbf{e}^S \cdot \mathbf{p}|^2}{32\pi^2\epsilon_0^2 c^3} \tag{1}$$

where $\mathbf{e}^S$ is the polarization unit vector of the emitted radiation. Due to quantized rotovibrational motion, the molecule polarizability $\boldsymbol{\alpha}$ is modulated at some internal frequencies ω_k (see below) and thus the induced electric dipole oscillates at frequencies ω^S which can be different from ω^I.

The expression for the (time-dependent) induced dipole of a molecule can be written, in very general terms, as a Taylor series in the field $\mathbf{E}$

$$p_\varrho = \sum_\sigma \alpha_{\varrho\sigma} E_\sigma + \frac{1}{2}\sum_{\sigma\tau}\beta_{\rho\sigma\tau} E_\sigma E_\tau + \frac{1}{3!}\sum_{\sigma\tau\upsilon}\gamma_{\rho\sigma\tau\upsilon} E_\sigma E_\tau E_\upsilon + \cdots \tag{2}$$

where p_ϱ is the ρth Cartesian component of the induced dipole and $\alpha_{\rho\sigma}, \beta_{\rho\sigma\tau}, \gamma_{\rho\sigma\tau\upsilon}$ are the elements of the (second rank) polarizability tensor and of the first and second hyperpolarizability tensors, respectively.

Raman scattering originates by the first term of this expression, namely, from molecular polarizability fluctuations; while in an isolated molecule the fluctuations are purely quantum mechanical effects, in a system of molecules in thermal equilibrium one also has to take into account thermodynamic fluctuations [5]. In the following we consider the expression (2) approximated to the first term, linear in the field. The time dependence of $\mathbf{p}$ is determined both by the time dependence of the tensor involved and from the time dependence of the electric field $\mathbf{E}$ of radiation. Combining the above formulae we obtain

$$I = \frac{(\omega^S)^4 |\mathbf{e}^S \cdot \boldsymbol{\alpha} \cdot \mathbf{e}^I|^2}{32\pi^2\epsilon_0^2 c^3}(E^I)^2 \tag{3}$$

The polarizability tensor $\boldsymbol{\alpha}$ is indeed a function of the nuclear coordinates. Neglecting rotations, for a vibrating molecule this dependence can be expressed in terms of the

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

set of all vibrational normal coordinates of the molecule $\{Q_k\}$ according to the relationship

$$\alpha_{\rho\sigma} = (\alpha_{\rho\sigma})_0 + \sum_k \left(\frac{\partial \alpha_{\rho\sigma}}{\partial Q_k}\right)_0 Q_k + \frac{1}{2}\sum_{kl}\left(\frac{\partial^2 \alpha_{\rho\sigma}}{\partial Q_k \partial Q_l}\right)_0 Q_k Q_l + \cdots \quad (4)$$

where the index 0 indicates the values at equilibrium geometry of the molecule in its electronic ground state. In the electric harmonic approximation we can write

$$\alpha_{\rho\sigma} = (\alpha_{\rho\sigma})_0 + \sum_k \alpha'_k Q_k$$

neglecting the higher order anharmonic terms in (4). The quantity $\boldsymbol{\alpha}'_k = (\partial \alpha_{\rho\sigma}/\partial Q_k)_0$ is called the derived polarizability tensor. According to the expressions above we can explicitly describe the time dependence of the fluctuating induced dipole resulting from the kth normal mode of vibration of the molecule irradiated by a plane monochromatic electromagnetic wave of frequency ω^I

$$\mathbf{p}_k = \boldsymbol{\alpha}_0 \mathbf{E}^I \cos(\omega^I t) + \boldsymbol{\alpha}'_k \mathbf{E}^I Q_{k0} \cos(\omega_k t + \delta_k)\cos(\omega^I t) + \cdots \quad (5)$$

Simple algebraic manipulation of the above equation allows us to identify three different contributions to $\mathbf{p}_k$ oscillating at different frequencies:

$$\mathbf{p}_k = \mathbf{p}(\omega^I) + \mathbf{p}(\omega^I - \omega_k) + \mathbf{p}(\omega^I + \omega_k) \quad (6)$$

$$\mathbf{p}(\omega^I) = \boldsymbol{\alpha}_0 \mathbf{E}^I \cos(\omega^I t) \quad (7)$$

$$\mathbf{p}(\omega^I \pm \omega_k) = \boldsymbol{\alpha}_k^{\mathrm{Raman}} \mathbf{E}^I \cos[(\omega^I \pm \omega_k)t + \delta_k] \quad (8)$$

$$\boldsymbol{\alpha}_k^{\mathrm{Raman}} = Q_{k0}\boldsymbol{\alpha}'_k$$

The dipole moment in Eq. (7) describes the elastic scattering of light (taking place at the same frequency as that of the exciting beam) called Rayleigh scattering. The fluctuating dipole moments in Eq. (8) are associated with inelastic Raman scattering. These random dipoles radiate two different monochromatic components oscillating, respectively, at lower frequency $\omega^I - \omega_k$ (Stokes line) and at higher frequency $\omega^I + \omega_k$ (anti-Stokes line) with respect to the frequency of the incident beam. The emission of the Raman light is a random process as it depends on the normal coordinate $Q_k = Q_{k0}\cos(\omega_k t + \delta_k)$, a function of time and of the random variables Q_{k0} and δ_k, the amplitude and the initial phase of a quantum harmonic oscillator obeying Bose–Einstein statistics [5].

Considering the Raman scattering from a polyatomic molecule (N atoms with $3N$-6 vibrational normal modes), we would expect to observe (at least in principle) $3N$-6 lines at negative frequency shifts with respect to the Rayleigh line (Stokes Raman spectrum) and $3N$-6 lines at positive frequency shift (anti-Stokes Raman spectrum). Usually the Raman spectrum is recorded in the Stokes region (see below) and Raman intensities are plotted versus Raman shifts, measured as wave numbers (units of cm^{-1}). In fact the number of Raman transitions which can be observed can be lower than that predicted by dynamics: indeed if the molecule has some symmetry selection rules may be active.

The simple derivation illustrated above holds in the hypothesis of strict electrical and mechanical harmonicity: the effect of anharmonicity is the appearance of lines at frequency shifts corresponding to near multiples of the fundamental vibrational frequencies (overtones) and at frequencies which are the sum or difference of fundamental vibrational frequencies (combination lines). Raman transitions corresponding to overtones and combination bands are usually referred to as second order Raman spectra.

According to Eqs. (8) and (3) we can state that the intensity of a Raman line (and thus the possibility of observing Raman scattering at a given frequency shift) is determined by the (i) vibrational dynamics of the molecules (i.e., normal vibrations and their frequencies: Q_k, ω_k); (ii) values of the Raman tensors $\boldsymbol{\alpha}'_k$; and (iii) orientation of the molecule with respect to the polarization directions of both the incident and scattered light beams.

This last point is important both in the discussion of the symmetry of the normal vibration responsible for Raman scattering and in the determination of the orientation of the molecule with respect to a space fixed reference system.

Analysis of (ii) can be further developed according to the following logical steps.

(a) Determination of the symmetry selection rules is done. The analysis of the molecular symmetry allows us to select (among the $3N$-6 molecular normal modes) those Q_k modes with $\boldsymbol{\alpha}'_k \neq 0$. This can be done without any specific knowledge of any numeric value of $\boldsymbol{\alpha}'_k$. For fundamental Raman transitions the selection rules can be easily stated: Q_k is Raman active if at least one element (uv) of the polarizability tensor fullfils the requirement $\Gamma_1 \subset \Gamma(Q_k) \otimes \Gamma(\alpha_{uv})$, where Γ_1 is the totally symmetric representation.

(b) The intrinsic intensity of a Raman allowed transition at ω_k (large or small Raman cross section) is determined by the electrical properties of the molecule considered (molecular polarizability and the extent of its change during the normal mode motion). Based on experimental findings it is usually accepted that Raman transitions associated with totally symmetric normal modes are stronger than those associated with vibrations belonging to non totally symmetric species.

(c) According to the simple classical theory illustrated in this section it is impossible to discuss one of the most important issues of Raman scattering. Raman intensities depend on the dispersion behavior of the Raman tensor while varying the frequency of the exciting electromagnetic wave (incident photon energy). Factors which determine this behavior are intimately related to the electronic structure of the molecule (material) and can be explicitly considered only in the frame of a quantum theory of Raman scattering [6–8]. The study of the evolution of the Raman spectrum with the exciting photon energy is of great importance, since it can give specific insight into the electronic properties of the material.

(d) A special case related to what is illustrated in (c) is the occurrence of resonance enhancement of a Raman line [6–9], when a suitable exciting energy is used matching the energy gap between energy levels

(usually the gap between the ground state and one selected electronic excited state) of the material. Resonance effects are responsible for a huge enhancement of Raman cross sections (which is often very selective for specific normal vibrations).

1.2. Raman Scattering in Extended Media

Many nanostructures can be thought as the assembly of elementary molecular units (conceptual bottom-up approach) and the scheme of the preceding section can be naturally extended. Yet, in some cases, a conceptual top-down approach is more convenient and the nanostructure is seen as a limited portion of an otherwise perfect infinite crystal. To this purpose we now generalize the classical theory of Raman scattering from single molecules in such a way that it can be applied to extended media and also to Brillouin scattering (inelastic light scattering from acoustic excitations). For the sake of simplicity we restrict ourselves to the case of optically isotropic solids (e.g., amorphous bodies or cubic crystals). Both Raman and Brillouin scattering can be treated introducing the instantaneous anisotropic dielectric susceptibility of the medium around the frequency ω^I, the tensor

$$\chi_{ij} = \chi\delta_{ij} + \delta\chi_{ij} \tag{9}$$

χ is the usual time-independent isotropic susceptibility, while the tensor $\delta\chi_{ij}$ is the anisotropic fluctuating part of the susceptibility due to the presence of phonon excitations modulating the optic response of the body. The corresponding fluctuating part of the dielectric constant is simply

$$\delta\epsilon_{ij} = \epsilon_0\delta\chi_{ij} \tag{10}$$

while the usual isotropic dielectric constant is $\epsilon = \epsilon_0(1+\chi)$. When the body is irradiated, there is a fluctuating part of the polarization vector radiating the scattered electromagnetic waves. This fluctuating continuous volume density of electric dipoles can be written as $\delta P_i = \delta\epsilon_{ij}E_j$ (in the following, the summation convention over repeated indexes is understood). Here

$$E_j = E_j^I + E_j^S \tag{11}$$

is the sum of the incident field and of the scattered field. Using the linearity of Maxwell equations the scattered electric field $\mathbf{E}^S$ radiated by the source $\partial\delta P_i/\partial t$ can be computed by means of first order perturbation theory (Born approximation), that is, assuming

$$\delta P_i \approx \delta\epsilon_{ij}E_j^I = \epsilon_0\delta\chi_{ij}E_j^I \tag{12}$$

The above equation is justified by the smallness of both the scattered field and $\delta\epsilon_{ij}$: in this way the excess polarization is seen as driven only by the incident electric field $\mathbf{E}^I$. Following the same scheme used above for molecules, it turns out that δP_i oscillates at the frequencies $\omega^S = \omega^I \pm \omega_\alpha$ as the time dependence of $\delta\epsilon_{ij}$ is $e^{-i\omega_\alpha t}$, where ω_α is one of the vibrational eigenfrequencies of the whole body, that is,

$$\delta\epsilon_{ij} = \epsilon_0\delta\chi_{ij}(\mathbf{r}|\boldsymbol{\alpha})e^{-i\omega_\alpha t} \tag{13}$$

$\delta\chi_{ij}(\mathbf{r}|\boldsymbol{\alpha})$ represents the space-dependent part of $\delta\chi_{ij}$ when only one normal mode of frequency ω_α is active. Solving the radiation problem in the far field approximation, as it is appropriate for most Raman and Brillouin scattering experiments, we find the fundamental formula [11]

$$\mathbf{E}^S(\mathbf{r},t) \approx \Re\left\{-\frac{E_0 e^{i(\mathbf{k}^S\cdot\mathbf{r}-\omega^S t)}}{4\pi\epsilon_r r}\mathbf{k}^S \times (\mathbf{k}^S \times \mathbf{G}_\alpha)\right\} \tag{14}$$

In the above equation the observation point $\mathbf{r}$ is at distance r from the scattering portion V of the irradiated crystal. The origin of the coordinate system is inside the scattering volume scanned by vectors $\mathbf{r}'$ and the direction of the vector $\mathbf{r}-\mathbf{r}'$, which is the same as that of the vector $\mathbf{k}^S$, has been confused with that of vector $\mathbf{r}$. The vector scattering integral $\mathbf{G}$ is given by the key expression

$$\mathbf{G}_\alpha = \int_V (\delta\boldsymbol{\chi}(\mathbf{r}'|\boldsymbol{\alpha}) \bullet \mathbf{e}^I)e^{-i\mathbf{Q}\cdot\mathbf{r}'}\,d\mathbf{r}' \tag{15}$$

where $\bullet$ denotes the product of a second rank tensor by a column vector generating again a column vector. The transferred wave vector $\mathbf{Q}$ is defined as

$$\mathbf{Q} = \mathbf{k}^S - \mathbf{k}^I \tag{16}$$

where $\mathbf{k}^S$ and $\mathbf{k}^I$ are the wave vectors of the incident and of the scattered photons, respectively. Equations (14) and (15) express the scattered electric field as a function of the random variables $\delta\chi_{ij}$. At thermodynamic equilibrium the thermal average $\langle\mathbf{E}^S\rangle_{th} = 0$ because $\langle\delta\chi_{ij}\rangle_{th} = 0$, but the mean square $\langle|\mathbf{E}^S|^2\rangle_{th}$, which depends on the finite mean square value of the phonon normal coordinate buried in $\delta\chi_{ij}$, does not vanish.

The measured outcome of any inelastic light scattering experiment is the power spectrum $S_{E^S}(\omega)$ of the scattered light, which is computed as

$$S_{E^S}(\omega) = \int_{-\infty}^{\infty}\langle E^S(t+\tau)E^{S*}(t)\rangle_{th}e^{i\omega\tau}\,d\tau \tag{17}$$

where $E^S = \mathbf{E}^S\cdot\mathbf{e}^S$ is the complex amplitude of the scattered field projected along a given analyzed polarization direction perpendicular to $\mathbf{k}^S$.

Defining

$$\delta\boldsymbol{\chi}_{\mathbf{Q}}^{\alpha} = \int_V \delta\boldsymbol{\chi}(\mathbf{r}'|\alpha)e^{-i\mathbf{Q}\cdot\mathbf{r}'}\,d\mathbf{r}' \tag{18}$$

as the Fourier transform of index $\mathbf{Q}$ of the spatial part of the susceptibility tensor fluctuation and using some straightforward manipulations, the power spectrum of the scattered light can be written as

$$S_{E^S}(\mathbf{Q},\omega) = \frac{|E_0|^2(\omega^S)^4}{8\pi\epsilon_r^2c^4r^2}\left\langle|\mathbf{e}^S\cdot\delta\boldsymbol{\chi}_{\mathbf{Q}}^{\alpha}\bullet\mathbf{e}^I|^2\right\rangle_{th} \times \delta[\omega-(\omega^I+\omega_\alpha)] \tag{19}$$

for anti-Stokes scattering. For Stokes scattering one has just to replace $\omega^I+\omega_\alpha$ with $\omega^I-\omega_\alpha$ in the argument of the delta function. In all practical applications the delta function is substituted by a Lorentzian line shape centered at $\omega^I\pm\omega_\alpha$ with a FWHM related to the lifetime of the phonon

and to instrumental and opacity broadening. Introducing the Raman susceptibility $\delta(\boldsymbol{\chi}_{\mathbf{Q}}^{\alpha})/\partial\xi_\alpha$ with respect to the amplitude ξ_α of normal mode α, Eq. (19) can be rewritten as

$$S_{E^S}(\mathbf{Q},\omega) = \frac{|E_0|^2(\omega^S)^4}{8\pi\epsilon_r^2c^4r^2}\left|\mathbf{e}^S\cdot\frac{\partial(\boldsymbol{\chi}_{\mathbf{Q}}^{\alpha})}{\partial\xi_\alpha}\bullet\mathbf{e}^I\right|^2\langle\xi_\alpha^*\xi_\alpha\rangle_{th} \times \delta[\omega-(\omega^I+\omega_\alpha)] \qquad (20)$$

Standard second quantization arguments [12] show that

$$\langle\xi_\alpha^*\xi_\alpha\rangle_{th} = \frac{\hbar}{2\omega_\alpha}n \qquad (21)$$

Again the above formula can also be applied to Stokes scattering provided $\omega^I+\omega_\alpha$ is replaced with $\omega^I-\omega_\alpha$ and n with $n+1$, where

$$n = \left(e^{\frac{\hbar\omega_\alpha}{k_BT}}-1\right)^{-1} \qquad (22)$$

is the Bose–Einstein statistical factor. Because of the statistical weight the Raman Stokes intensity is always higher than the anti-Stokes one and often only the Stokes signal is recorded.

In the case of a crystal the normal modes of vibration are lattice waves (or Bloch waves [13] called optic and acoustic phonons). As a consequence of space periodicity (translational invariance of the crystal) the mode eigenfrequencies $\omega_\alpha(\mathbf{q})$ of a given phonon branch α are periodic functions of the phonon wave vector $\mathbf{q}$ in the extended zone scheme [13], that is, $\omega_\alpha(\mathbf{q}+\mathbf{g}_i)=\omega_\alpha(\mathbf{q})$ for any reciprocal lattice vector $\mathbf{g}_i$. It is important to stress that the phonon wave vector $\mathbf{q}$ is, strictly speaking, a good mode label only for ideal crystals. Yet in real cases $\mathbf{q}$ also keeps a physical meaning for systems exhibiting a sufficient degree of long-range translational order (see the confinement theory below). In the limit $\mathbf{q}\to\mathbf{0}$ $\omega_\alpha(\mathbf{q})$ of acoustic phonons goes linearly to 0, the group velocity $\partial\omega_\alpha(\mathbf{q})/\partial\mathbf{q}$ becoming identical to the corresponding sound speed $\mathbf{v}_\alpha$. For optic phonons instead $\omega_\alpha(\mathbf{q})$ goes to a finite value $\omega_\alpha(\mathbf{0})$ (typically in the range of 100–4000 cm^{-1}) with a vanishing group velocity. While scattering from crystal optic phonons is still called Raman scattering, as for single molecules, scattering from acoustic phonons is called Brillouin scattering (see below).

If the phonon-induced susceptibility fluctuations are due to lattice waves, the main result one gets is that one-phonon scattering occurs only for $\mathbf{q}=\mathbf{Q}$ and $\mathbf{Q}\approx\mathbf{0}$ due to the low photon momentum transfer, in the visible or in the infrared, with respect to zone boundary wave vectors. The argument goes as follows. In a crystal with a basis, the basis atoms of cell $\mathbf{n}$ in the lattice position $\mathbf{r}_\mathbf{n}$ form a molecule with polarizability $\boldsymbol{\alpha}_\mathbf{n}$. Thus the crystal susceptibility fluctuation due to all individual polarizability fluctuations $\delta\boldsymbol{\alpha}_\mathbf{n}^\alpha$ caused by a phonon ($\alpha\mathbf{q}$) can be written as

$$\delta\boldsymbol{\chi}(\mathbf{r}|\alpha) = \sum_\mathbf{n}\delta\boldsymbol{\alpha}_\mathbf{n}^\alpha\delta(\mathbf{r}-\mathbf{r}_\mathbf{n})$$

Computing the Fourier transform in Eq. (18) we obtain

$$\delta\boldsymbol{\chi}_\mathbf{Q}^\alpha = \sum_\mathbf{n}\delta\boldsymbol{\alpha}_\mathbf{n}^\alpha e^{-i\mathbf{Q}\cdot\mathbf{r}_\mathbf{n}}$$

Because $\delta\boldsymbol{\alpha}_\mathbf{n}^\alpha$ is a Bloch wave with wave vector $\mathbf{q}$

$$\delta\boldsymbol{\alpha}_\mathbf{n}^\alpha = \delta\boldsymbol{\alpha}_0^\alpha e^{i\mathbf{q}\cdot\mathbf{r}_\mathbf{n}}$$

we get

$$\delta\boldsymbol{\chi}_\mathbf{Q}^\alpha = \delta\boldsymbol{\alpha}_\mathbf{0}^\alpha\sum_\mathbf{n}e^{-i(\mathbf{Q}-\mathbf{q})\cdot\mathbf{r}_\mathbf{n}} = \frac{1}{V}\delta\boldsymbol{\alpha}_\mathbf{0}^\alpha\sum_i\delta(\mathbf{Q}-\mathbf{q}-\mathbf{g}_i)$$

Due to the extremely low photon momentum (with respect to zone boundary wave vectors), we can assume $\mathbf{Q}\approx\mathbf{0}$. Thus only the optic phonons with $\mathbf{q}\approx-\mathbf{g}_i$ do contribute to Raman scattering in a crystal. These phonons have frequencies $\omega_\alpha(\mathbf{q}\approx-\mathbf{g}_i)=\omega_\alpha(\mathbf{q}\approx\mathbf{g}_i)=\omega_\alpha(\mathbf{q}\approx\mathbf{0})$. As the group velocity of optic phonons at small $\mathbf{q}$ is negligible we can put $\omega_\alpha(\mathbf{q}\approx\mathbf{0})=\omega_\alpha(\mathbf{0})$. This marks an important difference with respect to coherent Brillouin scattering and incoherent Raman scattering. In fact, in the case of Brillouin scattering, $\mathbf{q}$ is small but cannot be put exactly equal to $\mathbf{0}$ in the linear dispersion relationship of acoustic phonons: this would correspond to a rigid translation of the lattice as a whole.

Introducing the normal mode expansion for the displacement $\mathbf{u}\binom{p}{\mathbf{0}}$ of atom p of the basis, composed of s atoms, of the cell in $\mathbf{0}$:

$$u\begin{pmatrix}p\\\mathbf{0}\end{pmatrix} = \frac{1}{\sqrt{Nm_p}}\sum_{\mathbf{q}\alpha}\xi_\alpha(\mathbf{q})\mathbf{e}(p|\mathbf{q}\alpha)e^{i\mathbf{q}\cdot\mathbf{r}_p}$$

we can express $\partial\boldsymbol{\alpha}_\mathbf{0}^\alpha/\partial\xi_\alpha(\mathbf{0})$ as

$$\frac{\partial\boldsymbol{\alpha}_\mathbf{0}^\alpha}{\partial\xi_\alpha(\mathbf{0})} = \frac{1}{\sqrt{Nm_p}}\sum_{p=1,s}\frac{\partial\boldsymbol{\alpha}_\mathbf{0}^\alpha}{\partial\mathbf{u}\binom{p}{\mathbf{0}}}\bullet\mathbf{e}(p|\mathbf{0}\alpha)$$

where the $\mathbf{e}(p|\mathbf{0}\alpha)$ are the polarization eigenvectors of the basis excited by phonon $(\mathbf{0}\alpha)$. Now the power spectrum of the anti-Stokes Raman signal can be obtained from Eq. (19) just summing up over all phonons $(\mathbf{0}\alpha)$ with eigenfrequencies $\omega_\alpha(\mathbf{0})$, obtaining the fundamental result: $S_{E^S}(\mathbf{0},\omega)$ is proportional to

$$\sum_\alpha\left|\mathbf{e}^S\cdot\left[\sum_{p=1,s}\frac{\partial\boldsymbol{\alpha}_\mathbf{0}^\alpha}{\partial\mathbf{u}\binom{p}{\mathbf{0}}}\bullet\mathbf{e}(p|\mathbf{0}\alpha)\right]\bullet\mathbf{e}^I\right|^2 \times\langle\xi_\alpha^*(\mathbf{0})\xi_\alpha(\mathbf{0})\rangle_{th}\delta[\omega-\omega^I-\omega_\alpha(\mathbf{0})]$$

The quantity

$$\left|\mathbf{e}^S\cdot\left[\sum_{p=1,s}\left(\partial\boldsymbol{\alpha}_\mathbf{0}^\alpha/\partial\mathbf{u}\begin{pmatrix}p\\\mathbf{0}\end{pmatrix}\right)\bullet\mathbf{e}(p|\mathbf{0}\alpha)\right]\bullet\mathbf{e}^I\right|^2$$

governs the first order Raman selection rules in crystals [13, 214] through their symmetry properties combining the point-group symmetry of the basis with the translations group of the lattice [14]. In particular, the Raman tensor of odd-parity phonons in centrosymmetric crystals must vanish. In this case only even-parity phonons with $\mathbf{q}=\mathbf{0}$ can be Raman active.

To treat disordered bodies and finite size crystalline bodies (isolated nanocrystals and nanocrystalline systems), it is worth (see below) further generalizing the above formalism, using the fact that the square modulus of the Fourier

transform of some function $f(\mathbf{r})$ equals the Fourier transform of the autocorrelation function of the same function $\langle f(\mathbf{r})f(\mathbf{r}')\rangle$. In this way the Raman spectrum of a generic extended medium can be written in terms of direct space–direct time dependence of fluctuations as [43]

$$S_{E^S}(\mathbf{Q},\omega) \propto \Im \int d\mathbf{r} \int d\mathbf{r}' \int dt \langle \delta\chi(\mathbf{r},t)\delta\chi(\mathbf{r}',0)\rangle_{th} e^{[-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')-i\omega t]} \quad (23)$$

where, for the sake of simplicity, we dropped the tensorial notation.

The function $S_{E^S}(\mathbf{Q},\omega,)$ is called the dynamic structure factor, its $\omega = 0$ value being the static structure factor, governing elastic light scattering. The experimental investigation of the dynamic structure factor can be performed with different techniques depending on the time scale (i.e., frequency range). For frequencies in the range of some gigahertz, which is typical of Brillouin spectroscopy, interferometric techniques (Fabry–Perot interferometers) are necessary [14]. Higher frequencies, typical of Raman spectroscopy, require the use of diffraction gratings (monochromators).

As we have seen, the ingredients which determine the Raman spectrum are the normal modes of vibration (either molecular vibrations or optic phonons) and the derived polarizability tensors (with respect to normal modes).

Our discussion is organized according to the following questions.

- How are normal vibrations of a material affected by the confinement within material domains of a size of a few nanometers?
- How is the polarizability tensor affected? That is, how are Raman selection rules affected? How is dispersion behavior and/or resonant behavior affected?

In the following section we try to extract some answers to these questions. Some quite general predictions can be given (and also suitable formalisms can be used) if specific models for the description of nanostructures are developed: for this reason we need to introduce a classification of nanostructures, based on their peculiar structural characteristics.

2. CLASSES OF NANOSTRUCTURES

The vibrational dynamics and the sequence of Raman transitions of a periodic (infinite) system (crystal) are subjected to rules (and give rise to a formalism, i.e., that of phonons) determining a spectroscopic behavior deeply different from that of single molecules. On the other hand, nanostructures can be considered objects which occupy an intermediate position between a perfect, ideally infinite crystal and a giant molecule.

Thus it is useful to group nanostructures according to a similarity criterion with respect to either a crystal (in this case the nanostructure is seen as a very little crystal and the solid state physics approach is the suitable starting point for its description) or to a molecule (in this case the nanostructure can be seen as a very large molecule and the molecular approach is better).

As this classification is necessarily too rigid, a bridge between the two different approaches can be conveniently built in same specific cases.

We propose the following classification, based on pure geometrical properties.

1. **Nanoobjects.** These objects must necessarily be regarded as (large) molecular entities, which cannot be correlated to a parent crystal with higher perfection characteristics (first of all, large enough to be thought as infinite). Molecular normal modes and molecular point symmetry rule their Raman response. According to this definition clusters (e.g., fullerenes) are nanoobjects.
2. **Nanocrystals.** These objects can be obtained cutting a finite portion from an infinite crystal. This ideal procedure finds several examples in reality, for example, by ball milling crystalline materials can be reduced to a powder made of small crystallites of few nanometers in size. Moreover, in a large crystal, due to the presence of structural defects, translational order can also be lost over relatively few cells: in this case the material can be described as formed by crystalline subunits of reduced dimensions. Defected graphites are often regarded as assembly of finite size domains with the same structure of a graphite sheet.

   A specific example in which nature directly provides us with nanocrystals is the case of (semi)crystalline polymers where single crystals are born in forms of lamellae with a characteristic thickness of few nanometers.

   The relevant point characteristic of nanocrystals is the fact that they can be set in correspondence with a parent crystal (which can exist in reality, as in the case of Ge crystals described below, or which exist only ideally, as in the case of a polymer). This fact allows us to start from the phonon formalism for the treatment of the vibrational problem.
3. **Special nanostructures.** In this class we collect materials which, according to their geometry, would be classified as nanoobjects, but which, at least in some respect, can be also related to an ideal parent crystal. Carbon nanotubes belong to this class. Their structure can be described as originating by cutting and rolling a piece of matter from a single graphite sheet. A very long nanotube can be regarded as a graphite portion which has been confined in two dimensions (the radial and the tangential ones), giving rise to a one-dimensional 1D crystal. On the other hand, the rolling of the sheet gives rise to an object which is unique from the structural point of view (it is not a strip, but really a tube, which is even characterized by a given chirality) with respect to the parent crystal.

3. GENERAL CONSIDERATIONS ON VIBRATIONAL DYNAMICS AND RAMAN SCATTERING OF NANOSTRUCTURES

In this section we discuss the general criteria (relative to vibrational dynamics and spectroscopic selection rules) which can be derived for the three classes of nanostructures considered above.

1. Vibrational dynamics of nanoobjects must be treated on the usual basis of molecular dynamics (nanoobjects are indeed molecules in all the points of view). In this case selection rules are dictated by the point group symmetry of the molecule and can be derived applying the usual procedure described for molecules. Notice that, in the case of a very large object (made by N atoms) with low symmetry, at least in principle, all $3N$-6 vibrational modes can be observed in the Raman spectrum at spectral shifts corresponding to their eigenfrequencies.

   Also, when nanoobjects are assembled to form solid aggregates (e.g., crystals, as in the case of fullerite, made assembling C_{60} fullerenes in a f.c.c. lattice) their molecular nature still dominates and practically governs their spectroscopic behavior. This is especially true when intermolecular bonds are relatively weak (van der Waals interactions), as usually happens in molecular crystals. An exception to this rule seems to be the f.c.c. solid C_{20} in which interstitials carbon atoms could create stiff covalent bridges between different cluster units [28].

   The case of nanoobjects does not require any further discussion at this level. A detailed example is given in Section 5.1, where the Raman spectrum of C_{60}, the most famous fullerene, is described and commented upon.
2. In the treatment of the vibrational dynamics of nanocrystals one can start from the hypothesis that their normal modes would be described on the basis of the phonons of the parent perfect crystal. This hypothesis drastically reduces the complexity of the mathematical problem and has several immediate consequences.
   - The normal vibrations (of the nanocrystal) are described as stationary waves which can be easily built as a combination of two phonons of the parent perfect crystal traveling in the opposite directions, characterized by wave vectors **q** and −**q**, respectively.
   - According to this model, the vibrational frequencies of the normal modes obtained from the phonons of wave vector **q** are identical to those lying on the vibrational dispersion curves of the parent crystal (at the same **q**).
   - The choice of the suitable subset of **q** values is determined by boundary conditions dictated by the finite size, as in the case of the electromagnetic standing waves characteristic of a cubic cavity of edge length L. In the simplest case of a 1D crystal [15] this leads to the relationship $q = \pi j/L = j/(N + 1)d$, j being an integer ranging from 1 to N and N the number of unit cells of size d contained in $L = d(N + 1)$ (see Fig. 1). The 3D case is similarly treated.

Since in nanocrystals the translational symmetry is broken, all the frequencies selected with the procedure described above can (in principle) be observed in the Raman spectrum. However, selection rules determined by the point symmetry of the nanocrystal (thought of as a unity) will determine the activity of each normal mode.

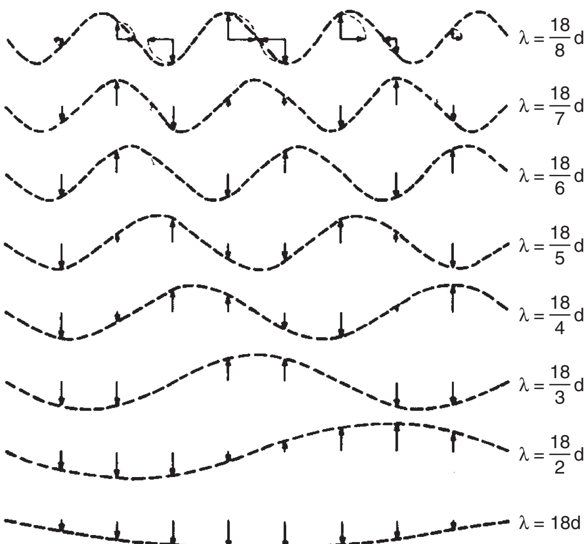


Figure 1. Longitudinal vibrations of a linear chain of eight point masses with fixed ends. The real displacements of the masses can be obtained by turning the vectors by 90° clockwise, as indicated for the mode with $\lambda = 18/8\ d$.

Under this point of view the finite size of a nanocrystal can be regarded as a very peculiar kind of defect of the parent crystal. As a consequence of the geometrical confinement, we can predict, as for any other kind of defect, the activation in the Raman (and/or infrared) spectrum of $q \neq \mathbf{0}$ phonons of the parent crystal. In a limiting case where the point symmetry of the nanocrystal is low we would expect that features observed in the Raman spectrum should be directly related to the density of phonon states, as is true to a first approximation for amorphous solids [29]. This is indeed what one expects for sufficiently large nanocrystals, where the spacing $\Delta q = 2\pi/(N+1)d$ between allowed $q's$ is small. However, in the simplest cases also, the experimental findings are often quite different from this ideal expectation.

The breakdown of the translational $\mathbf{q} = \mathbf{0}$ selection rule is usually effective only in the activation of phonons near to the zone center ($\mathbf{q} \approx \mathbf{0}$ phonons). This can easily be realized considering the simplest case of normal modes of a one-dimensional crystal: these modes can be classified according to the number of "nodes" of the associated stationary wave (Fig. 1).

The mode with the longest λ (lowest $q = 2\pi/\lambda$) allowed ($\lambda = 2L$) is characterized by two nodes at the boundaries (end of the chain): in this case the phase difference between amplitudes of vibration for atoms belonging to different repeating units of the chain is always less than π. As a result the contribution to the change of total polarizability by each unit sums up to that of the others (i.e., the effect is cooperative, by each unit). Modes with higher q values are described as an amplitude wave with more than two nodes. In this case we have several pairs of units along the chain which vibrate with the opposite phase, giving rise to contributions to the total polarizability which cancel each other. This mechanism of cancellation is more and more efficient as the number of nodes increases and as the dimension of the crystal increases.

On this basis it can easily be realized that for a relatively large nanocrystal only phonons in a narrow range of q values near zone center will be seen in the Raman spectrum. Another consideration must be added about confinement (for details see the case of small 3D Ge nanocrystals, quantum dots, as illustrated in Section 5.3.1). Confined phonons can be thought of as the application of an envelope factor to a plane wave. This amplitude modulation makes the phonon wave function significantly different from 0 only in a space region of linear size a. Thus a very general, though rough, evaluation of the size $|\Delta\mathbf{q}|$ of wave vector space involved is easily deduced from the uncertainty principle as $2\pi/a$, where a could be the nanoparticle radius. As $\mathbf{q}$ scans a volume $\approx|\Delta\mathbf{q}|^3$ around $\mathbf{q} \approx 0$, provided the group velocity of crystal phonons is significantly different from 0, one can observe the Raman contribution of several phonons leading to inhomogeneous broadening of the Raman line. If the group velocity is negative, one observes an average red shift; if it is positive, an average blue shift. In this way inhomogeneous broadening and either red or blue shifts are distinctive marks of vibrational confinement effects. Yet special interpreting care must be assumed, because small shifts in the Raman line can also be due to internal stresses via anharmonic effects.

Due to the extreme flatness of the optical phonon dispersion curves of polymers, it is practically impossible to reveal the small thickness of polymer lamellae by inspection of Raman bands corresponding to vibrational transitions which belong to optical branches. However, in the case of normal modes belonging to acoustic branches, new frequencies can be revealed. This is the case of longitudinal accordion motions, whose frequencies give reliable information about the dimension of the polymer lamellae (see Section 5.4).

4. CONFINEMENT EFFECTS

In Section 3 we have described the effect of confinement in nanocrystals solely as a pure geometrical constraint: all relevant dynamic properties were thought to remain the same as in the parent crystal and finite size boundary conditions ruled out the $\mathbf{q} = \mathbf{0}$ rule activating a whole set of wave vectors.

Unfortunately this first approximation shows some weakness in many cases. This can be ascribed to the following points:

(i) *Effects due to change in the vibrational potential energy* (*and then in vibrational force constants*). This effect is usually found in systems characterized by long-range interactions. It is the case of organic materials containing conjugated π electrons, with an electronic structure strongly sensitive to the delocalization path available (and then to the size of the crystal).

As a consequence, one finds a continuous change in the values of some force constants often accompanied by an increase in the distance of interaction as a function of the size of the crystal itself. The signature of this phenomenon in the Raman spectrum is a frequency shift of some characteristic Raman bands with respect to the frequency value of the parent crystal.

Systematic frequency and intensity dispersion of the strongest Raman active modes has been observed for linear polyenes with increasing chain length [18] (see Fig. 2). These frequencies do not reach the limit of the $q = 0$ phonon of polyacetylene, which can be estimated by extrapolation from the frequency trend observed for polyenes of increasing length. On the other hand, probing a polyacetylene sample [20] (see Fig. 3) with different exciting laser lines one can sweep through the range of energy gaps of polyacetylene. Thus by resonance it is possible to probe the poly-dispersity in conjugation lengths of a given sample (see Section 5.4).

The frequency dispersion observed in the case of polyenes cannot be simply explained considering the fact that normal modes of finite size chains (polyenes) correspond to polyacetylene phonons with small (but finite) q values. Due to the large frequency shifts observed, this would imply having very steep optical phonon branches (near $q = 0$) for the parent crystal. For this reason a change in some relevant force constant (a softening of the relevant force constant while increasing the chain length) has been

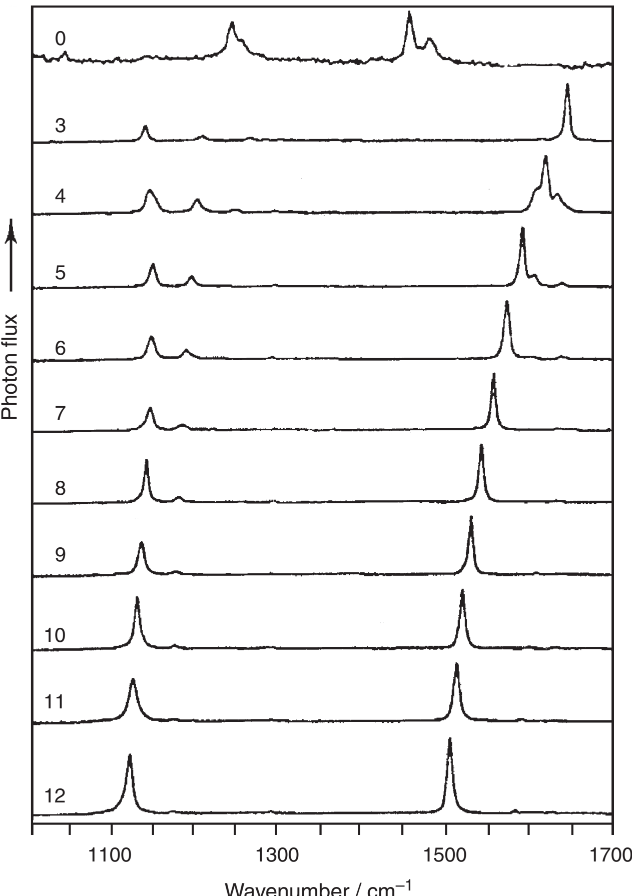


Figure 2. Raman spectra of t-butyl-capped polyenes $N = 3$–12 is the number of C=C bonds in the chain. $\lambda_{exc} = 647.1$ nm. The top spectrum is that of di (t-butyl) that is, $N = 0$ [18]. Reprinted with permission from [16], M. Gussoni et al., "Spectroscopy of Advanced Materials" (1991). © 1991, Wiley.

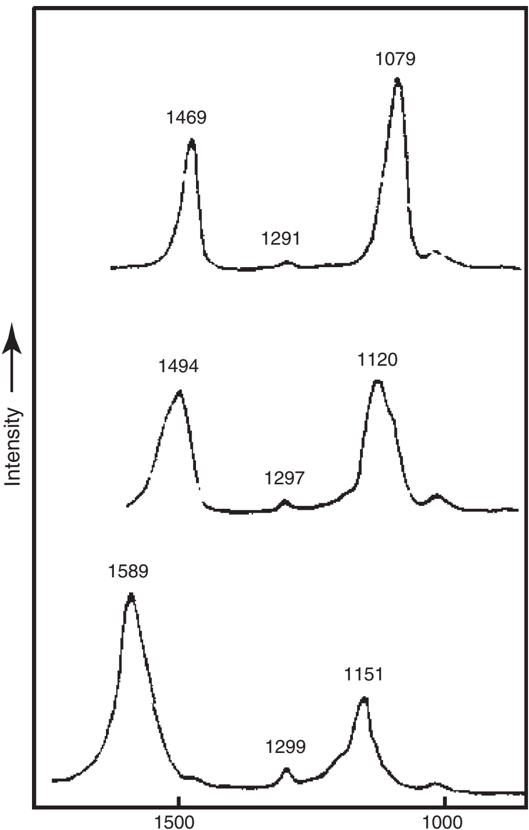


Figure 3. Raman spectra of *trans*-PA. The same sample has been examined with three different excitation lines: 647 nm (top); 488.0 nm (middle); 350.7 nm (bottom). Reprinted with permission from [16], M. Gussoni et al., "Spectroscopy of Advanced Materials" (1991). © 1991, Wiley.

proposed [16, 17, 19] in order to explain the observed phenomena.

(ii) *Structural relaxation infinite size crystal domains*. In the case of polyenes mentioned above it is possible to follow the effect of the changes in the electronic structure (due to the presence of conjugated π electrons) with chain length or conjugation length, also considering the ground state nuclear geometries. Theoretical calculations [21] of polyenes of different and increasing size show that the degree of bond alternation (i.e., the value of the difference between the equilibrium bond length of adjacent CC bonds along the polyene chain) decreases as the chain length increases, as a consequence of the larger delocalization of the p_z electrons. According to this trend one would expect that in the infinite crystal (polyacetylene) a perfect equalization of the CC bonds should be reached, thus making polyacetylene, a metal. However, due to the Peierls instability [22, 23] in the 1D crystal, this does not happen and the parent crystal keeps a structure similar to that of the oligomers (polyenes) with a sequence of quasi-simple and quasi-double bonds. The dimerization parameter is determined by the strength of the Peierls instability.

However, structural relaxation infinite size domains can bring much more drastic effects. This can be realized, for instance, if one compares the equilibrium nuclear geometry of graphite with that of molecular graphites [24] (very large polycyclic aromatic hydrocarbons (PAHs) [25]) (see Fig. 4). In Section 5.5 the Raman behavior of these molecular materials will be described and compared with that of graphite. What is relevant for the present discussion is the fact that the ground state geometries of PAHs do not show the presence of equalized CC bonds, as happens in the case of graphite. Theoretical predictions [24] indicate that PAHs start to show an equalized structure (but only in the very inner region) only when a very large size (about 10 nm of diameter) is reached. PAHs can be grouped according to three different canonical structures (Fig. 5). All the PAHs experimentally available belong to the all-benzenoid family, where the bonds are arranged in aromatic rings (sextets of short CC bonds) linked each other by longer CC bonds. According to these findings it can be concluded that the parent crystal of molecular graphites is not an infinite graphite sheet (with all CC bonds of identical length, two C atoms per hexagonal unit cell), but a 2D crystal with a relaxed structure, whose unit cell contains six C atoms arranged according to the same topology as graphite, but showing CC bonds of two classes, with different equilibrium bond length (Fig. 6). This fact has very important implications in the discussion of normal modes of PAHs in terms of the phonons of graphite.

Indeed a superlattice (with a threefold cell) has to be introduced as a reference lattice for PAHs (seen as nanocrystals cut from an infinite graphene sheet). The definition of this superlattice, implies that phonon dispersion curves obtained for a 2D graphite must be looked at after a zone folding procedure, following the introduction of a new reciprocal lattice. In Figure 7 it can be seen that this procedure brings K points of the first Brillouin zone (BZ) of graphite to coincide with

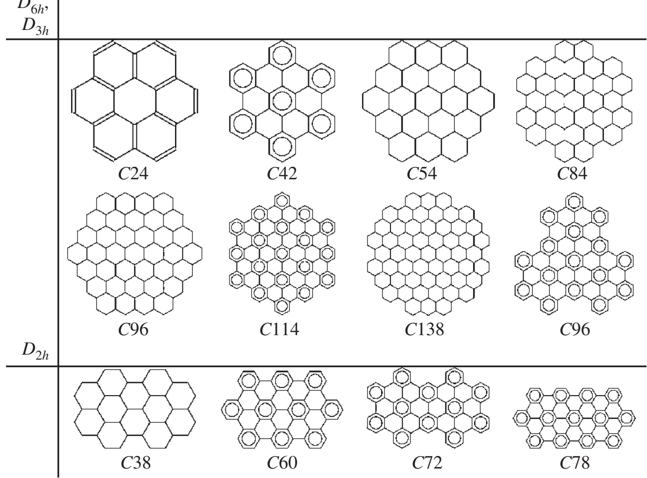


Figure 4. Structure of some PAH molecules. $C_{42}H_{18}$, $C_{114}H_{30}$, $C_{96}H_{30}$, $C_{60}H_{22}$, $C_{72}H_{26}$, and $C_{78}H_{26}$ are experimentally available thanks to the new synthetic route developed at Max Planck Institut of Mainz (D) [25].

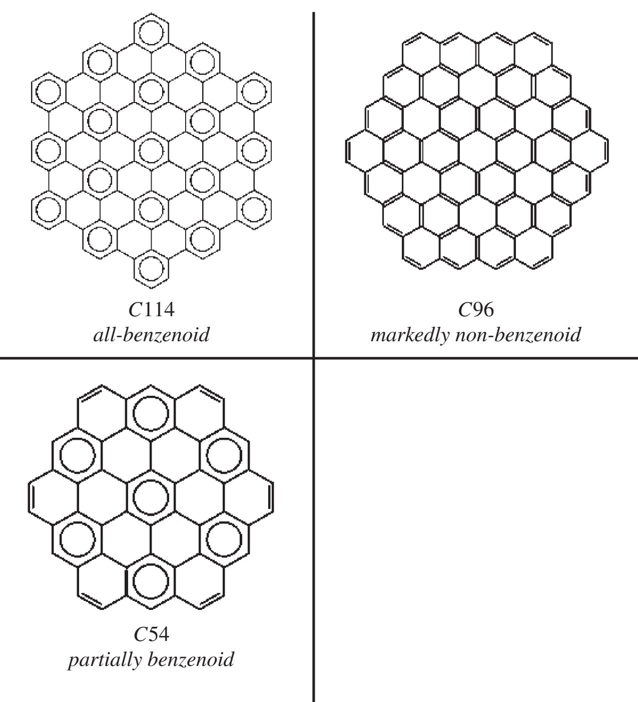


Figure 5. Schematic representation of the three classes of PAHs identified according to the ground state optimized geometry.

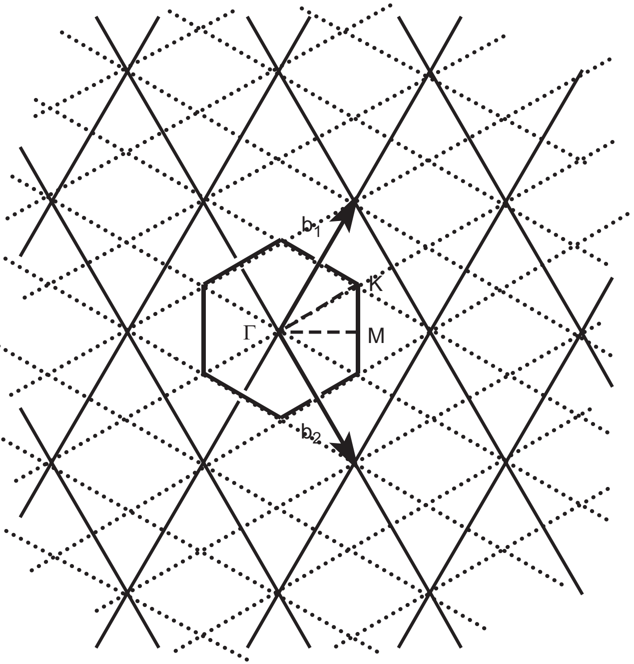


Figure 7. Reciprocal lattice of the supercrystal of Figure 6 (dotted lines). The reciprocal lattice and the first Brillouin zone of a perfect 2D lattice of graphite are shown (solid lines).

nodes ($\mathbf{q} = \mathbf{0}$ points) of the reciprocal lattice of the superlattice [26].

According to this finding we expect (see Section 5.5) that large molecular graphites show strong Raman transitions for those normal vibrations corresponding to graphite phonons with wave numbers $\mathbf{q} \approx \mathbf{0}$, where $\mathbf{q}$ is read according to the suitable reciprocal lattice (i.e., that of a threefold superlattice). As a consequence both phonons near Γ and near K points of the first BZ of graphite are predicted to be potentially active in the Raman spectrum.

This is what indeed happens in the case of large PAHs where two strong bands appear in the Raman spectrum which can be set in correspondence with the one Raman active optical phonon of 2D graphite at Γ point (the so-called G band) and to the totally symmetric phonon at K point (D band).

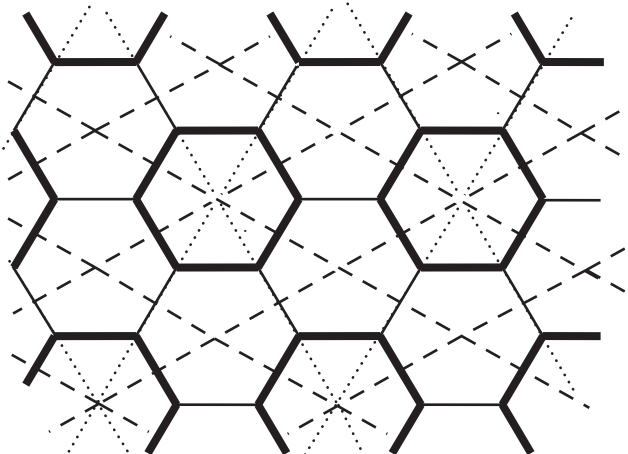


Figure 6. Perfect 2D graphite lattice (dashed lines) and superlattice (dotted lines) corresponding to the parent crystal of molecular graphites, showing an all-benzenoid structure.

The results obtained by the study of molecular graphite have been extrapolated to the case of disordered graphites and microcrystalline graphite in order to explain the activation of the D band in these materials [27] (see Section 5.5 for a discussion and comparison with different approaches).

The concept of zone folding is a quite general procedure which allows us to make use of the dynamics (and phonon dispersion curves) obtained for a given crystal structure in order to study systems characterized by different topologies and symmetries, but which are expected to keep the same (or a very similar) dynamic interactions (force constants) between atoms.

The frequencies of the active phonons of carbon nanotubes have indeed been predicted with success by this technique [109]. As expected, several ($q \neq \mathbf{0}$) points of the first BZ of a 2D graphite lattice are brought to coincide with Γ point, after zone folding is applied (Fig. 8). What is important is the fact that a given nanotube (characterized by a given diameter and a given chirality) requires the definition of its own superlattice: as a consequence the activation of different graphite phonons for nanotubes with a different structure can be predicted.

It can immediately be seen that this approach can be extremely useful for a structural diagnosis (in terms of tube diameters, chirality, and even distribution of these parameters) by Raman spectroscopy on nanotube samples (see Section 5.5 for a discussion).

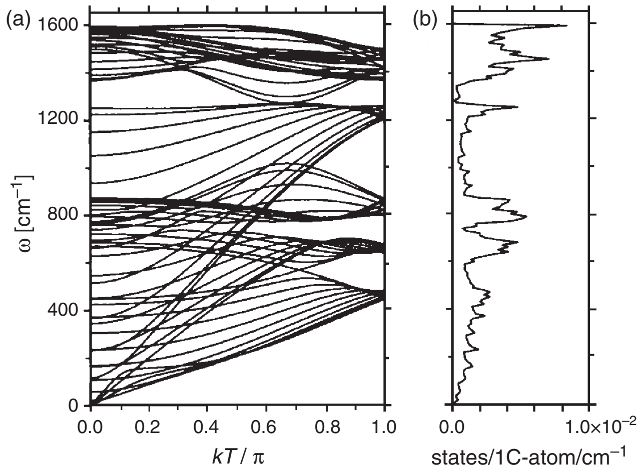


Figure 8. Phonon dispersion relations and phonon DOS for a (10, 10) nanotube, as obtained by a zone folding procedure applied to dispersion relations of a 2D graphite (left) and by dynamics calculations based on a force field adapted to the specific nanotube geometry. Reprinted with permission from [109], R. Saito and H. Kataura, "Topics in Applied Physics, Vol. 8, Carbon Nanotubes: Synthesis, Structure, Properies and Applications" (2001). © 2001, Springer-Verlag.

(iii) *Resonance effects*. As mentioned in the Introduction the dispersion behavior of the Raman intensity with the energy of the exciting laser line is a very important phenomenon, most of all for low band gap materials.

From a formal point view and similarly to the case of phonons, for the description of electronic states in nanostructures the band theory (as developed for an infinite perfect crystal) is also no longer exact. However, one expects to find discrete energy levels (which merge in a continuum of states as the size of the crystal increases) which approximately lie on the dispersion electronic branches of the parent crystal. This allows us to make use of the known electronic structure of the parent crystal in order to discuss effects due to interactions between phonons and electronic excited states.

According to this idea it is possible to predict important effects observed in the Raman spectra of nanostructures, as the resonance enhancement of scattering from $\mathbf{q} \neq \mathbf{0}$ phonons activated by the confinement (see ii above).

As an example, we can quote the case of graphites: in order to explain the activation of the *D* line in microcrystalline and nanocrystalline graphites, Thompsen and Reich [104] proposed a very elegant treatment based on a double resonance effect responsible of the very strong activation of a specific $\mathbf{q} \neq \mathbf{0}$ phonon. In the Thompsen model the use of the electronic band structure of graphite is explicitly introduced.

Moreover, the lowering in the dimensionality in going from graphite to nanotubes gives rise to Van Hove singularities in the density of electronic states. The implication of this phenomenon is remarkable: the intensity behavior of the Raman spectrum of nanotubes as a function of the exciting laser cannot be explained without the introduction of the concept of resonance enhancement by interaction with electronic states corresponding to these singularities [109, 116].

(iv) *End or surface effects*. When one considers finite size crystals it is immediately realized that the lack of a perfect translational invariance does not consist only in the fact of having a finite number of cells: while approaching the crystal boundaries unit cells cannot be considered identical due to effects of the (near) surface. When this effect is strong (i.e., it is a long-range effect) a perturbation of the vibrational potential can take place in the whole nanocrystal and possibly also changes in the nuclear equilibrium structure.

However, in all the cases (both in the presence or in the absence of long-range interactions) atoms lying at the surfaces have to be considered defects which probe an environment different than that probed by the atoms which belong to the bulk.

The case of polymers is meaningful: in any linear chain, for chemical reasons, end groups are chemically and structurally different from those of the central units. For instance polyethylene (chemical formula $(CH_2)_n$) can end on either side by methyl (CH_3) groups or by vinyl groups ($CH{=}CH_2$). Vinyl groups are also the ends of polyenes and polyacetylene (chemical formula $(CH)_n$).

As a consequence of the presence of end defects in polymers we expect to find specific signals in the vibrational spectra: characteristic group frequencies corresponding to normal modes localized on end atoms can indeed be detected in the case of relatively short chains. These signals are easily detectable in the infrared (they are routinely used for an estimate of the average chain length in low-molecular-weight polymers); however, they can be detected via Raman scattering only when the ratio between the number of end and bulk groups is high.

It is indeed an usual finding that in the Raman spectra collective vibrations are intrinsically much more active than localized modes.

5. SPECIFIC NANOSTRUCTURES

5.1. C_{60}: The Bucky Ball

Since the discovery of the main (most stable) fullerene, the C_{60} molecule (or cluster), also called the buckminsterfullerene (or bucky ball) [30–34], the detection and the study of its vibrational fingerprints has attracted a great deal of interest both from the experimental and the theoretical point of view. Due to the large number of papers that have appeared in the literature, it is a difficult task to give, in a necessarily short paragraph, a comprehensive summary of the work done on this subject during the last 15 years. For this reason, the reader interested in an overview of all the relevant aspects is referred to some books and specifically to chapters dealing with the spectroscopy of fullerenes [35–37].

The following discussion has merely the aim of pointing out some of the most relevant aspects of the Raman spectrum of (C_{60}) and its interpretation.

C_{60} can be described (with respect to its spectroscopic behavior) as a nanoobject. In other words, it is really a

molecule and the concepts of molecular spectroscopy can be applied to the interpretation of its vibrational spectra with success.

The structure of an isolated C_{60} unit is that of a closed carbon cage (fullerene), with the point group symmetry of a regular truncated icosahedron. C atoms in the cage are bounded together in an highly symmetrical arrangement of pentagonal (12) and hexagonal (20) rings (Fig. 9). The C atoms belonging to pentagonal rings are connected by longer (single) bonds ($R_{CC} = 1.455$ Å) and C atoms shared by adjacent hexagons are connected by shorter (quasi-double) bonds ($R_{CC} = 1.391$ Å) [38].

According to I_h point group symmetry, the structure of the representation in the vibrational space of C_{60} is

$$\begin{aligned}\Gamma_{vib} &= \Gamma_{vib}^{Raman} + \Gamma_{vib}^{infrared} + \Gamma_{vib}^{inactive}\\ \Gamma_{vib}^{Raman} &= 2A_g + 8H_g\\ \Gamma_{vib}^{infrared} &= 4F_{1u}\\ \Gamma_{vib}^{inactive} &= 3F_{1g} + 4F_{2g} + 6G_g + A_u + 5F_{2u} + 6G_u + 7H_{2u}\end{aligned} \tag{24}$$

According to this description we can immediately make some observations.

1. The very high symmetry of the molecule implies that the vibrational transitions which can be detected in vibrational spectra are few with respect to the number of degrees of freedom of the molecule. The vibrational degrees of freedom of an isolated C_{60} are 174, but they correspond to 46 different frequencies, due to the high degeneracy of some symmetry species (e.g., modes of H species have a degeneracy of 5). Moreover, among these 46 frequencies, only 4 are active with respect to infrared absorption and 10 are Raman active. On this basis we expect to obtain relatively simple Raman and infrared spectra from samples of C_{60}.
2. Another consequence of the high symmetry of the molecule is the fact that all normal modes of C_{60} involve the molecule as a whole (all C atoms are displaced during each normal mode). By virtue of this fact we expect that frequencies are strongly sensitive to the characteristic structure of C_{60} (on the whole): in this sense vibrational spectra are expected to behave as is usual in molecular vibrational spectroscopy as fingerprints of that specific molecule, with that specific structure.

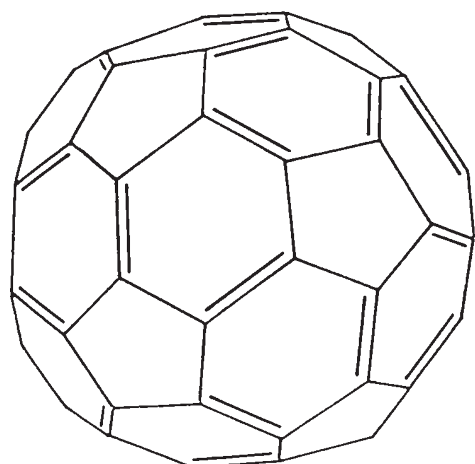


Figure 9. Sketch of the structure of buckminsterfullerene, C_{60}.

One of the first Raman experiments on purified C_{60} have been reported by Bethune et al. [39]. Spectra (see Fig. 10) have been taken by a micro-Raman from a C_{60} film deposited on a suprasil slide by a C_{60} solution.

According to these spectra the 10 Raman active fundamental transitions of C_{60} were clearly identified. Among these, the strongest lines at 1470 and 496 cm^{-1} have been assigned to the two A_g modes, according to their behavior in polarized Raman spectra.

In the same paper [39] vibrational spectra of C_{70} (the other abundant fullerene) are reported. In the case of C_{70}, due to the lower symmetry of the molecule, infrared and Raman spectra show a drastic increase in the number of observed frequencies. No coincidences of the frequencies of the main lines are found with observed frequency values of C_{60}. These last observations strongly support statements 1 and 2 above. On the other hand they make clear that Raman and infrared spectra of C_{60} provide evidence that C_{60}

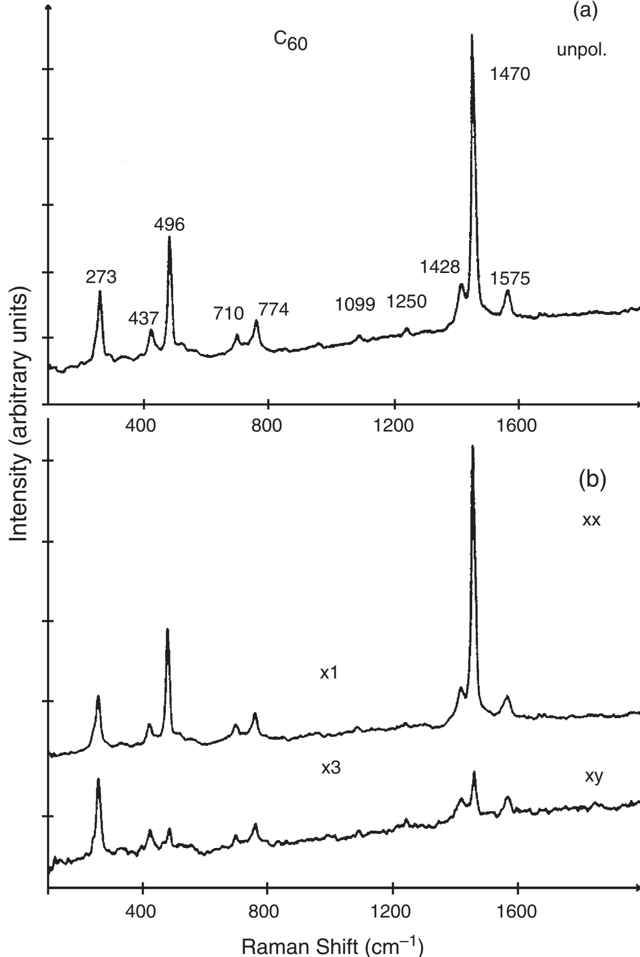


Figure 10. Raman spectra of a film of purified C_{60} on a suprasil substrate. (a) Unpolarized; (b) polarized. Reprinted with permission from [39], D. S. Bethune et al., *Chem. Phys. Lett.* 179, 181 (1991). © 1991,

belongs to the I_h symmetry group and that vibrational spectra can be safely used for the detection of C_{60} in unknown samples.

The identification and the vibrational assignment of C_{60} transitions proposed by Bethune et al. [39] were supported by theoretical predictions obtained by Stanton and Newton [40]. Theirs is only one of the many papers that has appeared in the literature dealing with the theoretical prediction of vibrational frequencies of C_{60}. In the more recent calculations, the accuracy of the results has been raised in such a way that the discrepancies found between computed and experimental frequencies are of the order of few wave numbers.

According to our opinion, two points have to be stressed regarding the development of theoretical models for the study of C_{60} vibrational dynamics.

(i) The possibility of obtaining reliable results is strongly influenced by the level of the theory used (which in turns depends on the available computational resources). In a recent paper by Schettino et al. [41]. DFT calculations using B3-LYP exchange and correlation functional and a large (6–31G*) basis set have been performed, thus reproducing the experimental frequencies of C_{60} with very good accuracy, with a single scaling factor (0.98) on the calculated frequencies. On the other hand, semiempirical methods based on suitable Hamiltonians have been shown to give excellent results. The use of an updated QCFF/PI Hamiltonian developed by Negri and Orlandi [42] made if possible to obtain an accuracy very close to that already reached by first principle calculations, at a fraction of the computational cost.

(ii) The choice of the best theoretical models and the critical comparison of the different approaches is based on the availability of as many experimental data as possible. In this respect the case of fullerene presents the problem that few vibrational frequencies give observable transitions. The problem of obtaining information about silent frequencies has been partially solved with the help of neutron scattering [33] data and partially making use of accurate data from C_{60} crystals and chemically substituted C_{60}.

As suggested at point ii, the difficulty arising from the lack of information about the frequencies of the inactive transitions can be overcame by some suitable molecule modification which lowers the symmetry of the molecule itself. A technique widely used in molecular spectroscopy is (asymmetric) chemical substitution with isotopes.

In the case of C_{60} nature itself provides a way to observe a C_{60} molecule with a lowered molecular symmetry. This happens when C_{60} molecules pack together into a fullerite crystal. I_h symmetry is not compatible with any crystal space group: in its low-temperature crystalline phase C_{60} forms a closely packed f.c.c. lattice with T_d symmetry.

As mentioned in Section 3, however, the interactions between pairs of C_{60} molecules in the crystal are weak: as a consequence the intramolecular vibrations are expected and found to be very weakly perturbed by such interactions and the Raman and infrared spectra of fullerite show a pattern which can be interpreted on the basis of the vibrations (and symmetry) of the single isolated molecule. On the other hand, the symmetry lowering of the molecular surrounding in the crystal causes (at least in principle) a breakdown of the Raman (or infrared) selection rules: weak transitions, corresponding to the silent frequencies of the isolated species, have been detected. Data from vibrational spectra of fullerite have been used by Schettino et al. [41] and Negri and Orlandi [42] for comparison with the predicted results.

According to these last observations, we can conclude that when the nanostructured materials (e.g., fullerite) are made by weakly interacting nanoobjects characterized by high symmetry, we can obtain from the Raman spectrum information essentially related to the molecular nature of the object (namely, its shape, symmetry, and intramolecular potential). On the contrary, relatively little information can be derived about the superstructure (molecular arrangement and interactions). On this subject we refer the reader to Eklund et al. [36] where a thorough discussion of optical phonon modes and their symmetry classifications in solid (fullerite) and doped C_{60} is presented.

5.2. 3D Nanocrystalline Materials

We have seen that the differential cross section for spontaneous light scattering $d^2\sigma/d\Omega d\omega_s$ is proportional to the dynamic structure factor $S_{E^S}(\mathbf{Q}, \omega)$. We now follow Nemanich et al. [43] and assume that in the presence of confinement effects also the susceptibility fluctuations are related to the phonon displacements $u(\mathbf{r}, t)$ to linear order by

$$\delta\chi_{ij} = A_{ij}u(\mathbf{r}, t) \tag{25}$$

In this way $S(\mathbf{Q}, \omega)$ (we drop the index E_S) becomes directly proportional to a phonon displacement autocorrelation function (see Eq. (23)).

In the case of nanocrystalline materials characterized by dimensions of particles or domains comparable with phonon wavelength, the description of the elementary excitations with (modulated) plane waves (Bloch waves) is no more justified. A wave packet whose spatial dimension is comparable to the crystallite size must be used: this introduces a spread or uncertainty in wave vectors. This uncertainty will be larger for smaller domain sizes because the wave packet becomes more localized in real space. Since the phonon dispersion curves are in general not flat, new frequencies will be introduced in the Raman spectra when contributions from phonons with wave vectors $\mathbf{q}$ significantly different from 0 start to play a role [44].

In general, assuming that the size reduction does not modify the phonon dispersion curves of the periodic reference system (parent crystal) and does not change the other Raman selection rules, Raman lines, now resulting from both $\mathbf{q} = \mathbf{0}$ and $\mathbf{q} \neq \mathbf{0}$ individual contributions, will exhibit an overall shift and asymmetrically broaden, the nature of the changes and the critical size where these changes take place depending on the dispersion curve of the material and on the intrinsic phonon lifetimes.

We can assume that phonons of the same branch of different **q** in small crystallites have a correlation function as in a perfect infinite crystal given by

$$\langle u^*(\mathbf{r}, t)u(\mathbf{r}', 0)\rangle_{th} = \frac{1}{V}\sum_{\mathbf{q}}\langle u^*(\mathbf{q})u(\mathbf{q})\rangle_{th} e^{[i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')+i\omega(\mathbf{q})t]} \tag{26}$$

where $\omega(\mathbf{q})$ are the intrinsic phonon dispersion curves. Thus [43] $S(\mathbf{Q}, \omega,)$ can be written as

$$S(\mathbf{Q}, \omega,) = \left(\frac{N}{V^2}\right)\frac{1}{V}\sum_{\mathbf{q}}|A_{ij}(\mathbf{q})|^2\delta(\omega-\omega(\mathbf{q}))\langle u^*(\mathbf{q})u(\mathbf{q})\rangle \times |F(\mathbf{Q}-\mathbf{q})|^2 \tag{27}$$

where N is the number of crystallites in the volume V and F is a integral restricted to a single crystallite:

$$F(\mathbf{Q}-\mathbf{q}) = \int e^{i(\mathbf{Q}-\mathbf{q})\cdot\mathbf{r}}\, d\mathbf{r} \tag{28}$$

then $S(\mathbf{Q}, \omega,)$ can be written as

$$\frac{N}{V^2}\left(\frac{n(\omega)+1}{\omega}\right)\frac{1}{V}\sum_{\mathbf{q}}C(\mathbf{q}, \omega)|F(\mathbf{Q}-\mathbf{q})|^2 \times \frac{\Gamma_{1/2}/4\pi}{[\omega-\omega(\mathbf{q})]^2+(\Gamma_{1/2}^2/4)} \tag{29}$$

where all coefficients are grouped in a single collective matrix element $C(\mathbf{q}, \omega)$, the boson occupation numbers have been introduced and a phonon lifetime has been included as a Lorentzian broadening $\Gamma_{1/2}$.

For a large crystal $F(\mathbf{Q}-\mathbf{q})$ becomes a delta function and the spectrum becomes, in general, a sum (over active phonon branches at Γ) of weighted Lorentzians; on the other extreme, for very small structures, the system can be considered amorphous and the whole phonon density of states becomes visible: the spatial integration will be limited to the local atomic environment and all the excitations can contribute to the scattering.

In the intermediate case, where crystallites of a few nanometers in diameter must be considered, the choice of $F(\mathbf{Q}-\mathbf{q})$ should be based on physical considerations to motivate a specific phonon confinement function.

For example, Nemanich and co-workers [43] have considered a confinement function for boron nitride microcrystals of the form

$$F(\mathbf{Q}-\mathbf{q}) = \prod_{i=x,y,z} 2\sin[(q_i-Q_i)L_i/2]/(q_i-Q_i) \tag{30}$$

Due to size distribution, this function should be more appropriately replaced with a Gaussian function. In reality a direct connection between the confinement function and the microscopic structure of the crystallites is not straightforward. Numerical fits of the experimental results are used to find the best confinement function. With this information the mean size of the crystallites can be extracted; however, this quantity can be exactly determined only for a single size ensemble of particles or if the size distribution and the phonon dispersion curves are known exactly.

5.3. Semiconductor Nanoparticles

Group-IV-based structures, consisting of Ge and Si nanocrystals embedded in matrices or free standing, exhibit visible photoluminescence that is sensitively dependent upon the particle size and size distributions. Despite considerable efforts having been concentrated on the characterization of these nanocrystalline systems and on their electronic structures, many aspects remain to be clarified. The confinement of elementary excitations in semiconductor nanocrystals has a strong influence on photoluminescence behavior and more generally on linear and nonlinear optical properties. For an unambiguous characterization of the role played by structure and dimensions in determining the quantum confinement effects, it is necessary to prepare well-characterized nanostructures and to be able to follow their evolution along the different steps of the synthetic routes.

Raman scattering is particularly suitable for the characterization of semiconductor low-dimensional systems: it is nondestructive, it can be applied to systems embedded in transparent matrices, and it can also be used for *in-situ* measurements with a lateral resolution of 1 μm. As we have already discussed, in many instances, a theory has been developed that successfully explains the observed spectra in terms of the structural properties of the particles. The effect of confinement on the phonon density of states can be quantitatively related to the size and distribution of the particles, making Raman scattering a powerful tool for the characterization of assemblies of semiconductor nanoparticles. Despite its potentiality a word of caution is in order concerning the different factors affecting the phonon density of states and hence the Raman spectra. As we discuss in more detail in the following, the interaction of the particles with the host matrix, the oxidation of the surface, the coexistence of different amorphous and crystalline phases in the same particle, and the size distribution of the particles can affect the Raman response. It is then difficult to disentangle the different contributions to the modification of the phonon density of states of a reference crystalline bulk system and to what extent confinement effects are the only responsible cause of the observed spectra features. This implies that an independent determination of the size distribution and of the structure of the particle and of their degree of crystallinity is highly desirable for a correct interpretation of the Raman spectra and an identification of the confinement effects.

5.3.1. Germanium Nanoparticles

Ge nanostructures embedded in different matrices have been extensively studied by Raman spectroscopy. A wide variety of techniques have been used to produce Ge nanocrystals, including gas evaporation [45, 46], rf magnetron sputtering [47–49, 238, 50], chemical vapor deposition [51], chemical reduction of metastable $Si_{1-x}Ge_xO_2$ [52–55], molecular and cluster beam deposition [56–58], oxide reduction in zeolite [59], inorganic solution phase synthetic routes [60] and ion implantation and annealing [61]. These techniques have to meet several requirements for the production of samples which can be unambiguously characterized: (i) Production of single sized particles or assemblies

of particles with a well-defined size distribution and (ii) control of the crystalline state of the particles and of their interaction with the embedding matrices.

The production of Ge nanocrystals embedded in a matrix such as SiO_2 is obtained by chemical reduction, precipitation, and annealing of the $Si_xGe_yO_z$ phase [52–55]. The influence of the host matrix on the structure of the embedded clusters is still an open question. The crystalline structure, the coexistence of amorphous and crystalline phases, and the nature of the interfaces are very difficult to characterize over macroscopic regions of the sample. X-ray scattering, TEM, and HRTEM analysis are usually employed for the determination of size distribution and for checking the crystalline structure of nanoparticles [46, 62]. In principle, with these techniques information about the structure and dimension of a particle can be achieved independently; however, an extrapolation over the entire sample is not always straightforward.

Raman scattering has been used as a diagnostic tool for the determination of size distribution; however, since Raman spectra are expected to be quite sensitive not only to quantum confinement, but also to disorder, strain, or in general to interface peculiarities and structural characteristics of the system, it is mandatory to have an independent way to characterize the crystalline nature of the particles and their size distribution. On the other hand the same problem is encountered in the study of the confinement effects themselves, since the size distribution and the interaction with the matrix can introduce effects which render the interpretation of the spectra in terms of confinement very difficult [46].

For bulk crystalline Ge, the Raman spectrum can be assigned to the $\mathbf{q} \simeq 0$, Γ_{25}^{+} optic phonon exhibiting a Lorentzian line at 300 cm^{-1} with an intrinsic width of $\approx$3 cm^{-1} [14, 57]. The $\mathbf{q} \simeq 0$ selection rule is a consequence of the full translational symmetry of a crystalline sample. In the presence of phonon confinement in the nanoparticles, the Raman intensity can be written as a superposition of Lorentzians centered at $\omega = \omega(\mathbf{q})$. This produces a symmetric broadening and red shift of the Raman line. In particular the red shift is caused by the fact the in germanium ω is a decreasing function of $|\mathbf{q}|$. Since it is very difficult to produce samples characterized by identical particles, the size distribution of the particles must be taken into account. This introduces a further asymmetry in the line which can be described by

$$I(\omega) \propto \int_0^\infty \rho(R) R^3 \, dR \int_0^1 \exp\left(-\frac{q^2 R^2}{4}\right) \times \frac{\Gamma/2\pi}{[\omega - \omega(q)]^2 + (\Gamma/2)^2} d^3 q \tag{31}$$

where the size distribution $\rho(R)$ should be measured independently [46].

Different authors used Raman spectroscopy to characterize the Ge particles in SiO_x matrices and to follow the change in size distribution due to thermal annealing. Kobayashi and co-workers produced an SiO_2/ncGe/SiO_2 structure [55] by depositing a Ge layer on a SiO_2 film and annealing it at a temperature between 1100 and 1300 K for 1 h in a dry O_2 ambient. Due to rapid diffusion of oxygen molecules through the Ge grain boundaries, the Ge become crystallized and embedded in the SiO_2 during high-temperature oxidation. Raman spectroscopy has been used to determine the size of Ge nanocrystals as a function of the annealing temperature. The Raman spectrum of the as-deposited amorphous Ge is characterized by a broad shoulder peaked around 270 cm^{-1}. The annealing, inducing a crystallization of the amorphous phase, causes the appearance of the phonon peak at 300 cm^{-1} and the disappearance of the shoulder. The asymmetric peak obtained at 1300 K has been fitted to obtain the average size of the particles.

Paine and co-workers performed Raman spectroscopy on $S_{1-x}Ge_x$ alloys oxidized and annealed under various conditions to promote the formation of Ge nanoparticles [52]. This work indicates that larger average particle size can be obtained with longer annealing times and higher temperatures. It should be remarked that no frequency shift of the Raman peak is observed contrary to the prediction of the theory. Samples deposited by rf sputtering in a glassy SiO_2 matrix and subsequently annealed showed a similar behavior, with, however, a shift of the peak position [238].

A systematic study of the growth of Ge nanocrystals in SiO_2 matrices by Raman spectroscopy and electron microscopy has been performed by Fuji and co-workers [48]. Rf cosputtering and annealing were employed to produce Ge nanoparticles. The dependence of the Raman spectra upon annealing temperature is similar to that of Kobayashi et al. [55], showing a crystallization for temperature higher than 900 K. The size distribution of Ge nanocrystals has been determined by HRTEM analysis and used to calculate the Raman spectra following the phonon confinement relation. They have found a qualitative agreement; however, the peak frequencies of experimental spectra are much lower than the calculated ones and no shift of the Raman peak is observed.

The discrepancies in the Raman shifts in apparently similar samples are difficult to account for. A possible explanation is the deviation of the lattice constant from the bulk value caused by cluster surface–matrix interaction. Fortner and co-workers have demonstrated the influence of the near-surface dangling bonds and of the bond–angle disorder on the phonon density of states [49].

Besides the particle–matrix interaction, several other factors can affect Raman response: Fujii and co-workers have suggested that the presence of compressive stress exerted by the matrix on the clusters can compensate for the downward shift of the Raman peak. However, no direct evidence for the existence of the stress has so far been provided. The coexistence of amorphous and crystalline phases in the same particle can also induce a departure of the Raman behavior from that predicted by a too naive theory taking only confinement into account. The effects of size distribution should also be taken into account explicitly. As already discussed the width of the Raman peak is determined from both homogeneous broadening caused by the phonon confinement and inhomogeneous broadening caused by size distribution. A strong size dependence of the peak frequency should indicate that the inhomogeneous broadening is nonnegligible.

In order to minimize the interaction of particles with host matrices, free particles can be grown from gas phase and in colloidal suspension [45, 60]. Hayash et al. [45] produced

oxidized germanium microcrystals by gas condensation and characterized by TEM and electron diffraction. Very small clusters show Raman spectra and electron diffraction typical of amorphous systems; a transition to crystalline structure is observed for relatively large sizes. The effect of surface should play an important role in determining the vibrational spectra.

In the case of colloidal suspensions of Ge nanocrystals, Raman spectra showed broadened bulk Ge features with virtually no dependence of the peak position on the particle dimensions. In order to reconcile the prediction of the phonon confinement model with TEM observations, Heath et al. [60] used a bulk homogeneous line width of the bulk peak of ≈1.85 cm^{-1} and proposed a very small amount of lattice contraction as the origin for the absence of frequency shift.

Raman spectroscopy of germanium nanoparticles has been performed in conjuction with TEM characterization and optical absorption measurements [46]. The samples were grown by an evaporation–condensation self-organization technique that allows obtaining particles with a truncated sphere shape, a size dispersion of $\sigma/R \le 20\%$ (R is the average radius), and with the possibility, in addition, of controlling their average radii in a range from about 1 up to 10 nm [186]. Samples are prepared by deposition of Ge vapor on amorphous Al_2O_3 substrates. The partially wetting character of the substrate and the high-temperature deposition allow the formation of particles with regular sizes and the minimization of interaction with the matrix and reduction of stresses.

The particle size distribution $\rho(R)$ has been obtained by TEM analysis and could be introduced to fit the Raman spectra. A deviation from the theory has been observed for the lowest size values. This can be due to a strong influence of the surface (dangling bonds and bond angle disorder [48, 49]); instrumental correction of TEM observations, needed to obtain the size distribution, tends to underestimate the percentage of larger particles with a high Raman cross section in favor of the more numerous smaller particles with a low cross section.

The relevance of confinement effects in the observed Raman spectra is confirmed by optical spectra which show a good crystallinity of particles, suggesting that matrix effects can be neglected. The coupling of different techniques allows us to distinguish different contributions affecting the vibrational spectra of small particles [46].

5.3.2. Silicon Nanoparticles

Technological interest has stimulated the study of silicon micro- and nanocrystals showing electronic and vibrational confinement [44, 63, 64]. More recently, a renewed interest has been motivated by the characterization of porous silicon, which promoted the use of Raman scattering as a technique for the determination of the size distribution of nanoparticles considered as building blocks of porous mesostructures.

Raman spectroscopy has been also used to characterize different silicon-based systems showing different degrees of order and crystallinity and ranging from amorphous silicon to polycrystalline and microcrystalline silicon obtained by laser annealing or ion implantation [44]. The Raman spectrum of crystalline silicon is characterized by a Γ_{25}^{+} optical phonon at 521 cm^{-1} with a natural width of ≈3.5 cm^{-1} [67]; in amorphous silicon the $\mathbf{q} \simeq 0$ selection rule does not apply and the Raman spectrum resembles the phonon density of states with a broad bump centered around 480 cm^{-1}. Samples characterized by incomplete or incipient crystallization show a Raman spectrum which is somehow intermediate between the two extreme cases. A broadening and red shift of the 521-cm^{-1} peak is observed together with the appearance of a broad feature around 480 cm^{-1} [63].

Microcrystalline silicon films, characterized by grain size of the order of a few tens of nanometers, show similar behavior, with the line shape and red shift of the 521-cm^{-1} line depending on the crystallite size.

Several authors interpreted the observed features in terms of a relaxation of the $\mathbf{q} \simeq 0$ vector selection rule and of the presence of confinement effects [44, 63, 64]. As in the case of germanium many experiments have shown that for microcrystalline samples a broadening and a red shift of the 521-cm^{-1} Raman peak is observed and those features that depend on the crystalline domain dimensions. Fauchet and Campbell [44] showed that phonon confinement theory can successfully be used to explain the Raman spectra of small silicon crystals and their dependence on crystalline size and shape. Considering spherical microcrystallites, Campbell and Fauchet have investigated the influence of the choice of the confinement factor $W(\mathbf{r}, L)$, the anti-Fourier transform of a generalization of the $F(\mathbf{Q})$ factor in Eq. (30) $W(\mathbf{r}, L) = F^{-1}\{F(\mathbf{Q})\}$ on the quality of the fit between theory and experiments. They have considered three types of functions:

$$W(\mathbf{r}, L) = \frac{\sin(2\pi r/L)}{2\pi r/L} \tag{32}$$

$$W(\mathbf{r}, L) = \exp(-4\pi^2 r/L) \tag{33}$$

$$W(\mathbf{r}, L) = \exp(-8\pi^2 r^2/L^2) \tag{34}$$

The first confinement function was chosen in analogy with the ground state of an electron confined to a hard sphere, the phonon amplitude being 0 at the boundary. Equation (33) can be written as $\exp(-\alpha r)$ and it has been chosen in analogy with a wave in a absorbing medium. Equation (34) can be written as $\exp(-\alpha r^2/L^2)$ (in agreement with [43, 63]). Equation (34) is found to lead to a good agreement with the experiments when the phonon is confined very strongly to the microcrystal. On this basis Raman spectroscopy can be applied for the characterization of silicon nanocrystalline systems.

A detailed study of silicon microcrystallites has been reported by Iqbal and Vepřek [64]. Microcrystalline silicon films have been prepared by chemical transport in a DC hydrogen plasma at low pressure; by varying the substrate temperature the dimensions of the crystalline domains have been varied. A shift and broadening of the Γ_{25}^{+} phonon peak together with the appearance of a second broad component around 480 cm^{-1} is observed. This component could be correlated with the feature observed in amorphous silicon. The origin of the amorphouslike feature has been related to a surfacelike mode involving Si–Si shearing motion of the Si surface groups [64]. Hayashi and Yamamoto [241] reported amorphouslike Raman spectra for silicon clusters smaller than 100 Å, in contrast with TEM measurements showing

that very small clusters present a crystalline structure. These discrepancies were tentatively attributed to surface effects: the amorphous contribution should be due to the outer disordered shell of the clusters surrounding a crystalline core.

It is known that below a crystallite size of ~3 nm silicon becomes thermodynamically unstable with respect to the amorphous phase [64]; it is thus reasonable to expect that when very small clusters are present they can contribute to the onset of the amorphouslike Raman spectrum. A size distribution extending toward small sizes may account for the observed features without invoking the presence of high-energy surface modes. An independent determination of the size distribution would improve the understanding of the different contributions to Raman spectra.

Raman scattering has been used to characterize SiO_2 films with embedded Si mesoscopic particles [241], very small clusters with diameters ranging from 10 to 1 nm have been produced by cosputtering and postannealing, as observed in germanium. The dimensions of the particles did not allow the use of TEM to determine in an independent way the size distribution. This prevented a quantitative determination of the size and of the structure of the clusters.

5.4. Nanomaterials in One Dimension: The World of Relevant Nanosized Molecular Sticks

A large class of organic chain molecules occurring in nature, or synthesized in the laboratory, develops anisotropic and directional interactions along the molecular backbone yielding systems whose equilibrium structure consists of long straight chains with a remarkable structural rigidity. On the whole, the intramolecular forces are larger than those which glue them together in the supramolecular clusters, which may reach, sometimes, a tridimensional crystalline order. Such world of one-dimensional nanosized objects cannot be neglected, especially since vibrational spectroscopy in general and Raman spectra in particular provide a tool for the understanding of their structural and dynamic properties which are directly connected with their extremely relevant functions in nature or in material science in general. As will be shown, phonon confinement does take place in these one-dimensional molecular sticks and can be nicely and relatively easily revealed by Raman spectroscopy, which becomes a unique probe of the structure and dynamics of these systems.

The world of chain molecules is very wide and includes (i) natural and synthetic linear polymers which need to be considered in this review since they are able to self-organize in crystalline lamellae with nanoscale thickness (Fig. 11 [65]) and (ii) long linear molecules which again self-organize in mesoscopic or crystalline lamellae [65]. The physics of the lamellar organization of such molecular nanosticks is still matter for extensive studies to which Raman scattering and other techniques (which probe their vibrational motions through infrared spectroscopy, inelastic neutron scattering, etc.) can substantially contribute.

A prototypical case of lamellar organization is that encountered when the process of self-organization of very long polymer chains takes place with the folding of the polymer chain into shorter segments (with lengths between 5 and 25 nm) and their tridimensional packing, which yields a large lamella (Fig. 11) (for a thorough discussion see Wunderlich [65]). The stacking of these lamellae, more or less free or bound to each other through tie molecules, is the basis for further self-assembling into various kinds of morphological architectures (Fig. 12) which determine the physicomechanical properties of any polymeric material. The important

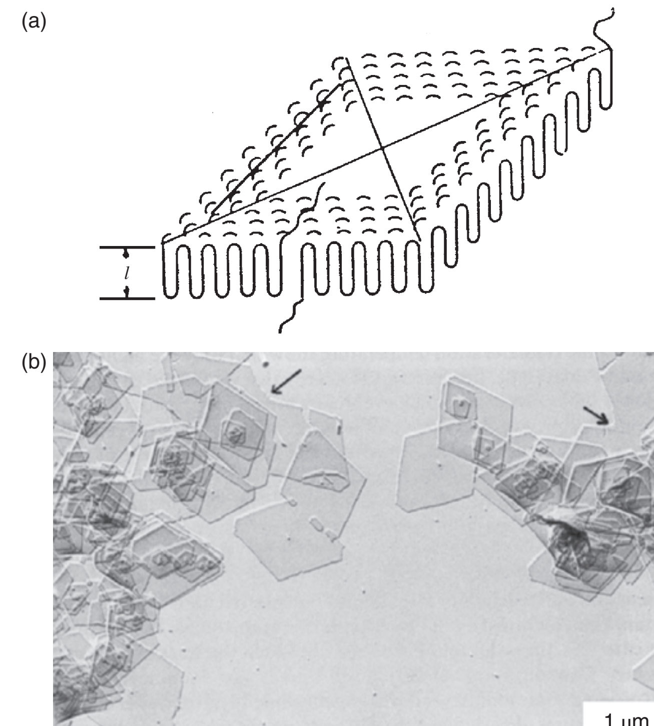


Figure 11. (a) Sketch of a polyethylene lamellar crystal the vertical lines represent segments of the molecular chains which fold at the surface. (b) Trasmission electron micrograph of single crystals (lamellae) of polyethylene, cystallized from a diluite solution.

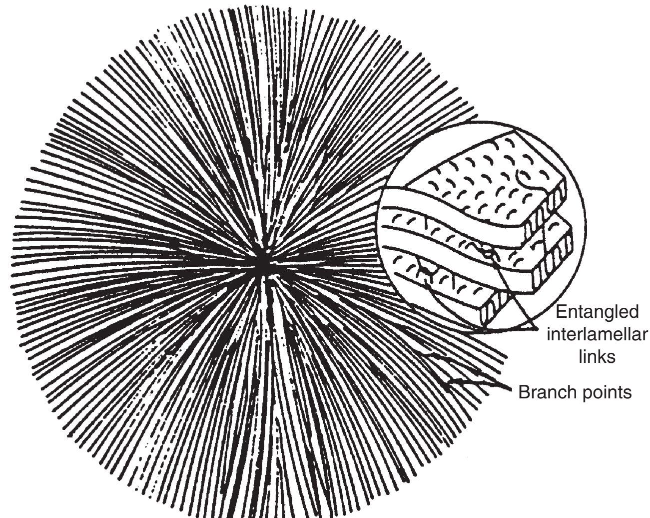


Figure 12. Polymer spherulite (grown from the melt). In the magnified section a stack of lamellae is shown.

concept in the physics of the organized soft matter consisting of regular polymer chains is the ubiquitous lamellar self-organization of molecules. The length of the segments, their topological distribution throughout the lamella (reached by suitable chain reentry as a consequence of some energetically plausible process of chain folding (see Fig. 11), and their local and/or collective mobility need to be precisely known).

The main concept underlying the treatment of the vibrational dynamics of segmented polymer chains is that the folding decouples the vibrations of each segment, thus allowing spectroscopic analysis of each segment as an isolated entity. The atoms forming the fold do perform their peculiar and characteristic vibrations (defect modes) which are fully localized in the fold itself [66] and need not considered or discussed in this review.

A second nonnegligible case of lamellar organization is found in nature when finite chain molecules of nanometric length self-organize in mesoscopic or crystalline lamellae, some of which are able to perform functions essential in the life of living organisms. The simplest case is the lamellar organization of the molecules of *n*-hydrocarbons with the general formula $CH_3{-}(CH_2)_n{-}CH_3$. In systems where the stacking of the lamellae has taken place small angle X-ray scattering experiments indicate the existence of long-range periodicities. When interchain packing is perfect wide angle scattering indicates short-range crystalline order inside the lamella, while the surface of the lamella may be ordered and/or disordered depending on the (temperature-dependent) molecular mobility of the chain. From spectroscopy [67, 68], STM experiments [69, 70], and numerical simulations by molecular mechanics [71] it has been shown that on the molecular scale the process of melting of such systems starts first with the generation of a surface disorder (surface melting) which proceeds into the lamella until the binding interchain forces of the ordered lamella have to yield to the incoming geometrical disorder (Fig. 13 [67]).

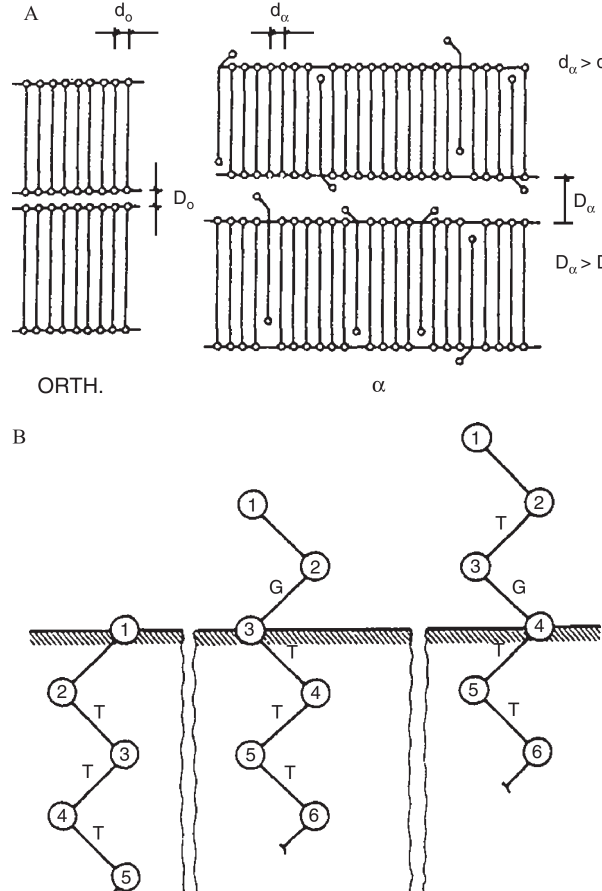


Figure 13. Sketch of the mechanism of phase transition of *n*-dodecane. (A) Schematic overall view; (B) detail on a molecular level.

The problem of order/disorder in conformationally flexible molecules self-organized into nanosized lamellae is basic in biological systems which perform essential functions in living organisms. First is the case of the intermolecular lamellar organization of long chain fatty acids or alcohols with the general formula $CH_3{-}(CH_2)_n{-}X$, where $X = -COOH$ or $-CH_2OH$. These molecules can also double their length because they can form dimers through the hydrogen bonding of the carboxylic or hydroxylic heads (e.g., $CH_3{-}(CH_2)_n{-}COOH \cdots HOOC{-}(CH_2)_n{-}CH_3$, Fig. 14).

If these kinds of molecules are tied as esters to phosphoric acids and derivatives the class of phospholipids is generated; such systems are universally found in all living cells and play a role essential in life, thanks to their hydrophilic or hydrophobic properties. The thickness of the membrane, the relative concentrations of straight sticks vs disordered matter (with a well-defined geometry) as function of temperature, ionic strength, attack by an external agent (e.g., a molecule of antibiotics), etc., are basic questions which can be answered by Raman spectroscopy studies of the phonons confined in such nanosized one-dimensional systems organized as lamellae (Fig. 15, [72]).

The list of other cases of nanosticks of relevance in the basic and applied science of nanomaterials could become longer (Langmuir–Blodgett films, monomolecular films deposited onto metallic surfaces relevant in microelectronics, soaps, lubricants, detergents), thus showing the relevance of the physics of nanosized soft matter at the molecular level studied through vibrational Raman spectroscopy.

In this chapter we will summarize the background concepts which may be applied and generalized to many other cases.

Let us first consider the phonon dispersion curves of a linear chain of polyethylene in its perfect conformational

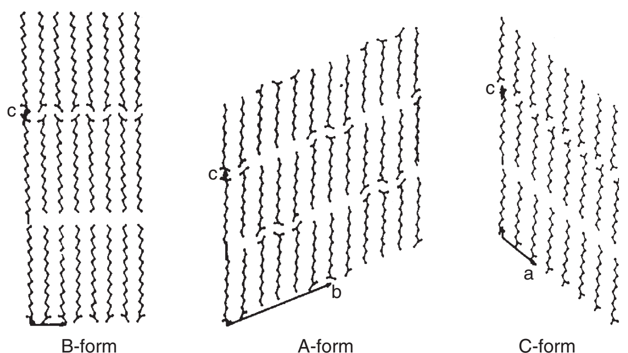


Figure 14. (A–C) Crystal form of fatty acids with an even number of carbon atoms.

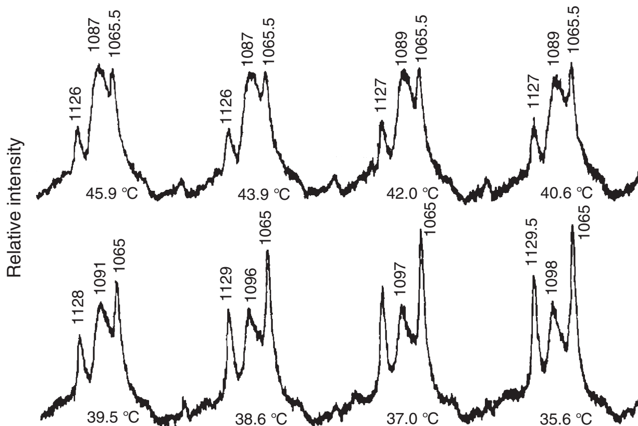


Figure 15. Alteration in the intensities of the bands attributed to the skeletal C–C stretching modes as a function of the temperature for a multilamellar dispersion of dipalmitoyl phosphatidylcholines (chain of 16 C atoms). Reprinted with permission from [72], I. W. Levin, "Advances in Infrared and Raman Spectroscopy" (1984). © 1984, Wiley.

structure (i.e., with a geometry corresponding to its absolute energy minimum) considered, so far, in the vacuum. The translational repeat unit of polyethylene as a 1D crystal consists of $N = 6$ masses associated with two equivalent chemical units (two CH_2 groups). The dynamics, group theory analysis, and optical selection rules can be and have been worked out in detail. We expect $3N$ phonon branches in the dispersion relation given by

$$M^{-1}(q)F(q)L(q) = L(q)\Lambda(q) \tag{35}$$

which represents the q-dependent (18×18) dynamic problem in terms of Cartesian coordinates [73, 74], where the matrix M^{-1} is the inverse masses tensor, $F(q)$ is the q-dependent tensor containing Cartesian harmonic force constants, $L(q)$ is the matrix made by the q-dependent Cartesian eigenvectors, and $\Lambda(q)$ the diagonal matrix of the frequency parameters $\lambda^s(q) = [(\nu^s(q))^2/4\pi^2c^2]$; the index s labels the sth phonon dispersion curve $\nu^s(q)$.

Since the vibrational harmonic force field for a chain of sp^3 carbon atoms (valence force field) is well known, highly reliable phonon dispersion curves are available and are routinely used (Fig. 16 [75]). It should be pointed out that consideration of a tridimensional crystalline system leaves the intramolecular potential practically unaltered, while it turns on weak intermolecular van der Waals type interactions which act only as small and often negligible perturbations of the phonons of the isolated chain (Fig. 17 [76]). This concept, which has been verified from both experiments and calculations for very many molecular systems, can be extended to all the "mesosticks" we are considering in this review and our discussion proceeds always referring to "single chains" with the only proviso that since for all polymer chains the single chain model neglects interchain forces it cannot describe the rotational (libration) motion about the chain axis, that is, the chain rotation about its axis is free. It follows that (i) while for tridimensional crystals we have the three acoustic branches (with a finite slope at $q = 0$ in the Brillouin zone), the single chain model adds an extra zero

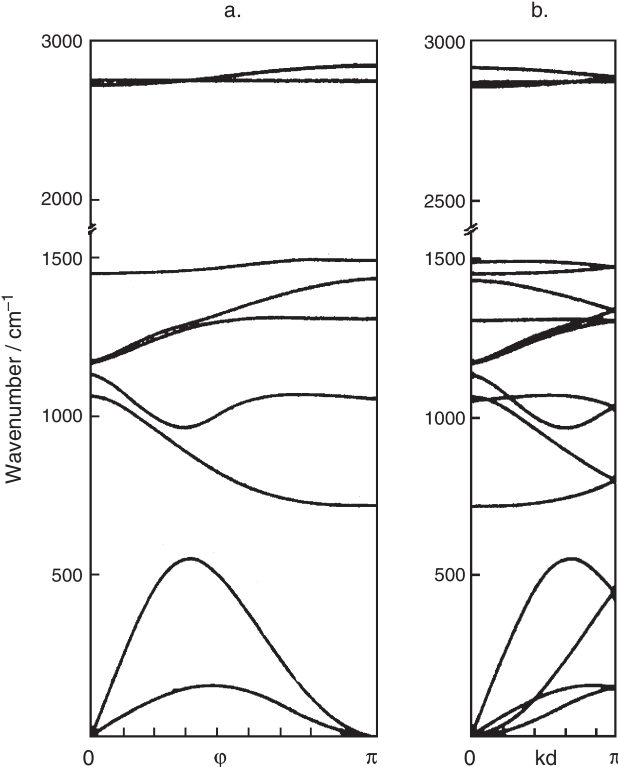


Figure 16. Phonon dispersion curves of a single polyethylene chain.

frequency (free rotation); (ii) the slope at $q = 0$ of the acoustic branches corresponding to the two "transversal" modes are wrong and the derived singularities near $q = 0$ in the density of vibrational states have no physical meaning. On the contrary, the longitudinal acoustic branch is fully reliable and gives the correct description of the low-frequency longitudinal accordion modes observed in the Raman spectra of nanosticks (see below).

A great contribution to the study of the physics of classical polymers and of the derived shorter nanoscopic chain segments was given as early as 1963 by Snyder and Shachtschneider (S, S) [77], who studied in detail the dynamics of polyethylene and of many shorter nanosized molecular models. If N is the number of atoms in the translational repeat units of polyethylene $3N$ are the branches in the phonon dispersion relations and $3N - 4$ are the nonzero phonon frequencies at Γ in terms of the single chain model.

For the sake of conciseness let us consider the longitudinal vibrations of a linear chain of n masses, each at a distance d, joined by weightless elastic springs; its standing waves can be characterized by the phase shift between vibrational amplitudes of adjacent units.

$$\varphi_j = q_j d \tag{36}$$

$$\varphi_j = \frac{\pi j}{n} \quad j = 0, 1, 2, \ldots (n-1) \tag{37}$$

$$\varphi_j = \frac{\pi j}{n+1} \quad j = 1, 2, \ldots n \tag{38}$$

where Eqs. (37) and (38) refer to a chain with free or fixed ends, respectively. Since in the case of polymers the

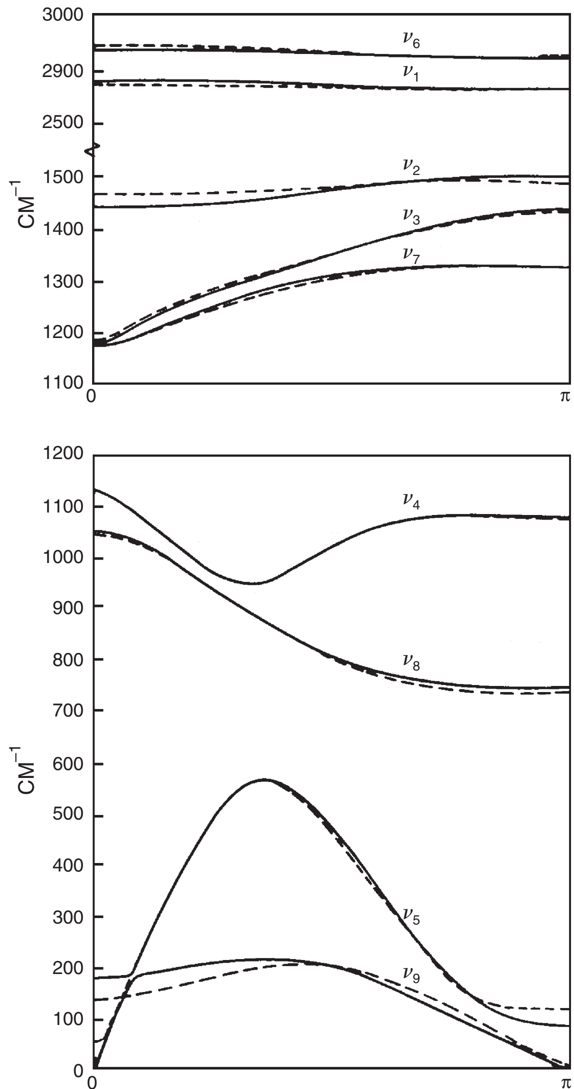


Figure 17. Phonon dispersion curves for crystalline (3D) polyethylene.

translational repeat unit is formed by several smaller chemical units related by a symmetry operation (e.g., through a screw axis if the polymer forms a helix), phonons are usually described with the phase shift between chemical units. For instance, in the case of polyethylene φ describes the amplitude shift between two adjacent CH_2 units. S,S showed experimentally (and later by dynamical calculations) that the vibrations of long chain hydrocarbons, tending to and reaching polyethylene, could be labeled in terms of the phase shift φ. Band progressions are observed mostly in infrared, with increasing intensity toward the $q = 0$ phonons of the infinite 1D crystal. For long chains, it can be shown that the finiteness of the chain does not give enough intensity to most of the $q \neq 0$ phonons to be observed in the Raman spectra and only $q \approx 0$ phonons are observed for most of the model molecules; the only exception are the phonons in the band progression for the longitudinal acoustic branch where all $q \neq 0$ gain remarkable intensities, with the lowest frequency components being the strongest line in the Raman spectra of each model molecule. From the spectroscopically observed (in infrared (IR) and or Raman) band progressions, experimental phonon dispersion curves were constructed. Only very recently has the new technology in inelastic neutron scattering [78] experiments the reaching of spectroscopic results similar to those obtained much earlier by S,S.

The dynamics of nanosized molecular sticks cannot avoid consideration of the so-called "end group modes," which derive by the fact that chain ends are different (in terms of masses and of force constants) from the repeat chemical units which make up the finite molecular entity. The number of end modes (not their location in the spectrum) can be predicted by Ledermann's theorem, which states that if in a Hermitian matrix the elements of r rows and their corresponding columns are modified in any way whatever, the number of eigenvalues which lie in any given interval cannot increase or decrease by more than $2r$ [79]. The locations of the end modes in the class of n-alkanes studied by S,S were analytically predicted with the method of the "transfer matrix" [80]. The present availability of powerful computers and of suitable algorithms for the construction of the dynamic matrices for each nanosized molecular stick of any chemical nature provides a quick and easy way to calculate and to describe the nature of all normal modes, including the end group modes. It has to be remarked that for chains made up of sp^3 carbon atoms the interactions between neighboring chemical units are very short range, reaching at most the second neighbor. It follows that for long nanosized sticks even if the size of the secular determinant to be solved becomes enormous the matrix is a typical band matrix consisting of a diagonal and a few codiagonals, while all other elements are 0 (the band width increases with the increasing of the molecular complexity of the chain). The numerical solution of the dynamic problem becomes extremely fast even for huge systems [74].

From the experimental viewpoint, in the case of chain segments with increasing length where the force field is not modulated by chain length (i.e., the numerical value of r is very small because long-range electronic interactions do not take place) end group modes are few and can be easily located in the spectrum because their relative intensity decreases when the number of repeat units in the chain increases.

The easy and unquestionable observation in the Raman spectra of finite molecular chains of phonons which lie on the longitudinal acoustic branch of the corresponding infinite polymer chain opens the way to further relevant and unique characterizations. Such vibrational modes are commonly referred to as "longitudinal accordion modes" (LAM) [81, 82] and correspond to stationary vibrational waves with one or more nodes along the chain backbone. For n-alkane molecules (n even), if j is the number of nodes, because of symmetry, only modes with odd j are Raman active and the Raman intensity strongly decreases when j increases. LAM-1 has received most of the attention since it gives rise to the strongest line in the Raman spectrum, thus becoming a very useful spectroscopic probe (Fig. 18). Mizushima and Shimanouchi [81] have shown that if a segment of a trans-planar polymethylene molecule is represented by an elastic rod of length l, density ρ, and Young modulus E, the LAM-1

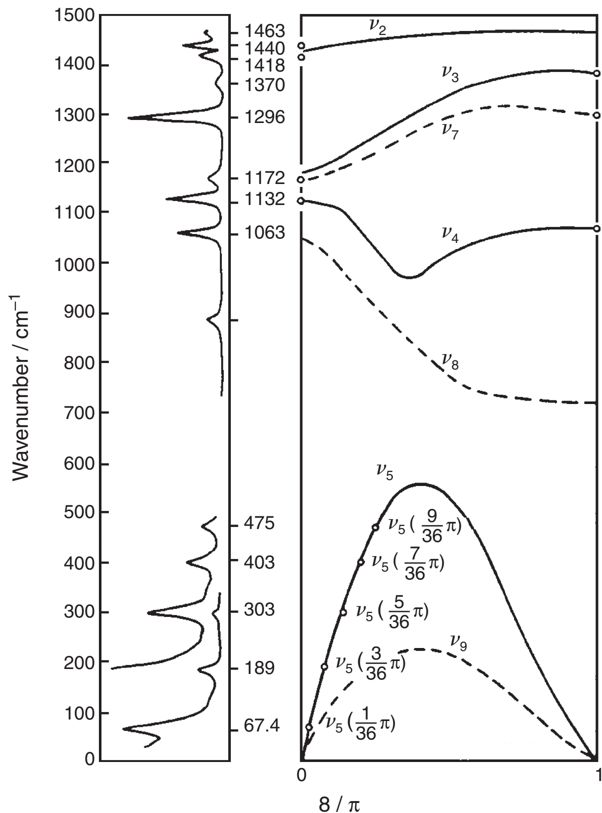


Figure 18. Frequency versus phase difference curves for acoustic (LAM) vibrations of a polymethylene chain with the corresponding observed Raman.

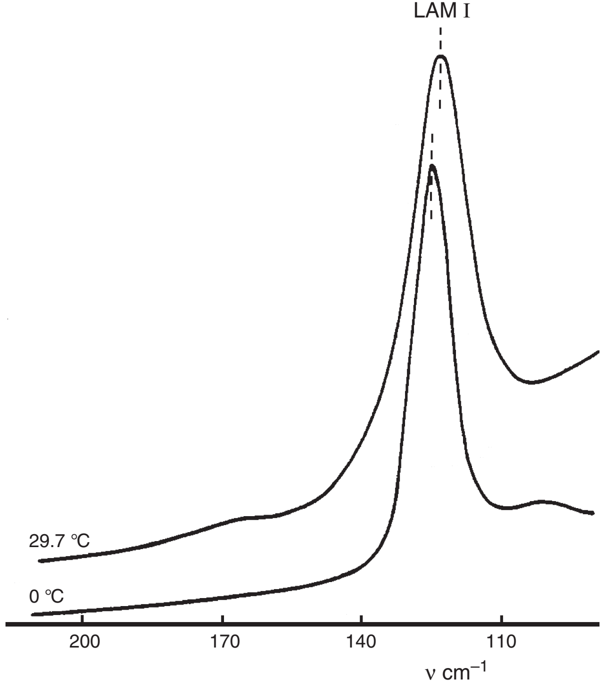


Figure 19. LAM band from the Raman spectrum of n-nonadecane at 0±C (orthorombic) and 29.7°C (α phase). Reprinted with permission from [67], G. Zerbi et al., *J. Chem. Phys.* 75, 3175 (1981). © 1981, American Institute of Physics.

frequency [cm^{-1}] can be related to l by

$$\nu(LAM1) = \frac{1}{2lc}\sqrt{\frac{E}{\rho}} \tag{39}$$

Once the numerical values of E and ρ are known for a given polymer the length l of the nanosized molecular sticks (which are chemically similar to that of the parent polymer) can be immediately derived from the observation in the Raman of their ν(LAM-1). On the other hand, if a sequence of several LAM-j modes can be experimentally observed for a given finite chain, part of the acoustical branch can be fitted on the experimental data and the elastic modulus of the polymer can be derived.

The same concepts form the basis of a recent study [83], in which the frequency behavior of the transverse acoustic modes (TAM) of polyacetylene has been predicted with a simple, classical model. In this model the molecular chain is represented by a continuous elastic beam with a flexural rigidity value extrapolated from *ab initio* computed TAM frequencies of polyenes of increasing length (ranging from C_4H_6 to $C_{22}H_{24}$).

LAM modes have been used in structural determinations of many nanosized molecular segments. A few interesting cases are quoted here as meaningful examples.

From what has been stated above the existence of straight nanosized sticks in a given material is relevant information in soft matter physics. Figure 19 shows the contribution of LAM spectroscopy to the study of phase transition (orthorhombic → pseudo-hexagonal → melt) of n-nondecane (taken as a prototype of a class of molecules) [67]. At $T < 0$ °C ν(LAM-1) = 124.5 cm^{-1} assures that n-nonadecane has 19 carbon atoms in a transplanar geometry, near $T = 27.78$ °C ν(LAM-1) = 123.2 cm^{-1} the chain is still straight, but the weak surface interactions at either end of the chain (interlamellar forces) further weaken, at $T > 32$ °C the nanostick collapses in a "disordered" structure and LAM-1 disappears, leaving a broad scattering near 220 cm^{-1}.

The nature of such broad scattering (observed in practically all systems consisting of long CH_2 chains in the liquid phase) has been the a center of attention for quite a while, suggesting that the material in the liquid phase may not consist of fully structurally collapsed and randomly coiled chains, but a nonnegligible fraction may consist of a distribution of short, but still straight segments. Statistical thermodynamics has been strongly involved in such studies. The observed broad scattering may result from the convolution of the LAM-1 transitions of short and "stiff" hydrocarbon segments. This view has refueled the relevant issue of the local order in amorphous polymers or systems with long alkyl residues in terms of short- or long-range reorientational correlations which may determine the process of crystallization. A contribution to these studies again comes from Raman LAM spectroscopy of short hydrocarbon chains, which provide evidence of short straight segments in liquid hydrocarbons as long as C_{12} [84, 85] thus supporting the idea that the model of nematiclike structure for liquid n-alkane melts is not applicable, while indications are collected

from LAM of the existence of weak orientational correlations characterized by a short correlation length and/or of a pretransitional ordering of very short segments as seeds of crystallization [85], in agreement with numerical simulations [71].

LAM has also contributed greatly to the study of the folding of polymer chains during the formation of lamellae. As early as 1978 Snyder and Scherer [86] were able to modify the existing Raman spectrometers in order to reach extremely small Raman shifts (reaching 3 cm^{-1}) for the recording of LAM modes of specimens of polyethylene as solid state extruded or as bulk-crystallized materials. From the shape of the observed LAM-1 scattering the authors were able to obtain a quantitative distribution of lengths of straight chain segments associated with polymer lamellae (Fig. 20).

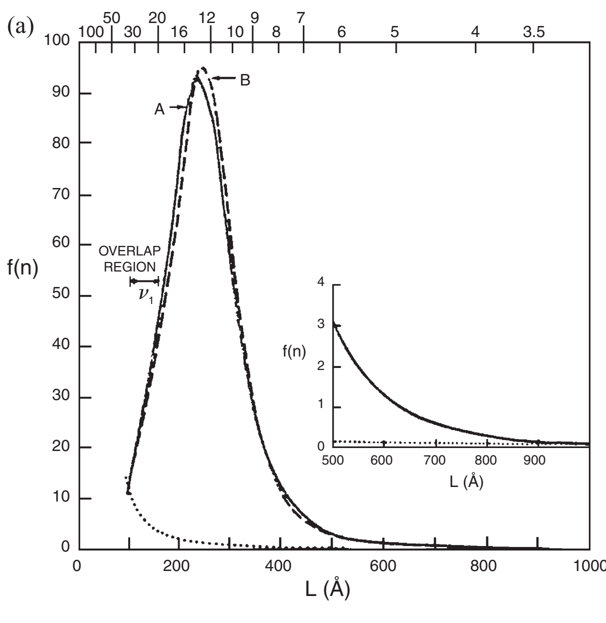
The use of Raman LAM spectroscopy for tackling structural and dynamic problems of complex biologically relevant nanosized molecular sticks has required the development of a simplified theoretical approach. Indeed it has been observed experimentally that heavy substituents at either end of chain molecules (e.g., monolayer comblike systems biochemically very common), weak or strong interactions between molecular chains facing each others in finite systems (e.g., hydrogen-bonded heads in fatty acid dimers), and interlamellar forces in crystals critically affect both LAM frequencies and intensities. A simplified model of coupled oscillators has been developed and applied to many cases offered by nature [87]. The theoretical results drastically modify the way experimental LAM frequencies and intensities have to be interpreted. The first direct consequence of such treatment is that the previous belief (based on a wrong interpretation of LAM modes) that the chains of fatty acids were not planar, but conformationally distorted had to be dismantled since evidence of fatty acid dimers as straight sticks has been unambiguously offered [87]. Finally, the coupled oscillator model has been extended to "composite" nanosized chain molecules, that is, linear chains made up of segments of different chemical natures (hence with different mechanical and optical properties) linked together. A classical case is that of semiflouorinated *n*-alkanes of the general formula $F(CF_2)_n - (CH_2)_mH$, with $m = n$ in the range 7 to 12 and with $n = 12$ and m raging from 1 to 20, produced and experimentally studied by Twieg and Rabolt [88] and interpreted by Minoni and Zerbi [89].

After the first synthesis of polyacetylene [90] (in this discussion we always refer only to *trans*-polyacetylene, hereafter indicated as PA) and after the discovery that it can be transformed in a highly conducting material by suitable doping [91] the interest of pure and applied chemistry, physics, and technology has grown dramatically and has branched out in many directions, including molecular electronics and photonics [92]. Actually, because of its poor chemical stability PA has been set aside and replaced by chemically more stable and workable oligomers and polymers of pyrrole, thiophene, paraphenylene vinylene, and many others.

The physical phenomenon which forms the basis of all these studies is that PA (and the many derivatives and homologues which have been produced) is a prototypical polymer molecule made up of carbon atoms in sp^2 hybridization. The p_z orbitals are delocalized along the molecular backbone, thus forming again a one-dimensional crystal. In the previous sections we have stressed that all polymers or short molecules considered were made up of carbon atoms in sp^3 hybridization, with the consequence that intramolecular interactions are short range and intermolecular forces are weak and almost negligible. In the case of polyconjugated polymers and short oligomers, because of delocalization, the HOMO–LUMO bandgap decreases substantially (thus changing colour of the material), the electronic interactions turn on the so-called Peierls distortion, strong electron/phonon coupling takes place, and intramolecular forces may extend long distances.

It is beyond the scope of this chapter to review the chemistry and physics associated with the existence of delocalized π electrons in polyconjugated polymers, again taken as 1D

(a)



(b)

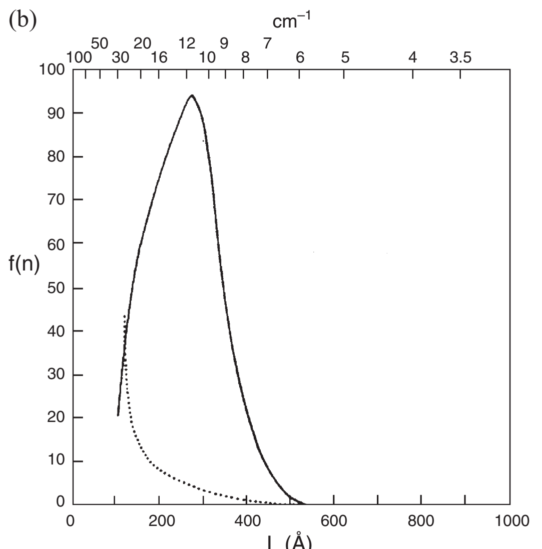


Figure 20. Distribution of extended segment lengths (a) for extruded polyethylene and (b) for annealed (1 h at 130°C) bulk crystallized polyethylene. The dotted curves represent the estimated error in f(n). Reprinted with permission from [86], R. G. Snyder and J. R. Scherer, *J. Polymer Sci.* 16, 1593 (1978). © 1978, Wiley Interscience.

crystals and in short molecular segment (oligomers) where both phonons and electrons become localized [16, 93]. As a conceptual continuation of the discussion started with chains and nanosticks made by sp^3 carbon atoms we feel obliged to point out that in polyconjugated systems with sp^2 carbon atoms all physical properties are conjugation length (CL) dependent. Since our present interest is Raman scattering of nanosized molecular chains we feel it necessary to focus on a few properties of the Raman spectra of these materials.

First is the observation that the Raman spectrum of a sample of PA is extremely simple, and similarly the spectra of nanosized polyene chains (ranging from polyenes to carotenoids etc.) show almost the same spectral pattern of two strong bands which shift by changing CL. It becomes obvious that if the electronic properties of these systems are CL dependent the molecular force field also changes with CL together with the derived dynamic and spectroscopic properties. We then expect a frequency dispersion and intensity dispersion of the Raman spectrum in terms of CL, as indeed is observed in various samples of PA and in very many nanosized chains of molecular models (see Figure 2).

Of particular interest is the observation for several polyconjugated materials of an extremely strong Raman transition (showing frequency and intensity dispersion with CL) in the range of 1600–1450 cm^{-1} which has been accounted for in terms of a collective totally symmetric vibrational mode which involves the in-phase stretching of all C=C bonds and the in-phase shrinking of all C—C bonds. These modes have been variously described in terms of the “dimerization coordinate” or the “effective conjugation coordinate” and form the basis for understanding the electronic properties and dynamics of the whole class of polyconjugated linear chains oligomers and polymers [16]. Since the mode probes the whole molecular chain it can be taken as a characteristic signal of each nanosized segment of a certain length.

Using these concepts, from Raman spectra of polyacetylene samples, bimodal or trimodal distribution of segment lengths has been derived [95].

5.5. Carbon-Based Nanostructured Materials

There are very many materials made of only carbon atoms, which show relevant deviations (from the viewpoint of structure and properties) from the two main allotropic forms of carbon, namely, graphite and diamond. Among carbon nanomaterials we find molecular objects with a well-defined structure and point symmetry (this is the case of fullerenes and other carbon clusters) and materials which can still be described starting from some parent structure (graphite or diamond), by the introduction of structural defects (for instance, the loose of the three-dimensional order as in turbostratic graphites, or the drastic lowering of the crystals size, as in micro- and nanocrystalline graphites). Sometimes the structural similarity with a parent, perfect crystal is restricted to very a short-range order, as in the case of amorphous carbons: here graphitic islands (characterized by a finite number of condensed hexagonal carbon rings) develop and increase their size (the characteristic diameters range from of a few to 10^1 nm) upon thermal treatment.

A very special case is that of carbon nanotubes, which can be regarded as nanoobjects but which obey a physics which can be written starting from concepts and calculations already developed for the case of a 2D graphite lattice.

5.5.1. Disordered Graphites

Under this name we can group a very wide family of materials (ranging from amorphous carbons, microcrystalline graphites, and graphite fibers to whiskers). All these materials are characterized by the presence of a nonnegligible number of sp^2 carbons which give rise to islands which can be correlated (in terms of structure and also in terms of their Raman response) to an ideally perfect crystal of graphite.

The Raman spectra of many carbon materials containing a variable number of sp^2 structures have been extensively studied [96–102, 155, 187]. Their first order spectra show two main signatures: first, a band located at about 1580 cm^{-1}, which is known as the graphite (G) band, since it is the only intense feature observed in the first order Raman spectrum of highly ordered, crystalline graphite. The second band (D band) appears at about 1350 cm^{-1} and is a characteristic feature of disordered samples of graphite (e.g., nanocrystalline and microcrystalline graphites) and amorphous carbon materials containing sp^2 graphitic islands.

In Figure 21 the first order Raman spectrum of a variety of disordered graphitic samples is shown. As can be immediately observed the different samples show

- Relatively modest frequency shift in the two main lines.
- A wide range of values of the intensity ratio between the two main lines. The ratio I_D/I_G ranges from the theoretical value for a perfect graphite crystal ($I_D/I_G = 0$) to values $I_D/I_G > 1$.
- Band shapes and especially bandwidths show sample dependence.

The fundamental experiments made by Póksic et al. [97] on microcrystalline graphites and those performed by Ferrari and Robertson [98] and Robertson and O’Reilly [99] on films of amorphous carbons showed a clear resonance behavior and frequency dispersion of the D line with respect

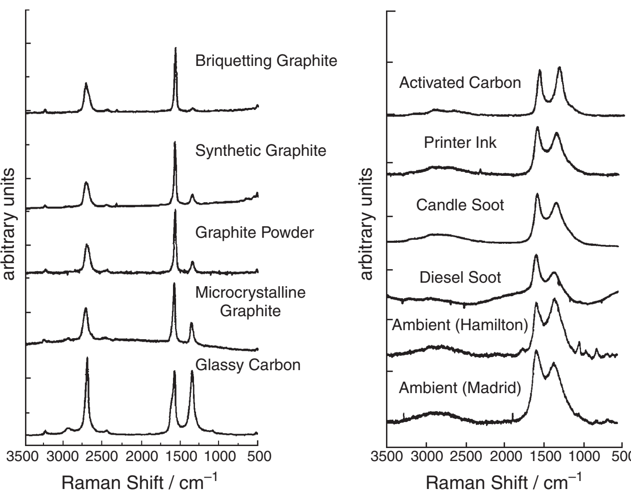


Figure 21. (Left) Raman spectra of commercial samples of graphitic materials. (Right) Raman spectra of “disordered” samples of various origins. Reprinted with permission from [101], R. Escribano et al., *Vibrational Spectrosc.* 26, 179 (2001). © 2001,

to the energy of the exciting laser line used in the Raman experiments. This behavior is illustrated in Figure 22.

Other peculiar features can be identified in Raman spectra of graphites, namely, that in several samples a second component in the high-frequency side of G band (referred as the G′ band) arises and several features ascribed to second order Raman spectra (overtones or combination bands) can be detected. All these features show a more or less marked sample dependence.

In what follows we restrict ourselves to the main features in the first order Raman spectrum of graphitic materials, namely, to the G and D lines.

The presence of the G band can be immediately related to the existence of structures like graphite in the samples. It corresponds to the only symmetry allowed phonon transition in the Raman spectrum of a perfect two-dimensional graphite lattice (the E_{2g} phonon at Γ point in the first Brillouin zone of graphite). This line is indeed the only one observed in the first order Raman spectrum of graphite crystals of high perfection (e.g., HOPG, highly oriented pyrolytic graphites). It is commonly accepted in the literature that the G line in disordered samples corresponds to the excitation of the same vibrational mode as in HOPG, confined into a finite size graphite domain. The small frequency shifts observed for this line in different samples can be ascribed to intrinsic changes in the force field (atom–atom interactions). Moreover, it can also be rationalized as due to the mere effect of the geometrical confinement which is well recognized in solid state physics, namely, the activation of phonons characterized by wave vector **q** different from **0**, in a narrow range around $\mathbf{q} = \mathbf{0}$ (see point 2, Section 3).

On the contrary, the understanding of the physical mechanism lying behind the activation of the D band has been, and is presently, a matter of debate in the literature.

A point is generally accepted, namely, the fact that the appearance of the D line is due to a transition corresponding to a $\mathbf{q} \neq \mathbf{0}$ graphite phonon which is activated by the presence of "disorder" in the sample. Independently of the kind of disorder, the measure of the intensity of the D line has been used by several authors to obtain an estimate of the degree of disorder. An empirical linear correlation

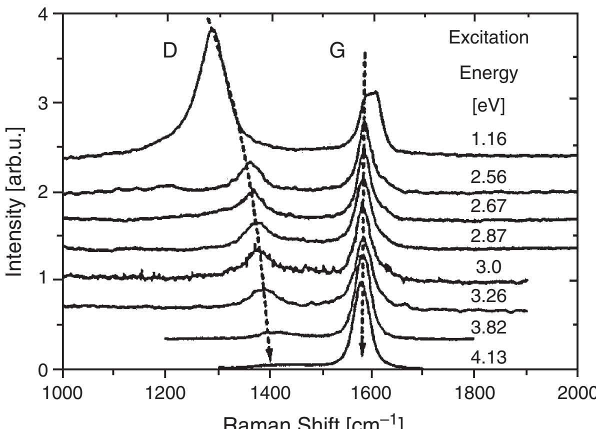


Figure 22. Raman spectra of a microcrystalline graphite sample, obtained with different exciting laser energies. Reprinted with permission from [97], Póksic et al., *J. Non-Cryst. Solids* 227, 1083 (1998). © 1998,

between the intensity ratio and the inverse of the crystal planar domain size L_a has been observed by Nemanich and Solin [187]. This correlation has been used for evaluating the average size of the graphite domains in both microcrystalline graphites and amorphous carbons. More recently, thanks to the systematic work of Ferrari et al. on several samples of amorphous carbons it was definitely shown that the I_D/I_G parameter does not follow a monotonic law as a function of L_a. It increases systematically in a regime were graphite domains of a few nanometers are present, thus reaching a maximum and than decreasing in samples with larger graphite domains (e.g., microcrystalline graphites). Moreover, it is now clear that the crystal size is not the only factor which determines the intensity of the D line: suitably induced defects (e.g., by ion irradiation [102]) have been shown to contribute to enhancement of the D line cross section. Other experiments reported by Wang et al. [103] showed that the D line intensity is independent of the crystalline size.

Despite the general agreement about the existence of a correlation between disorder and the D line, there are at least two fundamental problems which have to be clarified: (i) to explain the selectivity of the phenomenon and (ii) to give a description (e.g., in terms of structural parameters) of the nature of the defects responsible for activation.

The breaking of spectroscopic selection rules (by virtue of disorder) brings us (at least in principle) to the activation of the whole phonon density of states. What is surprising in this case is the fact that only one phonon (or a distribution of phonons in a very narrow range of energies) is activated by disorder. Moreover, the simpler explanation in terms of correspondence of the observed Raman frequency with a peak of the phonon density of states must be dismissed based on the observed frequency shift (Fig. 22) of the D line with excitation energy. Indeed a peak in VDOS lies definitely at a fixed frequency. All these observations are at the basis of the model proposed by Thompsen and Reich [104], which explains the appearance of the D line as due to a double resonant process involving (1) the creation of an electron–hole pair at the **k** point of the graphite BZ for a particular laser energy E_1 matching the energy difference between valence and conduction band at **k**; (2) the scattering of the electron to a state in a second band by a phonon with the proper phonon quasi-momentum; (3) the recombination of the electron assisted by elastic scattering from lattice defects.

Both processes 1 and 2 are affected by resonance behavior (double resonance). The model developed by Thompsen and Reich [104] has been analytically worked out, giving rise to equations which predict in a quantitative way, the experimentally observed dispersion relation of the D line frequency with respect to the energy of the exciting laser line (Fig. 22) [97].

An alternative interpretation of the origin of the D line in graphitic materials has been proposed [24, 26, 106] based on a molecular approach. This approach is built on the spectroscopic data obtained from a wide series of samples of PAHs. Very recently a new synthetic route has been developed [25, 107] and PAHs of very large size have become available. These large polycyclic aromatic hydrocarbons can be considered molecularly defined graphitic clusters, such as

those present in defected graphite samples and other carbon materials. Molecularly defined objects can indeed be investigated in order to provide explanations for some not yet well understood experimental behaviors observed for disordered carbon materials containing nanosized sp^2 domains.

Interestingly, the experimental spectra of PAHs [24, 26, 106] are also characterized by a few bands in approximately the same G and D regions (Fig. 23). This is expected, given the similarity of their structure to that of graphitic islands, and offers the advantage of using molecules with a well-known dimension and structure to probe defects of similar size and structure in carbonaceous materials.

To assess quantitatively the correlation experimentally found between the Raman bands of graphite and the Raman active modes of PAHs, we carried out a quantum–chemical study of vibrational force field and Raman intensities of PAHs of different sizes and symmetries [24, 26]. These calculations, performed at the BLYP/6-31G level of theory, confirmed the idea (earlier developed on the basis of classical vibrational dynamics calculations) that this correlation is due to the strong polarizability changes associated with two peculiar [182] collective vibrational displacements, characteristic of a graphitic cluster, namely, R^- and A vibrations.

The Raman active modes of PAHs in the 1600-cm^{-1} region correspond to vibrational eigenvectors with a large projection on the vibrational coordinate R^- and give rise to the G band. In a similar way the normal modes (usually two or three) giving rise to strong Raman transitions in the D band region show a large content of the vibrational coordinate A. In the case of a perfect two-dimensional crystal of graphite the vibrational coordinates R^- and A correspond to two phonons of the crystal: the only Raman active $\mathbf{q} = \mathbf{0}$ E_{2g} phonon and the totally symmetric A'_1 phonon at $\mathbf{q} = \mathbf{K}$ [26].

In Figure 24 the nuclear displacements associated with these phonons are shown and compared with the computed eigenvectors of a large PAH ($C_{114}H_{30}$). The eigenvectors displayed for $C_{114}H_{30}$ have been selected on the basis of their very large Raman cross sections.

The analysis in terms of local contributions to Raman intensities clarifies the origin of the large Raman cross section for the few bands at about 1300 cm^{-1}, observed (and calculated) for all the PAHs examined. It is shown that the values of the local stretching polarizability tensors ($\partial\alpha/\partial R_{cc} = \alpha'$ parameters) can be used as a very sensitive probe of the character of CC bonds in PAH molecules. More specifically, in the case of "all-benzenoid" PAHs, they allow easy identification of the presence of aromatic sextets. Moreover, the α' correlate with the equilibrium bond lengths that results from quantum–chemical calculations of equilibrium structure. This lies in the peculiar electronic structure of PAHs, which is induced by the confinement of

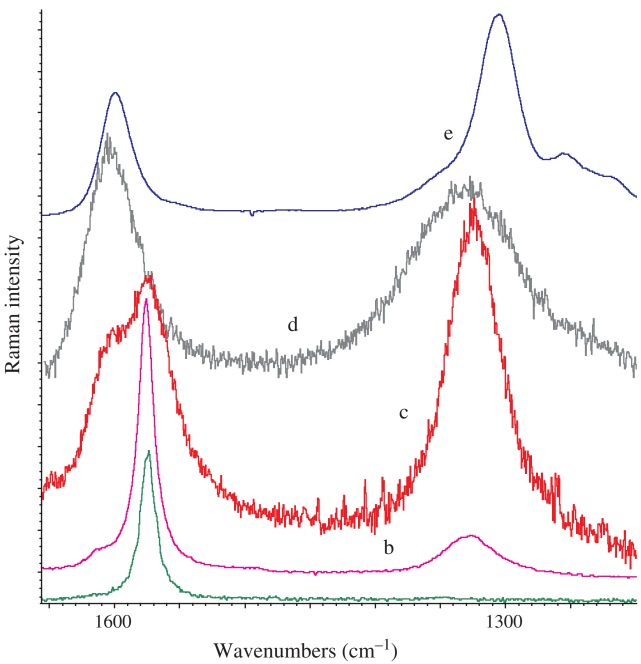


Figure 23. Raman spectra of different graphitic materials: (a) HOPG graphite ($\lambda_{exc} = 632.8$ nm); (b) disordered graphite ($\lambda_{exc} = 632.8$ nm); (c) multiwall nanotubes ($\lambda_{exc} = 632.8$ nm); (d) pyrolitic graphene sheet ($\lambda_{exc} = 632.8$ nm); (e) a D_{6h} PAH molecule containing 222 carbon atoms ($\lambda_{exc} = 632.8$ nm). Samples d and e were kindly provided by Professor K. Müllen [25].

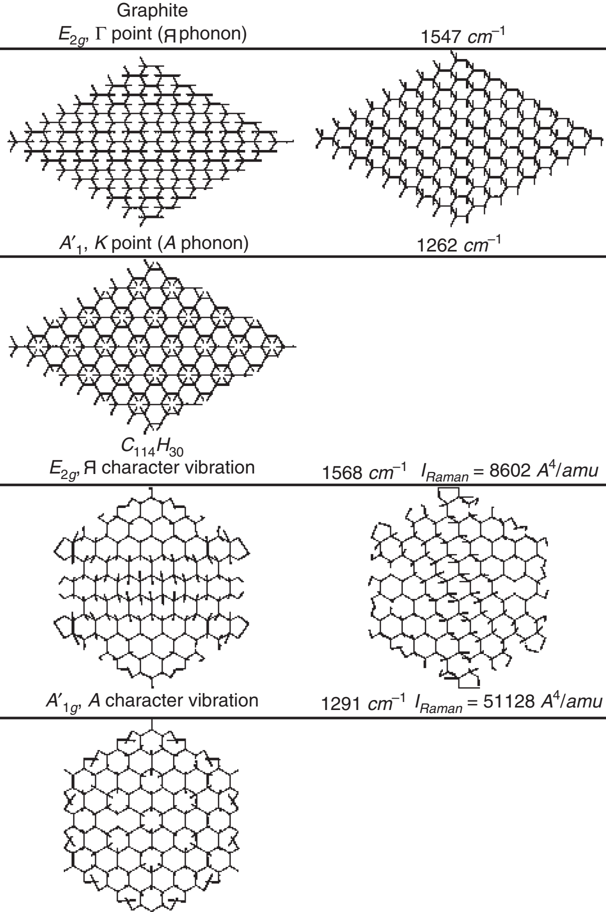


Figure 24. (Upper panels) Nuclear displacements associated to R^- and A phonons of a perfect two-dimensional lattice of graphite. (Lower panels) Comparison with selected normal modes of $C_{114}H_{30}$, as obtained from BLYP/6-31G calculations.

the π electron system into a finite size domain. This electronic structure is strongly affected by the molecular topology and it is reflected by the nuclear equilibrium geometry, characterized by nonequivalent CC bonds which can be described according to the canonical structures commented on in Section 4, point ii.

The relevant fact is that the existence of nonequivalent CC bonds (and, correspondingly, nonequivalent α' parameters) in PAHs is essential for the activation of normal modes in the D band region.

If these ideas are extended to the rationalization of the Raman spectra of carbon materials containing nanosized clusters of condensed aromatic rings, we have to conclude that in these cases the confinement of π electrons also induces a structural relaxation (like that described in Fig. 5), with respect to the structure of a perfect graphite crystal with all the CC bonds perfectly equivalent. The spectroscopic signature of this relaxation is just the appearance of the D line in the Raman spectrum. It must be noted that several years ago, Tuistra and Koenig [96], based on a simple symmetry consideration, proposed that the D line observed in disordered graphites was correlated to a totally symmetric phonon corresponding to the A mode sketched in Figure 24.

The success of the molecular approach consists in the fact that it makes it possible (1) to clearly identify the quasi-phonon involved in the Raman process as the A mode and (2) to describe the effect of disorder in terms of a well-defined structural relaxation, which is in turn correlated with effects of confinement of the π electrons. Effects of this kind can certainly be induced both by the presence of finite size graphitic domains and by the presence of edges, as in nanocrystalline graphites.

Moreover, as in the model of Thompsen and Reich [104], the molecular approach indicates that the appearance of the D line is a resonance phenomenon. It has indeed been shown (both experimentally and by means of theoretical calculations) that Raman intensity enhancement of the D band occurs in PAHs [27, 106].

In conclusion, we stress that the interpretation of the Raman spectra of materials containing graphitic domains of small size has been done with the help of models which, at first sight, may appear very different or even incompatible. However, both the models mentioned in this section make it possible to rationalize several experimental features and shine new light on the nature of nanostructured carbon materials. The challenge for future research is to find a bridge between the different approaches and to clarify the limitations, potentialities, and complementarities of the different points of view.

5.5.2. Carbon Nanotubes

Since their discovery [108] carbon nanotubes have attracted the interest of both theoretical and applied researchers due to their very peculiar physics and consequently to the many possible applications foreseen for these materials (e.g., in the fields of nanoelectronics, sensors, and energy storage).

Raman spectroscopy is a widely used technique for the structural diagnosis and study of electronic properties of carbon nanotubes. Among other things, vibrational spectra are useful for the identification in experimental samples of different nanotubes (characterized by different structures and properties). This is probably the reason for the large number of publications which have appeared in the literature in the last 10 years dealing with the prediction and interpretation of the vibrational spectra of carbon nanotubes. Due to the amount of literature on the subject we are forced to refer the reader interested to a deeper discussion to some selected reviews which contain a wide bibliography on the subject [109, 110].

A first classification of carbon nanotubes is based on the fact that they can be described as either made by a single (rolled) graphite sheet (single wall nanotubes, SWNT) or consisting of several concentric cylindrical graphite layers imbedded one inside the other (this configuration is traditionally called the "Russian-doll" model, and these nanotubes are referred as multiwall nanotubes, MWNT).

The interpretation of the Raman behavior of SWNT has now reached, thanks to the many experimental and theoretical works, a consistent set of reliable conclusions. On the contrary, the study of the spectroscopic behavior of MWNT presents much more complicated theoretical issues and has not yet reached the same degree of maturity. For this reason we deal in the following with just the case of SWNT.

The structure of a given SWNT can be specified in terms of the tube diameter and the chirality. These concepts are conveniently expressed in terms of a proper chiral vector $\mathbf{C}_h = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$ defined on the honeycomb lattice (see Fig. 25) and by a translation vector $\mathbf{T}$ which is in turn determined by the choice of the two integers, n_1, n_2. Accordingly, the pair (n_1, n_2) is the only information needed in order to fully specify the structure of the tube. The importance of the knowledge of nanotube structure consists mainly in the fact that it determines the electronic structure (and then the electrical behavior) of the tube: while varying (n_1, n_2) one can obtain a 1D metal or a semiconductor (and the energy gap value can be tuned by selecting (n_1, n_2)).

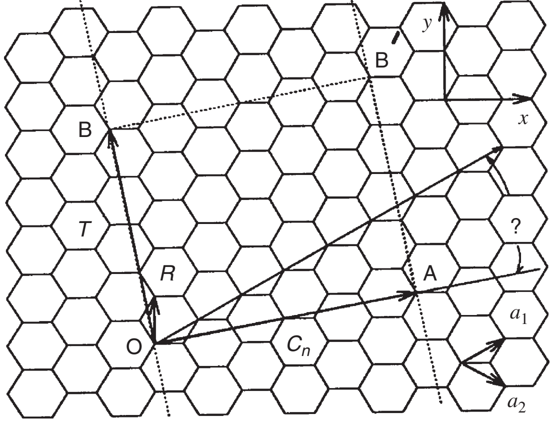


Figure 25. The unrolled honeycom lattice of a nanotube. Connecting sites O and A and sites B and B′ a nanotube can be constructed. C_h (the chiral vector) and T (translational vector) are indicated. The rectangle OAB′B defines the unit cell of a (4, 2) nanotube. Reprinted with permission from [109], R. Saito and H. Kataura, "Topics in Applied Physics, Vol. 8, Carbon Nanotubes: Synthesis, Structure, Properies and Applications" (2001). © 2001, Springer-Verlag.

Since the length of carbon nanotubes is usually between 10^2 and 10^3 nm, while diameters are of the order of 10 nm, the treatment of the electronic and vibrational properties of a single tube can be performed on the simplified hypothesis that they are infinite in the direction of the tube axis. This allows the avoidance of the explicit description of the nanotubes edges. However, as an example of a structurally well-defined nanotube end, in Figure 26, SWNTs are shown where each end of the cylinder is capped with half a fullerene.

Raman spectra of bundles of SWNT and, more recently, from isolated nanotubes [111] have been reported in the literature.

In Figure 27 [112] the first order Raman spectra recorded with different exciting laser lines for a purified SWNT sample are shown.

Despite the higher number of $\mathbf{q} = \mathbf{0}$ phonons of carbon nanotubes with respect to those of graphite, symmetry selection rules determine the activation of relatively few phonons in the Raman spectra.

The most important observed Raman features are the following: (1) the radial breathing mode (RBM), whose frequency varies according to the diameter of the tube; (2) the tangential G band (derived from the graphite $\mathbf{E}_{2g}$ Raman mode) in the range of 1550–1605 cm^{-1}; and (3) the "disordered induced" D band at about 1350 cm^{-1} and its second order harmonic (G′ band) at about 2700 cm^{-1}.

Another very relevant point is the fact that the Raman intensities of a given nanotube are strongly enhanced by resonance phenomena. Resonance effects are determined by matching between the laser energy and the energy of electronic transitions between Van Hove singularities in the valence and conduction bands. As already stressed, this is peculiar to one-dimensional structures.

The different locations of Van Hove singularities for SWNTs characterized by different diameters and chiralities offer the way to obtain (by suitable tuning of the laser energy) Raman spectra arising from nanotubes of a selected structure. This effect is clearly shown in Figure 27, where frequency dispersion of the main lines is observed while varying the excitation energy.

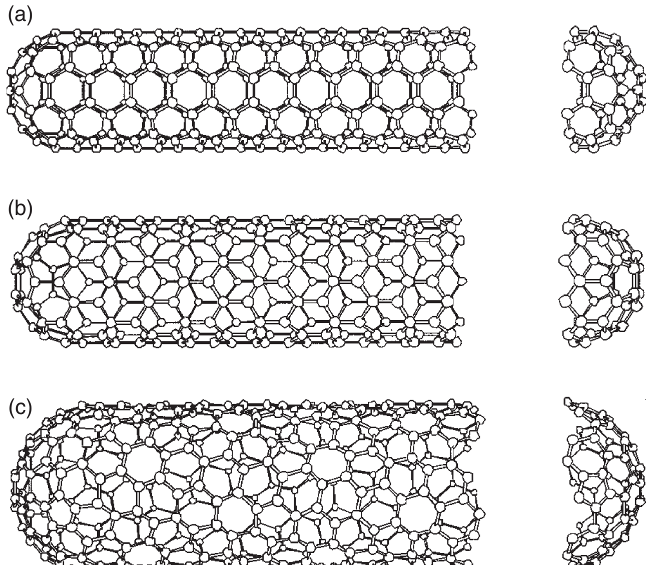


Figure 26. Schematic models for single-wall nanotubes capped with hemispherical fullerene-like caps: (a) armchair nanotube; (b) zigzag nanotube; (c) chiral nanotube.

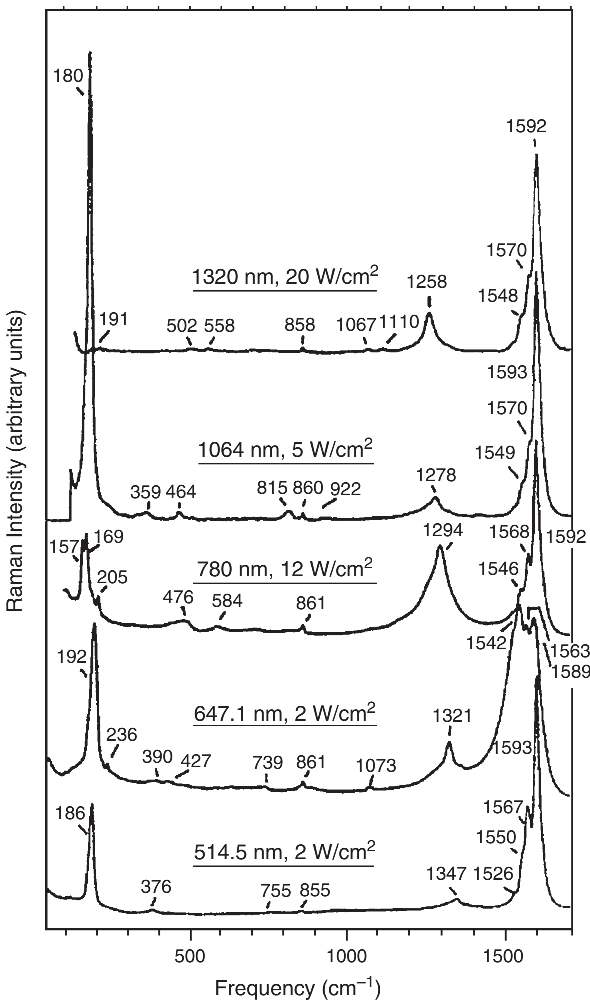


Figure 27. Experimental Raman spectra of purified single-wall carbon nanotubes excited at five different wavelengths; the laser wavelength and power density for each spectrum are indicated in the figure. Reprinted with permission from [110], M. S. Dresselhaus and P. C. Eklund, *Adv. Phys.* 49, 705 (2000). © 2000, Taylor & Francis.

The dispersion of the frequency of RBM has been reproduced by both empirical force constants [113] and *ab initio* calculations on model structures: these calculations showed a linear dependence of its value on the inverse of the tube diameter. Accordingly, the relationship ν (cm^{-1}) $= 248/d_t$ has been proposed for the practical determination of SWNT diameters from the Raman frequency of RBM.

As for the G band, the most dramatic features are the characteristic differences between the G bands for metallic and semiconducting nanotubes, as shown in Figure 28. The rationalization of these findings not only requires the description (frequency prediction by vibrational dynamics) of the Raman active phonons for a given (n_1, n_2) tube, but also calls for the understanding of the Raman intensity and line shape associated with the Raman transitions. These two features strongly depend on the characteristic electronic

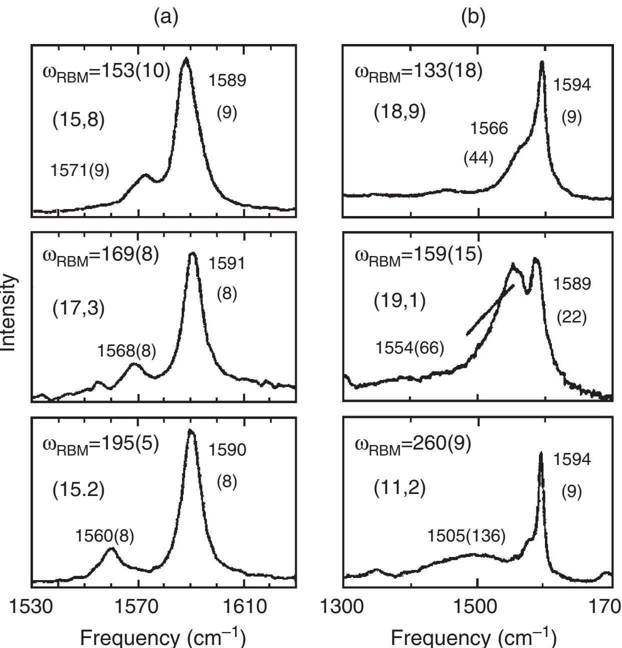


Figure 28. Profiles of the G band spectra for three semiconducting (a) and three metallic (b) nanotubes whose indices (n, m) are determined by their RBM spectra. Frequency values and line widths (in parentheses) are indicated in the figure. Reprinted with permission from [111], M. S. Desselhaus et al., *Carbon* 40, 2043 (2002). © 2002, Elsevier.

structure of the tube considered (see [111] for a discussion) and are the specific signature of the strong electron–phonon coupling in SWNTs.

As for the case of graphitic materials (see Section 5.2) the origin of the D band in Raman spectra of nanotubes is still ascribed to a double resonant mechanism [111] activated by the presence of disorder. In the discussion of the behavior of the D line in nanotubes, the existence of Van Hove singularities is explicitly taken into account in the treatment by Kürti et al. [116], giving rise to phenomena which are described by the authors as triple resonances.

The experimental study of the G band behavior and of the D line activation for nanotubes shines light on the fact that each (n_1, n_2) nanotube is characterized by its own unique electronic structure and that the strong coupling between electrons and phonons determines its Raman response. Conversely, the theoretical models developed in order to give a reasonable description of these interactions and to accurately predict the observed spectroscopic behavior are essential, at least for the following reasons: (i) developing the correct correlation for structural diagnosis (e.g., assignment of Raman lines to the corresponding (n_1, n_2) tube); (ii) extracting (from spectra) information about the electronic structure; and (iii) for a challenging test of different levels of theories and computational methods.

Point iii requires some comment. As mentioned in Section 2, the simplest way to treat the vibrational dynamics of nanotubes consists in applying the zone folding methods to the phonon dispersion curves obtained for an infinite graphite sheet. This technique has been followed by several authors who obtained in a relatively quick way good insight into the interpretation of SWNT Raman spectra. On the other hand, a more detailed comparison of the predictions with experimental data revealed nonnegligible weakness in the method. The main drawback of the zone folding treatment can be briefly summarized in the following two points.

1. The description of some modes of nanotubes cannot be made correctly on the basis of the nuclear displacements characteristic of the corresponding phonon of the graphene sheet (see Fig. 29). This happens, for instance, for the radial breathing mode of the nanotubes, which is put in correspondence with the out-of-plane acoustic mode of the graphene sheet (Fig. 29a). This last mode corresponds to a vanishing frequency (pure translation of the sheet) in the limit $\mathbf{q} \to \mathbf{0}$. It is immediately realized that the RBM of the nanotube is an optical mode, which does involve both the graphite out-of-plane (radial) force constants and the in-plane (tangential) force constants. As another example, nanotube modes exist for which the correct nuclear displacements can be obtained only by mixing of in-plane and out-of-plane modes of graphite (Fig. 29b), which cannot mix by reason of symmetry in the case of a graphene sheet.
2. Another problem arises from the fact that a force field transferred from graphite may be intrinsically different (different values of force constants) from the real field of a nanotube. This can happen for at least two reasons: (a) the curvature of the nanotube surface can modify, in principle, all radial and tangential force constants and (b) the presence of a finite sequence of CC bonds along the circumference of the tube is certainly responsible for confinement effects of the π electrons and of the resulting change in the intramolecular potential.

In a recent paper by Dubay and Kresse [117] state of the art calculations of phonon dispersion relations for nanotubes are presented and all the relevant references are quoted.

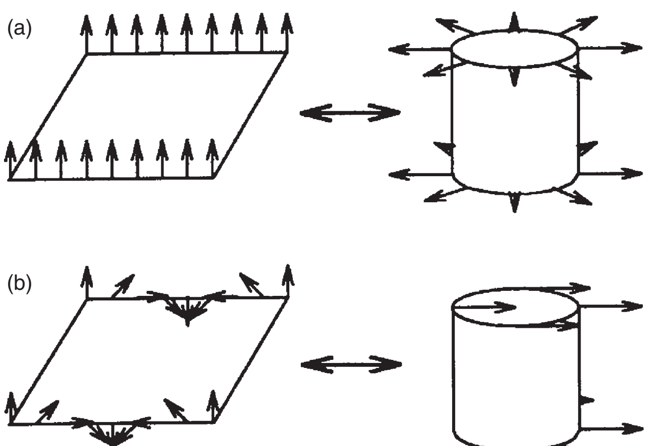


Figure 29. (a) The out-of-plane acoustic mode at $K = 0$ (left) in a single layer of graphite gives rise to a radial breathing mode in the carbon nanotube with nonzero frequency (right); (b) an acoustic mode of the nanotube corresponds to a linear combination of both in-plane and out-of-plane graphene modes. Reprinted with permission from [110], M. S. Dresselhaus and P. C. Eklund, *Adv. Phys.* 49, 705 (2000). © 2000, Taylor & Francis.

In this paper, accurate calculations for the phonon dispersion relations of single wall armchair (n, n) and zigzag (n, 0) nanotubes based on density functional theory are reported. The results are compared to those predicted by the zone folding procedure applied to graphite phonon dispersion branches as obtained with a consistent theoretical treatment at the same level of theory. According to this study an appreciable softening of the longitudinal A_1 mode in metallic nanotubes has been found with respect to frequency value obtained by the simple zone folding procedure: this effect has been correlated to a precise electronic mechanism taking place in metallic nanotubes.

5.5.3. Carbynes

Nanostructured carbon films produced by supersonic cluster beam deposition have been studied by *in-situ* Raman spectroscopy [135]. Raman spectra show the formation of an sp^2 solid with a very large fraction of sp-coordinated carbyne species showing a long-term stability under ultrahigh vacuum. Distinct Raman contributions from polyyne and cumulene species have been observed. The long-term stability and the behavior of carbyne-rich films under different gas exposures have been characterized showing different evolution for different sp configurations. These experiments confirm theoretical predictions and demonstrate the possibility of easily producing a stable carbyne-rich pure carbon solid. The stability of the sp^2–sp network has important implications for astrophysics and for the production of novel carbon-based systems.

Linear carbon chains with sp hybridization are considered fundamental constituents of the interstellar medium [118, 119]. They are abundant in circumstellar shells [118] and their presence can account for many features of the diffuse interstellar bands [118, 119]. The efforts to synthetise and study these species in the laboratory stimulated the discovery of C_{60} [120]. Several gas phase studies of carbon clusters with linear or cyclic structure have suggested that they can be elemental building blocks of three-dimensional fullerenes and nanotubes [121].

The sp carbon chains can present alternating single and triple bonds (polyyne) or only double bonds (polycumulene) [122]. Theoretical calculations suggest that polycumulenes are less stable than polyynes [122, 123]. Both species are characterized by an extremely high reactivity against oxygen and a strong tendency to interchain cross-linking [124], thus rendering the direct observation of a pure carbyne-assembled solid still a major challenge. Isolated carbynic species have been investigated in the gas phase in order to determine their geometry [121, 125]; their electronic and vibrational structure has been studied for non-interacting clusters embedded in matrices of cold rare gases [126].

The sp carbon chains have also been proposed as the elemental building blocks of a carbon allotrope called carbyne [124]. The carbyne story is very controversial: the first claims of direct observation of carbyne date from the sixties [124, 127]; however, up to now, no firm and unambiguous evidence of the existence of this allotrope has been provided [124]. Experimental results, mainly based on crystallographic recognition supporting the identification of crystalline carbyne, have been the object of strong criticism up to a complete rejection of the carbyne concept [124, 128].

Although the occurrence of crystalline carbyne is still a matter of debate, carbyne-rich or carbynoid solids have been produced by synthetic strategies mainly based on chemical routes. These include oxidative coupling reactions, dehydrohalogenation of polymers, polycondensation reactions of halides, electrochemical reductive carbonization, and condensation of end-capped chain molecules produced in the gas phase [129, 132]. In these systems polyyne chains, assembled together, are preserved against cross-linking and chemical decomposition by the presence of metal-based species and molecular groups terminating and separating the chains.

A different approach to the synthesis of carbyne is the high-temperature and high-pressure treatment of carbon-based materials. Resistive heating of graphite or high-energy laser or ion irradiation of carbon have been suggested as possible methods to produce carbynoid materials [133]. However, no definitive confirmation of the validity of these methods has yet been reported.

Raman spectroscopy is one of the techniques of choice for the study of carbon-based materials and for the identification of carbyne [124, 134]. The presence of a Raman band at roughly 2100 cm^{-1}, generated by the carbon triple bond, is admitted to be one of the strongest arguments in favor of the presence of carbynes [124, 131]. Raman spectra of carbynoid materials are characterized by G and D bands in the range of 1200–1700 cm^{-1} as in amorphous carbon and by the presence of the band in the 2100-cm^{-1} region [124, 131]. For carbyne-rich materials this band should have an intensity comparable to or even greater than the amorphous contribution.

Raman spectra from chemically stabilized carbynoid systems are reported in the literature, but due to the high reactivity and fast aging of the carbyne species, no Raman spectra are reported for pure carbon systems.

Ravagnan et al. [135] it have shown that it is possible to produce nanostructured carbon films rich in carbyne species by depositing pure carbon clusters from a supersonic beam. These films have been characterized by *in-situ* Raman spectroscopy determining the contribution to the spectra coming from the polyyne and cumulene species.

Cluster-assembled carbon films have been previously characterized *ex-situ* by Raman spectroscopy [139]. The films have an amorphous structure with a main sp^2 hybridization character. The G and D features reveal a substantial amount of distortion of bond lengths and angles [140]. A weak peak at about 2100 cm^{-1} was observed. Due to the high reactivity of carbyne this peak could be attributed to the residual presence of a primeval larger carbynoid population. Although an attribution based solely on *ex-situ* data cannot be considered definitive.

Figure 30 shows the comparison between the *ex-situ* and *in-situ* Raman spectra of two films deposited under the same conditions. The G and D bands are almost identical, whereas the peak at 2100 cm^{-1} has dramatically increased in intensity in the *in-situ* spectrum (we will refer to it as the C peak). This peak has a remarkably strong intensity and appears to be structured and composed of a main broad peak at about 2100 cm^{-1} and a weaker shoulder centered around 1980 cm^{-1}, as evidenced by a two-Gaussian fit. The relative C peak intensity, expressed by the ratio between its

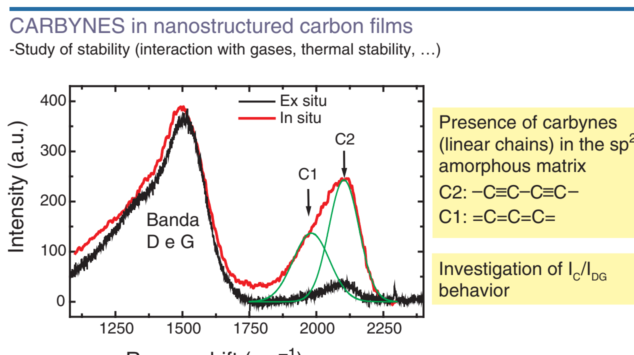


Figure 30. Raman spectra of nanostructured carbon films deposited by a supersonic cluster beam, measured *ex-situ* (gray) and *in-situ* in UHV conditions (red). For the second spectra we have indicated the two fitting curves identifying the cumulene (left) and polyyne (right) contributions.

integrated intensity and the D–G band integrated intensity (IC_{rel} = IC/ID, G) is roughly 45%.

Raman spectra confirm the formation of a carbynoid material with a substantial presence of sp linear structures among an sp^2-hybridized disordered network, although it is not possible to quantify the number of *sp*-hybridized C atoms, since the Raman cross section of this kind of vibration is unknown. Compared to other carbynoid systems reported in the literature the C peak at 2100 cm^{-1} presents a low-frequency shoulder at 1980 cm^{-1}. This is caused by the presence of both cumulenic and polyynic chains coexisting in the film and giving spectral contributions at different frequencies [141]. The broadening of the two components can be ascribed to the sp chain length distribution and to disorder [124].

These observations show that a carbyne-rich material can be formed by assembling carbon clusters at very low energies without high temperature and high pressure or chemical reactions inducing carbyne formation from polymers or hydrocarbons.

The experimental results reported here provide new elements for the interpretation of the carbyne puzzle and are of relevance for areas where carbon clusters play an important role, such as the composition of the interstellar medium and the existence of new allotropic forms of carbon. Low-energy cluster beam deposition in an oxygen-free environment appears to be a viable technique for the production of a pure carbon nanostructured material extremely rich in carbynoid species. The unexpectedly high stability of polyyne and cumulene species interacting with each other and with a fullerenelike type of clusters once deposited [138] suggest the existence of a cluster–cluster reaction mechanism leading to the formation of networks where the sp bonds are stabilized. These networks composed of sp^2 and sp linkages may be a disordered analogue of a new phase of carbon called graphyne and theoretically predicted by Baughman et al. [143]. The doping of this material with alkali metals or stabilizing molecular groups may lead to the observation of novel structural and functional properties in these systems.

The gas phase chemical models of interstellar clouds take into account the presence of different carbon nanoparticles, such as carbyne, fullerenes, and polycyclic aromatic hydrocarbons [118, 144]. These species are related by very complex reaction pathways to produce and destroy one species from another. Cluster–cluster interaction can produce large carbonaceous grains where carbyne species are present and stable.

5.6. Raman Scattering from Confined Acoustic Modes

5.6.1. Nanoparticles

As described in the preceding sections, the confinement of optical phonons in small particles can be studied by Raman spectroscopy. Raman scattering from confined acoustic phonons can also be observed and it can be used for the determination of particle size. Duval and co-workers have reported low-frequency Raman bands in nucleated cordierite glass [145]. The size of the particles has been determined by X-ray scattering and electron microscopy and the position of the low-frequency Raman peaks as a function of the inverse of particle diameter is reported in Figure 36. An analysis of Raman peak intensity and polarization have suggested the existence of a spherical source of scattering. This observation, together with the linear dependence of the Raman line shift from the inverse of nanocrystallites dimensions, suggested that vibrational surface modes of the particles are responsible for the observed Raman scattering [145].

A model that can account for the observed low-frequency scattering describes the nanocrystal embedded in matrices as an homogeneous elastic body of spherical shape vibrating elastically. Lamb [146] Tamura and Ichinokawa [147], and Fujii et al. [148] have theoretically studied the vibrational eigenfrequencies of a homogenous elastic sphere with a free surface (i.e., the eigenfrequencies of the confined acoustic modes). Two equations determining the spectrum are derived:

$$j_{l+1}(\eta) - \frac{l-1}{\eta} j_l(\eta) = 0 \tag{40}$$

$$2\left[\eta^2 + (l-1)(l+2)\left[\frac{\eta j_{l+1}(\eta)}{j_l(\eta)} - (l+1)\right]\right]\frac{\xi j_{l+1}(\xi)}{j_l(\xi)} - \frac{\eta^4}{2} + (l-1)(2l+1)\eta^2 + [\eta^2 - 2l(l-1)(l+2)] \times \frac{\eta j_{l+1}(\eta)}{j_l(\eta)} = 0 \tag{41}$$

where η and ξ are nondimensional eigenfrequencies and l is an angular momentum quantum number. The first equation describes, with $l \geq 1$, a torsional mode and the second, with $l \geq 0$, a spheroidal mode [148]. The torsional modes are vibrations without dilatation, which is instead present in spheroidal modes. The eigenfrequencies corresponding to different angular momentum quantum numbers can be written as

$$\eta_l^t = \frac{\omega_l^t d}{2v_t} \tag{42}$$

where ω_l^t is the angular frequency of the lth torsional mode, v_t is the transverse sound velocity, and d the diameter of the particle. For the spheroidal modes the eigenfrequencies are

$$\eta_l^s = \frac{\omega_l^s d}{2v_t} \tag{43}$$

$$\xi_l^s = \frac{\omega_l^s d}{2v_l} \tag{44}$$

where v_t is the transversal sound velocity.

With the above equations one can calculate the eigenvalues belonging to angular momentum $l\eta_{\text{ln}}^t$ and η_{ln}^s ($n+1$th eigenvalues). By introducing the transverse and longitudinal sound velocities with $n=0$, the Raman peak frequencies of the torsional and spheroidal modes can be reproduced by fitting the dimension of the crystallites d.

This model has been applied to different systems starting from the observation of Duval and co-workers [143]. Fujii et al. measured the low-frequency Raman spectra of small silver particles embedded in SiO_2; the agreement between theory and experiment is reasonable for particles with an average size $d \leq 40$ Å. This has been attributed to the fact that particles with a small average size have a narrow size distribution, whereas the growth of large particles cannot avoid a wider size distribution. Size distribution cannot be taken into account in the model [148]. Another possible reason for the discrepancy between experiments and calculations could be the effect of particle shape. The deviation from spherical shape can cause changes in the vibrational frequencies and Raman selection rules. TEM micrographs show that large particles have irregular shapes.

For Ag crystallites in SiO_2, the effect of the matrix seems to be negligible. This is true if the elastic properties of the matrix are very different from those of the particles. In this case the clusters can be considered under free surface boundary conditions. The assumption that the nanocrystal is a homogeneous elastic body is valid for Ag nanocrystals as small as 3 nm. For very small nanocrystals, the continuous assumption is not justified and lattice dynamics calculations should be performed [148].

The low-energy Raman scattering from silver particles in alkali halides was also investigated by Mariotto et al. [149]. From the energies of the observed modes the shape and dimensions of the particles were deduced. A good agreement with small angle X-ray scattering characterization of the samples has been found.

Films assembled by LiF clusters with diameters between 10 and 5 nm have been characterized by Raman spectroscopy [150, 151]. Bulk LiF does not show first order Raman scattering; the selection rules are relaxed in small crystallites because of the uncertainty of the phonon wave vector. The Raman spectra of LiF clusters show several sharp lines located between 28 and 1492 cm^{-1}. In particular two lines at 28 and 42 cm^{-1} have been attributed to localized acoustic vibrations of the clusters [151]. The interpretation of the origin of the other peaks observed at higher frequencies has been discussed in terms of surface optical modes. The role of size distribution and of cluster–cluster interaction in the film needs further investigation to be clarified.

Quasi-free silicon clusters deposited onto a porous silicon matrix also show low-frequency Raman lines that can be attributed to surface acoustic modes and to their combinations [152]. The clusters in porous silicon are only weakly bound and touching each other so that although the matrix is of the same material as the clusters, surface effects should not be suppressed. Comparison between theory and experiments suggests that surface modes are Raman active.

Low-frequency Raman scattering has also been used to detect acoustic modes in metallic nanocrystals and their behavior as a function of temperature for sizes in the confinement regime, giving new information about melting of the particles. The eigenfrequencies of the acoustic modes are in agreement with theoretical calculations of vibrations in a sphere. Approaching the nanoparticle melting temperature [153] the acoustic modes of the particles are progressively softened and damped; a sudden change in the mode frequency and damping is detected just below the previously measured melting temperature and is explained in terms of surface premelting; no acoustic peaks can be detected after melting. A broad central peak, whose intensity increases with temperature, was detected. The coupling of the particle acoustic modes with relaxing degrees of freedom, through a damping process also involving the matrix–particle interaction, produces overdamped fluctuations in dielectric function, thus explaining the observed behavior.

5.6.2. Nanotubes

As discussed above, SWNTs are formed by warping a single graphitic layer to form a seamless cylindrical object [155]. Their expected very high strength-to-weight ratio has stimulated an enormous interest in view of applications as reinforcement fibers in composite materials, high-strength cables, actuators, etc. [155, 160, 162, 163]. The quasi-static elastic properties of SWNTs (e.g., Young modulus) have been estimated by observing their bending with AFM [164] or their freestanding room-temperature bending vibrations with TEM [159, 163]. With these methods problems arise from the lack of precision in determining the SWNT diameter, length, temperature, and vibrational frequency [159] and the tendency of SWNTs to form bundles complicates the analysis. A Young modulus of the order of $E = 1.25$ TPa was found, even though, in the vibrating rod model, this quantity is related to a ill-defined quantity, the thickness of the tube walls. Moreover, different theoretical approaches have been proposed to model a SWNT with an elastic continuum model [159], or by *ab initio* calculations [156, 157, 160, 161] and the predicted Young moduli are scattered over a large interval. It is still a matter of speculation whether the curvature modifies the original properties of the graphene sheet [159, 165–168], since the Young modulus of SWNT's appears to be systematically higher than the in-plane elastic modulus of graphite, and no precise comparison could be made between SWNTs and MWNTs (see, e.g. [169, 170]). The higher value of their Young modulus could be attributed to an increased strength due to the wrapping [159] (stiffening of the graphite sheets is predicted for a radius lower than 0.5 nm [171]).

Bottani et al. [172] present the measurement by inelastic light scattering of confined longitudinal acoustic modes

(with a wavelength of the order of hundreds of nanometers) in aligned SWNT, whose dispersion relations depend on the 2D elastic constant of the tube walls, that is, of a single warped graphite sheet. This is a dynamic measurement, in contrast with AFM (static) and TEM (dynamic, flexural modes at lower frequencies). So far no direct dynamic measurement of these modes has been reported, even though theoretical calculations have been attempted. The analyzed sample is a freestanding film of pure SWNTs aligned in the plane of the film. Tubes are present in bundles or ropes (100–1000 tubes), with a preferential orientation.

The SWNT diameter was evaluated to be in the range of 12–14 Å by measuring the frequency of the breathing radial mode at about 170–180 cm^{-1} in the Raman spectrum, which is inversely proportional to the radius of the tube [155]. Raman spectroscopy also confirmed the pure SWNT nature of the samples. The absence of scattering for a scattering plane perpendicular to the tube direction confirmed the relative vibrational independence of the tubes and that the composite material is not an effective medium in which acoustic waves can propagate in all directions, at least at the scale of laser excitation wavelength λ_0. Nevertheless, the interaction between adjacent nanotubes probably plays some role and makes the theoretical description based on an isolated cylindrical shell only a first approximation to the problem of acoustic phonon propagation in this kind of material.

Depending on tube length compared to λ_0, the inelastic photon scattering from the acoustic excitations of a nanotube can be a new type of coherent line Brillouin scattering or an incoherent, Raman-like inelastic scattering by confined acoustic phonons (typically in the range of 0–10 cm^{-1} $\equiv$ 0–300 GHz [180]). The authors computed the scattering intensity going beyond the rod approximation and modeled the SWNT as an elastic cylindrical surface in the membrane approximation, i.e., as a 2D elastic object. This 2D model allows only dynamic 2D stress conditions and the nanotube elastic constants have the dimensions of force per unit length. Though the membrane approximation would be inadequate to describe strong bending modes of long tubes, it is precise enough to compute those normal modes along the tube axis, which have a frequency in the range of 1–300 GHz and are thermally excited. These modes are the only ones responsible for the scattering described here. The authors assumed that a surface elasto-optic effect couples the light to the tube thermal phononic strains which are then expanded in normal coordinates. This coupling produces a fluctuating surface polarization density that radiates the Stokes and anti-Stokes scattered photons. The computation has been carried out for the backscattering geometry.

The elasto-optic coupling [177] can be described by introducing the fluctuating part of the 2D dielectric susceptibility of the nanotube surface, the tensor $\delta\boldsymbol{\chi}$ whose three independent components (by symmetry arguments similar to those used for determining the strain tensor) are related to membrane displacements in cylindrical coordinates x, r, φ by the equations

$$\delta\chi_{xx} = \frac{K}{1-\xi^2}\left[\frac{\partial u_x}{\partial x} + \xi\left[\frac{\partial u_\varphi}{R\partial\varphi} - \frac{u_r}{R}\right]\right];$$

$$\delta\chi_{\varphi\varphi} = \frac{K}{1-\xi^2}\left[\xi\frac{\partial u_x}{\partial x} + \frac{\partial u_\varphi}{R\partial\varphi} - \frac{u_r}{R}\right];$$

$$\delta\chi_{x\varphi} = \left[\frac{K}{2(1+\xi)}\frac{\partial u_x}{R\partial\varphi} + \frac{\partial u_\varphi}{\partial x}\right] \tag{45}$$

where u_x, u_φ, and u_r are the longitudinal, radial, and tangential displacements, respectively, R is the tube radius, and K and ξ are the components of the 2D elasto-optic tensor $\mathbf{K}$ (having the same symmetry as the elastic constant tensor which, in turn, has only two independent components for a sixfold symmetry surface). Assuming traveling waves (W represents the normal coordinate and $q_\parallel$ the wave vector)

$$[u_x \quad u_\varphi \quad u_r] = \frac{1}{2}W[u_0 \quad v_0 \quad w_0]\exp i(q_\parallel x - \omega t + m\varphi) + c.c. \tag{46}$$

and using the elastodynamic equations of a cylindrical membrane we find the dimensionless eigenfrequencies (E_{2D} is the 2D surface Young modulus and ν is the Poisson ratio):

$$\Omega^2_{m,q_\parallel,j} \equiv (\rho_{2D}/E_{2D})(1-\nu^2)R^2\omega^2_{m,q_\parallel,j} \tag{47}$$

where j is a branch index. For an infinitely long tube the above traveling waves are the appropriate solution; for tubes of finite length L the proper boundary conditions affect the form of the solution and the corresponding frequencies, unless $L \gg 2\pi/q_\parallel$.

In our case, since $(R/\lambda_0) \ll 1$, the computation of the intensity of the light scattered from a single tube reduces to [177]

$$I(\omega) \approx \omega_0^4 \sum_{m,\mathbf{q}_\parallel,j} I_{m,\mathbf{q}_\parallel,j}\,\delta(\omega - \omega_0 \pm \omega_{m,q_\parallel,j}) \tag{48}$$

where the contribution of each normal mode $I_{m,\mathbf{q}_\parallel,j}$ is given by

$$I_{m,\mathbf{q}_\parallel,j} = \left\langle\left|\int_S e^{iQ_\parallel x}(\mathbf{e}^S \bullet \delta\boldsymbol{\chi} \bullet \mathbf{e}^I)_{m,q_\parallel,j}\,dS\right|^2\right\rangle_{th} \tag{49}$$

where S is the tube surface and $Q_\parallel$ is the component of the scattered wave vector parallel to the tube direction; in backscattering $\mathbf{Q} = \mathbf{k}_s - \mathbf{k}_i = -2\mathbf{k}_i$. An analysis of the above equations shows that only the longitudinal modes with $m = 0$ contribute to inelastic scattering with nonvanishing intensity, provided $Q_\parallel \neq 0$ (i.e., there must be a nonvanishing component of the transferred wave vector, i.e., a scattering plane not perpendicular to the SWNT orientation). In this case the scattered light remains polarized. When the tube length L is much longer than λ_0, the above equation gives

$$I_{m,q_\parallel,j} \propto \delta(Q_\parallel - q_\parallel) \tag{50}$$

In this case (infinite tube, i.e., length $L \gg \lambda_0$) the scattering would be a coherent line Brillouin scattering and only one longitudinal wave (determined by the selected value of $q_\parallel$) would contribute to the measured spectrum (see the surface Brillouin scattering below). The cross section is significant only if the parallel wave vector is conserved ($Q_\parallel = q_\parallel$), and

this conservation would produce a strong dependence of the peak position on the incident angle θ_i (dispersion of the modes) that was not observed. Instead the θ_i dependence of the peak position turned out to be negligible. This was interpreted taking into account that the free mean SWNT length in the samples was of the order of 1 μm or less ($L \approx \lambda_0$), and the tubes cannot be considered infinite with respect to the propagation of the modes selected by the scattering technique used. Assuming clamped ends boundary conditions, as suggested by AFM images, all longitudinal modes with $m = 0$ (stationary waves in this case) turn out to contribute to scattering, due to confinement effects when $L \approx \lambda_0$. The name usually given to this kind of scattering is low-frequency Raman scattering (e.g., in the case of scattering from vibrational eigenmodes of small spheres see [180]). This is the case with the inelastic scattering here reported. In Bottani et al. [172] a more detailed model for the description of the vibrational dynamics of the tubes was used (Donnel theory, see [178, 179]) and applied to a thin shell (thickness $h \ll R$), which considers the finite length L of the tube and takes into account shear forces and torques not belonging to the shell surface. Imposing clamped ends boundary conditions ($u_x = u_\varphi = u_r = du_x/dx = 0$), the normal modes can be written as

$$\begin{aligned} u_x &= u_0[-\sin(\mu/R)(L/2-x) + k\sinh(\mu/R)(L/2-x)] \\ &\quad \times \cos m\varphi \cos \omega t \\ u_\varphi &= v_0[\cos(\mu/R)(L/2-x) + k\cosh(\mu/R)(L/2-x)] \\ &\quad \times \sin m\varphi \cos \omega t \\ u_r &= w_0[\cos(\mu/R)(L/2-x) + k\cosh(\mu/R)(L/2-x)] \\ &\quad \times \cos m\varphi \cos \omega t \end{aligned} \tag{51}$$

The eigenmodes are characterized by a reduced wave vector $\mu = n(\pi/L)R$, with $n = 1.506, 3.5, 5.5, 7.5, \ldots$. The corresponding dimensionless eigenvalues can be written as in Eq. (47) and numerically computed.

It is found again that only longitudinal modes with $m = 0$ contribute to scattering, provided $Q_\parallel \neq 0$. Longitudinal modes with different values of n contribute to the scattered intensity with relative weights given by the above mentioned computation. The authors assumed $\rho_{2D} = 7.6 \cdot 10^{-7}$ kg/m^2 (surface density in SWNTs is assumed to be the same as that in graphite), $R = 6 \cdot 10^{-10}$ m (as obtained by Raman measurements), and Poisson's ratio $\nu = 0.2$ [160]. The dispersion relation of the longitudinal modes turns out to depend only very weakly on ν. Spectra corresponding to different tube lengths (0.5–1.5 μm) have then been convoluted. The 2D Young modulus being the only unknown, its value can be estimated to be the one that best reproduces the measured frequency position of the scattering features. The best correspondence between the computed and the measured peak positions is obtained for $E_{2D} = 110$ N/m. This value is of the same order of magnitude as the C–C atomic force constant computed for a graphite plane [182]. If one considers the thickness of a graphite layer, this corresponds to a Young modulus for an ideal isolated graphitic plane of the order of 1 TPa.

6. BRILLOUIN SCATTERING

6.1. Bulk and Surface Acoustic Waves and their Phonons

Bulk Brillouin scattering of laser light is the inelastic coherent scattering of a photon in a material from a long-wavelength traveling acoustic phonon. If instead the acoustic phonons are localized in nanostructures with sizes of $\delta < \lambda = 2\pi/|\mathbf{k}^I|$, λ being the photon wavelength, then the scattering from an ensemble of independent separated nanostructures is incoherent and it is better referred to as low-frequency Raman scattering (see the sections above on nanoparticles and nanotubes). To save space the reader is referred to Sandercock [206] for a detailed description of the experimental setup. Particularly interesting is the hybrid case of surfaces, thin films, and multilayers in which phonon localization can occur only in the depth z direction, while free propagation with wave vector $\mathbf{q}_\parallel$ takes place parallel to the surface in the (x, y) plane (surface Brillouin scattering). In view of the wavelength of the excitations involved, the vibrational dynamics responsible for every type of Brillouin scattering can be treated in the continuum elastodynamic approximation [195].

6.2. Bulk Waves

The elastodynamic equations can be written in terms of the dynamic displacement field $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$ mapping the equilibrium $\mathbf{r}$ configuration into the istantaneous dynamic configuration $\mathbf{r}' = \mathbf{r} + \mathbf{u}(\mathbf{r}, t)$, where $\mathbf{r}$ spans all the undeformed volume of the material. In the simplest case of a homogeneous isotropic body these equations read

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = v_t^2 \boldsymbol{\nabla}^2 \mathbf{u} + (v_l^2 - v_t^2)\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{u}) \tag{52}$$

where

$$v_l = \sqrt{\left(B + \frac{4}{3}\mu\right) \Big/ \rho}$$

is the longitudinal sound velocity and $v_t = \sqrt{\mu/\rho}$ is the transverse sound velocity. ρ is the mass density and B and μ the bulk and shear moduli, respectively [195]. Remembering that the most general deformation process is the superposition of a simple dilation $\mathbf{u}_l$ and of a simple shear $\mathbf{u}_t$, one is tempted to try a solution of the type $\mathbf{u} = \mathbf{u}_l + \mathbf{u}_t$, with $\boldsymbol{\nabla} \times \mathbf{u}_l = 0$ and $\boldsymbol{\nabla} \cdot \mathbf{u}_t = 0$. The second condition is identical to that holding for electromagnetic waves in a vacuum. This trick works perfectly (instead anisotropic materials require a more complex treatment [198]), leading to two decoupled wave equations for $\mathbf{u}_l$ and $\mathbf{u}_t$:

$$\frac{\partial^2 \mathbf{u}_l}{\partial t^2} = v_l^2 \nabla^2 \mathbf{u}_l \quad \text{and} \quad \frac{\partial^2 \mathbf{u}_t}{\partial t^2} = v_t^2 \nabla^2 \mathbf{u}_t \tag{53}$$

The fundamental bulk solution of Eq. (52) is then the superposition of three independent monochromatic plane waves, one longitudinal (LA) and two (mutually perpendicular) transverse (TA_1, TA_2), of the type

$$\mathbf{u}_\alpha = \Re\left\{Q_{\mathbf{q}\alpha} \mathbf{e}_{\mathbf{q}\alpha} e^{i[\mathbf{q}\cdot\mathbf{r} - w_\alpha(\mathbf{q})t)]}\right\} \tag{54}$$

where $\mathbf{q}$ is the wave vector, α is a branch index ($\alpha = l, t_1, t_2$), $Q_{\mathbf{q}\alpha}$ is the complex amplitude of the normal coordinate $\xi_\alpha(\mathbf{q}) = Q_{\mathbf{q}\alpha} e^{-i\omega t}$, and $\mathbf{e}_{\mathbf{q}\alpha}$ is a polarization unit vector ($\mathbf{e}_{\mathbf{q}l} \| \mathbf{q}$; $\mathbf{e}_{\mathbf{q}t} \perp \mathbf{q}$). Moreover: $\boldsymbol{\omega}_l(\mathbf{q}) = v_l|\mathbf{q}|$ and $\boldsymbol{\omega}_{t_1,t_2}(\mathbf{q}) = v_t|\mathbf{q}|$. The above classical description can be translated into the language of quantum mechanics of the systems of independent harmonic oscillators: the quanta of the fields $\mathbf{u}_\alpha$ are the long-wavelength acoustic phonons whose possible energies are

$$E_n(\mathbf{q}\boldsymbol{\alpha}) = \left(n_{\mathbf{q}\alpha} + \frac{1}{2}\right)\hbar\boldsymbol{\omega}_\alpha(\mathbf{q}) \tag{55}$$

with $n_{\mathbf{q}\alpha} = 0, 1, 2, 3, \ldots$.

6.3. Surface Waves

The prototypes of surface acoustic waves (SAWs) are Rayleigh waves (for a treatment of Rayleigh waves following the lines of the above section see [195]). When the medium is semi-infinite (let us assume the $z = 0$ plane is the surface, the medium is below it: $z < 0$) the translational symmetry is broken in the direction perpendicular to the surface. This produces new facts: (a) the reflection of bulk phonons (in general with branch conversion) and (b) the existence of surface waves [195, 199]. The simplest case of type (a) is that of a transverse (TA) bulk phonon propagating in the sagittal (xz) plane with wave vector $\mathbf{q} = q_\| \hat{\mathbf{e}}_x + q_\perp \hat{\mathbf{e}}_z$ polarized in the horizontal y direction (a shear horizontal phonon: SH) and impinging on the surface. If the surface is not subjected to external forces, applying the boundary conditions, one finds that the surface acts as a perfect acoustic mirror and that the parallel wave vector $q_\|$ of the reflected phonon is identical to that of the incident phonon. Then from the dispersion relation for TA phonons, we get

$$q_\perp = \sqrt{\frac{\omega^2}{v_t^2} - q_\|^2} \tag{56}$$

That is, once $q_\|$ is fixed, there is a lower frequency threshold $\omega_t = v_t q_\|$ for the existence of SH bulk phonons. Above this frequency the spectrum is continuous. If the incident phonon is either a longitudinal or a transverse one but polarized in the incidence plane (shear vertical, SV), there is both a longitudinal and a transverse phonon reflected at different angles; the parallel wave vector $q_\|$ is always conserved in the reflection. For the longitudinal components to be true bulk phonons the threshold is $\omega_l = v_l q_\| > \omega_t$. Otherwise the corresponding partial wave is evanescent in nature. The region of the continuous spectrum between ω_t and ω_l is that of the mixed modes. What happens in the spectral region $\omega < \omega_t$ when the polarizations are only L and SV? For a given $q_\|$ there exists a unique surface wave (the Rayleigh wave). In other words, there also exists a discrete spectrum with a single eigenvalue, $\omega_R = v_R q_\|$. Putting $\omega_R = \xi v_t q_\|$, v_R is found taking the sole real root of an algebraic sixth order equation in ξ resulting from the insertion of a linear combination of partial waves (one L and one SV) of the type

$$\boldsymbol{\psi}_\alpha(z|\omega_R, q_\|) = Q_{\mathbf{q}\alpha}\mathbf{e}_{\mathbf{q}\alpha} \exp(-\kappa_\alpha z) \exp[iq_\|(x - v_R t)] \tag{57}$$

in the boundary conditions and in (52). Considering the possible values of Poisson's ratio $\nu = (3B - 2\mu)/2(3B + \mu)$ [195], numerical computation shows that ξ can vary between 0.874 and 0.955 and that a good approximation is

$$v_R \simeq \frac{0.862 + 1.14\nu}{1 + \nu} v_t \tag{58}$$

Both the SV and the L components of a Rayleigh wave are nondispersive waves traveling parallel to the surface. Both waves decay exponentially with depth as $\exp(-\kappa_\alpha z)$, with

$$\kappa_t = q_\| \sqrt{1 - \frac{v_R^2}{v_t^2}} \quad \text{and} \quad \kappa_l = q_\| \sqrt{1 - \frac{v_R^2}{v_l^2}} \tag{59}$$

showing that the penetration (localization) depth $\delta_\alpha = \kappa_\alpha^{-1}$ is of the order of the wavelength $\lambda = 2\pi/q_\|$. This result, together with the existence of the transverse threshold, separating the discrete from the continuous spectrum, is a general feature of acoustic surface waves physics even in the presence of overlayers or of a more complex layered subsurface structure. In the general case the transverse threshold depends on the transverse sound velocity of the substrate material.

6.4. Layered Subsurface Structures: More General Surface Waves

The most compact way to face the case of a layered subsurface structure is followed generalizing the treatment of the previous paragraph to an elastic half space ($z \leq 0$) whose density ρ and elastic constants B and μ are functions of the depth z. Rigorously this method would imply again deriving the propagation equation which would take a form different from Eq. (52) [200–203]. For our limited purposes one can just think of the same Eq. (52) but with z-dependent coefficients. Excluding the case of very sharp interfaces between adjacent layers, this is also a tenable and useful approximation (for the alternative, standard approach, see, e.g. [204]).

Once we have found the surface wave spectrum of a layered structure by means of any numeric method, a very useful set of spectral functions can be constructed: the layer-projected phonon densities of LPPDS states. For given $q_\|$ and polarization α, considering the whole set of eigenfunctions of Eq. (52) together with all their boundary conditions,

$$\varphi_{n\alpha}(z|\omega_{n\alpha}, q_\|) \exp[iq_\| x - \omega_{n\alpha}(q_\|)t] \tag{60}$$

The spectral functions are defined as

$$g_\alpha(\omega, q_\||z) = \sum_n \langle|\varphi_{n\alpha}(z|\omega_{n\alpha}, q_\|)|^2\rangle_{th} \delta(\omega - \omega_{n\alpha}(q_\|)) \tag{61}$$

where $\langle\cdots\rangle_{th}$ is the thermal average and the summation is over the entire spectrum thought of as discrete. The continuous part of the spectrum can be made discrete using a slab approximation [200]. In a layered surface structure the Rayleigh wave is not the only true surface wave corresponding to a unique discrete eigenvalue and other surface waves, more or less confined in single or multiple layers, corresponding to several $\omega_{n\alpha}(q_\|)$, can exist (Sezawa, Stonley, and Love waves [204]). The contour lines of a smoothed

version of a specific $g_\alpha(\omega, q_\parallel|z)$ in the ω, z plane give a vivid and direct image of phonon localization [205]. The function $g_\alpha(\omega, q_\parallel|0)$ is the surface-projected phonon density of states which plays a major role in surface Brillouin scattering from opaque materials [205].

6.5. Volume Brillouin Scattering

The scattering kinematics can be described in a quite direct and intuitive picture. For the sake of simplicity we consider both the incident $(\mathbf{k}^I, \omega^I)$ and the scattered $(\mathbf{k}^S, \omega^S)$ photons only inside the medium. While a single acoustic phonon $(\mathbf{q}, \omega_\alpha = v_\alpha|\mathbf{q}|)$ is propagating within the material its periodic strain field sets up an anisotropic modulation of the otherwise isotropic dielectric susceptibility χ (Eq. (68)). This adiabatic modulation is viewed as a traveling diffraction grating by the incident electromagnetic wave. Then the scattering kinematics can be explained by the familiar wave concepts of Bragg reflection and Doppler shift. Equivalently, in the quantum language, the scattering process must obey the laws of conservation of momentum (wave vector) and energy. In the transition from the initial photon state $(\mathbf{k}^I, \omega^I)$ to the final (scattered) photon state $(\mathbf{k}^S, \omega^S)$

$$\mathbf{k}^S - \mathbf{k}^I = \pm\mathbf{q} \tag{62}$$

$$\omega^S - \omega^I = \pm\omega_\alpha(\mathbf{q}). \tag{63}$$

In Eqs. (62) and (63) the plus sign stands for the so-called anti-Stokes events, when a phonon is annihilated in the process and the scattered photon is more energetic, having gained energy from the medium, while the minus sign corresponds to the so-called Stokes events, when a new phonon is created in the process and the scattered photon is less energetic, having transferred energy to the medium. In the classical wave picture, Eq. (63) represents the Doppler shift of an electromagnetic wave dynamically diffracted by an elastic wave moving either in the direction of vector $\mathbf{q}$ or in the direction of vector $-\mathbf{q}$ with the same absolute sound velocity. To see how Brillouin inelastic scattering may be viewed as a Bragg reflection of the incident wave, let us consider Eq. (63) with the aid of the dispersion relations for photons $\omega^I = c|\mathbf{k}^I|$ and $\omega^S = c|\mathbf{k}^S|$ and phonons:

$$\frac{|\mathbf{k}^S| - |\mathbf{k}^I|}{|\mathbf{q}|} = \frac{v_\alpha}{c} \tag{64}$$

The order of magnitude of the ratio v_α/c could be typically 10^{-5} so that $|\mathbf{k}^S| \approx |\mathbf{k}^I|$. Because the energy of a photon is $c\hbar|\mathbf{k}|$ the scattering is quasi-elastic. Taking this into account and squaring Eq. (62) one easily obtains

$$2|\mathbf{k}^I|\sin(\phi/2) = |\mathbf{q}| \tag{65}$$

where ϕ is the scattering angle. The above equation can be given the form of Bragg law

$$2d\sin(\theta) = \lambda \tag{66}$$

$\theta = \phi/2$ being the Bragg angle. The grating spacing d is equal to the phonon wavelength $2\pi/|\mathbf{q}| = 2\pi v_\alpha/\omega_\alpha|$ and $\lambda = \lambda_0/n$ is the wavelength of the incident photon in the medium, while λ_0 is the corresponding quantity in a vacuum. From the above considerations it turns out that the measurement of the Brillouin shift $\Delta\omega = |\omega^S - \omega^I| = \omega_\alpha$ in a fixed scattering geometry (e.g., backscattering with $\phi = \pi$) leads to the possibility of measuring the sound velocity v_α, provided the index of refraction of the medium n is known, by means of the formula

$$v_\alpha = \frac{\lambda_0|\Delta\omega|}{4\pi n\sin(\phi/2)} \tag{67}$$

Equations (62) and (63) are only necessary conditions for the scattering events to take place. To have complete information one must compute the intensity of the scattered wave, or the elasto-optic scattering cross section. Once this is done, particular selection rules appear in connection with the scattering angle and the polarizations of both photons and phonons and their relative orientations with respect to the scattering plane (defined by the vectors $\mathbf{k}^I$ and $\mathbf{k}^S$); moreover, in the case of crystalline bodies, the dependence of v_α on the phonon propagation direction (elastic anisotropy) must be taken into account.

In the following we restrict ourselves to the case of optically isotropic solids. The key ingredient of bulk Brillouin scattering theory is the instantaneous anisotropic dielectric susceptibility of the medium around the frequency ω^I, the tensor

$$\chi_{ij}(\mathbf{r}, t) = \chi\delta_{ij} + \delta\chi_{ij}(\mathbf{r}, t) \tag{68}$$

where χ is the usual time-independent isotropic susceptibility, while the tensor $\delta\chi_{ij}(\mathbf{r}, t)$ is the anisotropic fluctuating part of the susceptibility due to the presence of thermal phonons. Then the fluctuating part of the polarization vector radiating the scattered waves can be written as $\delta P_i = \epsilon_0\delta\chi_{ij}E_j$, ϵ_0 being the vacuum dielectric constant. Here $E_j = E_j^I + E_j^S$ is the sum of the incident field and of the scattered field. Using the linearity of Maxwell equations, the scattered electric field E^S can now be computed by means of first order perturbation theory (Born approximation [207]):

$$\delta P_i \approx \epsilon_0\delta\chi_{ij}E_j^I \tag{69}$$

The above equation is justified by the smallness of both the scattered field and $\delta\chi_{ij}$. So far the treatment is analogous to that of Raman scattering in extended media. In the case of acoustic phonons the coupling between sound and light (the elasto-optic effect) can be written in terms of two elasto-optic constants a_1 and a_2 as

$$\epsilon_0\delta\chi_{ij} = a_1 e_{ij} + a_2\varepsilon_{ll}\delta_{ij} \tag{70}$$

and the phonon strains are very small. Sometimes a_1 and a_2 are written in terms of the Pockels coefficients p_{ij} as $a_1 = -\epsilon^2(p_{11} - p_{12})$ and $a_2 = -\epsilon^2 p_{12}$ [195, 206]. For a cubic crystal three independent coefficients are necessary to express the elasto-optic coupling. In Eq. (69) δP_i oscillates at the frequency $\omega^S = \omega^I \pm \omega_\alpha(\mathbf{q})$ because E_j^I is proportional to $e^{i\omega^I t}$ and $\delta\chi_{ij}$, being proportional to the strain field of a bulk phonon (see Eq. (54)), and oscillates as its normal coordinate $\xi_{\boldsymbol{\alpha}}(\mathbf{q}) = Q_{\mathbf{q}\boldsymbol{\alpha}}e^{\pm i\omega_\alpha(\mathbf{q})t}$ does. Solving the radiation problem (the Maxwell equations with a source proportional to

$\partial(\delta P_i)/\partial t$ [207]), the scattered electric field can be obtained. Once this is known the spectral intensity of the scattered light is computed as the power spectrum $S_{E^S}(\omega)$, which is essentially the measured outcome of a Brillouin scattering experiment:

$$S_{E^S}(\omega) = \int_{-\infty}^{\infty} \langle E^S(t+\tau)E^{S*}(t)\rangle_{th} e^{i\omega\tau}\, d\tau \tag{71}$$

where $E^S = E^S \cdot \mathbf{e}^S$ is the complex amplitude of the scattered field projected along a given analyzed polarization direction $\mathbf{e}^S$ perpendicular to $\mathbf{k}^S$. An explicit computation gives [208]

$$S_{E^S}(\mathbf{Q}, \omega,) \propto \frac{|E_0^I|^2}{\lambda_0^4} |\mathbf{e}^S \cdot \delta\boldsymbol{\chi}_{\mathbf{Q}} \bullet \mathbf{e}^I|^2 \langle |\xi_\alpha(\mathbf{q})|^2\rangle_{th} \times \delta[\omega - (\omega^I + \omega_\alpha(\mathbf{q}))] \tag{72}$$

for anti-Stokes scattering. For Stokes scattering one has just to replace $\omega^I + \omega_\alpha(\mathbf{q})$ with $\omega^I - \omega_\alpha(\mathbf{q})$ in the argument of the delta function. In the classical limit, valid for Brillouin scattering, $\langle |\xi_\alpha(\mathbf{q})|^2\rangle_{th} \propto k_B T/\omega_\alpha^2(\mathbf{q})$. In all practical applications the delta function is substituted by a Lorentzian line shape centered at $\omega^I \pm \omega_\alpha(\mathbf{q})$ (Brillouin doublet) with a FWHM related to the lifetime of the phonon and to instrumental and opacity broadening [206]; $\mathbf{e}^I$ is the polarization of the incident electric field. In Eq. (72)

$$\delta\boldsymbol{\chi}_{\mathbf{Q}} = \int_V \delta\boldsymbol{\chi}(\mathbf{r}')e^{-i\mathbf{Q}\cdot\mathbf{r}'}\, d\mathbf{r}' \tag{73}$$

is the Fourier transform of index $\mathbf{Q}$ of the spatial part of the susceptibility tensor fluctuation, the transferred wave vector $\mathbf{Q}$ being always defined as

$$\mathbf{Q} = \mathbf{k}^S - \mathbf{k}^I \tag{74}$$

Because the bulk phonon wave function (54) is a plane wave $\propto e^{i\mathbf{q}\cdot\mathbf{r}}$, computing $\delta\boldsymbol{\chi}_{\mathbf{Q}}$ we get

$$\delta\boldsymbol{\chi}_{\mathbf{Q}} \propto \int_V e^{-i(\mathbf{Q}-\mathbf{q})\cdot\mathbf{r}'}\, d\mathbf{r}' \propto \delta(\mathbf{Q}-\mathbf{q}) \tag{75}$$

that is, the wave vector conservation (62). The fact that $\mathbf{Q}$ must be equal to the phonon wave vector $\mathbf{q}$ to have a peak in the spectrum is typical of Brillouin scattering by bulk phonons in transparent materials. The partial relaxation of this rule in the case surface waves and/or of opaque media is one of the most characteristic points in the theory of surface Brillouin scattering [206]. Summarizing the main results that can be obtained from a detailed analysis of Eqs. (72), (18), and (70) (a) the light scattered by transverse phonons is completely depolarized; (b) there is no scattering by transverse phonons polarized in the scattering plane; (c) the intensity of scattering from transverse phonons goes to 0 in back scattering; (d) scattering by longitudinal phonons is fully polarized, that is, keeps the polarization of the incident wave.

6.6. Surface Brillouin Scattering

Surface Brillouin scattering (SBS) became a practical spectroscopic technique about 30 years ago due to the great progress made in Fabry Perot interferometry by Sandercock [206]. Because the formal theory of surface Brillouin scattering is rather complex and usually involves cumbersome computations both of the phonon density of states and of the scattering cross section [205], qualitatively only the main characteristic conclusions that can be drawn from Eq. (18) are outlined here. First of all, the finite penetration depth of light in the material (e.g., about 1000 nm in Si and only a few nanometers in Al at $\lambda_0 = 514$ nm, the typical wavelength of an argon laser) profoundly modifies the wave vector conservation law (62).

The skin depth of the light being δ, the effective scattering volume is limited in the direction z perpendicular to the surface within a length of the order of δ that can be less than the light wavelength $\lambda \approx |n\mathbf{k}^I|^{-1}$. We can then define a complex transferred wave vector $\mathbf{Q} = Q_\parallel \hat{\mathbf{e}}_x + (Q_\perp - i\delta^{-1})\hat{\mathbf{e}}_z$ to be used in Eq. (18). Moreover, taking into account that the phonon wave function of a true surface wave is proportional to

$$e^{iq_\parallel x} \sum_n b_n e^{-\kappa_{\alpha n} z}$$

with κ_α of the same magnitude as $q_\parallel$, from Eq. (18) it is found that

$$\delta\boldsymbol{\chi}_{\mathbf{Q}} \propto \delta(Q_\parallel - q_\parallel) \sum_n \frac{b_n}{\kappa_{\alpha n} + \delta^{-1} + iQ_\perp} \tag{76}$$

Due to the absence of $q_{\perp n}$ in the imaginary part of the denominator, only parallel wave vector conservation can occur even in the case of very low opacity. Introducing θ_i as the incidence angle and θ_s as the angle (positive in the clockwise direction) between the outgoing surface normal and the scattering direction and considering that $Q_\parallel = \pm q_\parallel$, we find

$$|\mathbf{k}^I|(\sin\theta_s - \sin\theta_i) = \pm q_\parallel \tag{77}$$

which is the fundamental kinematic relation of SBS. In backscattering, as $\theta_s = -\theta_i$, we obtain

$$\omega_{n\alpha} = v_{n\alpha}\frac{4\pi}{\lambda_0}\sin\theta_i \tag{78}$$

where $\omega_{n\alpha}$ is the Brillouin shift of a surface peak, $v_{n\alpha}$ is the phase velocity of the surface phonon responsible for the SBS event, and $q_\parallel = (4\pi/\lambda_0)\sin\theta_i$ is the allowed wave vector. Then, in backscattering, the SBS peak shifts of nondispersive SAWs (e.g., the Rayleigh wave) scale as the sine of the incidence angle, while the volume BS peak shifts of bulk phonons (Eq. (67)) do not depend on the angle. If the subsurface structure is not homogeneous but layered the surface waves are dispersive and $v_{n\alpha} = v_{n\alpha}(q_\parallel)$ in a nontrivial way for each different type of surface phonon [205]. The measurement of the dispersion relations is an effective way to measure the elastic constants of the surface once the density is known. The elasto-optic coupling is not the only SBS channel. A second scattering mechanism, the ripple effect,

can be operative if the phonons have a shear vertical polarization component [206]. In fact these phonons modulate the instantaneous shape of the free surface. The dynamically corrugated surface, depending on its reflectivity, scatters the incident light and, due to the Doppler effect, there appear scattered photons with frequencies shifted from ω^I. It turns out that the ripple scattering cross section is proportional to the power spectrum of the u_z displacement field computed at the surface ($z = 0$) and, thus, to $g_{SV}(\omega q_\parallel|0)$, the SV surface-projected phonon density of states [205, 208].

6.7. Thin Films and Multilayers: Localization of Surface and Pseudo-Surface Acoustic Phonons in Buried Layers

We have also seen that acoustic phonons, as optic phonons, can be confined (partially or totally) in limited regions of space in 3D and 2D. Here we focus our attention on 1D localization of surface acoustic phonons which can be measured by surface Brillouin scattering spectroscopy: in practice, this means phonon frequencies between 1 and 200 GHz. The characteristic length scale of such collective excitations falls in the range of a few hundreds of nanometers, a scale typical of VLSI microelectronics. In the particular case of shear horizontal (SH) surface acoustic phonons, the wave equation governing their displacement field produces a spectral problem showing strong analogies with the 1D quantum mechanics of spinless electrons. This case will be treated first.

The spectrum of long-wavelength surface acoustic phonons in opaque or semiopaque materials has been extensively investigated by means of Brillouin light scattering. Most of the studies have dealt with surface acoustic phonons polarized in the sagittal plane, defined by the surface phonon propagation wave vector, $\mathbf{q}_\parallel$, and the surface normal. For this type of excitation, light scattering occurs through the surface ripple and volume elasto-optic effect and the Brillouin scattering cross section, which depends on both these mechanisms, can show strong interference effects [218]. Until 1994 [200, 201] the case of SH phonons, that is, transverse phonons polarized in the plane of the surface perpendicularly to the sagittal plane, had received little attention, at least from the experimental point of view. Although a great deal of acoustic and geophysics literature exists about shear horizontal surface elastic waves [210] and although the corresponding mathematical treatment is easier than that needed to treat sagittal waves, SH thermal waves (phonons) have been considered sources of surface Brillouin scattering in a very limited number of papers and only in the case of supported films.

SH waves being polarized parallel to the surface, the mechanism of their interaction with the light is only the volume elasto-optic effect and no interference phenomena coming from scattering of the thermal surface ripple (as in the case of sagittal modes) can appear in the spectrum.

Calculations of Brillouin scattering cross sections for SH surface acoustic modes (in the discrete part of the spectrum these modes are called Love modes) have been performed by Bortolani et al. [220], by Albuquerque et al. [221], and Mayer [222], who performed a full quantum-mechanical computation of the cross section for dielectric layered media on a metallic substrate, but at that time these authors could not compare their theoretical predictions with any experimental data.

To our knowledge there are only two reported experimental works on the detection of SH surface acoustic phonons using Brillouin scattering. The first was conducted by Bell et al. [223] in Nb/Cu superlattices on a sapphire substrate. Those authors performed their measurements by detecting the tails of the acoustic modes in the sapphire substrate and gave no evidence of SH acoustic modes in the continuum part of the spectrum.

The second, by Bottani et al. [200], announced the observation of SH surface acoustic modes in both the discrete and continuum parts of the spectrum in silicon on insulator (SOI) structures (see Fig. 31). There the theory was limited to the computation of the LPPDS for the SH polarization and the Brillouin cross section was not evaluated.

Thin buried oxide (a-SiO_x) layers (BOX) have been prepared in recent years in crystalline silicon wafers by different techniques. Particularly interesting is the case of SOI-SIMOX structures (silicon on insulator structures obtained by separation of implanted oxygen techniques) analyzed by Ghislotti and Bottani [201].

In that paper the authors investigated surface Brillouin scattering in a silicon on insulator structure composed of a silicon dioxide layer buried in a Si(001) substrate. The structure was produced by low-dose oxygen implantation and subsequent high-temperature treatment of a Si(001) wafer.

The thicknesses of the buried oxide L and of the top silicon layer d were measured by cross-sectional transmission electron microscopy to be, respectively, either 110 and 350 nm, as in Bottani and Ghislotti [200], or 80 and 360 nm. Observations were conducted for phonons propagating both along $\langle 100 \rangle$ and $\langle 110 \rangle$.

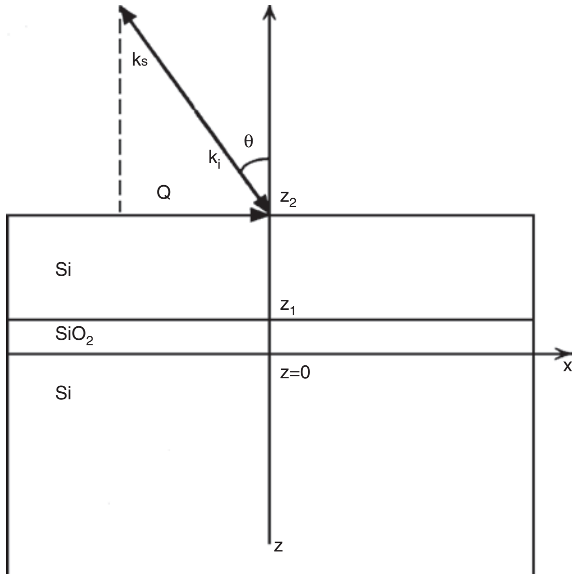


Figure 31. Silicon on an insulator structure: sample geometry and scattering configuration. The x axis is parallel to the surface phonon propagation wave vector, and the z axis is normal to the surface. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

As far as acoustic phonon localization is concerned, the key physical point is that the buried layer has lower sound velocities than silicon: thus either the BOX or the Si–SiO_x bilayer acts as an acoustic wave guide for surface phonons with respect to the bulk Si substrate and depth-dependent confinement effects can be detected.

As the treatment of SH localization is the easiest one still showing quite general characteristics, we treat it in great detail and just briefly discuss the sagittal phonon case at the end. In this way the general features of surface Brillouin scattering, summarized in the introductory part, are fully exemplified.

(i) We discuss the spectrum of surface SH phonons in the SOI-SIMOX structure presenting the LPPDS for SH polarization.
(ii) We show the fluctuating polarization vector field P_y produced by elasto-optic coupling when SH phonons modulate the dielectric function of the medium. The knowledge of P_y implies the computation of the incident p-electric field transmitted in the body.
(iii) The theoretical results are compared with the experimental Brillouin spectra and an extension is made to sagittal phonons and their confinement behavior.

6.7.1. The Layer-Projected Spectral Density of SH Phonons

We assume that the vacuum–Si(001) interface coincides with the $z = 0$ plane and that the z axis points downward in the medium. The first Si/SiO_2 interface is at $z = d$ and the second SiO_2/Si interface is at $z = L + d$. For SH waves propagating along any direction parallel to the $y = 0$ plane (which is both the sagittal and the scattering plane in the experiments) the relevant displacement field component is then u_y. In particular Love waves, that is, true SH surface waves belonging to the discrete spectrum and exponentially decaying in the Si substrate, propagate along the x direction. We consider only $\langle 100\rangle$ and $\langle 110\rangle$ propagation directions for which the sagittal motion is decoupled from the SH one. The wave equation for u_y when the x axis coincides with the $\langle 100\rangle$ direction is written as

$$\rho(z)\frac{\partial^2 u_y}{\partial t^2} = C_{44}(z)\frac{\partial^2 u_y}{\partial x^2} + \frac{\partial}{\partial z}\left[C_{44}(z)\frac{\partial u_y}{\partial z}\right] \qquad (79)$$

where the functions $\rho(z)$ and $C_{44}(z)$ take within each layer the constant value of the mass density and C_{44} elastic constant of the corresponding material.

Since the system is translationally invariant in the x direction parallel to the surface, we introduce the $(\omega, q_\parallel)$ Fourier component of the u_y SH displacement field, the parallel wave vector $\mathbf{q}_\parallel = q_\parallel \hat{\mathbf{e}}_x$, as

$$u_y(\omega, q_\parallel; x, z, t) = \xi(\omega, q_\parallel)\phi_y(\omega, q_\parallel, z)\exp[i(q_\parallel x - \omega t)] \qquad (80)$$

where $\xi(\omega, q_\parallel)$ is the normal coordinate of the SH phonon $(\omega, q_\parallel)$. Introducing (80) in the wave equation (79) we obtain the self-adjoint Liouville equation [226]

$$\frac{d}{dz}\left[C_{44}(z)\frac{d\phi_y(\omega, q_\parallel, z)}{dz}\right] + [\rho(z)\omega^2 - C_{44}(z)q_\parallel^2] \times \phi_y(\omega, q_\parallel, z) = 0 \qquad (81)$$

The mode z profiles $\phi_y(\omega, q_\parallel, z)$ are the real eigenfunctions of Eq. (81) corresponding to the real eigenvalues $\omega^2 = \omega^2(q_\parallel)$, the SH phonon eigenfrequencies. The $\phi_y(\omega, q_\parallel, z)$s play, in the continuum model adopted here, the same role of the polarization unit vectors in lattice dynamics. A similar Liouville equation is obtained for the $\langle 110\rangle$ case substituting $\frac{1}{2}[C_{11}(z) - C_{12}(z)]$ to $C_{44}(z)$ as the multiplying coefficient of $q_\parallel^2$ in (81).

Using a slab approximation [224] to solve the above spectral problem (in this way the whole spectrum becomes discrete) we impose the normalization conditions [226]

$$\int_0^h \rho(z)\phi_y^2(\omega, q_\parallel, z)\, dz = 1 \qquad (82)$$

the slab thickness being $h \gg L + d$.

Also, we assume stress-free boundary conditions which are written in terms of the vanishing of the z derivatives of the mode profiles at both the slab outer surfaces

$$\left(\frac{d\phi_y(\omega, q_\parallel, z)}{dz}\right)_{z=0} = \left(\frac{d\phi_y(\omega, q_\parallel, z)}{dz}\right)_{z=d} = 0 \qquad (83)$$

Together with (82) and (83) Eq. (81) constitutes a well-posed singular (at least in the case of sharp interfaces) Sturm–Liouville eigenvalue problem [227] with strong analogies with 1D quantum mechanics.

In principle the spectrum of SH long-wavelength acoustic phonons in a semi-infinite (when the slab thickness d goes to infinity) layered medium is the union of a discrete and a continuous part. The latter starts at the transverse threshold of the Si substrate $\omega_t = c_t q_\parallel$, where c_t is the shear horizontal sound velocity of Si in the appropriate propagation direction. In fact, at fixed $q_\parallel$, only for $\omega > \omega_t$ the partial plane waves in the unbounded Si substrate have a real perpendicular wave vector $q_\perp = (\omega^2/c_t^2 - q_\parallel^2)^{1/2}$ as it must be for a nondecaying bulk wave. Thus, strictly speaking, only the discrete eigenvalues correspond to true surface modes or Love waves (bounded states or resonances). Yet in the continuous spectrum (where the mode profiles are not decaying in the substrate, corresponding to SH bulk waves reflecting at the surface) one can also find considerable structure of surface character or pseudo-Love waves (quasi-resonances or pseudo-surface modes). These last vibrational states share some formal properties with virtual states in the continuum found in quantum scattering theory.

The adoption of the slab approximation produces only a discrete spectrum but the (still infinite but now numerable) eigenfrequencies separate into two classes: a limited number of eigenvalues with $\omega < \omega_t$ still correspond to the Love waves, while all the other eigenvalues tend to accumulate from above toward ω_t and build a quasi-continuous spectrum provided the slab is sufficiently thick. It is then possible to compute the density of phonon states and to simulate in this way the true continuous spectrum of a semi-infinite medium.

The more or less surface character of these collective excitations can be judged from inspection of the LPPDS for the SH polarization

$$g_{yy}(\omega, q_\parallel | z, z') = \sum_\alpha \sqrt{\rho(z)\rho(z')}\,\phi_y(\omega_\alpha(q_\parallel), q_\parallel, z) \times \phi_y(\omega_\alpha(q_\parallel), q_\parallel, z')\delta[\omega - \omega_\alpha(q_\parallel)] \qquad (84)$$

computed for $z = z'$

$$g_{yy}(\omega, q_\parallel|z, z) = \sum_\alpha \rho(z)\phi_y^2(\omega_\alpha(q_\parallel), q_\parallel, z)\delta[\omega - \omega_\alpha(q_\parallel)] \quad (85)$$

Although in general $g_{yy}(\omega, q_\parallel|z, z')$ in the case of Brillouin light scattering is not directly related to the cross section as it is instead in the case of low energy electron surface scattering [224], this quantity is, in any case, useful in investigating the surface localization of the phonon modes (see below).

Introducing the retarded and advanced Green functions of Eq. (81) $G_{yy}(\omega \mp i\epsilon, q_\parallel; z, z')$ as the solutions of the nonhomogeneous equation

$$\frac{d}{dz}\left[C_{44}(z)\frac{dG_{yy}}{dz}\right] + \left[\rho(z)(\omega \mp i\epsilon)^2 - C_{44}(z)q_\parallel^2\right]G_{yy} = \delta(z - z') \quad (86)$$

with the same boundary conditions as for the ϕ_y, a smooth representation $\bar{g}_{yy}(\omega, q_\parallel|z, z)$ of the LPPDS for SH phonons can be obtained as

$$\bar{g}_{yy}(\omega, q_\parallel|z, z) = \frac{\rho(z)}{2\pi i}[G_{yy}(\omega - i\epsilon, q_\parallel; z, z) - G_{yy}(\omega + i\epsilon, q_\parallel; z, z)] \quad (87)$$

Once a sufficiently high number of eigenfrequencies and eigenfunctions of Eq. (81) has been obtained numerically [200] the Green functions can be computed by means of the spectral expansion [226]

$$G_{yy}(\omega \mp i\epsilon, q_\parallel; z, z) = \sum_\alpha \frac{\phi_y^2(\omega_\alpha(q_\parallel), q_\parallel, z)}{(\omega \mp i\epsilon)^2 - \omega_\alpha^2(q_\parallel)} \quad (88)$$

6.7.2. Electromagnetic Wave Equations for the Incident and Scattered Fields for p–s Scattering

Neglecting thermal fluctuations silica and crystalline silicon have isotropic dielectric properties. The perturbation caused by long-wavelength acoustic phonons can be accounted for by means of an istantaneous anisotropic susceptibility (see above), a second order tensor the components of which are linear functions of the fluctuating elastic strains. Because of dispersion a simple constitutive law relating the polarization vector **P** to the electric field can be written only for monochromatic components. Taking into account the linearity of Maxwell equations and the smallness of thermal elastic strains the electromagnetic problem can be solved by first order perturbation theory (the Born approximation in scattering theory) [218].

At 0th order (neglecting the fluctuating part of the susceptibility and thus the scattered field) the electric field $\mathbf{E}^{\omega_0}(x, z)$ transmitted in the medium can be conveniently computed in terms of the y component of the magnetic induction field $B_y^{\omega_0}(x, z) = e^{i(k_\parallel x)}B(z)$, where $k_\parallel = (2\pi/\lambda)\sin\theta_i = (\omega_0/c)\sin\theta_i$ is the component parallel to the surface of the wave vector of the incident wave [228]. $B(z)$ satisfies the ordinary differential equation

$$\frac{d}{dz}\left(\frac{1}{\epsilon(z)}\frac{dB}{dz}\right) + \left(\frac{\omega_0^2}{c^2} - \frac{k_\parallel^2}{\epsilon(z)}\right)B = 0 \quad (89)$$

where $\epsilon(z)$ is the z profile of the complex relative dielectric function of the structure at frequency ω_0. In the ideal case $\epsilon(z)$ has a finite jump at each interface where the continuity of $B(z)$ and $\epsilon^{-1}(z)(dB/dz)$ is imposed and keep a constant value in each layer. In particular in the vacuum $(z < 0)$ above the surface $B(z) = (E_0/c)(e^{i(k_\perp z)} - r_p e^{-i(k_\perp z)})$, E_0 being the electric field amplitude of the incident p wave, $k_\perp = (\omega_0/c)\cos\theta_i$ is the component perpendicular to the surface of the wave vector of the incident wave and r_p the reflection coefficient. Once $B(z)$ has been computed $E_x^{\omega_0}(x, z)$ and $E_x^{\omega_0}(x, z)$ can be obtained as [228]

$$E_x^{\omega_0} = -i\frac{c^2}{\omega_0\epsilon(z)}\frac{\partial B_y^{\omega_0}}{\partial z} = -i\frac{c^2}{\omega_0\epsilon(z)}\frac{dB}{dz}e^{i(k_\parallel x)} \quad (90)$$

$$E_x^{\omega_0} = i\frac{c^2}{\omega_0\epsilon(z)}\frac{\partial B_y^{\omega_0}}{\partial x} = -\frac{c^2 k_\parallel}{\omega_0\epsilon(z)}B(z)e^{i(k_\parallel x)} \quad (91)$$

Now the next step is to compute the fluctuating part of the polarization vector in the medium. The first order result in perturbation theory is

$$P_i^{\omega_s} = \epsilon_0(\epsilon(z) - 1)E_i^{\omega_s} + \epsilon_0\delta\chi_{ij}(\omega_0| \pm \omega_\alpha(q_\parallel))E_j^{\omega_0} \quad (92)$$

In the above equation $\epsilon(z) - 1 = \chi(z)$ is the unperturbed isotropic susceptibility in the absence of phonons (because $\omega_\alpha(q_\parallel) \ll \omega_0$ we assume that $\epsilon(z|\omega_0 \pm \omega_\alpha(q_\parallel)) \approx \epsilon(z|\omega_0) = \epsilon(z)$ and $\delta\chi_{ij}(\omega_0| \pm \omega_\alpha(q_\parallel))$, the anisotropic fluctuating part of the susceptibility due the excitation of a single SH $(\omega_\alpha(q_\parallel), q_\parallel)$ phonon mode. The second term in the r.h.s. of Eq. (14) is responsible for the radiation of the Brillouin light, that is, of the scattered field $\mathbf{E}^{\omega_s}$ at frequencies $\omega_s = \omega_0 \pm \omega_\alpha(q_\parallel)$.

Taking explicitly into account the elasto-optic coupling one finds that the part of the fluctuating polarization vector radiating the scattered field has only one component $(P_y^{\omega_s})_R$ which can be written in terms of the fluctuating thermal elastic strains

$$u_{yx} = (1/2)[\partial u_y(\omega_\alpha, q_\parallel)/\partial x]$$

and

$$u_{yz} = (1/2)[\partial u_y(\omega_\alpha, q_\parallel)/\partial z]$$

as [200]

$$(P_y^{\omega_s})_R = \epsilon_0 k_{44}(z)[u_{yx}(\omega_\alpha, q_\parallel)E_x^{\omega_0} + u_{yz}(\omega_\alpha, q_\parallel)E_z^{\omega_0}] \quad (93)$$

for $\langle 100\rangle$, and as

$$(P_y^{\omega_s})_R = \epsilon_0\left[\frac{1}{2}(k_{11}(z) - k_{12}(z))u_{yx}(\omega_\alpha, q_\parallel)E_x^{\omega_0} + k_{44}(z)u_{yz}(\omega_\alpha, q_\parallel)E_z^{\omega_0}\right] \quad (94)$$

for ⟨110⟩, where the $k_{ij}(z)$ are the z profiles of the relevant elasto-optic coefficients. Thus Brillouin scattering of an incident p electromagnetic wave of a pure SH phonon produces a scattered s electromagnetic wave, that is, scattering rotates the polarization of 90°.

With the source of Eqs. (4.15) and (4.16) the radiation problem can be solved and the Brillouin cross section computed.

6.7.3. Experimental Results and Discussion of SH Phonon Localization

A Brillouin spectrum taken along ⟨100⟩ at $\theta = 40°$ is presented in Figure 32. Only fully depolarized light was measured (p–s scattering). Two peaks are clearly visible: they were identified as being due to scattering from SH acoustic modes. If only fully polarized light (p–p scattering) were measured a third different peak of sagittal character, due to the Rayleigh wave propagating at the free surface, would be observed alone in the same spectral region (see below [232, 233]). Only the elasto-optic coupling mechanism gives contribution to the p–s cross section.

The solid line in Figure 32 is the calculated p–s Brillouin differential cross section convoluted with a Gaussian of FWHM of 200 MHz to simulate the finite experimental spectral resolution. In the calculation the authors used the dielectric (ϵ), elastic (C_{ij}), and elasto-optic constants of bulk Si and SiO_2. The Si constants are $\epsilon = 18.5 + 0.52i$, $C_{11} = 166$ GPa, $C_{12} = 63.9$ GPa, $C_{44} = 79.6$ GPa, $K_{11} = 53.2$, $K_{12} = 25.0$, $K_{44} = 23.4$, $\rho = 2330$ kg^{-3}. The SiO_2 constants are $\epsilon = 2.16$, $C_{11} = 78.5$ GPa, $C_{44} = 31.2$ GPa, $K_{11} = -0.55$, $K_{44} = 0.345$, $\rho = 2200$ kg^{-3}. The theory does not contain any free parameter.

The experimental spectra show an increase in the magnitude of the lower frequency peak relatively to the higher one for decreasing incidence angles. This is in agreement with the calculated intensities and can already be explained by the analysis of plots of the LPPDS as a function of depth and frequency computed at the same angles of incidence (see below).

In Figure 33 a contour plot of $g_{yy}(\omega, q_\parallel|z, z)$ is shown for a parallel wave vector $q_\parallel \approx 0.0157$ nm^{-1}. Since in a surface Brillouin scattering event the total parallel wave vector of the photon–phonon system is conserved and these experiments where conducted in backscattering (see above), $q_\parallel = (4\pi/\lambda)\sin\theta_i$, where $\lambda = 514.5$ nm was the wavelength in a vacuum of the incident photon and θ_i the incidence angle. In the case of Figure 33 this corresponds to $\theta_i = 40°$. Inspection of Figure 33 shows clearly the nature of the guided mode of the Love wave below ω_t. The mode is highly localized within the silica buried layer whose transverse sound velocity is lower than that of silicon. Also clearly visible is the pronounced surface character of the pseudo-mode in the continuum above ω_t. In Figure 34a the surface-projected SH phonon density of states $SPPDS = g_{yy}(\omega, q_\parallel|0, 0)$ is shown for four different values of θ_i (or $q_\parallel$). It can be seen that the Love wave tends to become less confined in the silica layer for small $q_\parallel$ (its surface intensity increases), while the pseudo-Love wave has the highest surface intensity for $\theta_i \approx 40°$. In Figure 34b $g_{yy}(\omega, q_\parallel|z, z)$ is shown for $z = 400$ nm about in the middle of the silica layer. Only the peaks corresponding to the Love wave are clearly visible. One can see the increasing of the peak amplitude as $q_\parallel$ increases: the mode becomes more and more confined (guided) in the silica layer.

Figure 35 shows the mean square value of $(P_y^{\omega_s})_R$ corresponding to $\theta_i = 40°$ when either the Love mode is excited (the first phonon eigenvalue: the unique in the discrete spectrum for our geometry) or the pseudo-mode in the continuum at a frequency corresponding to the second peak in Figure 32.

Although the Love wave is essentially a wave guided in the silica buried layer, it is evident that the source of the Brillouin light is mainly located in the subsurface Si layer. This is due to the fact that the k_{44} elasto-optic coefficient of Si is much bigger than the corresponding quantity of silica

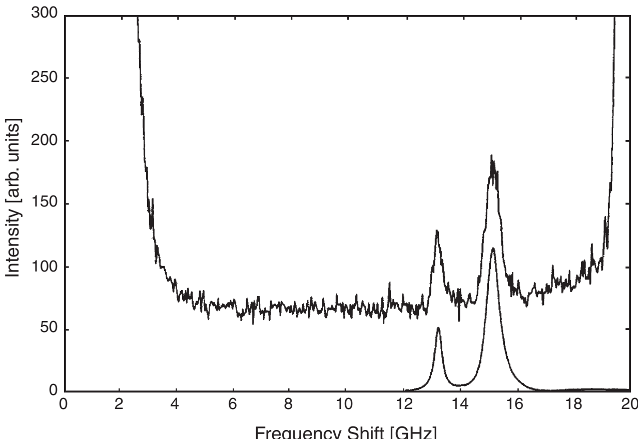


Figure 32. Measured p–s Brillouin anti-Stokes spectrum (top line) and theoretical p–s cross section (bottom line) along ⟨100⟩ for $\theta_i = 40°$. The peak tail immediately at the right of 0 shift belongs to the laser line, while that growing near 20 GHz is due to an instrumental ghost of the interferometer. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

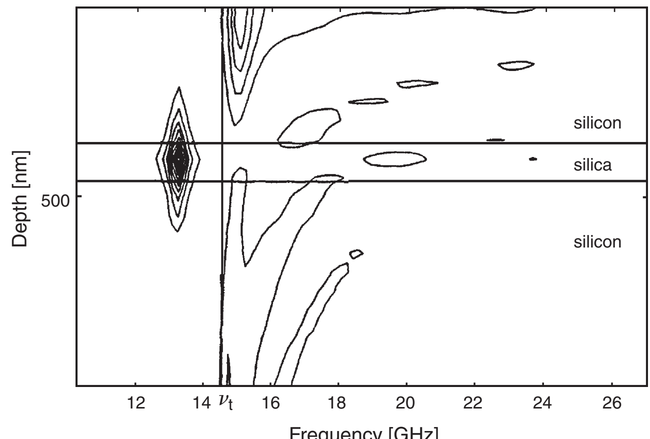


Figure 33. Contour plot of the LPPDS for the SH polarization as a function of depth and frequency for $\theta_i = 40°$ and propagation direction ⟨100⟩. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

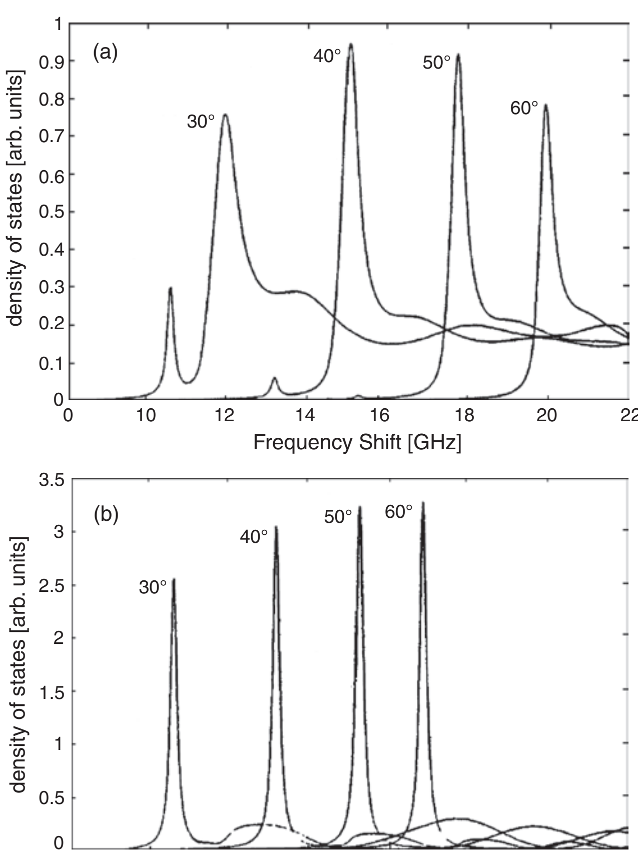


Figure 34. SH surface-projected phonon density of states (a) and LPPDS computed in the middle of the silica buried layer (b) for the same incidence angle as in Figure 3. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

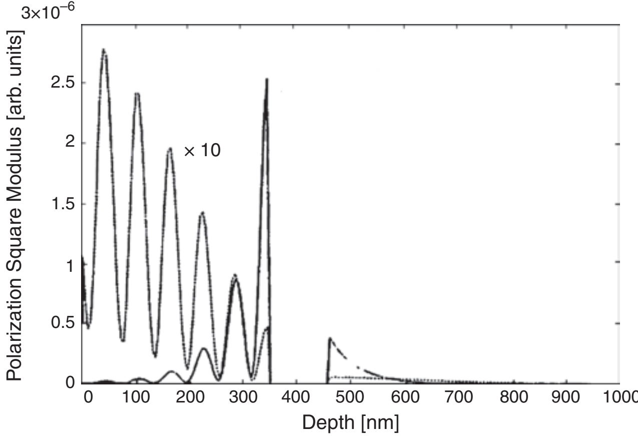


Figure 35. Mean square polarization radiating the Brillouin light vs z at the frequency of the Love (full line) and of the pseudo-Love (dashed line) waves for $\theta_i = 40°$ and propagation direction $\langle 100 \rangle$. It is seen that the source of the Brillouin light is confined in the top Si layer immediately below the surface. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

(see above) and because the incident electric field transmitted in the medium is maximum there. This is even more true for the pseudo-Love wave (in the continuous spectrum) the mode profile of which has a strong surface character. The exponential vanishing of the source for $z \to \infty$ in the Si substrate is simply explained in terms of the vanishing of $B(z)$ due to absorption and leads to the conservation of only the parallel total wave vector in the scattering integral (see below).

The theory presented above was generalized by Bottani and Caporali [202] and by Caporali et al. [203] to include the effect of smooth interfaces and, more in general, of layers with nonuniform acousto-optic properties.

6.7.4. Experimental Results and Discussion of SV and L Phonon Localization

The buried oxide and the bilayer made by top silicon and buried oxide (as well as any other system with similar acoustic properties) can also partially confine (localize) sagittal phonons visible in a p–p scattering experiment. Here the phenomenology is richer for, even along a high symmetry direction, the propagation of such phonons must be described with a two-component displacement vector (u_x, u_z) obeying a system of coupled wave equations. This rules out the analogy with simple scalar 1D quantum mechanics and requires a specific treatment of the spectral problem. The u_x component has a longitudinal (L) character, while the u_z component has a shear vertical (SV), transverse, character. Among sagittal pseudo-surface phonons a special role is played by the HFPSW (high-frequency pseudo-surface wave), or LR (longitudinal resonance), in the continuous spectrum and by the Rayleigh and Sezawa phonons in the discrete spectrum. The LR is a pseudomode of L character propagating along a free surface with a phase parallel velocity $\omega_{LR}/q_\parallel = v_L$ coinciding with the longitudinal bulk velocity. If the surface has a slower overlayer the LR tends to be confined in the overlayer and is called the longitudinal guided mode. The SOI-SIMOX structures have a bilayer on the top of the silicon wafer and the buried oxide is slower than both top and bulk silicon. This produces quite intricate Brillouin spectra in the continuum. The Rayleigh phonon is a true surface wave of both SV and L character. It exists beneath pure free surfaces and also survives in the presence of an overlayer: in this case it is called a modified Rayleigh wave. If the overlayer is slower than the substrate other guided true surface phonons can exist at higher frequencies: they are called Sezawa waves. We notice that for the SH polarization there are no true surface waves beneath pure free surfaces below the transverse threshold: a slower overlayer is required for their existence (the modified Love waves; see above). In SOI-SIMOX structures a mixing of Rayleigh and Sezawa waves can take place, giving rise to peculiar Brillouin spectra even in the discrete. The theory of Brillouin scattering from sagittal surface waves in buried oxides is due to Byloos et al. [231], while the first experiments are due to Nizzoli et al. [232] and Ghislotti et al. [233]. Figures 36 and 37 show experimental spectra and theoretical cross sections

Figure 36. (a) Anti-Stokes Brillouin spectra for different incidence angles taken along ⟨100⟩ in p–p polarization. (b) Theoretical cross sections evaluated at the same incidence angles. Reprinted with permission from [177], P. Milani and C. E. Bottani, “Handbook of Nanostructured Materials and Nanotechnology” (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

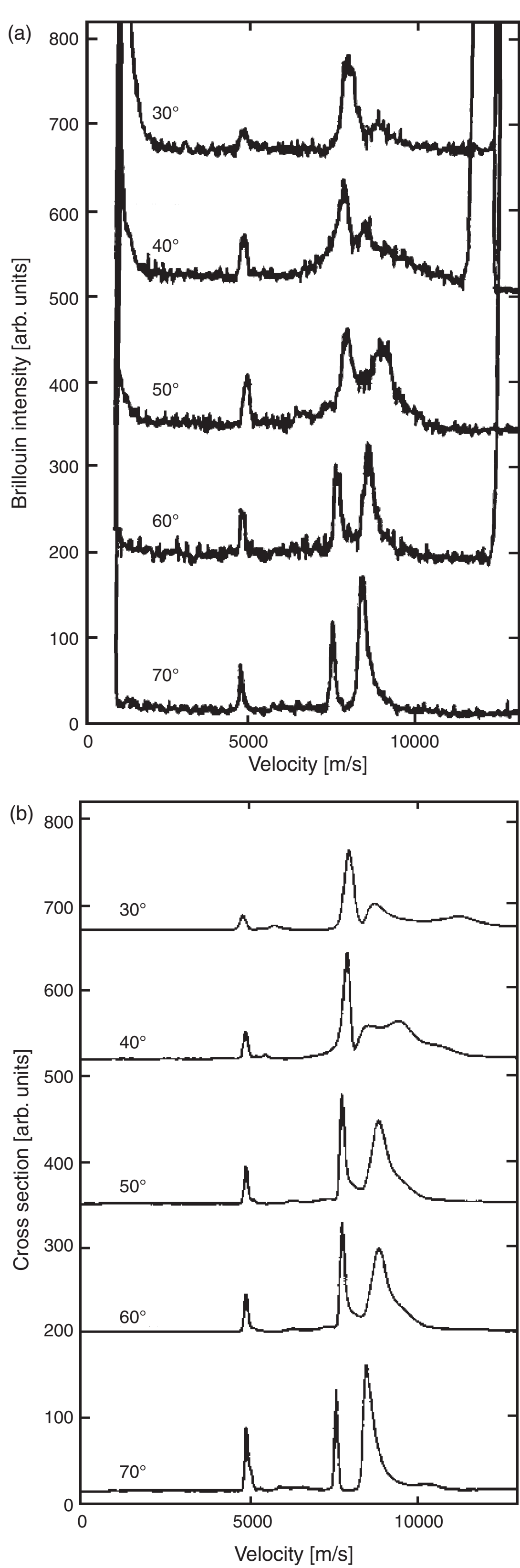
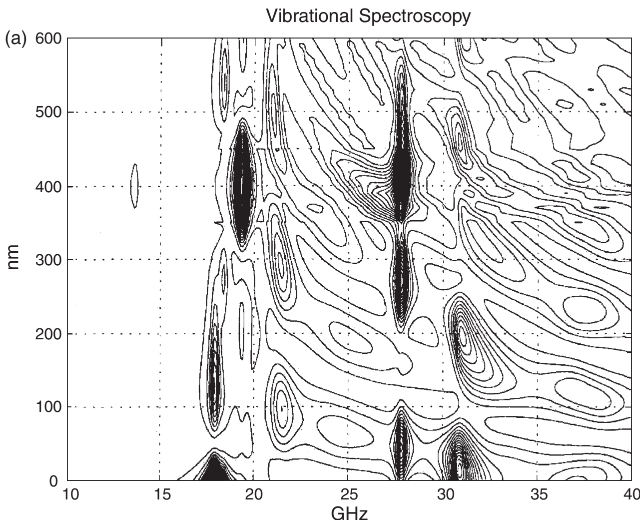


Figure 37. (a) Brillouin spectra along the ⟨100⟩ direction and (b) calculated cross sections for an incidence angle of 70° for samples of different top silicon layer thickness. Reprinted with permission from [177], P. Milani and C. E. Bottani, “Handbook of Nanostructured Materials and Nanotechnology” (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

at various incidence angles and for different BOX thicknesses. The localization of the observed phonons is shown for both SV and L polarization in Figure 38 by means of the spectral function LPPDS. In this figure it is evident that each mode (phonon) has a peculiar localization behavior: the confinement is more prominent in the top overlayer, or in the buried oxide, or in the substrate. This observation suggested the idea of using the measured dispersion relationships of each mode to get space-resolved information on real imperfectly layered structures. For a general review of Brillouin scattering in supported films the reader is invited to consult Nizzoli and Mills [215]. In the case of buried layers the space-selective propagation properties of surface acoustic phonons were used by Ghislotti et al. [234] and Beghi et al. [235] to detect the presence of silicon inclusions in the BOX and to measure their concentration in a completely nondestructive way. The results were confirmed by TEM.



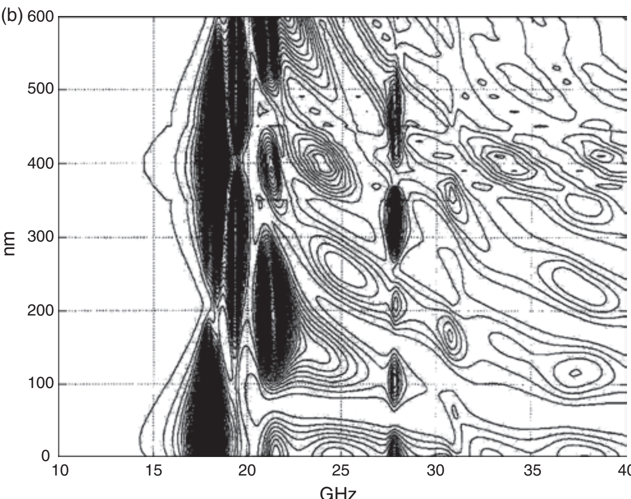


Figure 38. LPPDS for sagittal phonons of a SOI-SIMOX structure: the buried oxide layer is situated between 350 and 450 nm of depth. (a) Longitudinal polarization; (b) shear vertical polarization. Reprinted with permission from [177], P. Milani and C. E. Bottani, "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.) (2000). © 2000, Elsevier.

When examining the ability of surface Brillouin scattering to measure local elastic constants of thin films two major observations are in order. First, the quality of the results depends strongly on the independent knowledge of the film density. Second, the thickness of the film must be known. To keep the whole procedure completely nondestructive Beghi et al. made an integrated use of Brillouin scattering, X-ray reflectivity, and high-resolution X-ray diffraction [205]. In this way it was possible to measure with great precision the small effect of alloying low Cu and Ni concentrations in a thin Au film on silicon. It is in any case worth noting that, for films with a thickness less than 1 μm, surface Brillouin scattering is the only nondestructive technique that can be used to measure the elastic properties of the film.

6.8. Surface Melting in Nanoparticle Films

The abrupt change in velocity in surface acoustic waves in thin films of amorphous SiO_x containing nanometer-scale β-Sn crystals has been shown to be directly associated with the size-dependent melting of the nanoparticles, thus confirming preliminary experiments. High-resolution thin film powder diffraction using synchrotron radiation has shown that the abrupt red shift in the Brillouin spectra satellites occurs at the same temperature as the melting of the nanoparticles, evident for the loss of the Bragg peaks. Effective medium theory has been used to explain the origin of the anomaly. A central peak in the Brillouin spectrum, the intensity of which showed a maximum at the melting temperature, can be interpreted in terms of overdamped fluctuations in the dielectric function. The melting temperature as a function of particle size is in agreement with theoretical predictions. No evidence of strain could be found the X-ray diffraction profiles, the *a* and *c* axis thermal expansion coefficients being the same as those in bulk tin [236, 237].

6.9. Carbon Mesostructures

Experimental characterization and theoretical modeling of cluster-assembled materials have to face the problem of cluster coalescence and of their organization in structures spanning length scales from the nanometer up to the micrometer. The different structures in which the precursor clusters are organized need experimental probes sensitive to the different length scales typical of intracluster and intercluster interactions. For carbon-based materials, as we have already discussed, Raman spectroscopy can be used for characterization on a nanometer scale [193]. In order to study the organization of clusters on a scale of hundreds of nanometers, which is the typical scale of thermally excited long-wavelength acoustic phonons, Brillouin light scattering was used for the first time by Bottani et al. [209]. Films of graphite, polycrystalline diamond, diamondlike a-C:H, C_{60} (fullerite), and phototransformed C_{60} have also been studied by Brillouin scattering [193]. Normally SBS is applied to homogeneous compact films with perfect (atomically flat) surfaces and buried interfaces [205]. Particularly challenging is instead the attempt to get and interpret Brillouin spectra of films with a rough surface and/or a granular or

porous structure. The extraction of the elastic properties of these systems from the spectroscopic data can be based only partially on what is known in the case of good films and a complete theory is still lacking. Bulk and surface Brillouin scattering- signals have been obtained from films characterized by a complex structure from the atomic to the hundreds of nanometers level, showing that this technique can also be used for nanostructured materials with irregular surfaces. Bulk and shear modulus of the material have been determined giving information on the acoustic properties on a mesoscopic scale. This allows us to infer the nature of the bonding between the carbon aggregates.

Thick films (thickness, $\geq 1\ \mu$m) and thin films (down to a thickness of 20 nm) have been examined. The results depend strongly on the presence (or the absence) of a beam focalizer selecting (or not) only clusters of small size for deposition [210].

6.9.1. Films Grown by an Unfocused Beam

In thick films only damped bulk acoustic phonons with a typical wavelength λ_{ph} of the order of 170 nm have been detected. This indicates that for a length $d \geq \lambda_{ph}$ the film bulk can be modeled as a continuum [196] with approximate translational invariance and effective elastic constants, although structural disorder at smaller scales scatters the phonons significantly. The presence of a rather strong central peak in the spectra could be ascribed to nonpropagating (overdamped) or confined vibrational excitations, probably connected with different characteristic correlation lengths less than d. The most damped bulk acoustic phonons could be coupled to the confined modes by a relaxation mechanism. The surface phonons could be almost overdamped if not already replaced by surface fractons or by localized cluster modes [211]. Instead in thinner films, notwithstanding the high degree of surface roughness, which increases with film thickness, intense surface peaks can be detected up to a critical thickness of the order of 200 nm (Fig. 39). To make the SBS spectrum measurable the films must be grown on a ultrasmooth high-reflecting substrate to maximize the buried interface ripple scattering [212].

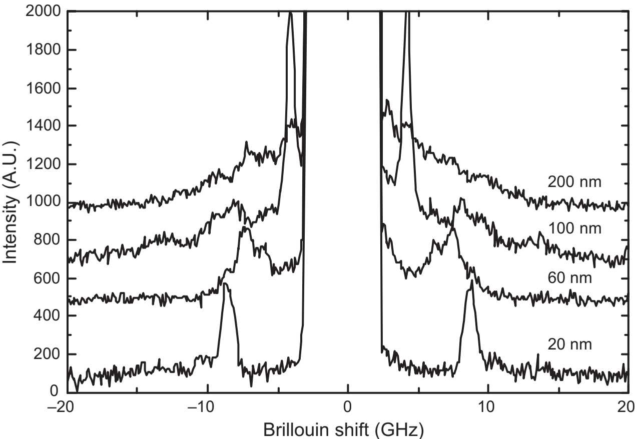


Figure 39. Brillouin spectra of unfocused beam grown films. Reprinted with permission from [286], C. E. Bottani, "Mechanical Properties of Low Dimensional Carbon Structures," "Nanostructured Carbon for Advanced Applications" (G. Benedek, P. Milani, and V. G. Ralchenko Eds.). 2001. © 2001, Kluwer Academic.

6.9.2. Films Grown by a Focused Beam

Films with a completely different nanostructure, more compact and with a smoother surface, can be grown using the focalization device. Typical spectra of a focalized thick film (1 μm) are shown in Figure 40. The study of a thick film with respect to a thin film gives an advantage: we can consider the film a semi-infinite medium and so the nature and the characteristics of the acoustic waves propagating in the system are not affected by the thickness (often not known with enough accuracy) and thus by the presence of the substrate.

Figure 40 shows the Brillouin spectra of a focalized thick film taken at two different angles of incidence, $\theta_i = 50°$ and $\theta_i = 70°$. The shift in the first two peaks scales with θ_i: both peaks belong to the surface spectrum. The first peak corresponds to a Rayleigh phonon (RW) and the second, the LR, is a general feature of the continuous spectrum whose shift scales as $v_l 4\pi \sin\theta_i/\lambda_0$. The LR is a so-called pseudo-surface phonon. It is a wave packet of bulk phonons with a strong surface localization of the longitudinal polarization behaving as a surface wave. Because of its longitudinal character it gives no contribution to ripple scattering. The third peak is due to a longitudinal bulk phonon whose shift is independent of θ_i and scales as $v_l 4\pi n/\lambda_0$. In the small inset boxes the fit of the $\theta_i = 50°$ spectrum is shown: the numerical simulation of the scattering total cross section (line) is superimposed to the experimental points in the right box, while in the left one the RW peak (discrete spectrum) was fitted separately with a pseudo-Voigt function. These spectra are interesting mainly for the high intensity and small width of the Rayleigh peak, never observed before in cluster-assembled carbon films. This feature supports the conclusion that long life acoustic phonons can propagate along the film surface; in other words, at a mesoscopic scale of observation this film appears as a homogeneous elastic continuum

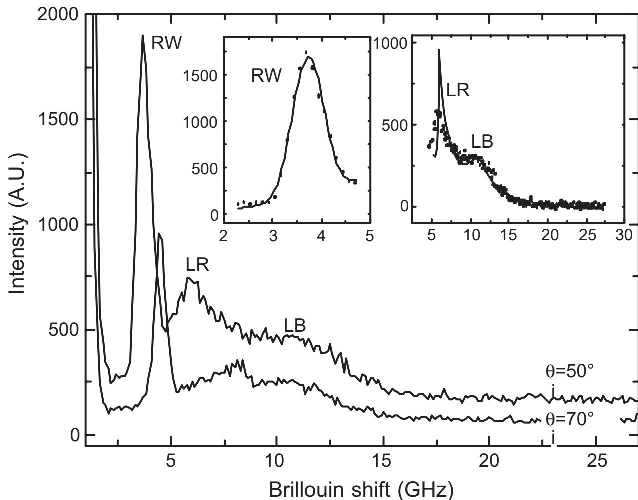


Figure 40. Brillouin spectra of a thick focused beam grown film. Reprinted with permission from [286], C. E. Bottani, "Mechanical Properties of Low Dimensional Carbon Structures," in "Nanostructured Carbon for Advanced Applications" (G. Benedek, P. Milani, and V. G. Ralchenko Eds.). 2001. © 2001, Kluwer Academic.

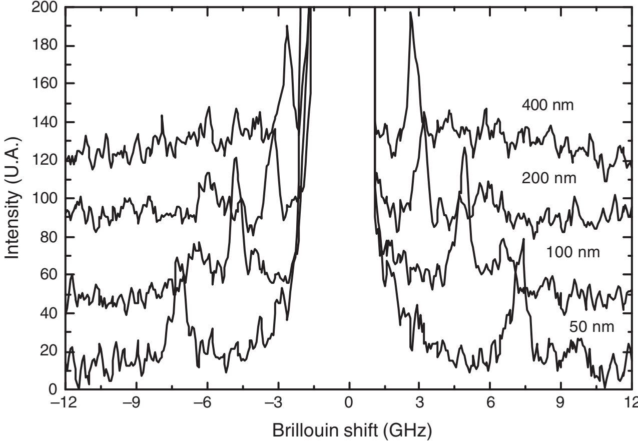


Figure 41. Brillouin spectra of focused beam grown films. Reprinted with permission from [286], C. E. Bottani, "Mechanical Properties of Low Dimensional Carbon Structures," in "Nanostructured Carbon for Advanced Applications" (G. Benedek, P. Milani, and V. G. Ralchenko Eds.). 2001. © 2001, Kluwer Academic.

and the relatively low surface roughness, acting as surface structural disorder, introduces only a small acoustic damping factor without hindering the phonon propagation (as in the case of nonfocalized films). This fact is consistent with a smooth surface film of compact material made of close-packed small clusters and it is important in comparison with the results of BS experiments on the deposited without beam focalization in which the surface roughness increases as the film thickness. These findings are confirmed measuring the SBS spectra of films of increasing thickness. The results are shown in Figure 41.

Here distinct surface peaks are clearly visible for thickness in the range of 50–400 nm. Measuring the dispersion relations of the surface phonons the elastic constants of the films can be estimated. The best results are obtained for the Young modulus E, as demonstrated by a detailed sensitivity analysis [213]. Using for the density ρ values coming from X-ray reflectivity [205] (all values are in the range 1–1.3 g/cm^3) the found elastic constants are, typically, of the order $c_{11} = 5.4$ GPa, $c_{44} = \mu = 2.5$ GPa, $E = 4.6$ GPa, $B = 1.9$ GPa, and $\nu = 0.1$. These values are comparable to those obtained from the thin films (thickness of about 100 nm or less) deposited with an unfocused beam [200]. In particular, they are near the values of the harder of this thin film series. The comparison with diamond elastic constants indicates a very soft and highly deformable material with a shear modulus (for shear between graphene planes) near that of graphite. These results can be compared with Raman measurements [208] which point out the mainly sp^2 carbon bonding present in the disordered granular structure of the films on a nanometer scale. Yet the film material is not elastically identical to nanocrystalline graphite: in fact the values of B (and, consequently, ν) are significantly different.

ACKNOWLEDGMENT

The authors thank the National Research Project FIRB Micro-strutture e nano-strutture a base di carbonio.

REFERENCES

1. D. A. Long, "The Raman Effect: A Unified Treatment of Raman Scattering by Molecules." Wiley, LTD, Chichester, UK, 2002.
2. L. A. Woodward, "Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy." Clarendon Press, Oxford, 1972.
3. J. Tang and A. C. Albrecht, in "Raman Spectroscopy" (H. A. Szymanski, Ed.), Vol. 2, p. 33, Plenum Press, New York, 1970.
4. J. D. Jackson, "Classical Electrodynamics." John Wiley & Sons, New York, 1975.
5. L. D. Landau and E. M. Lifshitz, "Statistical Physics—Part 1." Pergamon Press, Oxford, 1978.
6. A. C. Albrecht, *J. Chem. Phys.* 34, 1476 (1971).
7. P. P. Shorygin, *Dokl. Akad. Nauk. SSSR* 87, 201 (1952).
8. W. L. Peticolas, L. Nafie, P. Stein, and B. Fanconi, *J. Chem. Phys.* 52, 1576 (1970).
9. R. H. J. Clark and T. Dines, *Angew. Chem. Int. Ed. Engl.* 25, 131 (1986).
10. E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, "Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra." Mc Graw–Hill, New York, 1955.
11. L. D. Landau and E. M. Lifshitz, "Electrodynamics of Continuous Media." Pergamon Press, Oxford, 1970.
12. M. Cardona, in "Light Scattering in Solids II" (M. Cardona and G. Güntherodt, Eds.), p. 31. Springer–Verlag, Berlin, 1982.
13. J. P. Mathieu and H. Poulet, "Vibration Spectra and Symmetry of Crystals." Gordon and Breach, New York, 1976.
14. W. Hayes and R. Loudon, "Scattering of Light by Crystals." Wiley, New York, 1978.
15. R. Zbinden, "Infrared Spectroscopy of High Polymers." Academic Press, New York, 1964.
16. M. Gussoni, C. Castiglioni, and G. Zerbi, in "Spectroscopy of Advanced Materials" (Advances in Infrared and Raman Spectroscopy, R. J. H. Clark and R. E. Hester, Eds.), Chap. 5, p. 251. Wiley, New York, 1991.
17. S. Hirata, H. Torii, and M. Tasumi, *J. Chem. Phys.* 103, 8964 (1995).
18. H. E. Shaffer, R. R. Chance, R. J. Silbey, K. Knoll, and R. R. Shrock, *J. Chem. Phys.* 94, 4161 (1991).
19. A. Bianco, M. Del Zoppo, and G. Zerbi, *Synth. Met.*, in press.
20. L. Harada, Y. Furukawa, M. Tasumi, H. Shirakawa, and S. Ikeda, *J. Chem. Phys.* 73, 4746 (1980).
21. F. Zerbetto, M. Z. Zgieski, F. Negri, and G. Orlandi, *J. Chem. Phys.* 89, 3681 (1988); F. Negri, G. Orlandi, F. Zerbetto, and M. Z. Zgieski, *J. Chem. Phys.* 91, 6215 (1989).
22. R. E. Peierls, "Quantum Theory of Solids." Clarendon Press, Oxford, 1955.
23. E. J. Mele and M. J. Rice, *Solid State Commun.* 34, 339 (1980).
24. F. Negri, C. Castiglioni, M. Tommasini, and G. Zerbi, *J. Phys. Chem. A* 106, 3306 (2002).
25. M. D. Watson, A. Fechtenkötter, and K. Müllen, *Chem. Rev.* 101, 1267 (2001).
26. C. Castiglioni, F. Negri, M. Rigolio, and G. Zerbi, *J. Chem. Phys.* 115, 3769 (2001).
27. F. Negri, E. Di Donato, M. Tommasini, C. Castiglioni, G. Zerbi, and K. Müllen, submitted for publication.
28. I. Spagnoletti, Electron-Phonon Interaction in Carbon-Based Materials: C_{20} Fullerene, Schwarzites and Non-thermal Melting of C_{60}, Ph.D. Thesis, Università degli Studi di Milano, 2002.
29. R. Shuker and R. Gammon, *Phys. Rev. Lett.* 25, 222 (1970).
30. H. W. Kroto, J. R. Healt. S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
31. F. R. Curl and R. E. Smalley, *Science* 242, 1017 (1988).
32. H. Kroto, *Science* 242, 1139 (1988).
33. W. I. F. David, J. Tomkinson, R. C. Haddam, M. J. Rossiensky, and D. W. Murphy, in "The Fullerenes" (H. W. Kroto, J. E. Fisher, and D. E. Cox, Eds.). Pergamon Press, Oxford, 1993.

34. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, San Diego, 1996.
35. H. W. Kroto, J. E. Fisher, and D. E. Cox, Eds., "The Fullerenes." Pergamon Press, Oxford, 1993.
36. P. C. Eklund, P. Zhou, Kai-An Wang, G. Dresselhaus, and M. S. Dresselhaus, in "The Fullerenes" (H. W. Kroto, J. E. Fisher, and D. E. Cox, Eds.), pp. 221–243. Pergamon Press, Oxford, 1993.
37. J. R. Copley, D. A. Neumann, R. L. Cappelletti, and W. A. Katamitakara, in "The Fullerenes" (H. W. Kroto, J. E. Fisher, and D. E. Cox, Eds.), pp. 183–201. Pergamon Press, Oxford, 1993.
38. W. I. F. David et al., *Nature* 353, 320 (1991).
39. D. S. Bethune, G. Meijer, W. C. Tang, H. J. Rosen, W. G. Golden, H. Seki, C. A. Brown, and M. S. de Vries, *Chem. Phys. Lett.* 179, 181 (1991).
40. R. E. Stanton and M. D. Newton, *J. Phys. Chem.* 92, 2141 (1988).
41. V. Schettino, M. Pagliai, L. Ciabini, and G. Cardini, *J. Phys. Chem. A* 105, 11192 (2001).
42. F. Negri and G. Orlandi, *J. Phys. B* 29, 5049 (1996).
43. R. J. Nemanich, S. A. Solin, and R. M. Martin, *Phys. Rev. B* 23, 6348 (1981).
44. P. M. Fauchet and I. H. Campbell, *Crit. Rev. Solid State Mater. Sci.* 14, S79 (1988).
45. S. Hayashi, M. Ito, and H. Kanamori, *Solid State Commun.* 44, 75 (1982).
46. C. E. Bottani, C. Mantini, P. Milani, M. Manfredini, A. Stella, P. Tognini, P. Cheyssac, and R. Kofman, *Appl. Phys. Lett.* 69, 2409 (1996).
47. M. Zacharias, J. Blaesing, J. Christen, P. Veit, B. Dietrich, and D. Bimberg, *Superlattices Microstruct.* 18, 139 (1995).
48. M. Fujii S. Hayashi, and K. Yamamoto, *Jpn. J. Appl. Phys.* 687 (1991).
49. J. Fortner, R. Q. Yu, and J. S. Lannin, *Phys. Rev. B* 42, 7610 (1990).
50. Y. Kanemitsu, H. Uto, Y. Masumoto, and Y. Maeda, *Appl. Phys. Lett.* 61, 2187 (1992).
51. A. K. Dutta, *Appl. Phys. Lett.* 68, 1189 (1996).
52. D. C. Paine, C. Caragianis, T. Y. Kim, Y. Shigesato, and T. Ishahara, *Appl. Phys. Lett.* 62, 2842 (1993).
53. M. Zacharias, R. Weigand, B. Dietrich, F. Stolze, J. Blaesing, P. Veit, T. Druesedau, and J. Christen, *J. Appl. Phys.* 81, 2384 (1997).
54. C. Caragianis-Broadbridge, J. M. Blaser, and D. C. Paine, *J. Appl. Phys.* 82, 1626 (1997).
55. T. Kobayashi, T. Endoh, H. Fukuda, S. Nomura, A. Sakai, and Y. Ueda, *Appl. Phys. Lett.* 71, 1195 (1997).
56. S. Sato, S. Nozaki, H. Morisaki, and M. Iwase, *Appl. Phys. Lett.* 66, 3176 (1995).
57. Gonzalez-Hernandez, G. H. Azarbayejani, R. Tsu, and F. H. Pollak, *Appl. Phys. Lett.* 47, 1350 (1985).
58. M. Wakaki, M. Iwase, Y. Show, K. Koyama, S. Sato, S. Nozaki, and H. Morisaki, *Physica B* 219, 220, 535 (1996).
59. H. Miguez, V. Fornes, F. Meseguer, F. Marquez, and C. Lopez, *Appl. Phys. Lett.* 69, 2347 (1996).
60. J. R. Heath, J. J. Shiang, and A. P. Alivisatos, *J. Chem. Phys.* 101, 1607 (1994).
61. J. G. Zhu, C. W. White, J. D. Budai, S. P. Withrow, and Y. Chen, *J. Appl. Phys.* 78, 4386 (1995).
62. T. Kanata, H. Murai, and K. Kubota, *J. Appl. Phys.* 61, 969 (1987).
63. H. Richter, Z. P. Wang, and L. Ley, *Solid State Commun.* 39, 625 (1981).
64. Z. Iqbal and S. Veppek, *J. Phys. C: Solid State Phys.* 15, 377 (1982).
65. B. Wunderlich, "Macromolecular Physics," Vols. 1 and 2. Academic Press, New York, 1973.
66. G. Zerbi, in "Advances in Chemistry" Series, Vol. 203, p. 487. Am. Chem. Soc., New York, 1983.
67. G. Zerbi, R. Magni, M. Gussoni, K. Holland Moritz, A. B. Bigotto, and S. Dirlikov, *J. Chem. Phys.* 75, 3175 (1981).
68. G. Zerbi, Ed., "Modern Polymer Spectroscopy." Wiley–VCH, Weinheim, 1999.
69. L. Askadskaya and J. P. Rabe, *Phys. Rev. Lett.* 69, 1935 (1992).
70. R. Hentschke, B. L. Schurmann, and J. P. Rabe, *J. Chem. Phys.* 96, 6213 (1992).
71. M. L. Klein, *J. Chem. Soc. Faraday Trans.* 88, 1701 (1992).
72. I. W. Levin, in "Advances in Infrared and Raman Spectroscopy" (R. J. H. Clark and R. E. Hester, Eds.). Vol. 11, pp. 1–14, Wiley, New York, 1984.
73. L. Piseri and G. Zerbi, *J. Mol. Spectrosc.* 26, 254 (1968).
74. G. Zerbi, in "Advances in Infrared and Raman Spectroscopy" (R. H. J. Clark and R. E. Hester, Eds.), p. 301. Wiley, Heyden, London, 1984.
75. L. Piseri and G. Zerbi, *J. Chem. Phys.* 48, 3561 (1968).
76. M. Tasumi and S. Krimm, *J. Chem. Phys.* 46, 755 (1967).
77. R. G. Snyder and J. H. Shachtsneider, *Spectrochim. Acta* 19, 85 (1963); J. H. Shachtsneider and R. G. Snyder, *Spectrochim. Acta* 19, 117 (1963); R. G. Snyder and J. H. Shachtsneider, *Spectrochim. Acta* 21, 169 (1965).
78. S. Trevino and H. Boutin, *J. Macromol. Sci.* A1, 723 (1967).
79. A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Solid State Phys. Suppl.* 3, 1 (1963).
80. Machida, *Spectrochim. Acta.*
81. S. I. Mizushima and T. Shimanouchi, *J. Am. Chem. Soc.* 71, 1320 (1949); R. J. Shaufele and T. Shimanouchi, *J. Chem. Phys.* 47, 3605 (1967).
82. T. Shimanouchi and M. Tasumi, *Indian J. Pure Appl. Phys.* 9, 958 (1971).
83. M. Tommasini, C. Castiglioni, and G. Zerbi, *J. Mol. Struct.* (*Theochem.*) 500, 323 (2000).
84. G. Strobl, B. Ewen, E. W. Fisher, and W. Piesczez, *J. Chem. Phys.* 61, 5257 (1974).
85. L. Brambilla and G. Zerbi, submitted for publication.
86. R. G. Snyder and J. R. Scherer, *J. Polymer Sci.* 16, 1593 (1978).
87. G. Minoni and G. Zerbi, *J. Phys. Chem.* 86, 4791 (1982); G. Zerbi, G. Minoni, and M. P. Tulloch, *J. Chem. Phys.* 78, 5853 (1983); G. Minoni, G. Zerbi, and J. F. Rabolt, *J. Chem. Phys.* 81, 4782 (1984).
88. R. Twieg and J. F. Rabolt, *J. Polymer Sci. Polymer Lett. Ed.* 21, 901 (1983).
89. G. Minoni and G. Zerbi, *J. Polymer Sci. Polymer Lett. Ed.* 22, 533 (1984).
90. G. Natta, G. Mazzanti, and P. Corradini, *Atti Accad. Naz. Lincei, Rend. Sci. Fis. Mat. Nat.* 25, 2 (1958).
91. H. Shirakawa, E. J. Louis, A. G. McDiarmid, C. K. Chiang, and A. J. Heeger, *Chem. Commun.* 578 (1977).
92. T. A. Skotheim, Ed., "Handbook of Conducting Polymers," Vols. 1 and 2. Dekker, New York, 1986, 1999.
93. G. Zerbi, M. Gussoni, and C. Castiglioni, in "Conjugated Polymers: Novel Science and Technology of Conductive and Non Linear Optically Active Materials" (J. L. Brédas and J. Silbey, Eds.), p. 435. Kluwer, New York, 1991.
94. C. Castiglioni, M. Gussoni, J. T. Lopez Navarrete, and G. Zerbi, *Solid State Commun.* 65, 625 (1987).
95. G. P. Brivio and E. Mulazzi, *Phys. Rev. B* 30, 876 (1984); H. Kuzmany, *J. Phys* 44, C3-255 (1983).
96. F. Tuistra and J. L. Koenig, *J. Chem. Phys.* 53, 1126 (1970); Y. Kawashima and G. Katagiri, *Phys. Rev. B* 52, 10053 (1995).
97. Póksic, M. Hunfhausen, M. Koós, and L. Ley, *J. Non-Cryst. Solids* 227, 1083 (1998); M. J. Matthews, M. A. Pimenta, G. Dresselhaus, M. S. Dresselhaus, and M. Endo, *Phys. Rev. B* 59, 6585 (1999).
98. A. C. Ferrari and J. Robertson, *Phys. Rev. B* 61, 1 (2000).
99. J. Robertson and E. P. O'Reilly, *Phys. Rev. B* 35, 2946 (1987).
100. E. Riedo, E. Magnano, S. Rubini, M. Sancrotti, E. Barborini, P. Piseri, and P. Milani, *Solid State Commun.* 116, 287 (2000).
101. R. Escribano, J. J. Sloan, N. Siddique, N. Sze, and T. Dudev, *Vibrational Spectrosc.* 26, 179 (2001).

102. G. Compagnini, G. A. Baratta, R. S. Cataliotti, and A. Morresi, *J. Raman Spectrosc.* 26, 917 (1995).
103. Y. Wang, D. C. Aslemeyer, and R. L. McCreery, *Chem. Mater.* 2, 557 (1990).
104. C. Thompsen and S. Reich, *Phys. Rev. Lett.* 85, 5214 (2000).
105. C. Castiglioni, C. Mapelli, F. Negri, and G. Zerbi, *J. Chem. Phys.* 114, 563 (2001).
106. C. Castiglioni, F. Negri, M. Tommasini, E. Di Donato, and G. Zerbi, Proceedings of the XVI International Winther School "Electronic Properties of Novel Materials: Molecular Nanostructures," IWEPNM2002 (American Institute of Physics, Ed.), 2002, pp. 394–400.
107. M. D. Watson, A. Fechtenkötter, and K. Müllen, *Chem. Rev.* 101, 1267 (2001); A. Stable, P. Herwig, K. Müllen, and J. P. Rabe, *Angew. Chem. Int. Ed. Engl.* 34, 1609 (1995); A. Fechtenkötter, K. Saalwächter, M. A. Harbison, K. Müllen, and H. W. Spiess, *Angew. Chem. Int. Ed. Engl.* 38, 303 (1999).
108. Ijima, *Nature* 354, 56 (1991).
109. R. Saito and H. Kataura, Optical properties and Raman spectroscopy of carbon nanotubes, in "Topics in Applied Physics, Vol. 8, Carbon Nanotubes: Synthesis, Structure, Properies and Applications" (M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds.), p. 213. Springer-Verlag, Berlin, 2001.
110. M. S. Dresselhaus and P. C. Eklund, *Adv. Phys.* 49, 705 (2000).
111. M. S. Desselhaus, G. Dresselhaus, A. Jorio, A. G. Souza Filho, and R. Saito, *Carbon* 40, 2043 (2002).
112. A. M. Rao, E. Richter, S. Bandow, B. Chase, P. C. Eklund, K. W. Williams, M. Menon, K. R. Subbaswamy, A. Thess, R. E. Smalley, G. Desselhaus, and M. S. Dresselhaus, *Science* 275, 187 (1997).
113. R. A. Jishi, L. Venkataraman, M. S. Dresselhaus, and G. Dresselhaus, *Chem. Phys. Lett.* 209, 77 (1993).
114. J. Kürti, G. Kresse, and H. Kuzmany, *Phys. Rev. B* 58, 88 (1998).
115. D. Sanchez-Portal, E. Artacho, J. M. Soler, A. Rubio, and P. Ordéjon, Phys. *Rev. B* 59, 12678 (1999).
116. Kürti, Z. Zólyomi, A. Gruneis, and H. Kuzmany, *Phys. Rev. B* 65, 165433 (2002).
117. O. Dubay and G. Kresse, *Phys. Rev. B* 67, 35401 (2003).
118. W. W. Duley, *Astrophys. J.* 528, 841 (2000).
119. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Astrophys. J.* 314, 352 (1987).
120. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
121. J. Hunter, J. Fye, and M. F. Jarrold, *Science* 260, 784 (1993).
122. R. B. Heimann, J. Kleiman, and N. M. Salansky, *Nature* 306, 164 (1983).
123. M. J. Rice, S. R. Phillpot, A. R. Bishop, and D. K. Campbell, *Phys. Rev. B* 34, 4139 (1986).
124. For a review see: "Carbyne and Carbynoid Structures" (R. B. Heimann, S. E. Evsyukov, and L. Kavan, Eds.), Kluwer Academic, Dordrecht, 1999.
125. M. T. Bowers et al., *Science* 260, 1446 (1993).
126. J. P. Maier, *J. Phys. Chem. A* 102, 3462 (1998); J. Szczepanski et al., *Spectrochim. Acta A* 57, 775 (2001).
127. A. E. Goresy and G. T. Donnay, *Science* 161, 363 (1968).
128. A. G. Whittaker, *Science* 229, 485 (1985); P. P. K. Smith and P. R. Buseck, *Science* 229, 487 (1985).
129. Yu. P. Kudryavtsev, R. B. Heimann, and S. E. Evsyukov, *J. Mater. Sci.* 31, 5557 (1996).
130. L. Kavan, *Chem. Rev.* 97, 3061 (1997).
131. J. Kastner et al., *Macromolecules* 28, 344 (1995); L. Kavan et al., *Carbon* 33, 1321 (1995).
132. N.-R. J. Lagow et al., *Science* 267, 362 (1995).
133. J. A. Lenz et al., *J. Appl. Phys.* 89, 8284 (2001).
134. M. Nakamizo, R. Kammereck, and P. I. Walker, Jr., *Carbon* 12, 259 (1974).
135. L. Ravagnan, F. Siviero, C. Lenardi, P. Milani, C. Casari, A. Li Bassi, and C. E. Bottani, *Phys. Rev. Lett.* 89, 285506 (2002).
136. E. Barborini et al., *Rev. Sci. Instrum.* 73, 2060 (2002).
137. E. Barborini, P. Piseri, and P. Milani, *J. Phys. D: Appl. Phys.* 32, L105 (1999).
138. D. Donadio et al., *Phys. Rev. Lett.* 83, 776 (1999).
139. E. Barborini et al., *Chem. Phys. Lett.* 300, 633 (1999).
140. J. Robertson, *Prog. Solid State Chem.* 21, 199 (1991).
141. J. Kuerti et al., *Synth. Met.* 71, 1865 (1995).
142. K. Akagi et al., *Synth. Met.* 17, 557 (1987).
143. R. H. Baughman, H. Eckhardt, and M. Kertesz, *J. Chem. Phys.* 87, 6687 (1987).
144. R. P. A. Bettens and E. Herbst, *Astrophys. J.* 468, 686 (1996).
145. E. Duval, A. Boukenter, and B. Champagnon, *Phys. Rev. Lett.* 56, 2052 (1986).
146. H. Lamb, *Proc. London Math. Soc.* 13, 189 (1882).
147. A. Tamura and T. Ichinokawa, *J. Phys. C: Solid State Phys.* 16, 4779 (1983).
148. M. Fujii, T. Nagareda, S. Hayashi, and K. Yamamoto, *Phys. Rev. B* 44, 6243 (1991).
149. G. Mariotto, M. Montagna, G. Viliani, E. Duval, S. Lefrant, E. Rzepka, and C. Mai, *Europhys. Lett.* 6, 239 (1988).
150. F. Liu, M. Han, J. Zhao, X. Chen, Q. Wang, and G. Wang, *Appl. Phys. Lett.* 66, 523 (1995).
151. F. Liu, G. Wang, and G. Cheng, *J. Phys. Chem. Solids* 57, 507 (1996).
152. F. Liu, L. Liao, G. Wang, G. Cheng, and X. Bao, *Phys. Rev. Lett.* 76, 604 (1996).
153. C. E. Bottani, A. Li Bassi, A. Stella, P. Cheyssac, and R. Kofman, *Europhys. Lett.* 56, 386 (2001).
154. M. S. Dresselhaus and P. Eklund, *Adv. Physics* 49, 705 (2000).
155. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, 1996.
156. V. N. Popov, V. E. Van Doren, and M. Balkanski, *Solid State Commun.* 114, 395 (2000).
157. D. Kahn, K. W. Kim, and M. A. Stroscio, *J. Appl. Phys.* 89, 5107 (2001).
158. V. W. Brar, G. G. Samsonidze, M. S. Dresselhaus, G. Dresselhaus, R. Saito, A. K. Swan, M. S. Ünlü, B. B. Goldberg, A. G. Souza Filho, and A. Jorio, *Phys. Rev. B* 66, 155418 (2002).
159. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14013 (1998).
160. J. P. Lu, *Phys. Rev. Lett.* 79, 1297 (1997).
161. V. N. Popov, V. E. Van Doren, and M. Balkanski, *Phys. Rev. B* 59, 8355 (1999).
162. R. H. Baughman, C. Cui, A. A. Zakhidov, Z. Iqbal, J. N. Barisci, G. M. Spinks, G. G. Wallace, A. Mazzoldi, D. De Rossi, A. G. Rinzler, O. Jaschinski, S. Roth, and M. Kertesz, *Science* 284, 1340 (1999).
163. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996).
164. J.-P. Salvetat, G. A. D. Briggs, J.-M. Bonard, R. R. Bacsa, A. J. Kulik, T. Stöckli, N. A. Burnham, and L. Forró, *Phys. Rev. Lett.* 82, 944 (1999).
165. M. Grimsditch, *J. Phys. C* 16, L143 (1983).
166. M. Grimsditch, *Phys. Status Solidi B* 193, K9 (1996).
167. S. A. Lee and S. M. Lindsay, *Phys. Status Solidi B* 157, K83 (1990).
168. O. L. Blakslee, D. G. Proctor, E. J. Seldin, G. B. Spence, and T. Weng, *J. Appl. Phys.* 41, 3373 (1970).
169. E. W. Wong, P. E. Sheehan, and C. M. Lieber, *Science* 277, 1971 (1997).
170. J.-P. Salvetat, A. J. Kulik, J.-M. Bonard, G. A. D. Briggs, T. Stöckli, K. Méténier, S. Bonnamy, F. Béguin, N. A. Burnham, and L. Forró, *Adv. Mater* 11, 161 (1999).
171. C. F. Cornwell and L. T. Wille, *Solid State Commun.* 101, 555 (1997).
172. C. E. Bottani, A. Li Bassi, M. G. Beghi, A. Podestà, P. Milani, A. Zakhidov, R. Baughman, D. A. Walters, and R. E. Smalley, *Phys. Rev. B* 67, 155407 (2003).

173. A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tománek, J. E. Fischer, and R. E. Smalley, *Science* 273, 483 (1996).
174. A. G. Rinzler, J. Liu, H. Dai, P. Nikolaev, C. B. Huffman, F. J. Rodrfguez-Macfas, P. J. Boul, A. H. Lu, D. Heymann, D. T. Colbert, R. S. Lee, J. E. Fischer, A. M. Rao, P. C. Eklund, and R. E. Smalley, *Appl. Phys. A* (*Mater. Sci. Process.*) 67, 29 (1998).
175. D. A. Walters, M. J. Casavant, X. C. Qin, C. B. Huffman, P. J. Boul, L. M. Ericson, and E. H. Haroz, *Chem. Phys. Lett.* 338, 14 (2001).
176. X. Liu, T. Pichler, M. Knupfer, M. S. Golden, J. Fink, D. A. Walters, M. J. Casavant, J. Schmidt, and R. E. Smalley, *Synth. Met.* 121, 1183 (2001).
177. P. Milani and C. E. Bottani, Vibrational properties of mesoscopic structures, in "Handbook of Nanostructured Materials and Nanotechnology" (Hari Singh Nalwa, Ed.). Academic Press, New York, 2000.
178. Stefan Markus, "The Mechanics of Vibrations of Cylindrical Shells: Studies in Applied Mechanics." Elsevier, New York, 1988.
179. H. Kraus, "Thin Elastic Shells." Wiley, New York, 1967.
180. M. Fujii, T. Nagareda, S. Hayashi, and K. Yamamoto, *Phys. Rev. B* 44, 6243 (1991).
181. G. S. Duesberg, I. Loa, M. Burghard, K. Syassen, and S. Roth, *Phys. Rev. Lett.* 85, 5436 (2000).
182. C. Mapelli, C. Castiglioni, G. Zerbi, and K. Müllen, *Phys. Rev. B* 60, 12710 (1999).
183. N. Peyghambarian, E. Hanamura, S. W. Koch, Y. Masumoto, and E. M. Wright, in "Nanomaterials: Synthesis, Properties and Applications" (A. S. Edelstein and R. C. Cammarata, Eds.), p. 395. IOP, Bristol, 1996.
184. J. C. Parker, in "Nanomaterials: Synthesis, Properties and Applications" (A. S. Edelstein and R. C. Cammarata, Eds.), p. 573. IOP, Bristol, 1996.
185. J. M. Ballesteros, R. Serna, J. Solis, C. N. Afonso, A. K. Petford-Long, D. H. Osborne, and R. F. Haglund, Jr., *Appl. Phys. Lett.* 71, 2445 (1997).
186. E. Søndergård, R. Kofman, P. Cheyssac, and A. Stella, *Surf. Sci.* 364, 467 (1996).
187. R. J. Nemanich and S. A. Solin, *Phys. Rev. B* 20, 392 (1979).
188. P. Lespade, R. Al-Jishi, and M. S. Dresselhaus, *Carbon* 20, 427 (1982).
189. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996).
190. G. A. J. Amaratunga, M. Chowalla, C. J. Kiely, I. Alexandrou, R. Aharonov, and R. M. Devenish, *Nature* 383, 321 (1996).
191. J. P. Lu, *Phys. Rev. Lett.* 79, 1297 (1997).
192. P. Melinon, V. Paillard, V. Dupuis, A. Perez, P. Jensen, A. Hoareau, J. P. Perez, J. Tuaillon, M. Broyer, J. L. Vialle, M. Pellarin, B. Baguenard, and J. Lerme, *Int. J. Mod. Phys. B* 9, 339 (1995).
193. P. Milani and S. Iannotta, "Cluster Beam Synthesis of Nanostructured Materials." Springer-Verlag, Berlin, 1999.
194. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, San Diego, 1996.
195. L. D. Landau and Lifshitz, "Theory of Elasticity," 2nd ed. Pergamon Press, London, 1970.
196. I. A. Kunin, "Elastic Media with Microstructure," Vols. 1 and 2. Springer-Verlag, Berlin, 1982.
197. G. A. D. Briggs, Ed., "Advances in Acoustic Microscopy," Vol. 1. Plenum Press, New York, 1995.
198. C. Kittel, "Introduction to Solid State Physics," 3d ed. Wiley, New York, 1966.
199. J. D. Achenbach, "Wave Propagation in Elastic Solids." North-Holland, Amsterdam, 1990.
200. C. E. Bottani, G. Ghislotti, and P. Mutti, *J. Phys.: Condens. Matter* 6, L85 (1994).
201. G. Ghislotti and C. E. Bottani, *Phys. Rev. B* 50, 12 131 (1994).
202. C. E. Bottani and R. Caporali, *J. Phys.: Condensed Matter* 6, L791 (1994).
203. R. Caporali, C. E. Bottani, and G. Ghislotti, *Phys. Rev. B* 53, 4133 (1996).
204. G. W. Farnell and E. L. Adler, in "Physical Acoustics" (P. W. Mason and R. W. Thurston, Eds.), Vol. 9, pp. 35–121. Academic Press, New York, 1972.
205. M. Beghi, C. E. Bottani, P. M. Ossi, T. Lafford, and B. K. Tanner, *J. Appl. Phys.* 81, 672 (1997).
206. J. R. Sandercock, in "Light Scattering in Solids III" (M. Cardona and G. Güntherodt, Eds.), pp. 173–206. Springer-Verlag, Berlin, 1982.
207. M. Born and E. Wolf "Principles of Optics," VII exp. ed., Chap. XII. Cambridge Univ. Press, Cambridge, 1999.
208. P. Milani and C. E. Bottani, Vibrational spectroscopy of mesoscopic structures, in "Handbook of Nanostructured Materials and Nanotechnology" (H. S. Nalwa, Ed.), Vol. 2, pp. 213–269. Academic Press, New York, 2000.
209. C. E. Bottani, A. C. Ferrari, A. Li Bassi, P. Milani, and P. Piseri, *Europhys. Lett.* 42, 431 (1998).
210. P. Milani, Private communication, 1999.
211. V. V. Zosimov and L. M. Lyamshev, Fractals in wave processes, *Phys. Uspekhi* 38, 347 (1995).
212. C. E. Bottani, C. Casari, and A. Li Bassi, in press (2000).
213. R. Pastorelli, S. Tarantola, M. G. Beghi, C. E. Bottani, and A. Saltelli, *Surf. Sci.* (2000), in press.
214. A. Pinczuk and E. Burstein, "Light Scattering in Solids," in "Topics in Applied Physics" (M. Cardona, Ed.). Springer-Verlag, Berlin, 1975.
215. F. Nizzoli and D. L. Mills, in "Nonlinear Surface Electromagnetic Phenomena" (H. E. Ponath and G. I. Stegeman, Eds.), p. 445. North-Holland, Amsterdam, 1991.
216. R. Shuker and R. W. Gammon, *Phys. Rev. Lett.* 25, 222 (1970).
217. P. Mutti, C. E. Bottani, G. Ghislotti, M. Beghi, G. A. Briggs, and J. R. Sandercock, in "Advances in Acoustic Microscopy" (A. Briggs, Ed.), Vol. 1. Plenum Press, New York, 1995.
218. V. Bortolani, F. Nizzoli, G. Santoro, and J. R. Sandercock, *Phys. Rev. B* 25, 3442 (1982).
219. P. Malischewsky, in "Surface Waves and Discontinuities." Elsevier, Amsterdam, 1987.
220. V. Bortolani, A. M. Marvin, F. Nizzoli, and G. Santoro, *J. Phys. C* 16, 1757 (1983).
221. E. L. Albuquerque, R. Loudon, and D. R. Tilley, *J. Phys. C* 13, 1775 (1980).
222. A. P. Mayer, *J. Phys.: Condens. Matter* 1, 3957 (1989).
223. J. A. Bell, R. Zanoni, C. T. Seaton, G. I. Stegeman, W. R. Bennett, and C. M. Falco, *Appl. Phys. Lett.* 52, 610 (1988).
224. D. L. Mills, S. Y. Tong, and J. E. Black, in "Surface Phonons" (W. Kress and F. W. de Wette, Eds.), p. 200. Springer-Verlag, Berlin, 1991.
225. B. Laks and D. L. Mills, *Phys. Rev. B* 20, 4962 (1979).
226. R. Courant and D. Hilbert, in "Methods of Mathematical Physics," Vol. 1, p. 324. Interscience, New York, 1953.
227. H. Sagan, in "Boundary and Eigenvalue Problems in Mathematical Physics," p. 163. Dover, New York, 1989.
228. J. Lekner, in "Theory of Reflection," p. 5 and 157. Martinus Nijhoff, Dordrecht, 1987.
229. M. G. Cottam and A. A. Maradudin, in "Surface Excitations" (V. M. Agranovich and R. Loudon, Eds.), p. 114. North-Holland, Amsterdam, 1984.
230. J. R. Sandercock, in "Light Scattering in Solids III, Vol. 51, Topics in Applied Physics" (M. Cardona and G. Güntherodt, Eds.), p. 173. Springer-Verlag, Berlin, 1982.
231. C. Byloos, L. Giovannini, and F. Nizzoli, *Phys. Rev. B* 51, 9867 (1995).

232. F. Nizzoli, C. Byloos, L. Giovannini, C. E. Bottani, G. Ghislotti, and P. Mutti, *Phys. Rev. B* 50, 2027 (1994).
233. G. Ghislotti, C. E. Bottani, P. Mutti, C. Byloos, L. Giovannini, and F. Nizzoli, *Phys. Rev. B* 51, 9875 (1995).
234. G. Ghislotti, A. Gagliardi, C. E. Bottani, S. Bertoni, G. F. Cerofolini, and L. Meda, *Mater. Sci. Eng. B* 36, 129 (1996).
235. M. G. Beghi, C. E. Bottani, and G. Ghislotti, *Mater. Sci. Eng. B* 46, 20 (1997).
236. C. E. Bottani, A. Li Bassi, B. K. Tanner, A. Stella, P. Tognini, P. Cheyssac, and R. Kofman, *Phys. Rev. B* 59, R15601 (1999).
237. A. Li Bassi, C. E. Bottani, B. K. Tanner, A. Stella, P. Tognini, P. Cheyssac, and R. Kofman, *Eur. Phys. J. B* 18, 31 (2000).
238. Y. Maeda, *Phys. Rev. B* 51, 1658 (1995).
239. L. Y. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
240. R. Tsu, H. Shen, and M. Dutta, *Appl. Phys. Lett.* 60, 112 (1992).
241. S. Hayashi and K. Yamamoto, *J. Luminesc.* 70, 352 (1996).
242. M. C. Klein, F. Hache, D. Ricard, and C. Flytzanis, *Phys. Rev. B* 42, 11123 (1990).
243. See, for example, “Practical Raman Spectroscopy” (D. J. Gardiner and P. R. Graves, Eds.). Springer-Verlag, Berlin, 1989.
244. H. Muender, C. Andrzejak, M. G. Berger, U. Klemradt, H, Lueth, R. Herino, and M. Ligeon, *Thin Solid Films* 221, 27 (1992).
245. Z. Sui, P. P. Leong, I. P. Herman, G. S. Higashi, and H. Temkin, *Appl. Phys. Lett.* 60, 2086 (1992).
246. W. Theiss, *Surf. Sci. Rep.* 29, 91 (1997).
247. T. Schlief, J. Gross, and J. Fricke, *J. Non-Cryst. Solids* 145, 223 (1992).
248. J. Gross, J. Fricke, and L. W. Hrubesh, *J. Acoust. Soc. Am.* 91, 2004 (1992).
249. X. G. Zhang, *J. Electrochem. Soc.* 138, 3750 (1991).
250. E. Courtens, J. Pelous, J. Phalippou, R. Vacher, and T. Woignier, *Phys. Rev. Lett.* 58, 128 (1987).
251. J. Robertson, *Prog. Solid State Chem.* 21, 199 (1991).
252. J. Robertson, *Adv. Phys.* 35, 317 (1986).
253. M. W. Geis and M. A. Tamor, “Diamond and Diamondlike Carbon, Vol. 5 in The Encyclopedia of Applied Physics” (G. L. Trigg, Ed.), p. 1. VCH, New York, 1993.
254. D. R. McKenzie, *Rep. Prog. Phys.* 59, 1611 (1996).
255. M. A. Tamor and W. C. Vassell, *J. Appl. Phys.* 76, 3823 (1994).
256. D. S. Knight and W. B. White, *J. Mater. Res.* 4, 385 (1989).
257. R. J. Nemanich, J. T. Glass, G. Lucovsky, and R. E. Shroder, *J. Vac. Sci. Technol. A* 6, 1783 (1988).
258. R. E. Shroder, R. J. Nemanich, and J. T. Glass, *Phys. Rev. B* 41, 3738 (1990).
259. N. Wada, P. Gaczi, and S. Solin, *J. Non-Cryst. Solids* 35–36, 543 (1980).
260. D. Beeman, J. Silverman, R. Lynds, and M. Anderson, *Phys. Rev. B* 30, 870 (1984).
261. R. Dillon, R. Wolam, and V. Katkanant, *Phys. Rev. B* 29, 3482 (1984).
262. F. Tuinstra and J. L. Koenig, *J. Chem. Phys.* 53, 1126 (1970).
263. R. Alben, D. Weaire, J. E. Smith, and M. H. Brodsky, *Phys. Rev. B* 11, 2271 (1975).
264. F. Li and J. S. Lannin, *Appl. Phys. Lett.* 61, 2116 (1992).
265. F. Li, N. Lustig, P. Klosowski, and J. S. Lannin, *Phys. Rev. B* 41, 10210 (1990).
266. J. Schwan, S. Ulrich, V. Batori, H. Ehrhardt, and S. R. P. Silva, *J. Appl. Phys.* 80, 440 (1996).
267. V. I. Merkulov, J. S. Lannin, C. H. Munro, S. A. Asher, V. S. Veerasamy, and W. I. Milne, *Phys. Rev. Lett.* 78, 4869 (1997).
268. K. W. R. Gilkes, H. S. Sands, D. N. Batchelder, J. Robertson, and W. I. Milne, *Appl. Phys. Lett.* 70, 1980 (1997).
269. M. Yoshikawa, Y. Mori, H. Obata, M. Maegawa, G. Katagiri, H. Ishida, and A. Ishitani, *Appl. Phys. Lett.* 67, 694 (1995).
270. M. Yoshikawa, Y. Mori, M. Maegawa, G. Katagiri, H. Ishida, and A. Ishitani, *Appl. Phys. Lett.* 62, 3114 (1993).
271. P. Milani, M. Ferretti, P. Piseri, C. Bottani, A. Ferrari, A. Li Bassi, G. Guizzetti, and M. Patrini, *J. Appl. Phys.* (1997), in press.
272. P. Milani, P. Piseri, Application of Accelerators in Research and Industry (J. L. Duggan and I. L. Morgan, Eds.), AIP Conf. Proc., Vol. 392, p. 495. Woodbury, N.Y., 1997.
273. A. Canning, G. Galli, and J. Kim, *Phys. Rev. Lett.* 78, 4442 (1997).
274. H. Sjostrom, S. Stafstrom, M. Roman, and J-.F. Sundgren, *Phys. Rev. Lett.* 75, 1336 (1995).
275. G. Benedek and L. Colombo, *Mater. Sci. Forum* 232, 247 (1996).
276. A. W. Fung, Z. H. Wang, K. Lu, M. S. Dresselhaus, and R. W. Pekala, *J. Mater. Res.* 8, 1875 (1993).
277. C. E. Bottani, A. C. Ferrari, A. Li Bassi, P. Milani, M. Ferretti, and P. Piseri, *Carbon* (1997), in press.
278. M. Grimsditch, *J. Phys. C* 16, L143 (1983).
279. X. Jiang, J. V. Harzer, B. Hillebrands, C. Wild, and P. Koidl, *Appl. Phys. Lett.* 59, 1055 (1991).
280. X. Jiang, J. W. Zou, K. Reichelt, and P. Grünberg, *J. Appl. Phys.* 66, 4729 (1989).
281. D. Fioretto, G. Carlotti, G. Socino, S. Modesti, C. Cepek, L. Giovannini, O. Donzelli, and F. Nizzoli, *Phys. Rev. B* 52, R8707 (1995).
282. M. Manfredini, C. E. Bottani, and P. Milani, *Chem. Phys. Lett.* 226, 600 (1994).
283. A. Gri¢n, *Rev. Mod. Phys.* 40, 167 (1968).
284. O. L. Blakslee, D. G. Proctor, E. J. Seldin, G. B. Spence, and T. Weng, *J. Appl. Phys.* 41, 3373 (1970).
285. E. J. Seldin and C. W. Nezbeda, *J. Appl. Phys.* 41, 3389 (1970).
286. C. E. Bottani, “Mechanical Properties of Low Dimensional Carbon Structures,” in “Nanostructured Carbon for Advanced Applications” (G. Benedek, P. Milani, and V. G. Ralchenko Eds.). Kluwer Academic, Dordrecht, 2001.

www.aspbs.com/enn

Raman Spectroscopy of Quantum Wires and Quantum Dots

V. Wagner, J. Geurts, W. Kiefer

Universität Würzburg, Am Hubland, Würzburg, Germany

CONTENTS

1. Introduction
2. Fundamentals of Raman Spectroscopy
3. Experimental Setup
4. Analysis of Crystal Lattice Properties
5. Effects of Size Restriction by Interfaces
6. Nanostructures: Quantum Wires and Quantum Dots
7. Summary
References

1. INTRODUCTION

Raman spectroscopy is a realisation of inelastic light scattering. This implies that upon irradiation of a sample with monochromatic light the scattered light contains components with a characteristic frequency shift. This effect is due to the generation or annihilation of an elementary excitation (e.g., a lattice vibration) in the sample. It was predicted by Smekal (1923) and was first observed by Raman, Landsberg, and Mandelstam (1928). The first experiments were carried out using focused sunlight and filters and relied on the visual observation of color changes in the scattered light.

The Raman effect is an inherently weak effect, typically below 10^{-10} of the intensity of the exciting light. Therefore, in the early years, limited light source intensity and stability and the cumbersome detection of weak signals made Raman spectroscopy extremely difficult. It was not until the 1960s that the modern Raman renaissance started. This revival was triggered by the availability of commercial visible laser light sources (ion lasers with a series of laser lines at fixed wavelengths, and soon afterward also frequency-tunable lasers) and computer-controlled novel spectrometers with sensitive detection systems (photomultipliers, and more recently diode arrays and charge coupled devices).

In this way its sensitivity was substantially enhanced, and meanwhile Raman spectroscopy has achieved the ability to detect single monolayers (e.g., at surfaces and interfaces) and quantum dots, which are the focus of this chapter. The enhanced sensitivity also opened the path to a strongly improved lateral resolution: micro-Raman spectroscopy.

The main reason for the widespread utilization of Raman spectroscopy is its ability to analyze a broad variety of different sample properties. It is used, for example, for the identification of materials, to study their chemical composition, crystalline perfection, crystallographic orientation, strain, doping, and intermixing, as well as carrier-depleted layers and electric fields at surfaces and interfaces. The study of all these aspects essentially relies on the lattice dynamics. The eigenfrequencies, frequency broadening, and Raman scattering efficiencies of the vibrational eigenmodes are characteristically influenced by the properties listed above.

The analysis of all these properties is performed in a nondestructive way and without the requirement of a special sample preparation. The information depth can by tuned by applying different excitation wavelengths, allowing one to derive information from the sample surface as well as from deeper regions and interfaces. It will be shown below that this technique even can be applied for *in-situ* online analysis during epitaxial growth.

A general survey of theory and experiments of Raman scattering is given in various books and articles (e.g., [1–3]). A very detailed description of the achievements in light scattering in solids is presented in a series of eight volumes, edited by Güntherodt and Cardona [4]. It comprises, for example, a review on Raman scattering from surface phonons [5]. Moreover, it should be noted that besides its widespread application for research purposes, Raman spectroscopy also has revealed its broad potential for industrial applications [6].

The first part of this review of Raman spectroscopy on II–VI semiconductor interfaces and nanostructures (Sections 2 to 4) deals with general aspects required for appreciating

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

the discussion of interfaces and nanostructures in the second part (Sections 5 and 6). The general part starts with a brief survey of the fundamentals of Raman spectroscopy in Section 2 and a description of the experimental equipment (Section 3). Subsequently, Section 4 contains the discussion of the detection of bulk crystal lattice properties by Raman spectroscopy, with special emphasis to crystalline perfection and orientation, composition of mixed compounds, and the effects of doping and strain, which are relevant aspects for the analysis of interfaces and nanostructures.

In the second part of this review, first the consequences of size restriction by interfaces are discussed for single interfaces and (multi)-quantum wells in Section 5.

For single interfaces, first vibrational modes of interfaces are treated: chemical vibrational modes due to the specific interface bonds, as well as dielectric vibration modes due to the different dielectric constants of adjacent materials. The second single-interface phenomenon which can be studied by Raman scattering is electronic band bending due to charge trapping or exchange at interfaces of doped materials.

For multiple-layer systems such as (multi)-quantum wells and superlattices, qualitatively new phenomena are discussed:

(i) confined optical phonons, which give Raman spectroscopic access to the whole Brillouin zone;
(ii) folded acoustic phonons, which enable Raman scattering from the acoustic branch.

Section 6 deals with chemically produced dots, etched quantum wires, and dots, as well as self-organized dots. Common items of interest for all these systems are chemical composition, strain effects, and local strain gradients. Furthermore, essential properties are the average dot size and the size homogeneity, which also can be addressed by Raman spectroscopy. When studying these aspects, for very small dots also confinement effects have to bes considered.

In addition, quantum dot Raman studies of various items of fundamental physics are discussed, such as the size dependence of the electron–hole exchange interaction in the exciton, and the extension of the electronic wavefunctions beyond the nanostructure edges. Numerous examples are presented from binary, ternary, and quaternary II–VI compounds (e.g., ZnSe, $ZnSe_{1-x}S_x$, $Zn_{1-y}Mg_ySe_{1-x}S_x$, $Cd_xZn_yMg_{1-x-y}Te$), from interfaces between different combinations of II–VI compounds (e.g., ZnSe/BeTe, CdSe/BeTe, CdTe/ZnTe), as well as II–VI/III–V interfaces (e.g., ZnSe/GaAs and CdTe/InSb), and from various II–VI nanostructure systems (e.g., ZnCdSe/ZnSe wires, CdS, CdSe, and CdS_xSe_{1-x} dots, and fractional CdSe monolayers), produced with the different techniques mentioned above.

2. FUNDAMENTALS OF RAMAN SPECTROSCOPY

In this section the description of the theory of Raman scattering is confined to a brief treatment of its fundamentals, followed by several selected aspects, which are of direct relevance for its application to surfaces, interfaces, and quantum dots. Such aspects are, for example, symmetry-induced selection rules, resonant Raman scattering, analysis of intermixing, strain, and doping.

The fundamentals of the Raman scattering process can be described either by a macroscopic theory, based on the concept of a generalized dielectric susceptibility (see Section 2.1), or by a microscopic quantum-mechanical theory, based on electron–photon and electron–phonon coupling (see Section 2.2).

2.1. Susceptibility Theory

In the Raman scattering process a frequency shift of the scattered light with respect to the incident light occurs. The reason for this peculiar optical behavior of the sample can be of various natures, for example, a lattice vibration [mostly optical phonons: longitudinal-optical (LO) or transverse-optical (TO)], a lattice vibration coupled to a collective electronic excitation (plasmon–LO–phonon mode), etc.

The inelastic scattering of light can be described phenomenologically by a generalized dielectric susceptibility $\chi(\omega_i, \omega_s)$, which is also called the transition susceptibility. It connects the incident electric field $\mathbf{E}_{i\beta}(\omega_i)$ with the resulting polarization $\mathbf{P}_{s\alpha}(\omega_s)$,

$$\mathbf{P}_{s\alpha}(\omega_s) = \varepsilon_0 \chi_{\alpha\beta}(\omega_i, \omega_s)\mathbf{E}_{i\beta}(\omega_i) \tag{1}$$

where α and β denote the directions of the fields. It is the modulation of this susceptibility by, for example, a lattice vibration, which induces the inelastic light scattering. In the quasi-static approximation the susceptibility is defined for every instantaneous atomic displacement, and the modulation of $\chi(\omega_i, \omega_s)$ by lattice vibrations can be described by a Taylor series in terms of vibration normal modes j with amplitude Q_j, wavevector $\mathbf{q}_j$, and angular frequency ω_j [2]:

$$\chi_{\alpha\beta}(\omega_i, \omega_s) = \chi^0_{\alpha\beta}(\omega_i, \omega_s) + \chi^1_{\alpha\beta}(\omega_i, \omega_s) + \chi^2_{\alpha\beta}(\omega_i, \omega_s) + \cdots \tag{2}$$

where

$$\chi^0_{\alpha\beta}(\omega_i, \omega_s) = \chi^0_{\alpha\beta}(\omega_i)$$

$$\chi^1_{\alpha\beta}(\omega_i, \omega_s) = \sum_j Q_j \cdot \left(\frac{\partial \chi_{\alpha\beta}(\omega_i, \omega_s)}{\partial Q_j}\right) + \sum_j Q_j \cdot q_j \cdot \left(\frac{\partial^2 \chi_{\alpha\beta}(\omega_i, \omega_s)}{\partial Q_j\, \partial q_j}\right) + \sum_{j,\gamma} Q_j \cdot E_\gamma \cdot \left(\frac{\partial^2 \chi_{\alpha\beta}(\omega_i, \omega_s)}{\partial Q_j\, \partial E_\gamma}\right) + \cdots$$

$$\chi^2_{\alpha\beta}(\omega_i, \omega_s) = \sum_{j,j'} Q_j Q_{j'} \cdot \frac{1}{2}\left(\frac{\partial^2 \chi_{\alpha\beta}(\omega_i, \omega_s)}{\partial Q_j\, \partial Q_{j'}}\right) + \cdots$$

where E_γ is the component of a possible static electric field as an external perturbation. The main contribution comes from the first term χ^0, which represents the susceptibility of the unperturbed static crystal; this term leads to elastic light scattering with $\omega_s = \omega_i$, also denoted as Rayleigh scattering.

The subsequent terms represent the inelastic light scattering. The term χ^1 describes first-order Raman scattering (i.e., scattering from a single phonon). This scattering may occur due to the phonon-induced lattice deformation Q (first term of χ^1) or the simultaneous occurrence of symmetry-reducing effects, such as a finite phonon wavevector q (second term of χ^1) or, for example, the presence of a static electric field (third term of χ^1). The term χ^2 describes the two-phonon processes. The partial derivatives of the susceptibility in Eq. (2) constitute the Raman tensors, which will be elucidated below. The frequency values of the scattered light, which originates from the susceptibility term χ^1, are obtained from Eqs. (1) and (2):

$$\omega_s = \omega_i \pm \omega_j(\mathbf{q}_j) \tag{3}$$

Thus, the shift of the angular frequency of the light equals the angular frequency of the lattice vibration.

2.2. Microscopic Theory

In the quantum-mechanical model, the light frequency shift in the inelastic scattering process is explained in terms of an energy difference between the scattered photon and the incoming one. This difference occurs due to the creation or annihilation of an elementary excitation (e.g., a phonon) in the sample.

For the first-order phonon scattering, energy conservation reads

$$\hbar\omega_s = \hbar\omega_i \pm \hbar\omega_j(\mathbf{q}_j) \tag{4}$$

where the minus sign stands for the creation of a phonon and the plus sign for the phonon annihilation. The first scattering process is called a Stokes process, whereas the latter is denoted as an anti-Stokes process. Similar to the energy conservation, the wavevectors $\mathbf{k}_i$ of the incident laser photon and $\mathbf{k}_s$ of the scattered photon inside the sample are correlated by the momentum conservation law

$$\hbar\mathbf{k}_s = \hbar\mathbf{k}_i \pm \hbar\mathbf{q}_j + \hbar\mathbf{G} \tag{5}$$

where $\mathbf{G}$ is a reciprocal lattice vector.

Many solid-state applications of Raman spectroscopy deal with opaque materials, which implies experiments in backscattering geometry: the direction of the scattered light is approximately antiparallel to the incident one. In this case, the value of the phonon momentum in the Stokes process is given by

$$q_j = \frac{1}{c_0}[n(\omega_i)\omega_i + n(\omega_s)\omega_s] \tag{6}$$

The refractive index values of the II–VI compounds which are in the focus of this chapter are in the range between 2 and 3. Then the backscattering experiments with the available visible laser wavelengths yield scattering wavevectors in the range $q = 6$–$8 \cdot 10^5$ cm^{-1}. The reciprocal lattice vector $|\mathbf{G}|$ is of the order of 10^8 cm^{-1} and therefore the scattering process takes place near the center of the first Brillouin zone. The optical phonons of the materials of interest here show a negligible dispersion in this wavevector range, resulting in a well defined phonon frequency in the Raman spectrum, independent of the exciting wavelength. However, in the case of plasmons and plasmon–LO–phonon modes one has to consider that these modes can show dispersion in this wavevector range.

In the case of second-order Raman scattering the conservation laws impose

$$\hbar\omega_s = \hbar\omega_i \pm \hbar(\omega_j \pm \omega_{j'}) \tag{7}$$

$$\hbar\mathbf{k}_s = \hbar\mathbf{k}_i \pm \hbar(\mathbf{q}_j \pm \mathbf{q}_{j'}) + \hbar\mathbf{G} \tag{8}$$

Obviously, in the case of multiphonon processes one can observe combinations of excitations whose individual q-vectors may be of the order of the reciprocal lattice constant.

In order to calculate the probability of the annihilation of an incoming photon with energy $\hbar\omega_i$ and the creation of a scattered photon with energy $\hbar\omega_s$ accompanied with the creation or annihilation of an elementary excitation of energy $\hbar\omega_j$, the Raman scattering process can be described in a quantum-mechanical perturbation theory, which expresses explicitely the indirect photon–phonon interaction and resonance effects.

Raman experiments are usually performed in the visible region with photon energies in the eV range, whereas the elementary excitations, like phonons, have eigenenergies in the meV range. Due to this drastic energy mismatch, no direct photon–phonon interaction occurs. Therefore, the photon–phonon interaction is mediated by the electron system, whose excitation energies are in the range of the photon energies. The total process for phonon excitations can be described as a three-step combination of transitions [2, 7, 8], as schematically depicted in Figure 1 for Stokes scattering:

- An electron in the ground state $|0\rangle$ is excited to the state $|e\rangle$ with energy E_e due to the absorption of a photon with energy $\hbar\omega_i$; for this creation of an electron–hole pair the dipole operator p is involved.
- Due to the electron–phonon interaction, described by the operator H_{E-P}, the electron–hole pair is scattered from state $|e\rangle$ to another state $|e'\rangle$ with energy $E_{e'}$, accompanied by the creation of a phonon with energy $\hbar\omega_j$.
- The recombination transition from $|e'\rangle$ to $|0\rangle$ leads to the emission of a scattered photon with energy $\hbar\omega_s = \hbar\omega_i - \hbar\omega_j$.

The total scattering process arises from six time orderings of the three interactions in the perturbation treatment. The by far most dominant term is given by

$$\chi_{\alpha\beta}(\omega_i, \omega_s) \propto \sum_{e,e'} \frac{\langle 0|p_\beta|e'\rangle\langle e'|H_{E-P}|e\rangle\langle e|p_\alpha|0\rangle}{(E_{e'} - \hbar\omega_s)(E_e - \hbar\omega_i)} \tag{9}$$

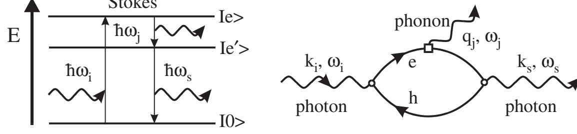


Figure 1. The Raman scattering process as a combination of electron–photon and electron–phonon interactions.

Momentum conservation is required for every single transition, while energy conservation is only required for the total scattering process. For interaction processes on a very short time scale the energy–time uncertainty relation allows the generation of virtual electron–hole pairs without energy conservation, for example, in the case of a below-bandgap excitation, which is used rather frequently for the investigation of wide gap II–VI semiconductors.

The anti-Stokes-to-Stokes intensity ratio in thermal phonon equilibrium far away from resonance (i.e., for equal susceptibility at the frequencies of Stokes and anti-Stokes scattered light) is given by [9]

$$\frac{I_{AS}}{I_S} \propto \left[\frac{\omega_i + \omega_j}{\omega_i - \omega_j}\right]^3 \cdot \exp\left[\frac{-\hbar\omega_j}{k_B T}\right] \tag{10}$$

This equation reflects that phonon states must be thermally occupied in order to appear as anti-Stokes line. For typical optical phonon wavenumber values $\tilde{\nu}_j = \omega_j/(2\pi \cdot c) > 150$ cm^{-1} (i.e., eigenenergy > 18.6 meV) the intensity ratio is strongly dependent on the sample temperature. For example, for ZnSe with $\tilde{\nu}_{LO} = 252$ cm^{-1} one finds at room temperature an anti-Stokes-to-Stokes intensity ratio of 0.32, whereas for $T = 80$ K one expects a ratio of about 0.011; only for very small wavenumber values $\tilde{\nu}_j \leq 50$ cm^{-1} are considerable anti-Stokes intensities observed. Therefore we will discuss only Stokes scattering throughout this work.

2.3. Scattering Mechanisms and Resonance

In the three-step Raman scattering process, different realizations of the electron–phonon interaction may occur, because a phonon can influence the electronic system in two complementary ways: it can modulate the electronic eigenenergies as well as the eigenfunctions.

Because the value of the energy gap depends on the momentary state of lattice deformation, a phonon leads to a modulation of the bandgap. A modulation of the energy eigenvalues (band shifting) means a phonon-induced transition between states $|e\rangle$ and $|e'\rangle$ in the same energy band (i.e., intraband scattering).

The modification of electronic eigenfunctions by the lattice deformation can be described by an admixture of states of adjacent bands; this means transitions between electronic states $|e\rangle$ and $|e'\rangle$ of different bands (i.e., interband scattering). In the susceptibility picture this implies a modification of the electronic interband oscillator strength. This effect is noticeable, for example, in semiconductors of zincblende structure with spin-orbit split valence bands.

The phonon property which causes the modulation of the electronic eigenvalues and/or eigenfunctions can be either the deformation-induced potential or the macroscopic electric field accompanied with the phonon. The former case is the dynamical analogue of the piezomodulation of the electronic band structure; the deformation-potential scattering at the E_0 gap of zincblende II–VI compounds has intraband and interband terms. In contrast to the TO phonon, the LO phonon has a macroscopic electric field, which has two consequences: (i) the corresponding additional restoring force for lattice distortions results in an enhanced frequency of the LO phonon with respect to the TO; (ii) the corresponding electric potential induces an additional light-scattering mechanism, denoted as Fröhlich interaction. This is the dynamical analogue of the electro-optic effect. The Fröhlich scattering for phonons at the Brillouin zone center $q = 0$ has only interband terms (i.e., a low efficiency), but for phonons with finite q values also the much more efficient intraband scattering can occur (symmetry-forbidden Fröhlich scattering). In the same way, the scattering intensity of the LO phonon can be enhanced in the presence of a static electric field (electric field induced Raman scattering), which is relevant for the analysis of interfaces [3]. The size dependence of the Fröhlich-induced electron phonon interaction in nanostructures is an actual item of interest in fundamental physics. It will be discussed in Section 6.1.5.

An item of extreme importance is the resonance behavior, as expressed by the denominators in Eq. (9), when the photon energy of the incident or scattered light closely matches the electron–hole energy. For transitions which do obey the energy conservation law, the contribution to the Raman cross section is enhanced resonantly. The matching of the incident or scattered photon with the electron–hole energy is referred to as incoming or outgoing resonance, respectively. By exploiting these resonances at the critical points in the electronic band structure (e.g., the E_0 or E_1 gap), the sensitivity in the Raman experiment can be drastically enhanced, which is of particular interest for the investigation of thin layers and constitutes a requirement for studying interfaces and nanostructures. Note that this sensitivity enhancement is material-specific, which allows focusing to, for example, interface regions with specific chemical bonds, although their thickness is far below that of the embedding layers.

The strength of resonant enhancement substantially depends on the scattering process: if the electronic states $|e\rangle$ and $|e'\rangle$ belong to different electronic bands (i.e., interband scattering), the energy values E_e and $E_{e'}$ in the denominator factors are considerably different, and a rather weak resonance occurs. However, for intermediate electronic states $|e\rangle$ and $|e'\rangle$ in the same band (i.e., intraband scattering) the energies E_e and $E_{e'}$ are comparable and therefore a much more pronounced resonance behavior is observed. An extremely strong resonance enhancement occurs in the special case when electronic states $|e\rangle$ and $|e'\rangle$ are available whose energy difference equals the phonon energy. Then the conditions for incoming and outgoing are fulfilled simultaneously, leading to double-resonant Raman scattering. This condition can be induced, for example, by the appropriate energetic positions of subbands in quantum wells, or by tuning the electronic energies by applying an electric or magnetic field [10–12].

3. EXPERIMENTAL SETUP

A main advantage of Raman spectroscopy as an optical method is its versatility from the viewpoint of sample preparation and surface conditions. No special treatment of the surfaces is required. Furthermore, Raman spectroscopy allows the investigation of very small samples and can also be applied for a locally resolved analysis with μm resolution: micro-Raman spectroscopy. The analysis may be performed

at low temperatures in a cryostat but also at room temperature in air or even at elevated temperatures. A recent experimental development is the application of Raman spectroscopy for *in-situ* studies during growth, for example, in an ultra high vacuum (UHV) vessel [13, 14] or in a gas ambient.

A typical setup for Raman spectroscopy consists of a laser light source, a spectrometer with a very good stray light suppression, and a photon detector with an extremely high sensitivity and low background level. The latter is especially relevant for interface and nanostructure analysis. These components are discussed in the next subsection. Subsequently, micro-Raman spectroscopy and the technique of *in-situ* growth analysis are treated in Sections 3.2 and 3.3, respectively.

3.1. Raman Spectrometer Components

3.1.1. Light Source

The most commonly used light sources for Raman spectroscopy are Ar^+ and Kr^+ ion lasers, which offer a series of discrete emission lines in the wavelength range from the near infrared to the near ultraviolet. This variety of laser lines is crucial for the analysis of interfaces and nanostructures, because it allows for the variation of penetration depth and for selectively taking advantage of the resonant interaction in the various semiconductor materials. Materials whose fundamental energy gap is in the visible (among the II–VI compounds, e.g., CdTe, CdSe, ZnTe, ZnSe) are transparent for some of the available laser lines and absorbing the others. The possibility of analysis with below-bandgap laserlines is typical for Raman spectroscopy while impossible for photoluminescence. Furthermore, in the absorbing range, the penetration depth can be varied between values beyond 10 μm and values around 10 nm at the wavelengths with maximum absorption. This variation of the depth of information by a factor beyond 1000 is extremely important for the analysis of buried interfaces: The region to be analyzed can be shifted from the surface and near-surface range to the underlying substrate.

The possibility of resonant interaction may be utilized to achieve a sufficient scattering intensity in extremely thin layers (down to single monolayers), but also for a selective enhancement of the scattering signal of one material with respect to others.

For a more versatile tuning of the resonant excitation, lasers with a continuously variable wavelength may be employed. For many years, various organic liquid dyes have been applied as a tunable laser medium. They cover the spectral region from below 400 nm to about 1000 nm. An important improvement was the introduction of a solid-state laser with continuously tunable wavelength. Its laser medium is a Ti–sapphire crystal. It covers the red and near-infrared spectral range between 600 nm and 1 μm and offers a more convenient operation and a constant efficiency over a wide power range, because in contrast to dye lasers no saturation effects occur.

3.1.2. Optics and Monochromator

The selective analysis of the different elements of the Raman tensor (e.g., for studying the sample symmetry properties and crystalline perfection) requires the variation of the polarization of the incident light. This is done by means of a $\lambda/2$ plate or a Fresnel rhombus.

The focusing of the laser beam on the sample is discussed in detail in the next section, with special emphasis on the optimization of lateral resolution in micro-Raman spectroscopy.

For collecting the Raman light a high-quality lens system with a large aperture is applied, which focuses the scattered light onto the entrance slit of the monochromator.

The separate analysis of the individual components of the Raman tensor requires the introduction of a polarization filter in front of the entrance slit.

Because of the extremely low intensity of the Raman scattered light compared with the diffusely reflected laser light that also enters the monochromator, the background reduction in the monochromator must be extremely good. Therefore, in contrast to photoluminescence experiments, for a sufficient separation of the Raman light from the undesired stray light background usually a system of two or even three grating monochromators is required. Several concepts are applied for the combination of these monochromators, varying between a pure addition of their dispersion and a partial use as stray light rejection filter [15]. Which concept is most preferable depends, for example, on the required spectral resolution, the width of the total Raman spectrum, and the intensity of the diffusely reflected laser light.

As a rather recent development, the employment of a single monochromator, together with a holographic notch filter for suppressing the diffuse laser light, should be mentioned. The main advantage of this combination is a high throughput for the Raman light.

3.1.3. Detector System

Until the late 1970s, the common detectors in Raman spectroscopy experiments were photomultipliers. They can achieve a very high efficiency of about 0.2 electrical pulses (counts) for each incident photon over the whole visible spectral range.

Since about 1980, multichannel detector systems have been commercially available, which allow the simultaneous detection of up to about 1000 spectral components of the Raman light on an array of photosensitive pixels. They caused a breakthrough in detecting spectral distributions with a very low intensity level [16] and are a prerequisite for the Raman analysis of interfaces and nanostructures. For optimizing the spectral resolution, even beyond the level that corresponds to the pixel size, the technique of scanning multichannel detection may be applied [17]. The most commonly applied types are CCD (charge coupled device) detectors [18] and, to a lesser extent, diode arrays.

All these detectors are cooled in order to reduce the electronic background. An extensive review was given by Tsang [19].

3.2. Micro-Raman Spectroscopy

In those Raman experiments which do not require microscopic lateral resolution (i.e., "macro"-Raman experiments), the laser beam is focused on the sample by a single spherical or a cylindrical lens, resulting in a point or line focus. The latter offers the advantage of a reduction of the local

power density by several orders, compared to a point focus. In this way heating effects, photoinduced chemical reactions, or screening of electric fields by excessive photocarriers can be avoided.

If, on the other hand, an analysis with a lateral resolution as good as possible is required, (e.g., for the assessment of a lateral strain distribution or compositional inhomogeneities, or for the study of patterned structures), a μm or even sub-μm focus is achieved by the application of a microscope objective: "micro"-Raman spectroscopy. This technique may also be applied for focusing on cleaved side-faces of heterostructures, which gives two new possibilities:

(i) due to the different crystallographic orientation of the side-faces with respect to the top surface, the symmetry selection rules are modified and different phonon modes may be observed: for example, on zincblende (110) surfaces the transverse optical phonon appears, in contrast to the longitudinal one on (001) surfaces.
(ii) a series of Raman spectra can be taken from a side face during a scan from the front edge to the substrate interface, which allows a depth-resolved characterization of the heterostructure with μm resolution.

When focusing down to the μm range, care has to be taken to avoid thermal damaging of the focus region due to the enhanced power density. Therefore, with decreasing focus area, the laser power has to be reduced accordingly (i.e., sub-mW intensities are appropriate).

In micro-Raman spectroscopy the focusing microscope is used as well for collecting the scattered light. Its aperture angle is much wider than that of long-focal lenses. This very efficient collection of the Raman light enables the detection of a sufficient signal in spite of the very small scattering area and the low incident light intensity.

Low-temperature micro-Raman spectroscopy is enabled by the use of extremely compact cryostats in combination with microscope objective lens systems with a working distance up to 8 mm. Of course, they have a low aperture angle, which leads to a reduced collecting efficiency.

Finally, it should be mentioned that a spatial resolution of 150 nm, which is distinctly below the laser wavelength, was reported [20], which was achieved by near-field Raman spectroscopy. This technique utilizes evanescent waves in the near-field region at the end of a glass fiber whose lateral size is below the wavelength of the employed light. Because of the extreme attenuation of these waves, this technique may only be applied for systems with a very high Raman scattering efficiency.

3.3. Online Growth Monitoring

Among the motivations for employing *in-situ* Raman spectroscopy during semiconductor growth processes are

(i) It has an enhanced depth of information as compared to electron-based techniques, which allows the observation of the development of a surface to an interface during overlayer growth.
(ii) During growth of mixed compounds, it allows the study of the composition with a typical accuracy of 0.5%.
(iii) From the anti-Stokes-to-Stokes intensity ratio the sample surface temperature can be deduced.
(iv) No UHV conditions are required, which enables the application during gas phase epitaxy, where no electron-based methods can be used.

For the real-time application of Raman spectroscopy in monitoring growth processes, the signal-to-noise ratio is a crucial aspect. While for steady-state applications no limits for the signal integration time exist, online applications require integration times short enough for monitoring the dynamics of the process under investigation. For epitaxial growth this means 1 to 10 seconds, which is realistic with today's Raman setups, equipped with multichannel detection systems, when resonant signal enhancement is exploited.

Optimizing the scattering efficiency is extremely important, because the input power density has to be kept sufficiently low to avoid influencing the growth process, which imposes a limit in the range from 10 to 50 mW, assuming a focus size in the 100 μm range.

In this context, the aperture angle for collecting the Raman light also plays an important role. In the growth vessel [e.g., an UHV chamber for molecular beam epitaxy (MBE) growth] the sample should be rather close to the window in order to achieve a sufficient light collection efficiency. For this purpose a reentrant window may be used. On the other hand, the distance between the sample and the window must be sufficient to avoid hampering of the layer deposition process.

An example of a setup design for *in-situ* Raman spectroscopy during MBE growth is shown in Figure 2. It is based on a conventional UHV vessel with various Knudsen cells, whose geometrical axes coincide at the substrate position. The extension for Raman spectroscopy consists of two additional windows, whose normal axes also coincide on the sample: one entrance window for the laser light and a second window for the scattered light. The position of the latter

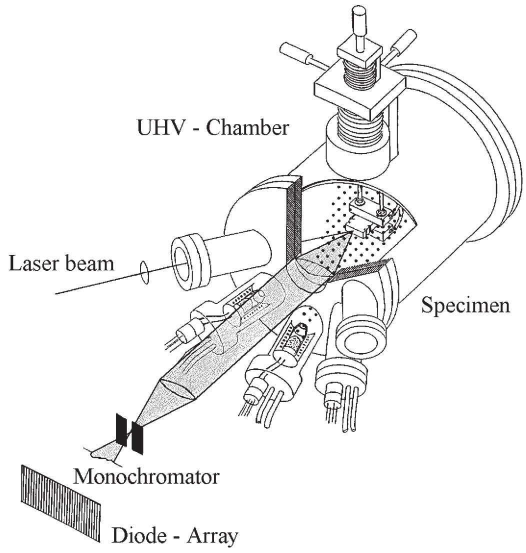


Figure 2. Experimental setup for monitoring MBE growth by Raman spectroscopy. The monochromator system is only indicated by its entrance slit. Reprinted with permission from [22], V. Wagner et al., *J. Raman Spectrosc.* 27, 265 (1996). © 1996, Wiley Interscience.

on the front flange, rather close to the sample, allows an aperture of 1:4 of the lens for collecting the Raman light. Several examples of *in-situ* Raman spectroscopy, which are discussed in this review, originate from this setup.

The development of the series of Raman spectra during the growth process is expected to show an increase of the overlayer signal, together with a decrease of the substrate signal when absorption takes place in the overlayer. However, for a quantitative evaluation, a modelling of the sample as a stack of layers with a specific index of refraction for each layer is required. In the calculation of the development of the Raman intensities with increasing overlayer thickness, transmission and reflectance at each interface, as well as absorption and Fabry–Perot interference in the overlayer, are considered. This procedure is similar to, for example, reflectance spectroscopy and ellipsometry, although in our case it has to be considered that the light is generated within each layer. It was first applied for normal incidence on a sample, consisting of a single layer on a substrate [21]. More recently, it was extended to oblique incidence and multilayer stacks [22].

As a typical application of *in-situ* online Raman scattering, the monitoring of epitaxial ZnS_xSe_{1-x} growth on GaAs(100) is shown in Figure 3. Resonant enhancement of the epilayer signal was obtained by employing for excitation the He–Cd laser line at 2.81 eV, which nearly coincides with the fundamental gap of the epilayer. The integration time per spectrum was 10 s. The starting spectrum only shows the LO phonon of the GaAs substrate at 291 cm^{-1}. During deposition, new peaks arise from the epilayer, which are identified as the vibrations of ZnS_xSe_{1-x} for $x = 0.15$ (for the systematics of compositional determination, see Section 4.2). The first pair of peaks, located at 249 and 319 cm^{-1}, belongs to first-order scattering, and the others belong to second- and third-order processes, which are quite strong due to the resonant excitation. Obviously, the development of the peak intensities with increasing overlayer thickness is essentially influenced by the Fabry–Perot interferences. Their consideration allows a quantitative description of the intensity development in terms of epilayer thickness. The maximum thickness in Figure 3 amounts to 120 nm, which corresponds to a growth rate of 0.4 nm/minute.

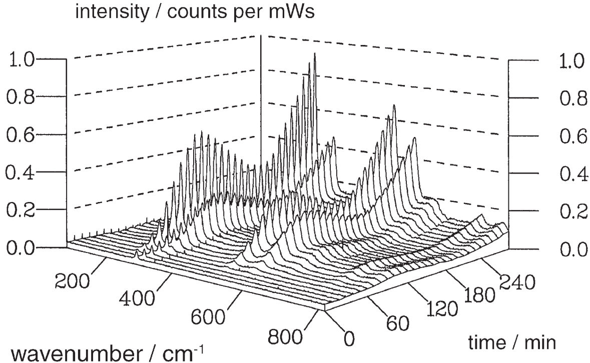


Figure 3. Online *in-situ* Raman spectra vs. deposition time during the growth of ZnS_xSe_{1-x} on GaAs(100) at 300 K. Adapted with permission from [22], V. Wagner et al., *J. Raman Spectrosc.* 27, 265 (1996). © 1996, Wiley Interscience.

Note that the GaAs substrate signal persists, because the wavelengths employed here are not absorbed in the epilayer. The GaAs LO intensity is even enhanced due to favorable interference at some epilayer thicknesses. This illustrates the possibility of enhancing the Raman sensitivity by appropriate choice of layer thickness [21].

Further results of *in-situ* Raman spectroscopy will be presented in Section 5. Besides direct layer growth, they will also deal with the *in-situ* observation of interface intermixing during annealing processes. Although a strongly enhanced temperature leads to a considerable thermal broadening of the phonon peaks in the Raman spectrum (see Section 4.1.2), Raman analysis can be performed up to temperatures far beyond 300 K.

4. ANALYSIS OF CRYSTAL LATTICE PROPERTIES

In this section experimental results of Raman scattering from lattice vibrations of II–VI layers are presented [e.g., ZnSe, epitaxially grown on GaAs(100) substrates]. These spectra are discussed in order to illustrate some aspects of relevance for the analysis of interfaces, surfaces, and quantum dots:

(i) sample properties such as lattice perfection, strain, composition of mixed compounds, and electronic properties in doped layers;
(ii) experimental aspects such as depth of information, resonant intensity enhancement, and polarization dependence.

At atmospheric pressure the binary and ternary II–VI semiconductors usually crystallize in the cubic zincblende-type structure [$F\bar{4}3m(T_d^2)$] or in the hexagonal wurtzite-type structure [$P6_3mc(C_{6v}^4)$]. In zincblende-type structures only phonon modes of F_2 symmetry are allowed. For finite values of wavevector q, these modes occur as TO and LO mode, whose eigenfrequencies are different due to the long range polarization field, which results from the partially ionic crystal bonding. In the wurtzite-type structure the F_2 mode is split in the hexagonal crystal field into A_1 (lattice deformation $\parallel$ c) and E_1 (lattice deformation $\perp$ c), whose eigenfrequencies are close to the F_2 frequency. Furthermore the $2E_2$- and $2B_1$-zone center modes are also allowed in the wurtzite-type structure [23, 24]. Among the wurtzite eigenmodes, A_1 and E_1 are both infrared (IR) and Raman active, E_2 modes are only Raman active, and B_1 are silent modes (i.e., neither IR nor Raman active). The A_1 and E_1 modes are polar modes and thus polarized parallel and perpendicular to the optical axis c, respectively. For a phonon propagation with wavevector $q \perp c$, the E_1 mode occurs as TO and LO mode, whose frequencies are different due to the long range polarization field, whereas A_1 only occurs as transverse mode: A_1(TO). For $q \parallel c$, the A_1 mode has a longitudinal character: A_1(LO) [25].

4.1. Crystalline Perfection and Orientation

Symmetry considerations determine for every crystal structure which component of the tensor $\tilde{\chi}$ may be modulated by a particular eigenmode: group theory provides the

information on which components of the corresponding Raman tensor (susceptibility derivative) have to be zero and which ones may be nonzero. These considerations lead to the so-called Raman selection rules, which serve as important guidelines for designing the scattering experiments and interpreting their results.

For compound semiconductors with zincblende structure (T_d point group symmetry) the Raman tensors for deformation-potential scattering from $q \approx 0$ phonons are given by

$$\mathbf{R}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & c \\ 0 & c & 0 \end{pmatrix} \qquad \mathbf{R}(y) = \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ c & 0 & 0 \end{pmatrix}$$

$$\mathbf{R}(z) = \begin{pmatrix} 0 & c & 0 \\ c & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{11}$$

where x, y, z are the directions of the lattice deformation. The above tensors for deformation potential scattering have only off-diagonal nonzero elements. In contrast, the Fröhlich scattering, which is symmetry-forbidden for $q = 0$ but can be observed, for example, due to a finite q-vector or the presence of a static electric field, leads to Raman tensors with only diagonal nonzero elements.

The Raman scattering intensity amounts to

$$I \propto |\mathbf{E}_{s\beta} \cdot \mathbf{R}_{\alpha\beta} \cdot \mathbf{E}_{i\alpha}|^2 \tag{12}$$

where E_i and E_s denote the polarization vectors of the incident and scattered light, respectively. For backscattering from a (100) surface (i.e., polarizations in the yz-plane), only the Raman tensor $R(x)$ leads to nonvanishing intensity. Since this implies a lattice deformation in the x-direction, which is also the phonon propagation direction, we deal with a longitudinal phonon (LO). The resulting selection rules for backscattering from a (100) surface for both scattering mechanisms are summarized in Table 1.

In contrast to the (100) surface, for backscattering from (110) surfaces only scattering from TO phonons is allowed, while LO and TO are allowed for backscattering from (111) surfaces; furthermore, each surface orientation has its specific selection rules [3]. The experimental configuration (i.e., directions of wavevectors $\mathbf{k}_i$; $\mathbf{k}_s$ and polarizations $\mathbf{E}_i$; $\mathbf{E}_s$ of incident and scattered light, respectively) is usually given in a compact notation from Porto as $\mathbf{k}_i(\mathbf{E}_i, \mathbf{E}_s)\mathbf{k}_{\bar{s}}$ [for backscattering from a (100) surface e.g., $100(010, 001)\bar{1}00$].

Table 1. Selection rules for first-order Raman scattering from $q \approx 0$ phonons for backscattering from a (100) surface of a T_d-symmetry crystal; DP means deformation-potential scattering and F means Fröhlich scattering.

	Polarization			
Surface	E_i	E_s	TO	LO
100	010	010	—	F
100	001	001	—	F
100	010	001	—	DP
100	011	011	—	DP & F
100	011	$0\bar{1}1$	—	—

As an illustration of the selection rules, Figure 4 shows a Raman spectrum (scattering intensity vs. wavenumber) of an undoped, strain-relaxed ZnSe layer, epitaxially grown on a GaAs(100) substrate. This spectrum was recorded in the backscattering configuration $100(010, 001)\bar{1}00$, as depicted schematically in the inset. The measurement was performed at a sample temperature $T = 80$ K with the excitation wavelength $\lambda = 476.5$ nm ($\tilde{\nu} = 20986$ cm^{-1}). For this wavelength the ZnSe layer is transparent. Therefore not only the ZnSe layer phonon modes but also the GaAs substrate phonon modes can be observed. According to the selection rules for a (100) surface, in the configuration $100(010,001)\bar{1}00$ only Raman scattering from LO phonons is allowed. Indeed, in the spectrum two pronounced peaks appear, which are attributed to the $q \simeq 0$ LO phonons of ZnSe and GaAs. The LO phonon of ZnSe is observed at a wavenumber of 256 cm^{-1} for the sample temperature $T = 80$ K; this corresponds to the phonon wavenumber of undoped ZnSe in the absence of strain. The LO wavenumber of the semi-insulating GaAs amounts to 294.7 cm^{-1}.

Beside the LO modes, two much weaker peaks are observed, which are identified from their wavenumbers as TO phonons. However, due to symmetry-selection rules, scattering from TO phonons is forbidden on (100) surfaces. This weak TO phonon scattering can be induced by crystal defects (e.g., dislocation lines, grain boundaries, etc.) but may also originate from experimental conditions: the exact backscattering geometry is not fulfilled due to (i) finite acceptance angle of the scattered light and (ii) deviation from normal incidence. This so-called *quasi*-backscattering geometry can induce a TO/LO intensity ratio that amounts to a few percent.

For a wurtzite material, like CdSe, the symmetry-selection rules are summarized in Table 2. Backscattering from planes perpendicular to the optical axis c should yield the $A_1(LO)$ and the E_2 mode for parallel polarization [configuration $z(y, y)\bar{z}$], while for crossed polarization only the E_2 mode should appear. In backscattering from surfaces which contain the optical axis c, the $A_1(TO)$ and the E_2 mode are

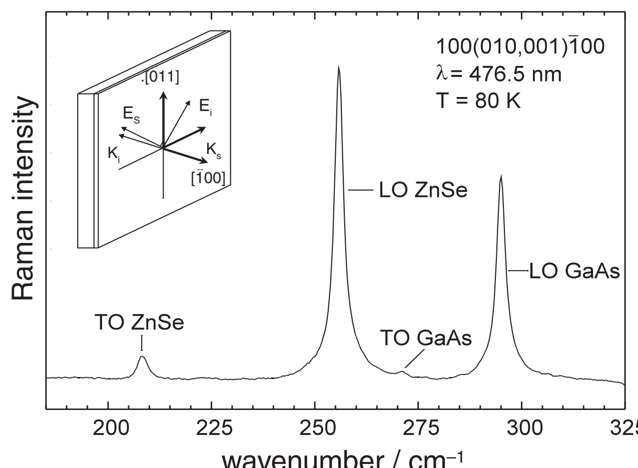


Figure 4. Typical Raman spectrum of a ZnSe/GaAs heterostructure at $T = 80$ K in the scattering configuration $100(010,001)\bar{1}00$. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

Table 2. Selection rules for 90° scattering and backscattering from various surfaces of a wurtzite crystal.

Geometry	q	Polarization parallel (HH)	Polarization crossed (VH)
Backscattering	$q \| c$	A_1(LO), E_2	E_2
Backscattering	$q \perp c$	A_1(TO), E_2	E_1(TO)
90° scattering		E_2	E_1(TO), E_1(LO)

allowed for parallel polarization [configuration $y(x,x)\bar{y}$], while for crossed polarization only the $E_1(TO)$ mode should be observed. Finally, for 90° scattering, the E_2 mode should occur for parallel and the $E_1(TO)$ and $E_1(LO)$ for crossed polarization. With these three scattering configurations it is possible to detect all zone center phonons of CdSe and to determine the lattice orientation. Because of the dependence of Raman signals on the wavevector and polarization of the incident and scattered light the rotation of the polarization of incident and scattered light leads to modulations of the peak intensities in the deformation potential allowed scattering. From these modulations the crystallite axes parallel to the crystal surface can be obtained with high accuracy.

4.1.1. Resonance and Depth of Information

As already mentioned in Section 2.3, the Raman scattering cross section can show a pronounced resonance when the excitation laser photon energy approaches an energy gap of the semiconductor material. Resonant Raman scattering constitutes the basis for (i) the detection of extremely thin layers and nanostructures and (ii) the selective analysis of one particular layer, embedded between others.

However, one should keep in mind that a variation of the excitation laser wavelength in many cases also leads to a variation of the penetration depth. Strong wavelength-dependent changes in the penetration depth occur in the neighborhood of the critical points in the combined density of the electron states, such as the fundamental electronic energy gap E_0. For ZnSe, the E_0 gap at 300 K amounts to 2.68 eV, which corresponds to 21614 cm^{-1}. Therefore, the Ar–ion laser excitation wavelength $\lambda = 476.5$ nm ($\tilde{\nu} = 20986$ cm^{-1}), which was employed in spectrum (4) is in the transparent range but rather close to the E_0 resonance. In contrast, the excitation lines at 19436 cm^{-1} (Ar–ion laser) and 24207 cm^{-1} (Kr–ion laser) give a comparably off-resonant condition in ZnSe; however, the former laserline leads to a negligible absorption coefficient resulting in a transparency of ZnSe whereas the latter line probes only the upper 95 nm of the ZnSe layer. In this way, the appropriate choice of laser line allows a focussing to the depth range of interest.

4.1.2. Temperature Dependence of Phonon Wavenumber and Linewidth

The phonon mode wavenumbers and linewidths may be influenced by a variety of effects, such as crystalline imperfections, a biaxial strain, free carriers, or a variation of the sample temperature T. One of the conditions for separating these effects is the knowledge of the intrinsic phonon eigenfrequencies and halfwidths as a function of the sample temperature.

Generally, the implication of temperature variation on the phonon properties is due to anharmonic effects in lattice dynamics. Therefore, the temperature dependence increases with increasing temperature, and its quantitative behavior depends on the material, especially on its phonon wavenumber value. For II–VI compounds the phonon wavenumbers are generally below 600 cm^{-1}, which corresponds to a phonon quantum energy <74 meV.

Here, the temperature dependence is discussed for the example material ZnSe, whose low-temperature phonon wavenumber is 256 cm^{-1} (≅31.7 meV). When T is increased, starting from 4.2 K, the phonon wavenumber is nearly constant up to about 80 K. A further T increase to 300 K leads to a wavenumber of 252 cm^{-1} (i.e., a decrease of about 4 cm^{-1}). Although this is a relative variation of only 1.5%, it cannot be neglected, because it is in the same order as strain-induced wavenumber changes.

Just like the phonon wavenumber, the spectral linewidth (full width at half maximum—FWHM) of the phonon is also almost constant in the temperature range $T < 80$ K. For samples with a high structural perfection its value is as low as 1.4 cm^{-1}. The spectral linewidth is directly connected with the phonon lifetime. For the LO phonons in ZnSe the observed intrinsic linewidth for $T < 80$ K corresponds to a lifetime $\tau \approx 4$ ps, in good agreement with the value of $\tau \approx 4.8$ ps determined directly by time-resolved Raman scattering [26]. For temperatures $T > 80$ K the FWHM clearly increases, up to factor of 4 at room temperature.

The corresponding decrease of the phonon lifetime with increasing sample temperature due to anharmonic processes in the crystal may be explained in terms of the decay of an optical phonon into acoustical phonons, for example, according to the Klemens model [27], whose basic idea is a stimulation of the decay into acoustical modes by the increasing T-induced occupation of the acoustical mode states (cf. stimulated emission).

These experimental results on the T-dependence of the optical phonon behavior imply that for high-accuracy experiments concerning, for example, disorder, local inhomogeneities, or strain, T-induced anharmonic effects should be avoided, which is achieved by cooling the sample to about 77 K by liquid nitrogen.

4.2. Composition of Mixed Compounds

Raman spectroscopy from mixed compounds is of special relevance for interfaces and quantum dots, because here intermixing effects are quite common, and Raman spectroscopy is one of the few methods for a nondestructive analysis. Therefore, fundamentals of lattice dynamics in ternary and quaternary compounds will be discussed here and Raman results from various II–VI mixed compounds will be presented.

For the $q \simeq 0$ optical vibrational modes of ternary compounds, such as $ZnSe_{1-x}S_x$, originating from the mixing of two binaries, we distinguish between two types of mode behavior:

- The one-mode behavior, which is phenomenologically quite similar to binary compounds. For every composition of the ternary $AB_{1-x}C_x$ only one TO and one

LO mode exist. With changing composition, the spectral position of these modes varies continuously and approximately linearly between the values of the constituting binary compounds. Usually, the intensity of these modes remains nearly constant.

- The two-mode behavior, which is quite distinct from binary compounds. In this case, for every composition $AB_{1-x}C_x$ two TO and two LO modes exist; they represent the A–B and A–C vibrations, respectively. The mode wavenumbers depend on the composition and the intensities are approximately proportional to the compositional fractions x and $1-x$. The LO–TO wavenumber splitting ($\simeq$mode oscillator strength) of the A–B and A–C vibration modes is approximately proportional to the relative abundance of A–B and A–C bonds, respectively. For very low concentrations of one type of bond ($x \approx 0$ or $x \approx 1$), the corresponding pair of TO and LO modes turns into a single local vibration mode.

The occurrence of either one- or two-mode behavior essentially depends on the mass difference of the constituent elements, which is directly reflected in the phonon wavenumbers of both binaries. For example, $ZnSe_{1-x}S_x$ belongs to the two-mode type, whereas $ZnSe_{1-x}Te_x$ belongs to the one-mode class. One of the most commonly used and cited models is the "modified random element isodisplacement" (MREI) model by Chang and Mitra [28], which is a modified and extended version of the "Random Element Isodisplacement" model by Chen et al. [29].

As an example for the compositional dependence of the phonon wavenumbers in the case of two-mode behavior, Figure 5 shows a series of spectra of (100) oriented $ZnSe_{1-x}S_x$ layers with different compositions ($x = 0$; $x = 0.06$; $x = 0.35$; $x = 0.73$; $x = 1.0$). Due to the GaAs substrates, all spectra show the GaAs LO peak at 292 cm^{-1}. The wavenumbers and intensities of all other peaks strongly depend on the ternary composition. The high-wavenumber mode corresponds to the Zn–S vibration. With decreasing S content its wavenumber continuously decreases from 350 cm^{-1} for pure ZnS ($x = 1$) to 297 cm^{-1} for $x = 0.06$. The latter case may be interpreted as the localized vibrational mode of S in ZnSe. Similarly, the low-wavenumber mode, which represents the Zn–Se vibration, starts at 256 cm^{-1} for pure ZnSe and decreases with increasing S content. The experimentally observed compositional dependence of the wavenumbers of both modes quantitatively agrees with the MREI model. Furthermore, the intensity ratio of both modes I_{Zn-S}/I_{Zn-Se} is approximately proportional to the compositional ratio $x/(1-x)$. The pronounced compositional dependence of the mode wavenumbers (e.g., 0.5 cm^{-1} per % for the Zn–S LO mode) allows an accurate determination of the ternary composition, although the extremely high accuracy of X-ray diffraction cannot be achieved by this technique. In the case of strained layers an additional strain-induced wavenumber shift will be superimposed (see Section 4.3), which in principle might lead to an inaccuracy in the determination of the composition. This problem is circumvented by evaluating the wavenumber difference between the Zn–S and the Zn–Se mode, which is strain-independent, since both modes exhibit the same strain-induced wavenumber shift.

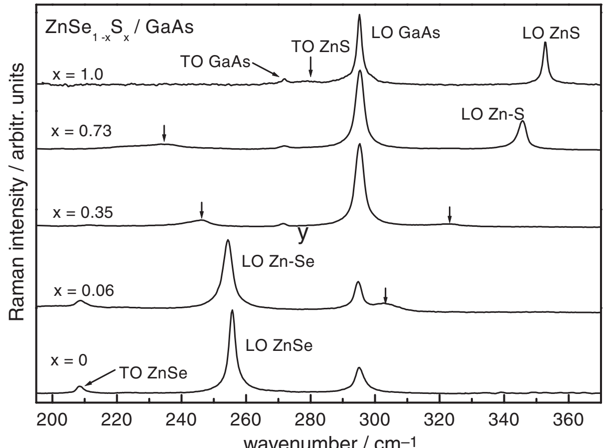


Figure 5. Raman spectra of $ZnSe_{1-x}S_x$ layers on GaAs in the composition region $0 < x < 1$ at $T = 80$ K. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

Besides the mode wavenumbers, also the peak shape and halfwidths are influenced by the mixing. With respect to the binaries a distinct broadening occurs. For the Zn–S mode, the maximum halfwidth amounts to 16 cm^{-1} for $x = 0.2$, as compared to 3 cm^{-1} for the binary ZnS. The broadening is accompanied by an increasingly asymmetric spectral lineshape. Both effects originate from the compositional disorder, which is inevitably induced by the random distribution of the S and Se atoms in the ternary compound. This disorder leads to a violation of the translation invariance in the crystal lattice and therefore to a relaxation of the $q_{Phonon} \approx 0$ condition in the Raman scattering process. Due to the dispersion of the optical phonon, the scattering for $q_{Phonon} \neq 0$ leads to contributions with reduced wavenumber, which broaden the peaks and make it increasingly asymmetric. The spectral width and lineshape can be described quantitatively in terms of a reduced correlation length, using the "spatial correlation model" [30], which is discussed in detail in Section 6.1.1. A further consequence of the disorder is a relaxation of the selection rules, resulting in the appearance of the symmetry-forbidden TO modes in the Raman spectrum from (100) surfaces.

Beside ternary mixed compounds, also quaternaries may originate spontaneously by intermixing at interfaces between binary layers AB and CD. In addition, they are of high relevance for technological application, because a controlled mixing of binaries AB and CD results in the mixed compounds $A_{1-y}C_yB_{1-x}D_x$, which offer the advantage of an independent tuning of the electronic bandgap and the lattice constant by the appropriate choice of the compositional ratios y of the cations A, C and x of the anions B, D. In this way, quaternary layers can be incorporated into heterostructure stacks without inducing strain. Raman spectroscopy from their vibrational modes is a useful technique for the identification of quaternary layers and for the determination of their composition. For a check of the compositional homogeneity lateral scans by micro-Raman spectroscopy can be applied.

An example of these quaternary compounds is $Zn_{1-y}Mg_ySe_{1-x}S_x$. Although MgSe and MgS crystals naturally have the NaCl structure, $Zn_{1-y}Mg_ySe_{1-x}S_x$ can be stabilized in the zincblende structure by growing the quaternary layers on GaAs substrates. The ternaries $Zn_{1-y}Mg_ySe$ and $ZnSe_{1-x}S_x$, which may be considered as the constituents of $Zn_{1-y}Mg_ySe_{1-x}S_x$, both show a two-mode behavior, as predicted by the MREI model. Therefore, four modes are expected for this quaternary. This was confirmed experimentally by the occurrence of four peaks in the Raman spectrum: the Zn–Se vibration in the range slightly below 250 cm^{-1}, Mg–Se and Zn–S near 320 cm^{-1}, and the Mg–S mode at about 360 cm^{-1} [31].

Another class of quaternaries consists of one anion element, combined with three cation elements, like $Cd_xZn_yMg_{1-x-y}Te$. The room-temperature Raman spectrum of the quaternary alloy $Cd_{0.4}Zn_{0.3}Mg_{0.3}Te$ (Fig. 6) shows three LO–TO mode pairs, which are identified as CdTe-like, ZnTe-like, and MgTe-like modes [32]. Just like in the ternaries discussed above, also in this case the TO phonon peaks are disorder-induced. For the same reason also additional features of the disorder-allowed acoustical modes LA and TA can be seen. The multimode behavior of this alloy can be explained on the basis of the MREI model.

4.3. Biaxial Strain Due to Lattice Mismatch

In most heterostructures strain due to lattice mismatch between adjacent layers plays a crucial role. On the one hand, strain may be favorable, because it is utilized to achieve shifts of electronic bandgaps or to induce the formation of self-organized quantum dots due to relaxation effects (e.g., in the systems InAs/GaAs or CdSe/ZnSe and CdSe/BeTe). They will be the subject of Section 6.3. On the other hand, in many cases strain will be disadvantageous or even detrimental, because it gives rise to the formation of lattice defects upon relaxation, which deteriorate the optical and transport properties. This occurs already for slightly lattice-mismatched material systems, such as ZnSe layers on GaAs. The relative lattice mismatch of only 0.27% induces the relaxation of the ZnSe layer when its thickness exceeds about 150 nm.

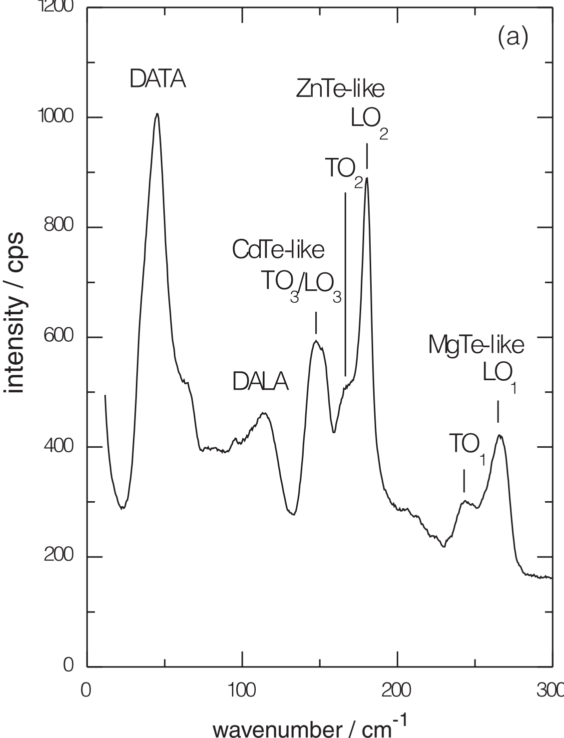


Figure 6. Raman spectrum of the quaternary $Cd_xZn_yMg_{1-x-y}Te$ showing the LO and TO phonons due to the three-mode behavior as well as the disorder-allowed TA and LA. Reprinted with permission from [32], R. Vogelgesang et al., *J. Raman Spectrosc*. 27, 239 (1996). © 1996, Wiley Interscience.

In quantum dots produced by sol/gel processes strain can be also caused by the surrounding matrix. The synthesis of these samples implies sintering processes or the implantation of the dots into the molten matrix material, whose thermal expansion constants are usually different from those of the semiconductor. As a consequence, the matrix can induce an effective pressure on the dot surface upon cooling the sample.

For epitaxial layers with sufficient thickness, strain and distortion can be characterized with an extremely high accuracy by X-ray diffraction. However, for layers in the monolayer range and for those cases where a lateral resolution in the μm range is required, optical methods are a very interesting alternative. Besides photoluminescence spectroscopy, utilizing the strain-induced modification of the excitonic features, Raman spectroscopy is also used for this purpose. The information depth for the usual Raman measurement on II–VI semiconductors with below-bandgap excitation is comparable to X-ray diffractometry. In addition, the information depth of the Raman experiment can be varied by changing the laser wavelength. The lateral resolution in the μm range with a microscope setup offers the interesting ability of investigating laterally structured semiconductor layers. However, the accuracy of the Raman method is distinctly below that of X-ray diffraction.

In Raman spectroscopy, the lattice strain of a thin epilayer due to a biaxial stress leads to small but significant wavenumber shifts and/or splitting of the phonon wavenumbers. Compressive strain induces an increase of the phonon wavenumber, while tensile strain leads to a decrease. In both cases the amount of wavenumber shift is proportional to the strain value. Strain-induced wavenumber shifts can mask other phonon wavenumber shifts (e.g., due to composition or doping). Therefore, it is necessary to separate the strain-induced part of the shift from the other parts.

For a quantitative evaluation of strain-induced phonon wavenumber shifts, results from pressure-dependent Raman experiments can be employed. With the knowledge of the amount of the shift per kbar, the effective pressure due to the lattice mismatch in strained quantum wells and superlattices can be calculated. Besides, pressure-dependent Raman experiments are also used for the investigation of phase transitions [33–37].

In zincblende bulk material, hydrostatic pressure induces a linear shift of all optical phonon wavenumbers, which is proportional to the pressure, while uniaxial as well as biaxial stress leads not only to a wavenumber shift, but also to a splitting of the degenerated TO band for backscattering

from planes containing both unstressed and stressed directions. From the amount of the shift and the splitting one can obtain experimental values for the phenomenological phonon deformation potentials (p, q, r), which describe the changes in the force constant of these phonons, and the mode Grüneisen parameter γ_i. The shift can be subdivided into a hydrostatic part $\Delta\omega_H$ induced by the change of volume and splittings $\Delta\omega$ due to the deformations in shape at constant volume [38–43]. For the important case of biaxial stress the threefold degenerated optical phonon at ω_0 (LO/TO splitting is neglected for the moment) is splitted into a singlet mode ω_s and twofold degenerated doublet mode ω_d:

$$\omega_s = \omega_0 + \Delta\omega_H + 2/3\ \Delta\omega \tag{13}$$

$$\omega_d = \omega_0 + \Delta\omega_H - 1/3\ \Delta\omega \tag{14}$$

with the hydrostatic part

$$\Delta\omega_H = \frac{(p+2q)(S_{11}+2S_{12})}{6\omega_0} \cdot 2\sigma \tag{15}$$

and the splitting

$$\Delta\omega = \begin{cases} \dfrac{(q-p)(S_{11}-S_{12})}{2\omega_0} \cdot \sigma & \text{for } z\|[001] \\ \dfrac{r\,S_{44}}{2\omega_0} \cdot \sigma & \text{for } z\|[111]. \end{cases} \tag{16}$$

If the phonon wavevector is chosen either parallel or perpendicular to the unstressed direction z, as is usually the case, the LO/TO splitting introduces no additional complexities and ω_0 can be simply replaced by ω_{LO} or ω_{TO}, respectively. For pseudomorphic epitaxial layers the in-plane lattice constants have the value of the substrate lattice. For zincblende materials with the [001] growth direction this leads to a tetragonal deformation of the cubic lattice. Due to this deformation, the epilayer lattice constant in the growth direction is also modified. Its value is determined by the Poisson number of the epilayer material. This situation corresponds to a biaxial strain. The relative lattice mismatch f of the epitaxial layer with respect to the substrate is defined as

$$f = \frac{a_s - a_l}{a_l} \tag{17}$$

where a_s and a_l are the bulk lattice constants of the substrate and epilayer, respectively. This mismatch determines the strain component in the layer parallel to the interface,

$$\varepsilon_\| = \varepsilon_{xx} = \varepsilon_{yy} = (a_\| - a_l)/a_l \tag{18}$$

where $a_\|$ is the lattice constant of the pseudomorphic epitaxial layer in the interface plane. Realistic mismatch values for pseudomorphic systems are in the order of some tenth of a percent. The corresponding strain values can be converted into an equivalent stress through the elastic compliance tensor values S_{ij} of the epilayer material:

$$\sigma = \varepsilon_\|/(S_{11}+S_{12}) \tag{19}$$

This results in equivalent stress values σ in the kbar range. For (001) heterostructures one obtains in backscattering for the symmetry-allowed LO phonon of the epilayer by using Eqs. (13), (15), (16), and (19):

$$\omega_{LO} = \omega_{0,LO} + \frac{q\,\varepsilon_\|}{\omega_{0,LO}} + \frac{p\,\varepsilon_\| S_{12}}{\omega_{0,LO}(S_{11}+S_{12})} \tag{20}$$

where ω_{LO} and $\omega_{0,LO}$ are the LO wavenumbers for the strained epilayer and unstrained layer material, respectively. For a pseudomorphic ZnSe epilayer on GaAs(100), one obtains by inserting the p, q, and S_{ij} values of ZnSe a wavenumber shift for the ZnSe LO phonon with respect to its bulk value of about 1.15 cm^{-1}. Thus, although for this slightly mismatched material system ($f = 0.27\%$) the relative LO wavenumber shift for pseudomorphic growth is only about 0.45%, it is by far sufficient for detection by Raman spectroscopy.

At the onset of strain relaxation, the employment of laser lines with different penetration depths allows a depth-dependent strain analysis. In this way, for ZnSe layers on GaAs(100), grown by metalorganic vapor phase epitaxy, the gradual progression of the relaxation from the surface to the epilayer–substrate interface was studied [31].

4.4. Coupled Plasmon–Phonon Modes in Doped Materials

A discussion of doping effects in Raman spectroscopy is of relevance here, because interfaces may induce, for example, carrier trapping, which results in carrier-depleted interface regions between doped layers, which will be the subject of Section 5.1.2.

Doping leads to new eigenmodes of the sample, which are observed in the Raman spectrum. The origin for this modified behavior is the collective charge density oscillation of free carriers, denoted as plasmon, whose wavenumber depends on the carrier concentration. With this charge density oscillation an oscillating macroscopic electric field is associated. This field leads to a coupling of the plasmon with the LO-phonon, because the latter also exhibits a macroscopic electric field. The coupling of these modes results in two new eigenmodes, denoted as plasmon–LO–phonon (PLP) modes, which replace the former LO and plasmon mode. The PLP modes are labeled as Ω_- and Ω_+, and their wavenumbers depend on the carrier concentration. Both PLP modes have a longitudinal character and therefore they may occur only at those wavenumbers for which the dielectric function $\varepsilon(\omega)$ is zero.

For a doped polar semiconductor the dielectric function can be written as the sum of the susceptibility contributions $\chi(\omega)$ of valence electrons (VE), phonons (Ph), and free carriers (FC) [44]:

$$\varepsilon(\omega) = 1 + \chi_{VE}(\omega) + \chi_{Ph}(\omega) + \chi_{FC}(\omega) \tag{21}$$

In the low-wavenumber range, which is relevant for the Raman shifts, the χ_{VE} contribution is constant. Therefore it

is considered formally in the constant $\varepsilon_\infty := 1 + \chi_{VE}$. In contrast, the contributions of the phonon and the free carriers show a pronounced angular-frequency dependence:

$$\chi_{Ph}(\omega) = \frac{\Delta\chi \cdot \omega_{TO}^2}{(\omega_{TO}^2 - \omega^2) - i\omega_{\tau,Ph} \cdot \omega} \tag{22}$$

where ω_{TO} is TO phonon resonance angular frequency, $\Delta\chi$ is the phonon oscillator strength, which originates from the bond polarity, and $\omega_{\tau,Ph}$ represents the phonon damping.

The free carriers experience no restoring force and therefore the eigenfrequency of their oscillator is zero; the free carrier susceptibility can be described by the Drude–Lorentz model

$$\chi_{FC}(\omega) = \frac{-\omega_p^2}{\omega^2 + i\omega_{\tau,FC} \cdot \omega} \tag{23}$$

where ω_p is the plasma angular frequency and $\omega_{\tau,FC}$ is the scattering angular frequency of the free carriers.

When damping effects are neglected, one obtains from the $\varepsilon(\omega) = 0$ condition for longitudinal modes the eigenfrequencies $\Omega_\pm$ of the coupled PLP modes Ω_- and Ω_+:

$$\Omega_\pm^2 = \frac{1}{2}\left(\omega_{LO}^2 + \omega_p^{*2}\right) \pm \frac{1}{2}\left[\left(\omega_{LO}^2 + \omega_p^{*2}\right)^2 - 4\omega_p^{*2}\omega_{TO}^2\right]^{1/2} \tag{24}$$

where

$$\omega_p^* = \left[\frac{\omega_p^2}{\varepsilon_\infty}\right]^{1/2} = \left[\frac{1}{\varepsilon_\infty}\frac{ne^2}{\varepsilon_0 m^*}\right]^{1/2} \tag{25}$$

is the plasma resonance angular frequency. The coupling of the LO phonon with the plasmon leads to a renormalization of the zeros of the dielectric function; note that the plasma resonance angular frequency depends on the free carrier concentration.

Figure 7 shows the doping dependence of the eigenfrequencies of the coupled plasmon–LO–phonon modes of *n*-ZnSe. For low carrier concentrations ($n < 0.5 \cdot 10^{17}$ cm^{-3}), the coupling is very weak: the PLP modes nearly coincide with the original LO and plasmon mode. Strong coupling occurs between 10^{17} and 10^{18} cm^{-3}. For higher concentrations the Ω_+ mode approaches the plasmon mode. Note, however, that the Ω_- mode asymptotically approaches the TO phonon wavenumber instead of the LO, an effect due to the screening of the electric field of the LO phonon by the free carriers.

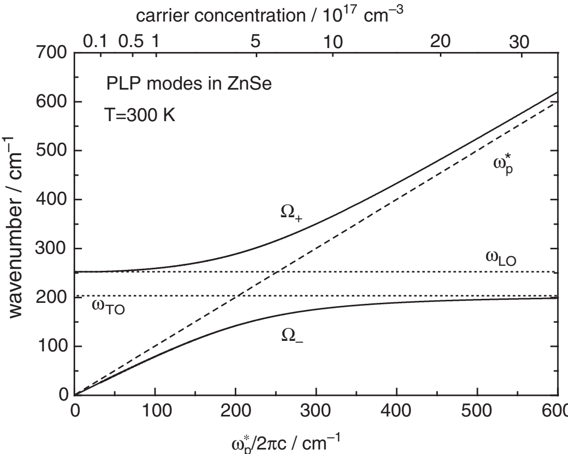


Figure 7. Doping dependence of the eigenfrequencies of the coupled plasmon–LO–phonon modes of *n*-ZnSe. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

The coupling also affects the polarization selection rules of the PLP modes: a near-phonon PLP mode is observable in an off-diagonal scattering configuration whereas a near-plasmon mode is observable for diagonal scattering. The peakwidth of both modes drastically increases with increasing distance from the phonon frequency, since the damping of the electrons is much higher than that of the phonon.

From Figure 7 it can be concluded that the Ω_+ mode is best suited for doping analysis: it shows the strongest doping dependence, and for our example (*n*-ZnSe) it can be distinguished from the LO phonon for $n > 10^{17}$ cm^{-3}. In the low-doping range, the Ω_- mode often appears quite weakly. Therefore, the doping analysis by Raman spectroscopy in *n*-ZnSe is limited to the range $n > 10^{17}$ cm^{-3}. When comparing different materials, the relevant parameter for the sensitivity of Raman spectroscopy for doping analysis is the effective mass m^* of the free carriers, which appears in Eq. (25). The lower m^*, the higher the Raman sensitivity.

For a quantitative doping level analysis from the PLP modes the finite wavevector transfer in the Raman scattering process has to be taken into account [see Eq. (6)]. The finite q-value of the PLP mode implies a periodic carrier concentration gradient, resulting in an additional restoring force, which means an enhanced eigenfrequency. This spatial dispersion can be taken into account by the hydrodynamical theory [45] as an extension of the Drude model, which yields for the free carrier contribution

$$\chi_{FC}(\omega, q) = -\frac{\omega_p^2}{\omega^2 - \frac{3}{5}v_F^2 q^2 + i\omega_\tau \omega} \tag{26}$$

where v_F is the Fermi velocity. A calculation of the wavevector dependence of the PLP frequencies shows that for *n*-ZnSe the dispersion is negligible for $n < 5 \cdot 10^{17}$ cm^{-3}, due to the relatively large effective carrier mass.

P-doped samples with a rather low carrier concentration constitute an additional challenge because of the very high effective hole mass and the quite low carrier mobility (e.g., for *p*-ZnSe they amount to $m_p^* = 0.6\,m_0$, and $\mu_p = 60$ cm^2/Vs, respectively).

Nevertheless, for nitrogen-doped ZnSe with $p = 2.4 \cdot 10^{17}$ cm^{-3} at 300 K, the comparison of Raman spectra, taken at $T = 12$ and 300 K, showed the thermal activation of *p*-carriers. At 300 K the emerging PLP mode shows a distinct additional broadening compared to the LO, due to the free carrier scattering, although the PLP wavenumber still nearly coincides with the LO [31]. In this way, the sensitivity of PLP Raman scattering for low concentrations together with high carrier masses and low mobilities is demonstrated.

5. EFFECTS OF SIZE RESTRICTION BY INTERFACES

With decreasing layer thickness in heterostructures, and especially with the decreasing size of nanostructures, interfaces and surfaces play an increasingly important role. Crucial aspects of interfaces are their structural, chemical, and doping properties, and as a direct consequence their electronic behavior, which, for example, governs the electrical characteristics of heterostructure devices. In layer and heterostructure growth (e.g., for subsequent nanopatterning), surface analysis has attracted an increasing interest, because the surface structure and quality are decisive for the further growth process, either homo- or heteroepitaxial.

Among the established methods for the experimental analysis are transmission electron microscopy at interfaces and electron spectroscopy methods at surfaces. The former technique yields information about structural perfection and intermixing with a spatial resolution in the atomic range, while the latter techniques are employed, for example, for probing the surface chemistry and for an extremely sensitive adsorbate detection.

Besides these electron-based techniques, recently methods of optical spectroscopy have found increasing interest for interface and surface analysis. At first, this may be surprising, because their rather high information depth (generally $\geq$10 nm) seems to be a drawback for surface and interface analysis. However, if a sufficient sensitivity and selectivity is achieved (i.e., if signals from surfaces and interfaces can be detected and separated from the remaining signal), the relatively high information depth turns into an advantage. During overgrowth it enables the monitoring of the development from surface to interface (*in-situ* growth control; see also Section 3). Furthermore, it allows the selective analysis of buried interfaces in a nondestructive way without a sophisticated and troublesome sample preparation.

Raman spectroscopy fulfills the requirements of sensitivity and selectivity for many material systems. A sufficient sensitivity for single atomic layers was achieved by instrumental progress during the last decades, for example, by tunable lasers for resonant enhancement of the scattering efficiency, and by the employment of extremely sensitive CCD multichannel detectors (see Section 3).

The selectivity for signals from interfaces and surfaces is based on

(i) Characteristic eigenfrequencies of surface and interface vibrational modes.
(ii) A specific polarization behavior of the light scattering process from these modes, due to the surface and interface symmetry properties.
(iii) The tuning of the laser wavelength to an electronic interband resonance of the surface or interface, which may selectively enhance its scattering efficiency by several orders of magnitude.

Our discussion of the consequences of size restriction by interfaces in Raman spectroscopy will start with a treatment of single interfaces and surfaces (Section 5.1). Here we will consider vibrational modes of abrupt and reacted interfaces, which reflect their chemical and structural properties, and we will also discuss electronic band bending, which influences the transport properties at interfaces.

Subsequently, we will draw our attention to quantum wells, multiquantum wells, and superlattices in Section 5.2. These multilayer structures give rise to new phenomena in the Raman spectrum, such as confined optical phonons modes from the individual layers, and folded acoustical phonons from the total stack.

5.1. Single Interfaces and Surfaces

5.1.1. Vibrational Modes

The subjects of this section are spectral eigenmodes which find their origin in a single interface or in a surface. When considering such interface-derived Raman signals, one has to distinguish between two different possible origins:

(i) The characteristic chemical bonds at the interfaces give rise to vibration modes whose wavenumber values are different from the bulk values of the interfacing materials. These modes are denoted as chemical vibrational modes.
(ii) The different dielectric properties of both materials induce vibrational modes which are centered at the interface. They are referred to as dielectric interface modes.

Both types of interface modes will, of course, occur not only at single interfaces but also in multiple layer systems, such as multiquantum wells and superlattices, which are the subject of Section 5.2. In these systems the multiplicity of interfaces gives rise to an enhanced intensity of the interface modes in the Raman spectrum, or even to a modified dynamics due to mode coupling between adjacent interfaces. However, the interface modes are discussed in this section on single interfaces, because their occurrence does not necessarily require multiple interfaces as, for example, folded acoustical phonons do.

The occurrence at interfaces of unique chemical bonds which do not belong to either of the constituting materials is easily understood, for example, for stacks of Si and Ge layers, in which the Si–Ge bonds occur uniquely at the interfaces and lead to new chemical vibrational modes in the Raman spectrum. These modes also may arise for heterostructures consisting of two binary compounds which have no element in common (i.e., systems AB/CD). For growth in [100] direction, the interfaces may contain either A–D or B–C bonds, depending on the switching procedure during growth. For these two types of interfaces bonds, generally quite different vibration wavenumbers are expected, especially when there is a signifcant mass difference between both cations A and C, or between both anions B and D.

The vibrational wavenumbers of chemical interfaces modes can be calculated from a semiempirical linear chain model, whose input parameters are the atomic masses and the bond strengths, which may be derived from the phonon wavenumbers of the corresponding bulk materials. However, in this calculation the strain in the interface bonds has to be taken into account, because at AB/CD interfaces generally strain values up to several percent will occur for the A–D or B–C bonds, even if the compounds AB and CD are lattice matched.

The linear chain calculation yields the eigenfrequency and the eigenmode pattern (i.e., the displacement amplitudes of the involved atoms). The mode pattern shows an extremely strong mode localization at the interface if the mode frequency is distinctly beyond the vibration frequencies of the constituting layers. For AB/CD systems this occurs if the atoms with the lowest masses build the interface (e.g., for ZnSe/BeTe, a strongly localized mode is expected for Be–Se interface bonds). A heavy-atom interface (Zn–Te bonds for ZnSe/BeTe) has a much lower eigenfrequency, which generally merges into the phonon wavenumber intervals of the constituing materials. Therefore this low-wavenumber chemical vibrational mode will penetrate much further into the adjacent layers.

Many material systems do not allow ideally abrupt interfaces but tend to intermixing. This gives rise to layers of reacted compounds whose thickness is in the range of at least several monolayers. The resulting ternary or even quaternary compounds have phonon wavenumbers which are different from those of the original materials. Therefore Raman spectroscopy is very well suited for their direct identification.

Utilizing resonant Raman scattering is an essential experimental technique for achieving a sufficient sensitivity to detect interfacial bond vibrations and to identify reacted interface layers. However, it should be mentioned that a quantitative determination of interface layer thicknesses from Raman spectroscopy is extremely difficult, since the absolute scattering efficiency usually is not known.

Finally, for the case of ideally abrupt interfaces the extremely low thickness of the interface layers (only one monolayer) results in a modified symmetry with respect to the bulk [46]. These symmetry properties should be reflected in the polarization dependence of the interface Raman peaks. Therefore polarization studies are an additional tool in Raman analysis of interfaces and surfaces.

In addition to the chemically induced vibrational modes arising from the different bonds at abrupt or intermixed interfaces, a further class of interface-induced vibration modes may be observed which originates from the different dielectric properties of the constituent layers. Maxwell's equations give rise to the existence of vibrational modes which are centered at the interfaces and show an exponential decay of the vibrational amplitude into the layers on both sides of the interface. For these modes the materials at both sides of the interface must have an opposite sign of the dielectric function $\epsilon(\omega)$ at the angular frequency ω of the dielectric modes. Therefore, a condition for their existence is that the ϵ-value of one of the materials must be negative in this frequency region. This implies that at least one of both materials must show a dielectric resonance (i.e., a phonon with an accompanying dipole moment due to noncovalent atomic bonds). The dielectric interface mode wavenumber then lies between the TO and LO phonon wavenumber of this material. If both constituting materials have noncovalent bonds and corresponding lattice vibrations with a dielectric resonance, two interface modes are expected. For stacks whose individual layer thicknesses are below the decay lengths of the dielectric modes, a coupling of the modes from the different interfaces may occur. This behavior is in contrast to the interface-bond-induced chemical vibration modes, which usually decay already within one or two monolayers. Dielectric interface modes not only occur at interfaces between two materials but also at the interfaces between a material and vacuum or air. Then they are referred to as dielectric surface modes.

Chemical Vibrational Modes from Abrupt Interfaces

The first Raman studies of chemical vibration modes from abrupt interfaces of semiconductor layers were performed on the III–V material systems InAs/GaSb and InAs/AlSb, whose interfaces may consist either of In–Sb or of Ga–As or Al–As bonds, respectively. In Raman spectra from InAs/AlSb samples with In–Sb interfaces the InSb vibration mode was observed near 190 cm^{-1} [47], which is very close to the LO wavenumber value of bulk InSb (196 cm^{-1}). The comparison of InAs/AlSb and InAs/GaSb revealed that intermixing was virtually absent for In–Sb interfaces but took place for Ga–As and was strongest for Al–As [48]. Furthermore, the modified symmetry-selection rules for the interface layers were confirmed by polarization-dependent experiments [46]. The strong dependence of the interface quality on the type of bonds (A–D vs. B–C), which originates from the different chemical bonding strength and diffusion probability, is a commonly observed phenomenon, as can also be seen from examples of II–VI interfaces, discussed below.

The strong correlation between the choice of interface bonds (A–D vs. B–C) and the resulting interface quality was also observed for ZnSe/BeTe superlattices. Here X-ray diffraction, transmission electron microscopy, and Raman spectroscopy unanimously confirm intermixing in the case of Be–Se bonds, whereas abrupt interfaces occur when Zn–Te bonds are employed [49]. In the latter case, confined optical and folded acoustical phonons in the Raman spectrum confirm the high degree of structural perfection and chemical abruptness of the interfaces (see Section 5.2). However, no chemical interface mode is observed, which may seem surprising but is explained by the correlation between the mode wavenumber and the degree of localization, as discussed above. The Zn–Te mode is expected at about 200 cm^{-1}. This is clearly within the wavenumber range of ZnSe vibrations, whose dispersion curve extends from zero to about 252 cm^{-1} (LO-wavenumber) and shows hardly any wavenumber gap between the optical and acoustical branch, because of the nearly equal atomic masses of Zn and Se. Therefore the interface-derived mode merges into the phonon continuum. For interfaces of the type Be–Se, which fulfills the wavenumber requirement for a localized mode (Be–Se vibrational wavenumber $\approx$576 cm^{-1}), its occurrence is hampered by the inferior quality due to intermixing.

Among the II–VI compounds, BeTe takes a special position, for example, with regard to the possible observation of chemical interface vibration modes. This compound, which consists of the smallest group-II element ($m_{Be} = 9$ amu) and the second-heaviest group-VI element ($m_{Te} = 127.6$ amu), has unique properties in two respects: (i) its bond polarity and (ii) its lattice dynamics.

Because of the combination of the row-II cation Be with the row-V anion Te, the electronegativity difference between cation and anion is strongly reduced with respect to other

II–VI compounds, and the bond polarity of BeTe is by far the lowest of all II–VI compounds [50]. This is expected to result in an improved lattice stability, an aspect that has made BeTe and other binary and ternary Be-chalcogenides promising candidates for II–VI optoelectronics [51].

Here we focus to its lattice dynamics, which shows two peculiar features:

(i) Due to the low atomic mass of Be, the optical phonon wavenumbers are very high: they amount to 502 cm^{-1} for the LO and 461 cm^{-1} for the TO vibration, respectively [50].

(ii) The very strong mass asymmetry ($m_{Te} = 14.2 \cdot m_{Be}$) leads to a quite broad interval of forbidden wavenumbers between the dispersion branches of the optical and the acoustical modes ("phonon gap"). This reduces the possibility for interface modes to penetrate into the BeTe layers and, consequently, gives rise to more favorable conditions for the observation of localized interface vibrational modes.

These favorable conditions are fully exploited in CdSe/BeTe heterostructures, consisting of CdSe monolayers, embedded in BeTe layers [52]. Here, no layers of a second compound with bulklike properties are involved. Thus, the BeTe phonon dispersion curve is the only relevant one.

An important motivation for this strongly lattice-mismatched material system ($\Delta a/a = 7\%$) is the formation of spontaneous CdSe quantum dots due to strain relaxation for layer thicknesses beyond about 3 ML. However, the CdSe epilayers of only 1 ML, which are discussed here, should be fully strained.

Note that also in this stack two interface configurations may be chosen: The interface (i.e., the transition bond between the BeTe and the CdSe monolayer) may be either Be–Se or Cd–Te. Just like for the ZnSe/BeTe system, pronounced intermixing occurred for Be–Se interfaces (i.e. for the atomic layer sequence ⋯–Te–Be–Se–Cd–Se–Be–Te–⋯), while the highest abruptness and structural perfection was obtained for Cd–Te interfaces (i.e., for the sequence ⋯–Be–Te–Cd–Se–Cd–Te–Be–⋯). This behavior is deduced from the Raman spectra, shown in Figure 8. Apart from the common GaAs substrate peaks at 265 and 292 cm^{-1}, the upper spectrum (Be–Se interfaces) and the lower one (Cd–Te interfaces) are quite different:

(i) While the sample with the Cd–Te interfaces shows a well defined BeTe LO phonon at the expected wavenumber of 505 cm^{-1}, for Be–Se interfaces a wavenumber-shifted, weaker, and broadened peak appears, which may originate from the ternary $BeSe_xTe_{1-x}$.

(ii) In the wavenumber range between 150 and 220 cm^{-1}, which is the domain of Cd-related vibrations, the sample with Be–Se interfaces only shows a single, very broad peak, while for Cd–Te interfaces a much better resolved structure appears.

The CdSe monolayer-related structure in the Raman spectrum of the heterostructures with Cd–Te interfaces is explained as a superposition of vibrational eigenmodes, whose wavenumbers may be derived from the linear chain model. For the fundamental vibration mode (i.e., maximum

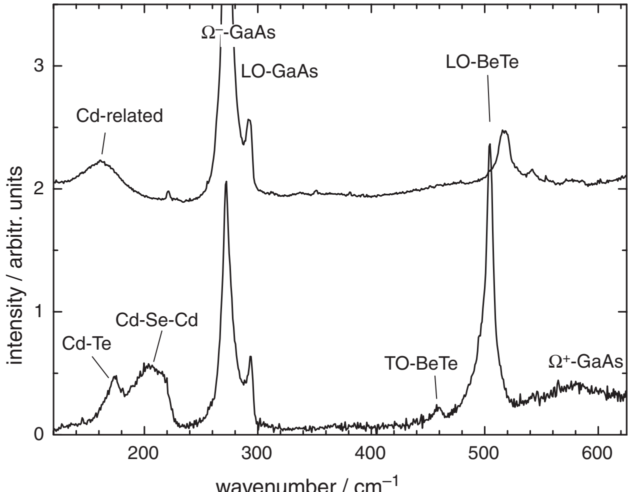


Figure 8. Raman spectra of a stack of 5 × 1 ML CdSe well in BeTe. Upper curve: Be–Se interfaces, lower curve: Cd–Te interfaces. Reprinted with permission from [52], V. Wagner et al., *Appl. Surf. Sci.* 175, 169 (2001). © 2001, Elsevier Science.

amplitude for the central Se atom, lower amplitudes for the next neighbor Cd atoms, strongly reduced amplitudes for the second-next neighbors Te, etc.), the wavenumber value 216 cm^{-1} is obtained, while the next mode (i.e., zero amplitude for the central Se atom, and maximum amplitude but opposite directions for the neighboring Cd atoms, followed by a comparable amplitude for the second-next neighbors Te, etc.) should appear at 169 cm^{-1}. These calculated wavenumbers give a reasonable explanation of the experimental result; however, for a quantitative agreement inhomogeneities have to be invoked, which constitute the onset of quantum dot formation. For this reason they will be discussed in detail in Section 6.3.

For strained CdS/ZnSe superlattices the successful observation of a localized Zn–S interface mode was reported [53]. Here high-quality Zn–S interfaces occur which give rise to a localized mode. The linear chain model predicts a wavenumber of 317 cm^{-1}, which is in very good agreement with the experimental observation. From the atomic masses, a wide separation between the Zn–S interface mode wavenumber and the CdS confined optical phonons would be expected. However, due to the strain-induced blue shift of the CdS phonons, they appear very close to the Zn–S interface mode.

Raman signatures of monolayers at interfaces were also observed for the group V elements on III–V or group IV substrates. Sb on InP(110) or GaAs(110) leads to an epitaxial monolayer, in which the Sb atoms form zigzag chains. The corresponding Raman spectra show very sharp peaks of the monolayer vibration modes, originating from Sb chains [54]. Raman peaks due to molecular vibrations were observed for As and Sb, adsorbed on Si(111) [55].

Chemical Vibrational Modes from Reacted Interfaces

Raman spectroscopy has been successfully applied for the study of reacted interfaces for a wide variety of material systems, for example, metal–semiconductor interfaces, such as Pt/Si [56, 57], and oxidized semiconductor surfaces, such as GaAs [58], InAs [59], and InSb [60]. Here we will focus on

reacted semiconductor–semiconductor interfaces. Interfaces between different III–V epitaxial layers have been investigated, for example, for InGaAs/InP, grown by MOVPE (metal organic vapor phase epitaxy). Here, intermixing-induced InAsP and InGaAsP interface layers were detected, whose thickness showed a systematic dependence of the gas switching procedure at the interfaces [61–63].

II–VI epilayers on III–V substrates were studied, for example, for the lattice-matched systems CdS/InP(110) [24], ZnSe/GaAs(100) [64], and CdTe/InSb(100) [65]. Here, for many growth conditions interlayers of the corresponding III–VI compound were detected: Ga_2Se_3 (Raman peak, e.g., at 155 cm^{-1}) at the ZnSe/GaAs interfaces, and accordingly, at the interface CdTe/InSb elementary Sb (Raman peaks at 113 and 152 cm^{-1}) and In_2Te_3 were observed due to the reaction of the deposited Te with In from the substrate. This reaction was considerably suppressed by starting the CdTe layer growth with an enhanced concentration of Cd in order to induce preferentially Cd–Te bonds instead of In–Te ones [65].

By *in-situ* online Raman spectroscopy (see Section 3.3) the formation of islands, interlayers, or other reaction products at the interface can be observed when recording the temporal evolution of Raman signals either during overlayer deposition or during the annealing of samples with previously abrupt interfaces [22]. Both techniques have been applied for group III elements, deposited onto Sb(111) substrates. It was observed that for deposition of Ga or In at 300 K GaSb or InSb was formed instantaneously; however, while the InSb growth mode is two dimensional, GaSb shows three-dimensional and disordered growth, and for an improved crystalline quality a deposition at elevated temperature (beyond 200 °C) was required [66]. Group III deposition at low temperature (170 K) resulted in abrupt interfaces. Subsequently, intermixing was induced by annealing, while monitoring the compound development by Raman spectroscopy, exploiting the ability to observe the buried interface region below the group III overlayer. The evaluation of the temperature dependence of the III–V formation allowed the determination of the relevant diffusion parameters for InSb, GaSb, and AlSb [66].

Dielectric Interface Vibrational Modes For a dielectric interface vibration with angular frequency ω_{diel} at the interface between the materials A and B Maxwell's equations impose the condition $\epsilon_A(\omega_{diel}) = -\epsilon_B(\omega_{diel})$ [67]. For example, for a ZnSe epilayer on a GaAs substrate, this means at the ZnSe/GaAs interface $\epsilon_{ZnSe}(\omega_{diel}) = -\epsilon_{GaAs}(\omega_{diel}) = -12.9$, and at the ZnSe/air interface $\epsilon_{ZnSe}(\omega_{diel}) = -1$. Because of its harmonic oscillator behavior with a resonance at the TO wavenumber and zero value at LO, the dielectric function of ZnSe fulfills both conditions in the wavenumber interval between TO and LO: the ZnSe/GaAs interface mode should appear about midway between TO and LO, while the ZnSe/air mode (surface mode) is expected very close to the LO, since $\epsilon(\omega_{LO}) = 0$. Thus, a qualitative assignment of interface modes can be made.

As an illustration of a dielectric interface mode, Figure 9 shows a Raman spectrum of a ZnSe layer on GaAs, with a thickness of $d_{ZnSe} = 920$ nm; in this enlarged view in the spectral region of the TO and LO phonon of ZnSe, a weak peak appears at about 218 cm^{-1} (i.e., between the TO and LO wavenumber). Furthermore, a detailed peak shape analysis of the ZnSe LO peak (not shown here) reveals a shoulder on its low-wavenumber side, corresponding to an additional weak Raman peak whose wavenumber is close to the LO. Interestingly, for different ZnSe layer thicknesses these two peaks appear with constant intensity, while, as expected, the TO and LO intensity are proportional to the ZnSe thickness. Therefore, the wavenumber and intensity behavior of the peaks are strong indications of dielectric interface vibrations.

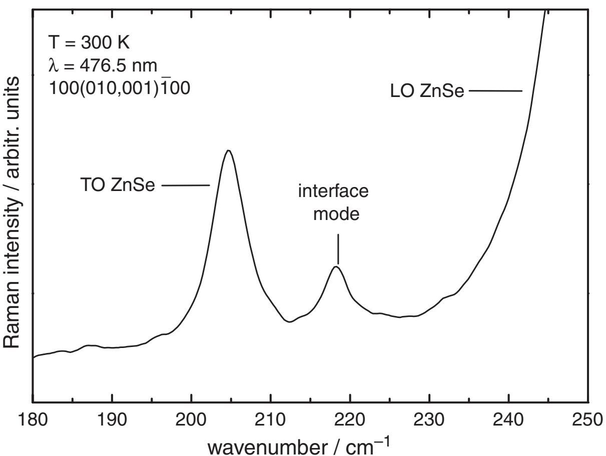


Figure 9. Enlarged view of a Raman spectrum of a ZnSe/GaAs heterostructure, showing an interface mode between the TO and LO phonon of ZnSe. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

However, for a quantitative interface mode assignment the theory of dielectric modes has to be considered in more detail: (i) their dispersion properties and (ii) the coupling between the modes from the upper interface (epilayer–air) and the lower one (epilayer–substrate). With regard to the dispersion effects, it has to be considered that the interface modes, as derived from Maxwell's theory, propagate parallel to the interface. This implies that they should not appear in the Raman spectrum for ideal backscattering geometry but are only observed due to finite wavevector components $\mathbf{q}_\parallel$ parallel to the interface because of the *quasi*-backscattering geometry and because of surface roughness. These effects lead to $\mathbf{q}_\parallel$ values in the range of some percent of the total q-transfer, that is, about $\mathbf{q}_\parallel \approx 40{,}000$ cm^{-1}, which is relevant not only for the mode wavenumbers but also for their exponential decay length in the ZnSe layer. As a result, theory predicts a coupling between the surface and interface mode for ZnSe layer thicknesses below about 0.3 μm, which induces noticeable wavenumber shifts.

The calculated mode wavenumbers are shown in Figure 10 and compared with experimental results for different ZnSe layer thicknesses between 0.06 and 0.62 μm. The quantitative agreement between theory and experiment confirms the assignment of the observed Raman peaks to the coupled dielectric modes, centered at the ZnSe surface and the ZnSe/GaAs interface.

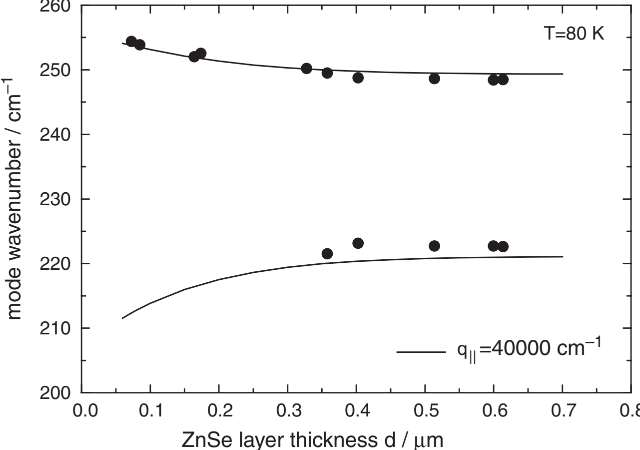


Figure 10. Comparison between the measured and calculated wavenumbers for the coupled dielectric interface modes of ZnSe/GaAs vs. ZnSe film thickness. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

5.1.2. Electronic Band Bending

At semiconductor surfaces and at crystallographically nonideal interfaces usually unsaturated electronic bonds are present which may tend to capture free electrons from the underlying bulk material and trap them at the surface or interface, respectively. This leads to an accumulation of negative charge, which results in a repulsive potential for the remaining free electrons. As a consequence, the region immediately behind the surface or interface will be depleted from free carriers (depletion layer), which may strongly influence its transport properties. Furthermore, for quantum dots, especially those produced by reactive ion etching, such near-surface regions may reduce the luminescence efficiency. The depletion layer depth increases with increasing trap density and is reduced with increasing doping level, since for high doping already a thin slab provides enough electrons to fill the trap levels. Realistic values for depletion layer depths vary from the nm to the μm range.

The depletion layer carries a net positive charge density which corresponds to the doping level, since in the layer the electric charges of the ionized dopant atoms are uncompensated due to the absence of free electrons. Therefore the interface and the depletion layer together form a dipole bilayer with an electric field inside, which decays linearly with distance from the interface. According to Poisson's equation the depletion layer is described by a quadratic increase of the electrostatic potential toward the interface. This phenomenon is referred to as electronic band bending. It is treated quantitatively, for example, by Rhoderick and Williams [68].

The development of electronic band bending at surfaces due to gas adsorption or the start of an overlayer growth is usually analyzed by photoemission spectroscopy [69]. However, this method, which requires UHV conditions, is limited to overlayer thicknesses of only some monolayers. Therefore it cannot be applied for the analysis of buried interfaces, which are relevant in heterostructures.

Raman spectroscopy, as an optical method, gives access to buried interfaces. Besides, it is sensitive to doping, as was shown in Section 4.4. Therefore Raman spectroscopy is a powerful tool to analyze changes in the free carrier behavior due to electronic traps at interfaces.

Two features in the Raman spectrum are sensitive to band bending variations, namely the coupled PLP modes and the LO phonon. For a doped material one would expect only the PLP peaks Ω_+ and Ω_-. However, because in a depletion layer no free carriers are present, the coupled PLP modes can exist only in the bulk region behind the depletion layer, while in the depletion layer only the pure LO phonon without any coupling to free carriers can exist. Therefore, in the presence of band bending the LO phonon modes as well as the PLP modes are expected in the Raman spectrum, provided the light penetration depth is sufficient to reach the doped region behind the depletion layer. The intensity ratio $I_{\mathrm{LO}}/I_{\mathrm{PLP}}$ is a measure of the depletion layer width and therefore of the density of trap states at the interface or surface [70]. These phenomena are described in detail and with a quantitative evaluation in ref. [3].

As an example of the abilities of Raman scattering for the analysis of electronic trap states at interfaces here scattering results from a ZnSe/GaAs(100) interface are discussed. The GaAs substrate was n-doped with a free carrier concentration $n = 2 \cdot 10^{18}$ cm^{-3}. The influence of the interface was deduced from the Raman spectrum of the GaAs PLP modes and LO phonon. The laser excitation wavelength was 476.5 nm, implying a negligible absorption in the ZnSe layer and a penetration depth of 70 nm in the GaAs substrate. The polarization configuration was off-diagonal: $100(010,001)\bar{1}00$. In order to trace the effects of the ZnSe relaxation, which starts at about 180 nm (see also Section 3.3), Raman spectra were taken for different ZnSe layer thicknesses up to 1.6 μm.

Some representative experimental results are plotted in Figure 11. The Raman spectrum of the GaAs substrate without ZnSe layer shows a broad structure, corresponding to the lower PLP mode Ω_-, at about 265 cm^{-1}, that is, near the

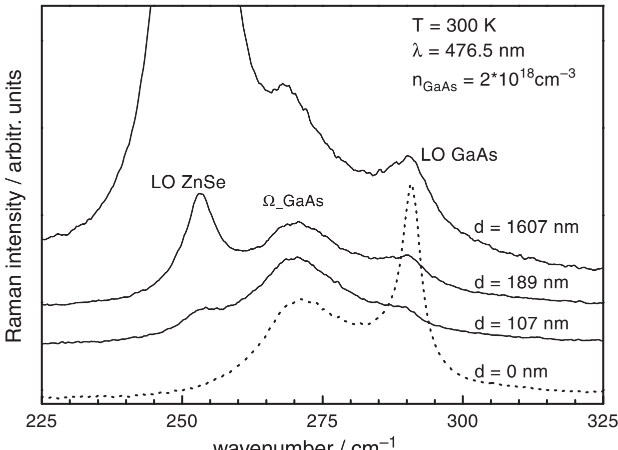


Figure 11. Raman spectra of ZnSe/n-GaAs heterostructures for different ZnSe layer thicknesses $d = 0 \cdots 1.6$ μm. Doping level $n_{\mathrm{GaAs}} = 2 \cdot 10^{18}$ cm^{-3}. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

TO-wavenumber, as is expected for this high doping level. In addition, the LO phonon is observed at about 292 cm^{-1}, which is typical for undoped GaAs. The simultaneous observation of both features is due to the presence of a depletion layer, which means electronic band bending. It is well known that for a clean GaAs(100) surface with the doping level of our sample the band bending amounts to 0.7 eV, which leads to a depletion layer width of 22 nm. Using this value of the layer width the measured LO/Ω^- intensity ratio can be interpreted in terms of the ratio of the Raman scattering efficiencies for LO and Ω_- without the need to know their absolute values.

After deposition of ZnSe, the Raman spectra in Figure 11 show besides the appearance of the ZnSe LO phonon also a change of the LO/Ω_- intensity ratio of GaAs. First, the LO intensity shows a strong decrease compared to the Ω_- intensity, as plotted for a layer thickness $d = 107$ nm. For larger ZnSe thicknesses the ZnSe LO phonon intensity, of course, increases very strongly, but also the GaAs LO/Ω_- intensity ratio increases again. This behavior reveals a strong dynamics in the depletion layer width, which means a thickness dependence of the electronic trap density at the ZnSe/GaAs interface.

A quantitative evaluation of the Raman spectra in terms of electronic band bending is shown in Figure 12. For this evaluation the ratio of the scattering efficiencies of LO and Ω_- was used, which was obtained from the uncovered surface. Besides the results derived from the spectra of Figure 11 the plot also shows band bending values obtained from spectra with another laser line, taken at 80 K. The results are consistent within ± 50 meV.

The extremely strong reduction of the band bending for ZnSe layer thicknesses below 150 nm proves the very high quality of the ZnSe/GaAs interface. The pseudomorphic ZnSe layer saturates about all the electronic traps at the surface and near-flatband condition is achieved. This very sensitive test underscores the excellent quality of the interface for this MOVPE-grown heterostructure. Its quality is comparable with MBE-grown samples which were analyzed by the same technique [71, 72].

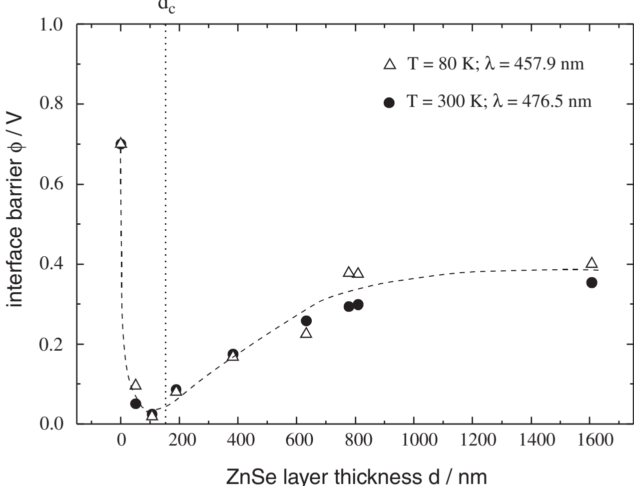


Figure 12. Band bending at the ZnSe/n-GaAs interface for different ZnSe layer thicknesses d, derived from the Raman results in Figure 11. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

The onset of the renewed increase of the band bending is explained in terms of the strain relaxation beyond the critical layer thickness d_c, which was obtained from ZnSe LO phonon data from the same samples. The gradual increase reflects how during the process of strain relaxation more and more relaxation-induced dislocations reach the heterointerface. The saturation thickness of about 1 μm agrees well with the value obtained from the strain-induced shift of the ZnSe LO phonon (see Section 4.3). The saturation band bending value of 0.4 eV corresponds to the defect density at the interface which is expected from the lattice mismatch between ZnSe and GaAs.

5.2. (Multi)-Quantum Wells and Superlattices

For II–VI materials like ZnSe and CdTe the condition for quantum wells to be thin enough to exhibit quantization effects in the electron wavefunctions implies thicknesses below about 20 nm. If the thickness is further decreased, these thin layers may also exhibit specific vibrational features. The optical phonons, whose wavenumbers are very characteristic for each material, usually cannot penetrate into the neighboring materials because of the wavenumber mismatch. This gives rise to the confinement of the optical phonons in the quantum well layer, which is reflected in specific confined optical phonon (COP) peaks in the Raman spectrum. They are discussed in Section 5.2.1. For periodic thin-layer sequences of two alternating materials in multiquantum wells (MQWs) and in superlattices (SLs), consisting of even thinner layers, these COPs usually occur in both materials.

For acoustical phonons the situation is completely different. Due to the nearly equal sound velocity in most solids, these waves may travel through the total stack of layers in a MQW or SL. However, the periodicity which is superimposed due to the layer stacking leads to a modification of the translational invariance, and as a consequence for Raman scattering also to a modification of the possible values of the momentum transfer in the scattering process. As a result, in contrast to bulk material spectra, in the Raman spectrum of MQW and SL acoustical phonons may be observed. These modes are referred to as folded acoustical phonons (FAPs) and are discussed in Section 5.2.2.

Confined optical and folded acoustical phonons can be utilized for the analysis of, for example, interface sharpness and layer thicknesses. Several excellent extensive reviews exist of the basic aspects of these features and the wide spectrum of experimental results, especially from III–V systems [67, 73–78]. Therefore here only a very short overview of the principles of confined optical and folded acoustical phonons is presented, together with some examples from different II–VI material systems. Of course, the interfaces in MQW and SL may also give rise to dielectric and chemical vibrational modes. These will not be discussed in this section, because they have been treated extensively within the framework of interface-induced phenomena in Section 5.1.1.

5.2.1. Confined Optical Phonons

The amplitude envelope functions of confined optical phonon modes in a MQW or SL can be visualized as the eigenmodes of a vibrating string (basic mode and overtones) in the slabs of one material, essentially unable to penetrate into the adjacent layers of the other material, since these layers have different vibration eigenfrequencies. Therefore, for a MQW or SL which consists of a sequence of alternating layers of materials A and B the allowed wavelengths λ of these confined modes can be described in a good approximation by the condition $d_A = m \cdot \lambda/2$ for those vibrational modes which are confined in material A, and correspondingly for the modes in material B, where m is an integer number, and d_A and d_B are the slab thicknesses of material A and B, respectively [79].

For a slab consisting of n monolayers with a thickness a_0 per monolayer (half of the [001] lattice constant in a zincblende structure), this condition yields for the wavevector q of the confined phonon [67]

$$q = \frac{\pi m}{(n+1)a_0} \qquad \text{with } 1 \leq m \leq n \tag{27}$$

This means that the COP eigenmodes belong to a set of equidistant q-values in the Brillouin zone. The factor $(n+1)$ (instead of the monolayer number n) takes into account the very small but finite decay length of the vibrational amplitude beyond the interfaces. Due to the optical phonon dispersion curve the modes with different index m correspond to different eigenfrequencies, through which they may be distinguished in the Raman spectrum. The negative dispersion of the optical phonons implies a decreasing eigenfrequency with increasing mode index m. For materials whose bulk phonon dispersion curves already were known from neutron scattering experiments, Raman-derived data show a quantitative agreement with the neutron data [67]. This opens up the path to derive phonon dispersion curves from COP wavenumbers in Raman experiments on superlattices (e.g., for materials of which no bulk samples for neutron scattering are available). Beside the possibility of an optical determination of the phonon dispersion curve throughout the whole Brillouin zone, COPs have two additional relevant consequences:

(i) They are a very sensitive nondestructive probe for the lateral homogeneity and interface sharpness of extremely thin layers. Thickness variations and intermixing induce a smearing of q-values and consequently an inhomogeneous broadening of the COP-peaks, especially of those with higher-order number m.
(ii) The confinement-induced wavenumber shift has to be taken into account when analyzing strain or composition-induced peak shifts.

For these reasons the wavenumber shifts of the various confined LO phonon modes of ZnSe ($m = 1, \ldots, 6$), calculated from Eq. (27) and based on the bulk ZnSe LO dispersion, are plotted in Figure 13 vs. the ZnSe layer thickness.

Obviously, the first-order confined mode shows a detectable wavenumber shift only for a layer thickness $d <$ 5 nm, which is, interestingly, far below the onset of electronic confinement near $d = 20$ nm. However, for $d = 2$ nm, the shift of the first-order mode is already 1 cm^{-1}, which is comparable to typical strain-induced shift values.

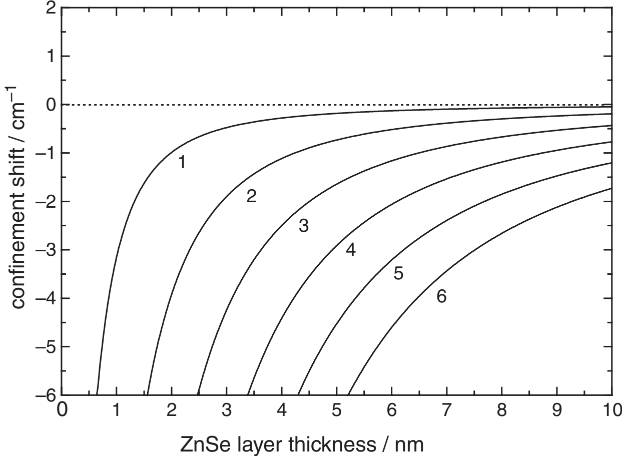


Figure 13. Confinement-induced wavenumber shift of the ZnSe LO phonon, plotted as a function of the ZnSe layer thickness; index: confinement order. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

Finally, for a correct mode-number assignment of the COP peaks, it should be considered that the odd-order modes ($m = 1, 3$, etc.) appear in the depolarized scattering configuration $z(x, y)\bar{z}$ with x, y, z the main crystal axes, whereas the even-order modes occur in the polarized $z(x, x)\bar{z}$ configuration.

As a first example of confined optical phonons in II–VI materials ZnSe/BeTe superlattices are considered, which exhibit negligible lattice mismatch. Furthermore both materials are almost lattice matched to a GaAs substrate with $\Delta a/a = +0.26\%$ and -0.48%, respectively and can therefore be grown with very high quality. Balance of the residual strain is obtained for a thickness ratio $d_{\text{ZnSe}}/d_{\text{BeTe}} \approx 2$ of the individual layers. In this case the strain-induced shift of the phonon wavenumber in ZnSe is as small as $+1.15$ cm^{-1}. In small period superlattices ($d_{SL} = 40$ Å) consisting of 5 ML BeTe and 9 ML ZnSe layers confined phonons in ZnSe up to the sixth order can be observed [80, 81], which proves the excellent sharpness of the interfaces. More detailed analysis of the influence of interface quality on the confined phonon spectrum is given in ref. [82].

In cubic CdS/ZnS superlattices, the optical phonon structures are dominated by the strain in the individual layers because of the extremely high relative lattice mismatch of 7%. Probably due to this heavy strain, which is likely to induce lattice imperfections, no higher-order confined modes are observed. The essentially strain-induced shift of the ZnS LO phonon amounts to -21.9 cm^{-1}, which agrees very well with the wavenumber shift, expected from the compliance and phonon deformation potential values [83].

For CdSe/CdS superlattices with wurtzite structure on GaAs(111), confined phonons were observed in far-infrared reflectance. Under oblique incidence (angle of incidence = 45°) confined longitudinal modes occured for light polarization parallel to the plane of incidence, while confined transverse modes appeared for polarization perpendicular to this plane [84]. The mode wavenumbers are distinctly shifted due

to the strain in the individual layers, which originates from the lattice mismatch of 3.9%.

CdS/ZnSe superlattices offer the opportunity to stabilize CdS, which naturally exhibits a hexagonal structure, in the cubic zincblende modification, imposed by the embedding ZnSe layers [85]. The layers in the resulting stack are heavily strained due to the very high relative lattice mismatch of 2.64%, which limits the maximum thickness of the individual layers to some monolayers. The strain leads to a strong shift of the ZnSe confined phonon modes. The observed COP wavenumbers of CdS as well as ZnSe are described consistently by a symmetrized strain (i.e. the SL adopts a strain state, which corresponds approximately to that of a freestanding SL). This implies for the ZnSe layers a redshift of 4 cm^{-1} due to tensile strain, while the COP peaks in the CdS layers are blueshifted by 6 cm^{-1} due to compressive strain [53].

5.2.2. Folded Acoustical Phonons

Generally, in light scattering experiments from bulk samples and simple heterostructures acoustical-phonon peaks exhibit a very low wavenumber shift from the laser line (≈4 cm^{-1} for the LA phonon) because of the small momentum transfer in light scattering. Therefore they are the domain of Brillouin scattering. MQWs and SLs, however, exhibit a periodicity length D along the growth direction of the stack, which is strongly enhanced with respect to the bulk periodicity: $D_{MQW} = n_A \cdot a_{0,A} + n_B \cdot a_{0,B}$, which may be one or two orders of magnitude beyond a_0. The enhanced periodicity in the growth direction implies a reduced periodicity in reciprocal space, which means a reduction of the Brillouin zone length to π/D_{SL} instead of its bulk value π/a_0.

For the acoustical-phonon dispersion branch, the reduction of the Brillouin zone length may be considered as a folding [67]. The slope of the dispersion branch (i.e., the SL sound velocity) is obtained from the sound velocities in the materials A and B by a weighted reciprocal averaging [see Eq. (28)]. In the inset of Figure 14 the folding is shown schematically for two superlattices [full line: SL (a), dashed line: SL (b)] with different periodicities $D_a > D_b$. The folded acoustical phonons may appear in the Raman spectrum as peaks at those wavenumbers for which the q-vector of the folded phonon branches equals the momentum transfer vector q_s, which was introduced in Section 2.2 and is plotted as here as a vertical dash-dotted line. The intersections of this line with the phonon dispersion curve yield the wavenumbers of pairs of FAP peaks. The corresponding FAP spectra are shown in Figure 14 for two $ZnSe/ZnSe_{0.9}S_{0.1}$ SLs of 120 periods with periodicities $D_a = 17$ nm and $D_b = 9.8$ nm. Both show a pair of well resolved peaks as proof of a very homogeneous periodicity. Moreover, the wavenumber shift between both pairs indicates the sensitivity of the peak position for changes of the periodicity.

The exact position of the FAP peaks in the Raman spectrum not only depends on the sample periodicity but also on the laser wavelength, which determines the q-transfer. For a quantitative evaluation, the FAP dispersion curve has to be considered in detail. Its behavior can be described according the Rytov continuum model [86], which yields for longitudinal acoustical modes propagating in the growth direction

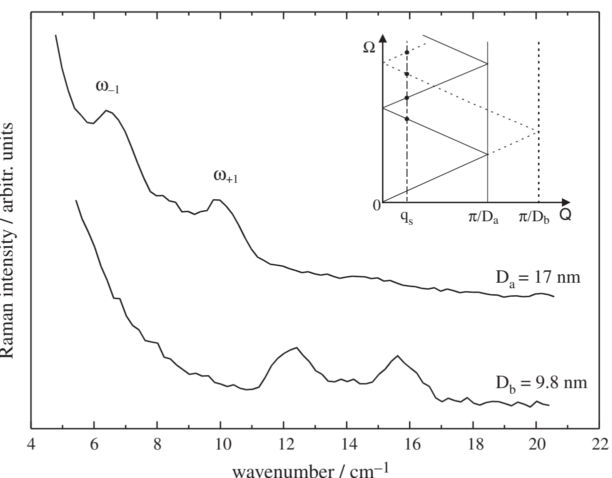


Figure 14. Low-wavenumber Raman spectra for $ZnSe/ZnSe_{0.9}S_{0.1}$ superlattices with different periodicity lengths. The inset shows schematically their wavenumber dependence on the periodicity. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

of a (100) superlattice, of compounds A and B, with thicknesses d_A, d_B, sound velocities v_A, v_B, and mass densities ρ_A, ρ_B

$$\cos(Q(d_A + d_B)) = \cos\left[\omega\left(\frac{d_A}{v_A} + \frac{d_B}{v_B}\right)\right] - \left(\frac{\eta^2}{2}\right) \cdot \sin\left(\frac{\omega d_A}{v_A}\right) \cdot \sin\left(\frac{\omega d_B}{v_B}\right) \tag{28}$$

In this expression, η is the normalized difference of the acoustical impedance $v \cdot \rho$ of the two constituting bulk materials

$$\eta = \frac{v_A\rho_A - v_B\rho_B}{\sqrt{v_A\rho_A v_B\rho_B}} \tag{29}$$

The first term of Eq. (28) describes the folding of the "average" dispersion curve due to the periodicity of the superlattice and therefore reflects the geometry of the structure. The second term describes the acoustical modulation due to the different acoustical impedance and can lead to a wavenumber splitting of the modes at the Brillouin zone center and edge. The splitting is usually very small because the acoustical impedances of the common semiconductor materials are comparable and therefore $\eta^2/2 \sim 10^{-2}$.

The calculated periodicity dependence of the FAP wavenumbers is shown in Figure 15 for $ZnSe/ZnSe_{0.9}S_{0.1}$ and for ZnSe/ZnS. The $m = 0$ line denotes the bulk phonon, whose wavenumber is of course independent of the superlattice period, and its value is only 2 cm^{-1}. The $m = \pm 1$ lines represent the first pair of FAP peaks. Their increasing slope with decreasing periodicity indicates the increasing sensitivity for thickness fluctuations in very short periodicity SLs, which goes down to fluctuations in the monolayer range.

It should be noted that the appearance of the lowest order folded acoustical phonons just proves a regular in-depth periodicity over the lateral range which is covered by the laser focus, without allowing the assessment of the interface sharpness. For the latter purpose the intensities of the

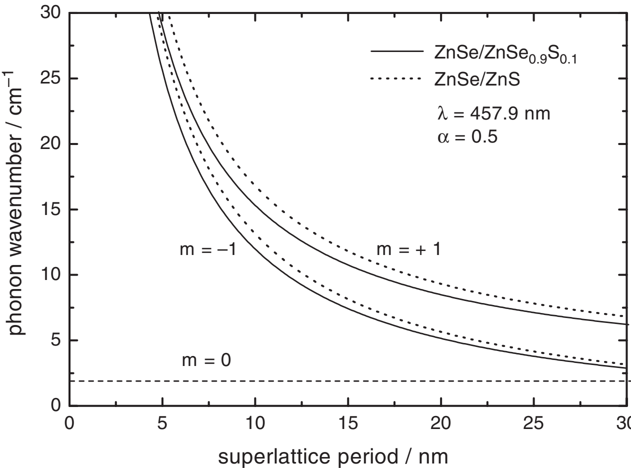


Figure 15. Wavenumber of the folded acoustical phonons as a function of the superlattice periodicity D. Adapted with permission from [31], J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, 1996. © 1996, Augustinus Publ. Co.

higher order FAP pairs ($m = \pm 2$; ± 3, etc.) and especially the confined optical phonons are a sensitive criterion.

For CdS/ZnSe superlattices, FAP modes up to the third order were observed [53] whose wavenumbers agree very well with those calculated with the Rytov model. For a SL with individual layer thicknesses of 1.7 and 1.4 nm, respectively, the intensity ratio of the third-order modes with respect to the first ones amounts to 0.38, which agrees rather well with the theoretically predicted value of 0.33 for a symmetrical SL. In this way, a very good lateral and in-depth homogeneity of the SL is confirmed, as well as a clear interface abruptness. Similarly, for CdS/CdSe strained SLs FAP modes were observed whose wavenumbers are described very well by the Rytov model [83].

Besides cubic SL hexagonal systems were also applied for FAP studies. The observation of FAP modes was reported for hexagonal CdSe/CdS superlattices, which are heavily strained because of the extremely high lattice mismatch of 7% between CdS and CdSe. However, for the evaluation of FAP mode wavenumbers, the strain is of no significant relevance, because it is symmetrical: tensile in one material and compressive in the other. Therefore it has hardly any net effect for the wavenumber of the acoustical phonons, which are eigenmodes of the total stack [85]. However, the strain-induced piezoelectric effect in the individual CdS and CdSe layers enables the observation of FAPs by measuring infrared reflectance spectra at oblique incidence [84].

A quite original application of FAP modes was reported for BeTe/ZnSe superlattices [82]. Here very pronounced FAP peaks were observed at an average wavenumber of 35.8 cm^{-1} and a splitting of 3.1 cm^{-1} for superlattices, whose periodicity consisted of 1.3 nm BeTe and 2.6 nm ZnSe. This result enabled the determination of the hitherto unknown sound velocity of BeTe, since all other input parameters of the SL sound velocity (i.e., the layer thicknesses and the sound velocity in ZnSe) were known. The resulting value of $v_s^{\text{BeTe}} = 4180$ m/s $\pm$ 5%, which is very similar to the v_s value for ZnSe, corresponds to an elastic constant $c_{11} = 8.9 \pm 0.9 \cdot 10^{10}$ N/m^2 and agrees well with first-principle density functional theory calculations, based on the local density approximation [87].

6. NANOSTRUCTURES: QUANTUM WIRES AND QUANTUM DOTS

This section covers the main item of this chapter, that is, the application of Raman spectroscopy to nanostructures such as quantum wires and dots. In these studies on nanostructures the various abilities of Raman spectroscopy are employed, which have been discussed above for bulk materials and for layered structures and interfaces, such as confinement effects and the determination of strain and composition. The examples are classified according to the production method of the nanostructures.

First, Raman results on chemically produced dots in glass matrices will be presented in Section 6.1. For these systems, besides the chemical composition, also the average dot size and the size distribution are main issues. Dot sizes are determined from low-wavenumber Raman spectroscopy, while the chemical composition is derived from the optical phonon wavenumbers. Furthermore, the electron–phonon coupling and the extension of the electronic wavefunction beyond the dot edges are studied.

Section 6.2 deals with wires and dots which were generated lithographically by patterning and etching of quantum well layers. Here strain distributions and the size dependence of strain relaxation are the main subject of the investigations.

As an alternative to the lithographic production of dots, the spontaneous dot formation by three-dimensional growth of strongly lattice-mismatched systems has gained much attention. These self-organized dots are discussed in Section 6.3. Here Raman spectroscopy can provide information, for example, on the onset of dot formation, their chemical composition, stability, and interface sharpness.

6.1. Chemically Produced Dots

6.1.1. Phonon Confinement in Quantum Dots

Quantum dots (QDs) fabricated by precipitation methods are usually embedded in an amorphous matrix, like glass or organic polymers, to stabilize them from aggregation. Theoretical investigations predict that among the optical phonons TO, LO an infinite series of surface optical modes (SO) exist in the nanocrystallite. Because of the eigenfrequency mismatch between the dots and the amorphous matrix the lattice vibrations cannot penetrate into the matrix: the phonon wavefunctions are confined within the dots. In the Raman scattering process, this spatial confinement results in a relaxation of the momentum conservation, which increases with decreasing crystallite radius. This leads to an enhancement of transition probability between the vibrational states and therefore to a shortening of the lifetime of the excited state. Finally, this shortening of the lifetime results in a size-dependent increase of the homogeneous linewidth. The relaxation can be observed in the Raman spectra as a redshift of the LO wavenumber, caused by the negative dispersion of the LO phonons, as well as an asymmetrical broadening of the peakwidth [88–92]. Recently, the appearance of SO phonons as well as the size-dependent

broadening of both LO and SO was accounted for the observed broadening. The broadening should be due to the boundary relaxation at the dot–matrix interface, which is caused by the reduction of the mean free path $L = kr$ in QDs [93]. The lineshape of the confined LO phonon modes can be calculated with the modified model for bulk phonons. In this model the localization of the phonon wavefunctions within crystallites with spherical shapes (or disc- or tube-shaped) and a dispersion function due to the dispersion of the dot sizes in the sample are considered [90, 94]. The calculations result in a dispersion of the phonon momentum, which depends on the crystallite radius r [89, 92, 94]:

$$I(\omega) \sim \int \frac{|C(0, \mathbf{q})|^2}{(\omega - \omega(\mathbf{q}))^2 + \left(\frac{\Gamma_0}{2}\right)^2} d\mathbf{q} \tag{30}$$

For the Fourier coefficient of the confinement function usually a Gaussian function is taken:

$$|C(q)|^2 = \exp\{-q^2 d^2/16\pi^2\} \tag{31}$$

In a simple model the dependence of the redshift of the LO phonon on the crystallite radius r can be deduced from the phonon dispersion relation $\omega(q)$ of the bulk mode, when q is replaced by [38, 95–97]

$$q_n = n\pi/(r + a/2) \tag{32}$$

$\omega(q)$ is given by a fit of the phonon dispersion in form of

$$\omega^2(\mathbf{q}) = A + \{A^2 - B[1 - \cos(qa)]\}^{1/2} \tag{33}$$

Here, A and B are fit parameters. The discussed model is valid only for off-resonance Raman excitation. In resonance Raman spectroscopy there are other factors that can also broaden the linewidth of the LO phonon peak and are not considered in this model. Many groups have already observed the asymmetrical broadening in Raman experiments on II–VI semiconductor dots (e.g., [94, 98]), but the observed amount of the redshift lies below the expected one. This can be caused by different effects. The most important one is that the investigated dots are not small enough: The dot radius should be smaller than the exciton Bohr radius of the respective semiconductor material to obtain measurable shifts (strong confinement regime). Furthermore, matrices, especially those which form bonds with the surface atoms of the crystallite, can induce strain on the crystallite surface. The strain can cause a blueshift of the phonon wavenumbers, which may obscure the confinement redshift [97]. A further explanation of the missing shift supported by some groups [33, 92, 99] is the existence of defects, which are formed during the preparation and which cause the LO wavenumber to shift to higher values. For CdSe clusters the exciton Bohr radius a_0 is 5.6 nm [100, 101]. Thus, for cluster radii below this value the confinement shift should be observable.

This effect was experimentally demonstrated for LO phonons of CdSe clusters in GeO_2 glass matrix [89]. For different samples containing nanocrystals with average diameters between 9.6 and 3.8 nm, the redshift of the LO phonon signal as well as the broadening of the linewidth due to the phonon confinement was clearly observed. The resulting dependence of the peak wavenumbers on the nanocrystal diameter showed a good agreement with the results of calculations from Eqs. (30) and (31) [89].

The size dependence of the LO lineshape can also be proven experimentally by wavelength-dependent Raman investigations on a quantum dot sample with large size inhomogeneities. When changing the excitation wavelength the resonating cluster sizes change, too. Thus, the whole size distribution of a quantum dot sample can be probed and size-dependent Raman spectra can be obtained from one and the same sample. For higher excitation wavelengths the LO phonon line shows the characteristics of confinement: the asymmetrical broadening as well as the appearance of the surface optical mode [102].

6.1.2. Size Determination of Nanocrystallites by Low-Wavenumber Raman Spectroscopy

Because of the strong size dependence of the optoelectronic properties of nanocrystalline samples, the determination of the crystallite sizes is a crucial point in the characterization of nanocrystallites. Besides methods like high resolution transmission electron microscopy and X-ray diffractometry, low-wavenumber Raman spectroscopy was established as a fast method to determine the average dot size. The modes in the low-wavenumber range of quantum dot samples can be attributed to the torsional (depolarization ratio $\rho = 3/4$, $\omega^T_{l,n}$) and spheroidal acoustical modes (depolarization ratio $\rho = 0$, $\omega^S_{l,n}$). The amplitude of the torsional mode is largest at the dot surface. Therefore, size-dependent changes of the interface conditions are expected to influence the intensity of this mode. Thus, the torsional modes should show the highest intensity for the particles with the smallest radius and therefore weak particle matrix interaction [103, 104]. In contrast, the amplitude of the spheroidal mode is displaced within the particle and should have a weaker dependence on the matrix. Effects, arising from the matrix, strain, or shape, mainly have an influence on the eigenfrequencies of the higher wavenumber modes, whereas the influence on the lowest eigenfrequency mode is normally rather small [104, 105]. Theoretical considerations have shown that the spheroidal modes (even l) can be Raman active, while the torsional modes (odd l) are Raman inactive [106]. A further criterion for the identification of the low-wavenumber modes is that spheroidal modes with $l = 0$ are observed only in the polarized geometry, whereas the quadrupolar modes with $l = 2$ are observed in polarized as well as in depolarized geometry. The spheroidal as well as the surface modes show resonance enhancement when the laser energy matches the region of the lowest absorption band [107]. With the knowledge of the elastic constants of the matrix and the dot shape, it is possible to determine the dot radius from the wavenumber of the spheroidal acoustical mode [108–110]: The quantum dots are considered as elastic bodies, which are vibrating with eigenmodes, whose frequencies are proportional to their diameter. The angular eigenfrequencies of the torsional and the spheroidal modes are then given by [103–105]

$$\omega^T_{l,n} = \eta^T_{l,n} v_l / d \tag{34}$$

$$\omega^S_{l,n} = \eta^S_{l,n} v_l / d \tag{35}$$

v_l is the longitudinal sound velocity [m/s]; the factor η depends on the crystallite shape and the matrix. For spherical cluster shapes the following equation for the dependence of the lowest frequency eigenmode ν_{LFR} on the cluster diameter d is obtained [103–105, 107, 108]:

$$d = \xi \frac{v_i}{c\nu_{LFR}} \tag{36}$$

c is the speed of light. The numerical factor ξ depends on the ratio between longitudinal and transversal sound velocity. For CdS and CdSe it is 0.9. The wavenumber positions of the low-wavenumber modes of quantum dots show a dependence on the excitation wavelength. This is due to the fact that each excitation wavelength induces resonance for a specific cluster size [103]. From the shape of the low-wavenumber band this size distribution can be calculated by fitting the peak with a distribution function [Gaussian (symmetric) or Lifshitz–Slyozov (asymmetric)] [108]. An example of low-wavenumber Raman spectra of different semiconductor-doped glasses (Schott glasses, CdS_xSe_{1-x} quantum dots) is shown in Figure 16 [111]. On both Stokes and anti-Stokes sides, the bands appear typically around 10–30 cm^{-1}. The derived diameter for example of OG550 glass was 7.0 ± 0.4 nm.

6.1.3. Composition Determination from Optical Phonons

For the determination of the composition of nanocrystallites, which are embedded in a glass matrix material and often occupy a volume fraction as low as 1%, X-ray diffraction seems to be extremely difficult. Raman spectroscopy offers the advantage of the resonant enhancement of the phonon signal intensity when exciting at an electronic transition of the semiconductor material. In these resonance Raman spectra the mixed-compound phonon peaks appear,

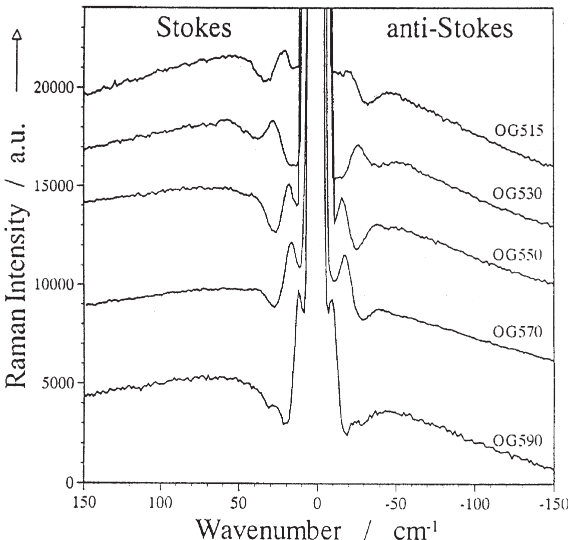


Figure 16. Nonresonant low-wavenumber Raman spectra of different semiconductor doped glasses recorded with laser excitation at 647 nm. Reprinted with permission from [111], T. Bischof et al., *J. Raman Spectrosc.* 27, 297 (1996). © 1996, Wiley Interscience.

which were discussed for bulk material in Section 4.2. Now the nanocrystallite composition is determined from a direct comparison of the nanocrystallite mode frequencies with the known composition dependence for the bulk material. A consideration of possible size effects reveals that these are expected for nanocrystallite sizes below 4 nm. Many relevant systems consist of larger clusters, for example, the CdSSe microcrystals in Schott glass long-pass filters, whose cluster sizes are generally beyond 6 nm. Here resonant Raman spectroscopy was applied successfully for composition determination [111].

Raman spectroscopy from optical phonon modes was also applied for the study of the effects of thermal sintering up to 500 °C on CdTe/CdS "core-shell" clusters (i.e., CdTe cores, embedded in CdS shells) in colloids and films [112, 113]. Comparison of the peaks in the cluster spectra with the well-known LO phonons of CdTe (170 cm^{-1}) and CdS (305 cm^{-1}) revealed that sintering between 100 and 200 °C initiates the CdTe cluster growth. Sintering temperatures of 300–400 °C lead to a release of tributylphosphine-capping ligands and to the formation of bare CdS and CdTe nanocrystals. Finally, above 400 °C the CdTe part of the nanostructures sublimates, leaving behind nearly pure CdS nanocrystallites.

6.1.4. Pressure Effects

In quantum dots produced by sol/gel processes strain can be also caused by the surrounding matrix. The synthesis of these samples implies sintering processes or the implantation of the dots into the molten matrix material. The thermal expansion constants of this matrix are usually different from the constants of the semiconductor. As a consequence, the matrix can induce an effective pressure on the dot surface upon cooling the sample.

For CdS and CdSe quantum dots pressure-dependent Raman experiments resulted in a shift of the LO phonon wavenumber and in a decrease of the LO phonon intensity with increasing pressure. Finally, the LO phonon peak disappears. This disappearance is caused by the transition of the lattice structure from the direct bandgap wurtzite structure into the indirect bandgap rocksalt structure, in which first-order Raman scattering is symmetry forbidden [33, 42, 43, 99, 114–117].

In pressure-dependent Raman spectra of CdS nanocrystals with an average diameter of 8 nm deposited in glass matrix, a linear blueshift with increasing pressure at a rate of 0.47 ± 0.01 and 0.98 ± 0.02 cm^{-1}/kbar, respectively, was observed for the wavenumbers of the LO fundamental and its first overtone [33]. The pressure coefficients are similar to those of the bulk CdS, which is 0.5 ± 0.03 cm^{-1}/kbar for the LO fundamental [33, 37, 117–119]. The peak intensities initially showed an increase with increasing pressure due to resonance enhancement. Beyond 12 kbar the intensities decreased strongly until at 60 kbar the intensity has decreased to 8% of the intensity at 1 bar. In contrast to measurements on bulk CdS, where a complete phase transition occurs above 27 kbar [33, 100, 116–121], the LO intensity did not vanish completely [33, 116]. The initial increase due to resonance effects is also much smaller than in bulk CdS. This reflects the fact that the energy levels in the CdS-glass composites are distributed over a large energy region, caused by the size inhomogeneity of the sample.

Pressure-dependent Raman experiments on CdS colloids with an average diameter of 6 nm show the same pressure coefficients as the glass composites. The pressure for the highest resonance intensity in this sample is about 20 kbar. With further increasing pressure the Raman intensity decreases strongly but shows no sharp cutoff like the bulk [99, 119]. This cutoff is a sign for a lattice reorganization to the rocksalt structure [33]. In some cases the phase transition is incomplete and a weak LO signal due to the mixed rocksalt/wurtzite structure can be observed even at high pressures [99, 118]. A further difference in the observations in the bulk is the hysteresis in the pressure dependence of the LO frequency shift, which is about 5 kbar in QDs compared to 15 kbar in the bulk [35]. Other high pressure studies with microcrystalline CdS show that there is a complete phase transition at pressures below 40 kbar. This difference in the phase transition behavior can be caused by defects on the crystal surface. The defects induce a large strain field in the crystal, which shifts the local phase transition to higher pressures. A second factor, which may cause a shift, is the effect of the surface. The surface tension and reconstruction energy result in a contraction of the lattice constants in small particles [122].

6.1.5. Electron–Phonon Coupling

In the brief treatment of Raman scattering mechanisms in Section 2.3 it was already discussed that in II–VI compounds longitudinal-optical phonons are accompanied by a strong electric field that induces a strong Fröhlich coupling of the phonons to the electric field of the optically excited excitons, which constitute the intermediate states of the Raman scattering process. Therefore, for these materials the Fröhlich mechanism dominates the electron–phonon coupling in resonant Raman scattering.

In order to understand the impact of nanostructures for the electron–phonon coupling efficiency, the Fröhlich interaction mechanism is considered here in more detail. The origin of the electric field of the exciton is the difference in spatial extent of the electron and the hole wavefunction. In bulk materials the electron is much less localized than the hole. From this incomplete overlap and because of the difference in the effective masses, a carrier distribution results in a positive net charge in the central region and a negative net charge on the edges. The Fröhlich mechanism describes the coupling of this electronic charge inhomogeneity with the polar lattice [38, 123–126]. The magnitude of the electron–phonon coupling depends on the shape of the phononic and excitonic carrier distributions and on the polarizability of the excitonic state. Its maximum value occurs for those phonons whose wavelengths match the exciton dimension [38, 123].

Experimentally, the electron–phonon coupling (EPC) efficiency can be derived from resonant Raman scattering and its excitation profile, i.e., the dependence of the scattering efficiency on the incident and scattered photon energy. In this way, incoming and outgoing resonance profiles are obtained. For this purpose, not only first-order Raman scattering but also scattering from phonon overtones is employed. The electron–phonon coupling efficiency is derived from the relative scattering intensities of these modes [125, 126]. For example, in bulk CdS, up to eight overtones are observed in resonant Raman spectra because of the very efficient coupling between excitons and strongly polar lattice vibrations [127, 128].

In small clusters the situation is different in various aspects: In contrast to the bulk the conservation of momentum in quantum dots is relaxed and first-order Fröhlich-induced Raman scattering becomes allowed. The bulk model for calculating the EPC cannot be applied for nanocrystals since it does not imply boundary conditions of the carrier and vibrational confinement. To calculate the size dependence of the EPC, a new model has to be considered, which includes the confinement and also surface effects, like the trapping of the hole at localized surface states. Many authors performed experimental and theoretical investigations of the size-dependent coupling to the optical phonons (LO; TO; SO) of nanocrystallites [95, 96, 123, 125, 129–132]. Different approaches were applied for the calculation of the exciton wavefunction and for boundary conditions for the coupling to the LO and SO phonons. Despite the numerous investigations, up to now no consistent picture was obtained [38, 94, 124, 130, 133–135].

Generally, there are two compensating size effects on the EPC: The first effect is the confinement-induced increasing spatial overlap between the electron and hole wavefunction of the exciton, which is expected to reduce the EPC because of the reduced polarizability of the exciton by the lattice vibration [38, 39, 95, 106, 123, 136]: For sufficiently small clusters, the electron and hole wavefunctions coincide. For these clusters the excitonic polarizability vanishes and so should the EPC [123, 129]. In addition, the effect of the crystallite surface has to be taken into account: The hole can be localized in traps like defects or vacancies on the crystallite surface, for example, in uncoordinated Se-*p*-orbitals. By this trapping the wavefunction overlap is reduced and therefore the polar character of the excitonic charge distribution increases. This charge distribution causes a lattice distorsion which produces an enhancement of the Raman cross section of the LO mode and its overtones [38, 123, 128, 131].

Taking into account both effects one can conclude that with decreasing cluster size there must be first an increase of the EPC due to the trapping-induced decreased overlap between electron and hole (weak confinement regime), followed by a decrease of the EPC due to the increasing overlap between electron and hole with further decreasing crystallite radius (intermediate confinement regime) [127]. This decrease can be observed in the vibrational spectra in which a decreasing number of overtones as well as a decrease of the ratio of the integrated LO-overtone and fundamental intensities are observed [126]. As an example, resonance Raman spectra of CdS bulk and three different CdS nanocrystallite samples with average sizes of 10, 6, and 4 nm showed a clear decrease of the ratio of the integrated intensities of overtone to fundamental with decreasing crystallite size [125]. Further Raman experiments with even smaller CdS crystallites confirmed this decrease [38]. The observed excitation wavelength dependence of the 2LO/1LO ratio, which is due to the inhomogeneous size distribution in these samples, also agrees with this size dependence.

In the weak confinement regime, a decrease of the EPC is also found for higher temperatures: with increasing temperature the trapped hole can be thermally activated across the potential barrier on the surface and therefore becomes delocalized. The result is an increased overlap of electron and hole and thus a strong decrease of the EPC coupling with increasing temperature [127].

Recent calculations, involving all kinds of optical phonons or based on the nonadiabaticity of the electron–phonon system, resulted in an increasing overtone/fundamental ratio with decreasing dot radius in accordance with experiments on CdSe and CuBr QDs [135, 132].

6.2. Lithographically Patterned Wires and Dots

For heterostructure systems, which are the origin of nanostructures such as quantum wires and dots, generally lattice mismatch between the layer materials occurs. This mismatch induces strain, which affects the electronic band structure and consequently the optical properties. In addition, this strain also affects the vibronic properties, as discussed in Section 4.3. The optical as well as the vibronic properties of nanostructures are influenced not only by strain but also by, for example, composition and confinement effects. Fortunately, for the vibronic properties these influences have different features than for the optical behavior. Therefore, mostly a combined analysis by Raman spectroscopy and photoluminescence allows the separation of these effects.

For combinations of materials with a small or moderate lattice mismatch, pseudomorphic growth occurs (i.e., a strained epitaxial layer growth) up to a critical thickness. From this thickness onward, the strain decreases because of the formation of misfit dislocations. The appearance of misfit dislocations can be observed in the Raman spectra as an increase in the intensity of the symmetry-forbidden TO phonon peak as well as a broadening of the LO phonon peak.

The following section treats the pressure dependence of the vibrational properties as well as the strain effects in quantum wells and superlattices. Finally, the last part of this chapter deals with the strain relaxation in etched quantum wires and dots.

6.2.1. Strain Effects in Nanostructures Due to Lattice Mismatch in (Multi)-Quantum Wells

In the MBE growth of layered systems the lattice mismatch between barrier and well induces biaxial strains in the interface regions. These strains cause shifts of the LO phonon wavenumber, which are proportional to the strain. For compressive (tensile) strain a blueshift (redshift) of the phonon wavenumber can be observed [43, 115, 137–139]. The amount of this shift may depend on the layer thickness [39, 140, 141]. For ZnSe on GaAs, from the beginning of the epilayer growth large elastic strains occur parallel to the interface due to the lattice mismatch. This strain causes an distinct blueshift of the phonon wavenumber. At a critical thickness of $h_0 = 80$ nm, the first stable dislocation pairs can be formed [142–144]. At a thickness d_c of 300 nm other arrangements of dislocations are formed [143] and total relaxation by the formation of arrays of misfit dislocations takes place. For thicknesses beyond 600 nm the wavenumber gradually approaches the value of the stress-free LO phonon. The appearance of misfit dislocations and stacking faults leads to a strong decrease of the lattice quality. Therefore, for the growth of a high-quality layer it is necessary to keep the layer thicknesses below the critical thickness. The strains in these layers lead to changes in the vibrational properties, which can be estimated by Raman spectroscopy. By these experiments the dependence of the strain on the layer thickness can be evaluated in one and the same sample [22, 145]. In ZnTe/ZnSe superlattices with a lattice mismatch of about 7% and thicknesses of 0.5–5.0 nm the superlattice is not destroyed by diffusion of Se or Te even for very small layer thicknesses. The Raman spectra of these samples show an asymmetrical LO phonon peak with increasing width for increasing layer thickness [146]. This increasing peakwidth is caused by inhomogeneous strain, whose amount is increasing with increasing layer thickness. The LO wavenumber position depends rather on the ratio d_{Se}/d_{Te} than on the absolute thicknesses of the layers [142]. For these superlattices the LO phonon position shifts with decreasing ratio d_{Se}/d_{Te} from the wavenumber position of pure ZnSe to the wavenumber position of pure ZnTe. The TO phonon peak, which appears because of the strain effects, shows a shift of the wavenumber position, which is linear to $d_{Te}/(d_{Se} + d_{Te})$. The experimental shifts are in good agreement with the shifts calculated from the equations, given in Section 4.3 [22, 39, 137, 145, 147–156]. Biaxial strain in the layer perpendicular to [001] makes the sample quasi-two-dimensional and the pressure-dependent shift is then given by Eqs. (13) and (14) [39, 42, 43, 146].

Besides the strain caused by lattice mismatch, another contribution exists which gains in significance when the sample is cooled to very low temperatures. This thermally induced strain is caused by the different thermal expansion coefficients. At room temperature the amount of the thermal strain is about one order smaller than the lattice mismatch induced strain, whereas for low temperatures this strain can become dominant [155]. The rather small strains for very thin layers can be explained by the growth of a kind of island structure, which can be used for the MBE fabrication of quantum dots (Stranski–Krastanov model: see refs. [157–161]). In contrast to the lattice for a complete layer, the lattice at the edges of these islands is relaxed, which leads to a reduced strain shift. For the investigation of inhomogeneous strains (i.e., local strain fields like quantum islands), the second-order Raman signals have to be used. Due to the momentum conservation, the optical phonons involved in a first-order process originate from the center of the Brillouin zone. Because of this, they have quite long wavelengths, on the order of ≥ 100 nm, so only homogeneous strains with extensions larger than the phonon wavelength can be obtained. The phonons participating in second-order processes mostly originate from the zone boundary and therefore obtain shorter wavelengths in the range of 1 nm. Therefore, small structures, like quantum islands or interface clusters, can give rise to observable shifts of the second-order signals, whereas the first-order signal is an average over a larger area [162]. A discrepancy in magnitude of the

strains calculated from the shift of the first-order signals on the one hand and from the splitting of the second-order bands on the other hand is a sign for inhomogeneous strain fields. This inhomogeneity can be explained when assuming large low strain areas and small high strain areas at the interface, as is the case for small Ga_2Se_3 islands (1–10 nm) at the GaAs/ZnSe interface [162, 163].

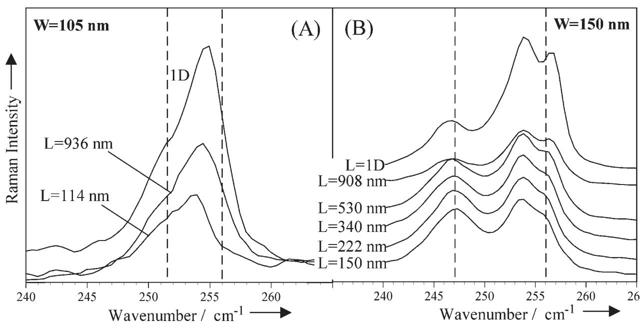
6.2.2. Strain Relaxation in Quantum Wires and Etched Quantum Dots

In the fabrication of etched quantum wires and quantum dots, the strain caused by lattice mismatch in MBE grown layers is partially reduced by the etching process. At their lateral surfaces these structures have the opportunity for relaxation. This ability is increasingly hampered toward the center of the wires or dots. Therefore (e.g. for wires) the overall degree of relaxation depends on the ratio of the relaxed wire edges to the total volume of the wire. Thus, for a reduction of the wire dimensions, like the reduction of the wire width or the wire length, a reduction of the total amount of the strain is possible. As already mentioned above, the first-order Raman signals do not distinguish local strains, so only one signal for the whole wire or dot can be obtained, which consists of the signals of the strained center as well as the relaxed edges. But even in the first-order spectra the strain relaxation can be observed as a shift of the LO phonon wavenumber toward the value of the unstrained material.

The Raman analysis of edge relaxation will be discussed here for the system $Cd_xZn_{1-x}Se$/ZnSe [164–166]. The investigated structures consisted of $Cd_xZn_{1-x}Se$ quantum wells (thickness ≈5 nm), embedded in ZnSe layers on GaAs(100), which had been patterned to wires and dots by e-beam lithography and subsequent wet etching. Wire widths and dot sizes were down to about 15 nm. The Raman spectra are expected to show the LO phonon peaks of $Cd_xZn_{1-x}Se$ and ZnSe. In the size range of these structures no phonon confinement effects are expected, as shown in Figure 13. Thus, the $Cd_xZn_{1-x}Se$ LO phonon wavenumber is fully determined by the composition of the layers and their strain state. The Cd content of all structures was chosen ≤35% in order to limit the lattice mismatch, thus avoiding the onset of relaxation processes before the patterning, which would deteriorate the crystalline perfection. The composition-dependent vibronic behavior of $Cd_xZn_{1-x}Se$ is described by the two-mode model (see Section 4.2). The Zn–Se vibration mode is expected to be much more intense than the Cd–Se mode, because $x \ll 0.5$. Therefore in the following the ZnSe-like mode will be analyzed. With increasing Cd content x, its wavenumber decreases, starting from LO_{ZnSe} (=256 cm^{-1}) with a downshift of about 0.26 cm^{-1} per percent Cd, which results in $LO_{CdZnSe} = 247$ cm^{-1} for $x = 0.35$ [114], thus enabling the spectral separation between the LO peaks of the CdZnSe wires and the ZnSe barriers. However, for lower Cd content ($x \leq 0.2$) this separation will become increasingly difficult.

The Raman spectra for two different compositions and various wire widths and lengths are shown in Figure 17. The Cd content was $x = 0.2$ for spectra (A) and $x = 0.35$ for spectra (B), while the wire widths were either beyond 100 nm (upper spectra) or below 40 nm (lower spectra). These values had been chosen because side-face relaxation effects become observable for wire widths below about 100 nm and full relaxation is achieved at about 40 nm [167]. The wire length was varied between 100 μm (denoted 1D) and the wire width W, thus ending up with a quantum dot geometry. The most pronounced changes occur in the Raman spectra (B) of samples with a Cd content $x = 0.35$, because here the peaks are well separated. The spectra B show the ZnSe LO peak at 256 cm^{-1} (= bulk wavenumber) and additional peaks due to the CdZnSe wires. As a common feature for all wire lengths the 247 cm^{-1} peak occurs, whose wavenumber corresponds to bulk $Cd_{0.35}Zn_{0.65}Se$. Therefore it is explained in terms of the relaxed regions at the side faces and the end faces of the wires. Obviously, its relative intensity increases with decreasing wire length. Finally, especially in the spectrum for the wider wires ($W = 150$ nm), a third feature appears in between, which is attributed to the central region of the wires. Here the ZnSe-like LO phonon of the CdZnSe wires is blueshifted with respect to bulk CdZnSe due to compressive strain. This interpretation is underscored by the decrease of the relative weight of this structure with decreasing wire size.

For the spectra (A), due to the low Cd content of $x = 0.2$, no separation between the ZnSe and CdZnSe phonon peaks is obtained. However, in the superimposed peak a clear shift and shape change is observed, which can be explained along the same line as for the samples with the higher Cd content [164].



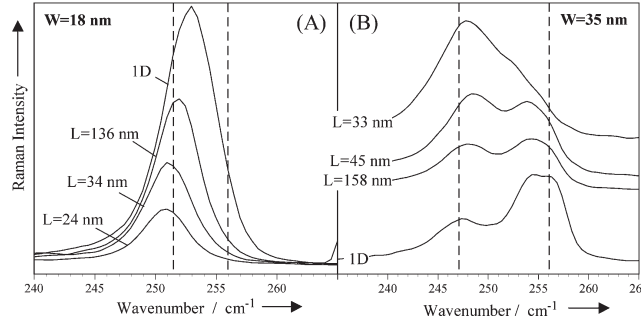


Figure 17. Resonance micro-Raman spectra of $Cd_xZn_{1-x}Se$/ZnSe quantum wires with different wire width W and wire length L. Cd content: $x = 0.2$ for spectra (A) and $x = 0.35$ for spectra (B). Reprinted with permission from [164], B. Schreder et al., *J. Appl. Phys.* 88, 764 (2000). © 2000, American Institute of Physics.

Small shifts of the sum peak when changing the excitation wavelength are due to the change in the resonance conditions of the well and the barrier, which leads to changes in the relative intensities of the two peaks. With decreasing wire width x the LO phonon position of the ZnSe-like phonon shows a redshift toward the wavenumber of the unstrained material, whereas the wavenumber of the ZnSe LO phonon remains constant [168].

In order to evaluate these data as detailed as possible, the expected phonon wavenumbers and their strain-induced local variation were calculated. In this procedure, for each point of the wire cross section the strain was minimized with respect to the nearest neighbors by applying the finite element method [168–170]. In this way, every point of the wire cross section is characterized by a particular phonon wavenumber, dependent on the stress in this point. More detailed calculations of the strain field in short wires or dots result in shear strains at the corners of the wires which also have to be considered in the calculation of the wavenumber shift [171]. The calculated position-dependent wavenumbers show a strong lateral inhomogeneity as well as a clear height dependence. It can be seen from the model that the wire edges are totally relaxed and therefore obtain the bulk phonon wavenumber, whereas the strain in the wire centers is comparable with the strain in the well for the thick wires but decreases with decreasing wire width. Total strain relaxation is predicted for widths below 10 nm.

A comparison of the expected LO phonon wavenumbers, which may be obtained by averaging over the wire height z and the wire width, with experimental results shows that experimentally full relaxation is obtained already for widths below 30 nm. The averaging over wire regions with different strain leads to a theoretically predicted redshift of the position of the resulting wire peak with decreasing wire width due to the increasing influence of the relaxed regions. This phenomenon was experimentally observed [167–170, 172] for samples with various Cd contents and resulted in a temperature-independent wavenumber difference between the LO peaks of CdZnSe and ZnSe [167, 168]. Furthermore, the spectra in Figure 17 show distinct peaks from the relaxed and the strained range. These results indicate a sharp change in the lattice constants between the center and the edge of the wire rather than a step by step relaxation.

For a quantitative evaluation of the Raman peak intensities and shapes in terms of the relative contributions from differently strained regions, the local scattering efficiency has to be taken into account. The resonant enhancement at the electronic bandgap, which was discussed in Section 2.3, may lead to a preferential selectivity for regions with a specific strain value, because the electronic bandgap value is also strain-dependent. This phenomenon was confirmed experimentally by experiments with different laser lines, which resulted in different peak shapes [168, 173]. Systematic investigations of the excitation profiles in resonant Raman scattering revealed that for the system $Cd_xZn_{1-x}Se$/ZnSe the wavelength dependence of the peak shape increases with increasing Cd content and with increasing wire width. This is attributed to more pronounced inhomogeneities [174].

6.2.3. Oxidation at Open Sidewalls

Etching of quantum wires and dots out of $2d$-grown quantum well structures (mesas), which is the topic of this section, inherently creates new interfaces at the sidewalls of those dots and wires. The cleanliness of the side faces is crucial for the optical properties of these systems. The increased surface recombination rate at the sidewalls creates dead layers. The process technology applied for the etching process as well as exposure of the sidewalls to air ambient crucially affects the properties of the dead layers [175–177]. Additionally, oxide layers are expected to affect the sidewall strain relaxation of a strained active layer in very small structures (<50 nm). By using a UHV-compatible process (e.g., dry etching), the influence of sidewall oxidation due to air ambient can be investigated. By coupling the sample preparation via a UHV transport system to a UHV Raman setup, the intentional oxidation of the sidewall can be followed by Raman spectroscopy. So MBE-grown CdZnSe quantum wells embedded in ZnSe grown on GaAs(100) substrates exhibit a systematic downshift of the ZnSe–LO mode upon air contact. This downshift, which does not exist for $2d$ reference mesas and increases with decreasing wire width (e.g., -0.6 cm^{-1} for 40 nm wire width) is attributed to the additional strain component due to the sidewall oxidation [14].

6.3. Self-Organized Dots

Self-organized dots are formed by atomic rearrangement due to strain relaxation in strongly mismatched epilayers (i.e., lattice mismatch in the order of percent) when the epilayer thickness exceeds a critical value, usually only few monolayers. This three-dimensional growth mode is referred to as Stranski–Krastanov growth. For achieving nanostructures, it constitutes a very interesting alternative to lithographically defined dots, because it does not require a highly sophisticated patterning technology. Moreover, it allows direct overgrowth of the dots without interruption of the UHV conditions. Of course, a crucial point is the dot size homogeneity.

These self-organized dots have attracted much attention because they offer an interesting potential as active layers in semiconductor laser diodes. II–VI dots are considered promising candidates for the realization of diodes in the blue and blue-green spectral range.

Raman spectroscopy is applied for studying, for example, the onset of dot formation, their chemical composition, interface sharpness, and stability. The Raman analysis of self-organized dots constitutes a special challenge from the viewpoint of sensitivity, because we deal with extremely low scattering volumina. The dot thickness can be as low as only a few monolayers, and the lateral coverage factor is below one. These drawbacks are overcome by resonant excitation at the electronic transitions of the dots, which selectively enhances their Raman signals. However, one has to avoid the obscuring of the Raman peaks by photoluminescence (PL). This may be achieved either by excitation at the low-energy edge of the PL peak or by utilizing the resonant Raman enhancement at a nonfundamental electronic transition (e.g., $E_0+\Delta_0$).

6.3.1. Onset of Dot Formation

For CdSe/BeTe, Raman studies were performed at the very beginning of quantum dot formation (i.e., thickness fluctuations as low as one monolayer). In this very strongly lattice-mismatched material system ($\Delta a/a = 7\%$) self-organized dots are known to occur for layer thicknesses beyond about 3 ML. In the study which is discussed here, the initial stage of the dot formation process was conserved in CdSe layers of nominally one ML by embedding them between BeTe layers. For reasons of Raman intensity, a stack of five CdSe single MLs was grown, separated by BeTe barrier layers with a thickness of 4.5 nm each. For the interfaces Te–Cd bonds were chosen because they give the best abruptness, as shown in Section 5.1.1. The resulting Raman spectra were partly presented already in Figure 8. Their explanation in terms of CdSe-ML vibrations, (i.e., eigenmodes of the sequence $\cdots$–Be–Te–Cd–Se–Cd–Te–Be–$\cdots$) resulted in a quantitative agreement of the experimental peak positions with the relative positions according to the linear chain model but could not quantitatively explain the peak shape. For obtaining this detailed agreement, additional vibration modes have to be invoked which require the assumption of laterally distributed inclusions of two MLs (i.e., the sequence $\cdots$–Be–Te–Cd–Se–Cd–Se–Cd–Te–Be–$\cdots$). The atomic displacements of the resulting additional eigenmodes are shown in Figure 18, together with the corresponding eigenwavenumbers, as derived from the linear chain model. The essential modifications as compared to the one ML modes are a slight downshift of the Te–Cd interface-centered mode by 5 to 164 cm^{-1} and a splitting of the most strongly confined mode, the former Cd–Se–Cd mode (216 cm^{-1}), yielding two new wavenumbers: 204 and 224 cm^{-1}. Thus, the 190–220 cm^{-1}-structure in Figure 8 is fitted quantitatively as a superposition of three peaks: the central one from the one ML regions and the outer ones from the two ML inclusions.

A further confirmation of this model was obtained from resonance effects in Raman spectra for various laser lines. Due to the different electronic transition energies of the one and two ML regions, a variation of the relative contribution of these regions is expected when changing the incident photon energy. This should result in changes of the peak shape. Indeed, the spectra in Figure 19 show a systematic trend: in the 190–220 cm^{-1}-structure, the edge contributions, which are attributed to the two ML vibrational modes, are more pronounced for the 2.5 eV laser line, while an increase of the laser energy to 2.71 eV results in an increase of the central part, which is attributed to the one ML mode. This trend of an increased contribution of the one ML regions with increasing laser photon energy corresponds to the expected resonance behavior, since the confinement-induced enhancement of the electronic transition energy for one ML regions should exceed the two ML value.

These Raman spectroscopy results, concerning the abruptness of interfaces with Te–Cd bonds and fluctuations of the CdSe layers, which result in one and two ML regions, are fully consistent with transmission electron microscopy results. The latter reveal within each layer a regular distribution of alternating one and two ML regions and, moreover, a correlation between neighboring layers, which leads to a chessboardlike pattern of one and two ML regions.

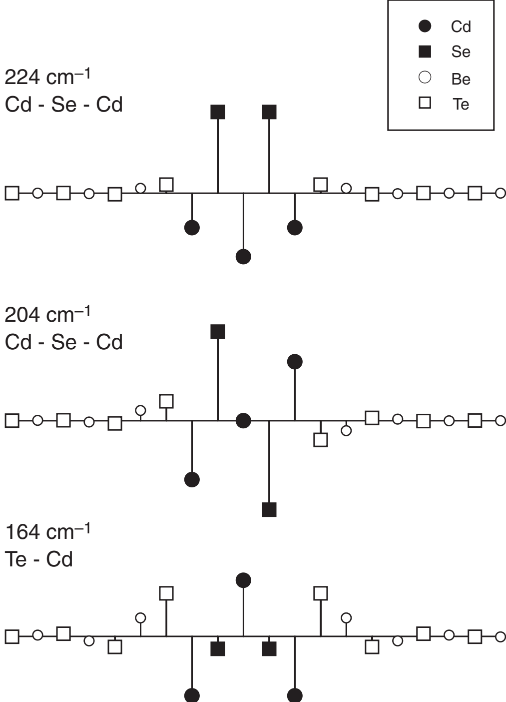


Figure 18. Displacement vectors and wavenumbers of vibration modes, related to two ML CdSe, embedded in CdTe with Cd–Te interface bonds. These modes were obtained by a linear chain calculation with next neighbor interactions. Adapted with permission from [52], V. Wagner et al., *Appl. Surf. Sci.* 175, 169 (2001). © 2001, Elsevier Science.

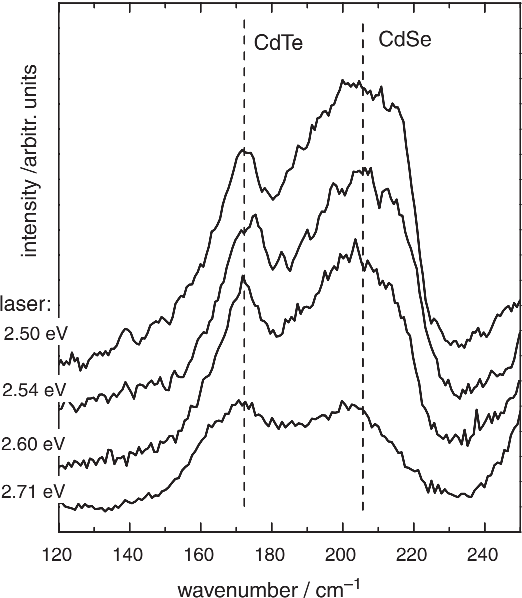


Figure 19. Resonance effects in the Raman spectra of CdSe monolayers, embedded in BeTe. Variation of the incident photon energy from 2.71 to 2.50 eV results in a change of the peak shape due to an increasing contribution of two ML regions. Reprinted with permission from [52], V. Wagner et al., *Appl. Surf. Sci.* 175, 169 (2001). © 2001, Elsevier Science.

6.3.2. Chemical Composition and Stability

CdSe Dots on ZnSe For CdSe growth on ZnSe, the Stranski–Krastanov growth mode is induced by the lattice mismatch of 7%. However, it does not simply result in binary CdSe dots. Instead, transmission electron microscopy (TEM) studies revealed that ternary ZnCdSe dots are formed with a Zn content up to 30%, on top of a ZnCdSe wetting layer with a Zn content of about 15%. In addition, smaller planar Zn-rich quantum islands appear, with a Cd concentration of only 30% [178].

In Raman spectroscopy, the latter can be excited resonantly by a wavelength of 488 nm. They give rise to a ZnSe-like LO phonon mode at 246.4 cm^{-1}, thus confirming the TEM results [179]. The resonant excitation also allows the observation of the 2LO and 3LO multiphonon peaks. In micro-Raman experiments (spot diameter $\approx 0.7\ \mu m$), this situation is stable for sufficiently low input power (about 0.2 mW). However, upon increasing the input power to 5 mW, the mode intensities strongly decrease with irradiation time on a time scale of hundreds of seconds due to a shift of the electronic resonance energy. At the same time, the LO wavenumber is blueshifted by 1.7 cm^{-1}. These results are explained consistently by an outdiffusion of Cd from the planar $Cd_{0.3}Zn_{0.7}Se$ islands, which is induced by the local heating which results from the resonant absorption in the islands [179]. This example underscores on the one hand the caution which has to be taken in micro-Raman spectroscopy to avoid any unintentional changes of the sample; on the other hand it illustrates the possibilities of a selective modification by optical irradiation.

6.3.3. Excitonic Confinement Effects, Probed by Spin Flip Raman Scattering

Because of the excitonic character of its intermediate states, Raman scattering allows the analysis of excitonic states in quantum dots. The most favorable type of scattering for this purpose is spin flip Raman scattering (SFRS). In this scattering process the elementary excitation involved in the sample is a change of the spin orientation (e.g., of an electron or a hole in an external magnetic field B). The corresponding change in photon energy is given by the Zeeman splitting $g \cdot \mu_B \cdot B$, where g is the gyromagnetic ratio (g-factor) of the involved electron or hole, and μ_B is the Bohr magneton. For nonmagnetic semiconductors SFRS induces wavenumber shifts on the order of 1 cm^{-1} per Tesla. These very low values lead to an extremely pronounced resonance behavior, because now $\omega_s \approx \omega_i$, and both factors in the denominator of Eq. (9) vanish nearly simultaneously. This extremely strong resonance offers two major advantages:

(i) The sensitivity for very small scattering volumina (e.g., in interface regions or nanostructures).
(ii) A pronounced material selectivity in heterostructures.

The Raman shift of the SFRS peak (e.g. of the electron spin flip) varies linearly with the magnetic field, and the slope of this line allows the evaluation of the corresponding g-factor with high accuracy. Interestingly, in quantum dots the g-factor which is determined in this way may deviate considerably from the bulk g-value of the dot material and show a shift toward the g-value of the underlying or embedding material. This shift allows an assessment of the extention of the wavefunction beyond the dot.

Beside electrons or holes, the optically excited excitons themselves may also undergo a spin flip [180, 181]. These SFRS processes allow the analysis of the exchange interaction between the electron and the hole in the exciton and, moreover, of its modification in confined systems, such as quantum dots [182]. Due to the exchange interaction Δ_{ex} the field dependence of the excitonic SFRS peak exhibits for low B-fields a positive deviation from the linear dependence, leading to a $\sqrt{\Delta_{ex}^2 + (g\mu_B B)^2}$ field dependence. Therefore the extrapolation of the field dependence to $B = 0$ results in a finite Raman shift, whose energy value yields the excitonic exchange energy Δ_{ex}. Typical bulk values for II–VI materials are of order 0.1 to 1 meV, but for quantum dot sizes below the bulk exciton size a considerable increase of the exchange energy is expected.

Examples of excitonic SFRS in nanostructures have been reported for self-organized III–V quantum dots in the material systems InP/InGaP and InAs/GaAs [183], as well as for II–VI dots: chemically produced CdS dots in a glass matrix [184].

The excitonic SFRS studies which are most relevant for our discussion of self-organized nanostructures deal with CdSe fractional monolayers (≤ 0.5 ML), embedded in ZnSe epilayers [185, 186]. They form islands of various sizes (i.e., with different confinement effects). Moreover, the lateral dimensions and separation of the islands are smaller than the exciton radius. In this laterally inhomogeneous system, the pronounced resonance behavior of SFRS allows one to address each island size separately by the appropriate choice of the laser wavelength.

In this way, the confinement dependence of the exciton exchange energy has been studied. The experimental results reveal for decreasing island size an increase of Δ_{ex}, which is proportional to the exciton confinement energy. The Δ_{ex} range goes from 0.5 meV (bulk value) to 1.2 meV for the smallest islands, whose exciton confinement energy amounts to 30 meV.

The electronic SFRS in these islands shows the expected linear dependence on the B-field. However, its g-factor amounts to 1.15 ± 0.02, which is much closer to ZnSe than to CdSe. Thus, it reflects the penetration of the wavefunctions into the ZnSe, in which the islands are embedded.

7. SUMMARY

In this chapter an overview was presented of the abilities of Raman spectroscopy for the analysis of interfaces, surfaces, and nanostructures, such as quantum wires and dots. Within this framework, the application of *in-situ* Raman spectroscopy for growth monitoring was also discussed.

The required sensitivity and selectivity were obtained by resonant excitation, that is, by the appropriate choice of the laser line, combined with sophisticated multichannel detection equipment, such as a CCD.

It was shown that information is obtained on a wide variety of aspects. Examples were presented for the identification of materials, the study of their crystalline perfection, the composition of mixed compounds, and strain or doping.

Several interface- or surface-induced phenomena were discussed, such as chemical and dielectric vibrational modes and electronic band bending. Examples of interfaces between different II–VI materials as well as II–VI/III–V interfaces were presented. For ZnSe/GaAs(100), the coupling of the dielectric vibrations at the surface and the interface was analyzed, and it was shown that from Raman spectra of the epilayer LO phonon together with the coupled plasmon–LO–phonon mode of the substrate a consistent picture was obtained of the epilayer strain relaxation and the electronic band bending on the substrate side due to the interface defects. Furthermore, examples were presented of the influence of the growth conditions on the interface chemistry and the crystallographic perfection of heterostructures (e.g., for embedded monolayers of CdSe in BeTe and for the onset of CdTe growth on InSb substrates).

For the study of (multi)-quantum wells and superlattices new phenomena in the Raman spectrum were exploited, such as confined optical and folded acoustical phonons. As an example, for the lattice-matched system BeTe/ZnSe the confined optical phonon behavior was evaluated in terms of interface perfection, showing that Zn–Te interfaces are far superior to Be–Se ones. In this material system the folded acoustical phonons were used to determine the elastic constant c_{11} and the sound velocity of BeTe. For heavily strained systems, such as cubic CdS/ZnS and hexagonal CdS/CdSe superlattices, the strain-induced shift of the confined phonon modes was studied.

The Raman studies on nanostructures which were presented here cover lithographically patterned wires and dots, as well as self-organized dots due to Stranski–Krastanov growth and dots in amorphous matrices, produced chemically (e.g., by the sol–gel technique). For these systems, common issues studied by Raman spectroscopy were their composition, their size, and their strain behavior.

For lithographically patterned strained wires and dots, consisting of ZnCdSe embedded in ZnSe layers, the strain relaxation after patterning was studied. In this case, strained and relaxed regions were observed by separate peaks in the Raman spectrum. From their intensity ratio the degree of relaxation was determined, which revealed that relaxed regions can be observed for wire widths below 100 nm and full relaxation occurs below 15 nm.

The onset of the formation of self-organized dots was studied for CdSe, embedded in BeTe. Here thickness fluctuations as low as one ML were detected by selective intensity enhancement of the different regions in resonant Raman scattering. Furthermore, for self-organized ZnCdSe islands in ZnSe, the compositional stability during laser irradiation was studied. Here a distinctly enhanced diffusion during resonant irradiation was observed.

The size of dots in matrices was determined from low-wavenumber Raman spectroscopy, for example, for Schott glasses, consisting of glass matrices, doped with CdS_xSe_{1-x} quantum dots in the range of about 10 nm. For a quantitative evaluation of size distributions the size dependence of the resonance excitation profile had to be taken into account.

The composition of mixed-compound dots was determined from their optical phonon wavenumbers. It was shown that for crystallite sizes beyond 5 nm no confinement effects had to be considered. As an example, the sintering behavior of CdTe/CdS "core-shell" clusters was discussed.

Moreover, the pressure-dependent behavior of quantum dots was addressed. Here, pressure-induced shifts of optical phonon wavenumbers were employed (e.g., for studying the zincblende-to-rocksalt phase transition in CdS colloids with a particle diameter of 8 nm in a glass matrix).

Finally, spin flip Raman spectroscopy was applied for studying the size dependence of the electron–hole exchange interaction and the extension of the electronic wavefunction beyond the edges of nanostructures. The electron–hole exchange in CdSe islands in ZnSe was shown to increase by a factor of about 3 with decreasing island size, while in the same material system the electronic g-factor of the islands was close to the ZnSe value, thus confirming the penetration of the electronic wavefunctions into the embedding ZnSe.

ACKNOWLEDGMENTS

Part of the reported material was obtained within a research project sponsored by the Deutsche Forschungsgemeinschaft (SFB 410, Projects B7 and C3). We highly appreciate the financial support by the DFG as well as by the Fonds der Chemischen Industrie.

REFERENCES

1. W. Hayes and R. Loudon, "Scattering of Light by Crystals." Wiley, New York, 1978.
2. W. Richter, "Resonant Raman Scattering in Semiconductors," Springer Tracts in Modern Physics, Vol. 78. Springer, Berlin/Heidelberg/New York, 1976.
3. J. Geurts, *Surf. Sci. Rep.* 18, 1 (1993).
4. M. Cardona and G. Güntherodt, "Light Scattering in Solids," Topics in Applied Physics, Vols. 8, 50, 51, 54, 66, 68, 75, 76. Springer, Berlin/Heidelberg/New York.
5. N. Esser and W. Richter, "Light Scattering in Solids VIII," Topics in Applied Physics, Vol. 50, p. 96. Springer, Berlin/Heidelberg/New York, 2000.
6. I. R. Lewis and H. G. M. Edwards, "Handbook of Raman Spectroscopy; From the Research Laboratory to the Process Line." Dekker, New York/Basel, 2001.
7. R. Loudon, *Proc. Roy. Soc. London Ser. A* 275, 218 (1963).
8. E. Burnstein and A. Pinczuk, "Fundamentals of Inelastic Light Scattering in Semiconductors and Insulators," Topics in Applied Physics, Vol. 8. Springer, Berlin/Heidelberg/New York, 1975.
9. A. Compaan and H. J. Trodahl, *Phys. Rev. B* 29, 793 (1984).
10. F. Cerdeira, E. Anastassakis, W. Kauschke, and M. Cardona, *Phys. Rev. Lett.* 57, 3209 (1986).
11. T. Ruf, C. Trallero-Giner, R. T. Phillips, and M. Cardona, *Phys. Rev. B* 41, 3039 (1990).
12. F. Agullo-Rueda, E. Menendez, and J. Hong, *Phys. Rev. B* 38, 12720 (1980).
13. V. Wagner, D. Drews, N. Esser, D. Zahn, J. Geurts, and W. Richter, *J. Appl. Phys.* 75, 7330 (1994).
14. V. Wagner, J. Wagner, W. Faschinger, G. Bacher, J. Geurts, B. Schreder, and W. Kiefer, "Proc. of the XVIIth Int. Conf. on Raman Spectroscopy," 2000.
15. N. Esser and J. Geurts, "Raman Spectroscopy, Optical Characterization of Epitaxial Semiconductor Layers," Springer, Berlin/Heidelberg/New York, 1996.
16. R. K. Chang and M. B. Long, Light Scattering in Solids II, Topics in Applied Physics, Vol. 50, p. 179. Springer, Berlin/Heidelberg/New York, 1982.

17. V. Deckert and W. Kiefer, *Appl. Spectrosc.* 46, 322 (1992).
18. R. B. Bilhorn, J. V. Sweedler, P. M. Epperson, and M. B. Denton, *Appl. Spectrosc.* 41, 1114 (1987).
19. J. C. Tsang, "Light Scattering in Solids V," Topics in Applied Physics, Vol. 66, Springer, Berlin/Heidelberg/New York, 1989.
20. S. Webster, D. N. Batchelder, and D. A. Smith, *Appl. Phys. Lett.* 72, 1478 (1998).
21. M. Ramsteiner, C. Wild, and J. Wagner, *Appl. Opt.* 28, 4017 (1989).
22. V. Wagner, W. Richter, J. Geurts, D. Drews, and D. R. T. Zahn, *J. Raman Spectrosc.* 27, 265 (1996).
23. E. K. Suh, A. K. Arora, A. K. Ramdas, and S. Rodriguez, *Phys. Rev. B* 45, 3360 (1992).
24. D. R. T. Zahn, C. Maierhofer, A. Winter, M. Reckzügel, R. Srama, U. Rossow, A. Thomas, K. Horn, and W. Richter, *Appl. Surf. Sci.* 56, 684 (1992).
25. A. K. Arora and A. K. Ramdas, *Phys. Rev. B* 35, 4345 (1987).
26. W. E. Bron, J. Kuhl, and B. K. Rhee, *Phys. Rev. B* 34, 6961 (1986).
27. P. G. Klemens, *Phys. Rev.* 148, 845 (1966).
28. I. F. Chang and S. S. Mitra, *Phys. Rev.* 172, 924 (1968).
29. Y. S. Chen, W. Shockley, and G. L. Pearson, *Phys. Rev.* 151, 648 (1966).
30. H. Richter, Z. P. Wang, and L. Ley, *Solid State Comm.* 39, 625 (1981).
31. J. H. W. E. Hermans, Ph.D. Thesis, RWTH Aachen, Augustinus Publ. Comp., Aachen, 1996.
32. R. Vogelgesang, A. J. Mayur, M. D. Sciacca, E. Oh, I. Miotkowski, A. K. Ramdas, S. Rodriguez, and G. Bauer, *J. Raman Spectrosc.* 27, 239 (1996).
33. X. S. Zhao, J. Schroeder, P. D. Persans, and T. G. Bilodeau, *Phys. Rev. B* 43, 12580 (1991).
34. A. K. Arora, E. K. Suh, U. Debska, and A. K. Ramdas, *Phys. Rev. B* 37, 2927 (1988).
35. B. F. Variano, N. E. Schlotter, D. M. Hwang, and C. J. Sandroff, *J. Chem. Phys.* 88, 2848 (1988).
36. C. Yu, I. L. Spain, and E. F. Skelton, *Solid State Comm.* 25, 49 (1975).
37. F. Cerdeira, B. L. Buchenauer, F. H. Pollack, and M. Cardona, *Phys. Rev. B* 5, 580 (1972).
38. J. J. Shiang, S. H. Risbud, and A. P. Alivisatos, *J. Chem. Phys.* 98, 8432 (1993).
39. E. K. Suh, D. U. Bartholomew, A. K. Ramdas, S. Rodriguez, S. Venugopalan, L. A. Kolodziejski, and R. L. Gunshor, *Phys. Rev. B* 36, 4316 (1987).
40. S. Venugopalan and A. K. Ramdas, *Phys. Rev. B* 8, 717 (1973).
41. H. J. Lozykowski and V. K. Shastri, *J. Appl. Phys.* 69, 3235 (1991).
42. A. Yamamoto, Y. Yamada, and Y. Masumoto, *Appl. Phys. Lett.* 58, 2135 (1991).
43. A. Yamamoto, Y. Yamada, and Y. Masumoto, *J. Cryst. Growth* 117, 488 (1992).
44. P. Grosse, "Freie Elektronen in Festkörpern." Springer, Berlin/Heidelberg/New York, 1979.
45. U. Nowak, W. Richter, and G. Sachs, *Phys. Status Solidi B* 108, 131 (1981).
46. S. G. Lyapin, P. C. Klipstein, N. J. Mason, and P. J. Walker, *Phys. Rev. Lett.* 74, 3285 (1995).
47. J. Spitzer, A. Höpner, M. Kuball, M. Cardona, B. Jenichen, H. Neuroth, B. Brar, and H. Kroemer, *J. Appl. Phys.* 77, 811 (1995).
48. J. Schmitz, J. Wagner, F. Fuchs, N. Herres, P. Koidl, and J. D. Ralston, *J. Cryst. Growth* 150, 858 (1995).
49. T. Walter, A. Rosenauer, D. Gerthsen, F. Fischer, T. Litz, A. Waag, and G. Landwehr, *Inst. Phys. Conf. Ser.* 157, 315 (1997).
50. V. Wagner, J. J. Liang, R. Kruse, S. Gundel, M. Keim, A. Waag, and J. Geurts, *Phys. Status Solidi B* 215, 87 (1999).
51. A. Waag, F. Fischer, H. J. Lugauer, T. Litz, T. Gerhardt, J. Nürnberger, U. Lunz, U. Zehnder, W. Ossau, G. Landwehr, B. Roos, and H. Richter, *Mater. Sci. Eng. B* 43, 65 (1997).
52. V. Wagner, J. Wagner, T. Muck, G. Reuscher, A. Waag, J. Geurts, N. Sadchikov, S. V. Sorokin, S. V. Ivanov, and P. Kopev, *Appl. Surf. Sci.* 175, 169 (2001).
53. A. Dinger, M. Göppert, R. Becker, M. Grün, S. Petillon, C. Klingshirn, J. Liang, V. Wagner, and J. Geurts, *Phys. Rev. B* 64, 245310 (2001).
54. M. Hünermann, J. Geurts, and W. Richter, *Phys. Rev. Lett.* 66, 640 (1991).
55. H. Wilhelm, W. Richter, D. R. T. Zahn, U. Rossow, D. A. Woolf, D. I. Westwood, and R. H. Williams, *Surf. Sci.* 251, 556 (1991).
56. J. C. Tsang, Y. Yokota, R. Matz, and G. W. Rubloff, *Appl. Phys. Lett.* 44, 430 (1984).
57. R. J. Nemanich, R. T. Fulks, B. L. Stafford, and H. A. V. Plas, *Appl. Phys. Lett.* 46, 670 (1985).
58. G. P. Schwartz, B. Schwartz, D. Distefano, G. J. Gualtieri, and J. E. Griffiths, *Appl. Phys. Lett.* 34, 205 (1979).
59. R. L. Farrow, R. K. Chang, S. Mroczkowski, and F. H. Pollak, *Appl. Phys. Lett.* 31, 768 (1977).
60. R. L. Farrow, R. K. Chang, and S. Mroczkowski, *Appl. Phys. Lett.* 31, 768 (1977).
61. J. Finders, M. Keuter, D. Gnoth, J. Geurts, J. Woitok, A. Kohl, R. Müller, and K. Heime, *Mater. Sci. Eng. B* 21, 161 (1993).
62. J. Geurts, D. Gnoth, J. Finders, A. Kohl, and K. Heime, *Phys. Status Solidi A* 152, 211 (1995).
63. J. Geurts, J. Finders, J. Woitok, D. Gnoth, A. Kohl, and K. Heime, *J. Cryst. Growth* 145, 813 (1994).
64. A. Krost, W. Richter, D. R. T. Zahn, K. Hingerl, and H. Sitter, *Appl. Phys. Lett.* 57, 1981 (1990).
65. D. R. T. Zahn, W. Richter, T. Eickhoff, J. Geurts, T. D. Golding, J. H. Dinan, K. J. MacKey, and R. H. Williams, *Appl. Surf. Sci.* 41, 497 (1989).
66. V. Wagner, W. Richter, and J. Geurts, *Appl. Surf. Sci.* 104, 580 (1996).
67. B. Jusserand and M. Cardona, "Light Scattering in Solids V," Topics in Applied Physics, Vol. 66, p. 49. Springer, Berlin/Heidelberg/New York, 1989.
68. E. H. Rhoderick and R. H. Williams, "Metal Semiconductor Contacts." Clarendon, Oxford, 1988.
69. M. Cardona and L. Ley, "Photoemission in Solids," Topics in Applied Physics, Vol. 26, p. 1, Springer, Berlin/Heidelberg/New York, 1978.
70. A. Pinczuk, A. A. Ballman, R. E. Nahory, M. A. Pollack, and J. M. Worlock, *J. Vac. Sci. Technol.* 16, 1168 (1979).
71. D. J. Olego, *Appl. Phys. Lett.* 51, 1422 (1987).
72. D. J. Olego, *Phys. Rev. B* 39, 12743 (1989).
73. J. Sapriel and B. D. Rouhani, *Surf. Sci. Rep.* 10, 189 (1989).
74. M. V. Klein, *IEEE J. Quantum Electron.* 22, 1760 (1986).
75. B. Jusserand and D. Paquet, "Semiconductor Heterojunctions and Superlattices," p. 108. Springer, Berlin/Heidelberg/New York, 1986.
76. M. Cardona, "Lectures of Surface Science," p. 2. Springer, Berlin/Heidelberg/New York, 1987.
77. R. Enderlein, D. Suisky, and J. Röseler, *Phys. Status Solidi B* 165, 9 (1991).
78. A. Fasolino and E. Molinari, *Surf. Sci.* 228, 112 (1990).
79. G. Kanellis, J. F. Morhange, and M. Balkanski, *Phys. Rev. B* 28, 3406 (1983).
80. V. Wagner and J. Geurts, *Phys. Status Solidi A* 184, 29 (2001).
81. V. Wagner, M. Becker, M. Weber, M. Korn, M. Keim, A. Waag, and J. Geurts, *Appl. Surf. Sci.* 166, 30 (2000).
82. V. Wagner, J. Geurts, T. Gerhard, T. Litz, H.-J. Lugauer, A. Waag, G. Landwehr, T. Walter, and D. Gerthsen, *Appl. Surf. Sci.* 123, 580 (1998).
83. A. Dinger, M. Hetterich, M. Göppert, M. Grün, C. Klingshirn, B. Weise, J. J. Liang, V. Wagner, and J. Geurts, *J. Cryst. Growth* 200, 391 (1999).

84. M. Göppert, M. Hetterich, A. Dinger, C. Klingshirn, and K. O. Donnel, *Phys. Rev. B* 57, 13068 (1998).
85. A. Dinger, R. Becker, M. Göppert, S. Petillon, M. Hetterich, M. Grün, C. Klingshirn, J. J. Liang, B. Weise, V. Wagner, and J. Geurts, *Phys. Status Solidi B* 215, 413 (1999).
86. S. M. Rytov, *Sov. Phys. Acoust.* 2, 68 (1956).
87. V. Wagner, S. Gundel, J. Geurts, T. Litz, H.-J. Lugauer, A. Waag, G. Landwehr, R. Kruse, and C. Becker, *J. Cryst. Growth* 184, 1067 (1998).
88. A. P. Alivisatos, T. D. Harris, L. E. Brus, and A. Jayaraman, *J. Chem. Phys.* 89, 5979 (1988).
89. A. Tanaka, S. Onari, and T. Arai, *Phys. Rev. B* 45, 6587 (1992).
90. H. Richter, L. Ley, and Z. P. Wang, *Solid State Comm.* 39, 625 (1981).
91. R. Rossetti, S. Nakahara, and L. E. Brus, *J. Chem. Phys.* 79, 1086 (1983).
92. A. Tanaka, S. Onari, and T. Arai, *J. Phys. Soc. Japan* 61, 4222 (1992).
93. Y. N. Hwang, S. H. Park, and D. Kim, *Phys. Rev. B* 59, 7285 (1999).
94. I. H. Campbell and P. M. Fauchet, *Solid State Comm.* 58, 739 (1986).
95. M. C. Klein, F. Hache, D. Ricard, and C. Flytzanis, *Phys. Rev. B* 42, 11123 (1990).
96. D. J. Tannor and E. J. Heller, *J. Chem. Phys.* 77, 202 (1982).
97. R. Merlin, G. Güntherodt, R. Humphreys, M. Cardona, R. Suryanarayanan, and F. Holtzberg, *Phys. Rev. B* 17, 4951 (1978).
98. T. Baron, K. Saminadayov, S. Tatarenko, H. J. Lugauer, and A. Waag, *J. Cryst. Growth* 185, 415 (1998).
99. U. Venkateswaran, M. Chandrasekhar, and H. R. Chandrasekhar, *Phys. Rev. B* 30, 3316 (1984).
100. A. I. Ekimov, F. Hache, M. C. Schanne-Klein, D. Ricard, C. Flytzanis, I. A. Kudriavtsev, T. V. Yazewa, A. V. Rodina, and A. L. Efros, *J. Opt. Soc. Am. B* 10, 100 (1993).
101. J. Puls, V. Jungnickel, F. Henneberger, and A. Schülzgen, *J. Cryst. Growth* 138, 1004 (1994).
102. A. M. de Paula, L. C. Barbosa, C. H. B. Cruz, O. L. Alves, J. A. Sanjurjo, and C. L. Cesar, *Appl. Phys. Lett.* 69, 357 (1996).
103. A. V. Baranov, Y. S. Bobovich, and V. I. Petrov, *J. Raman Spectrosc.* 24, 767 (1993).
104. A. Tamura, K. Higeta, and T. Ichinokawa, *J. Phys. C* 15, 4975 (1982).
105. F. Viras and T. A. King, *J. Noncryst. Solids* 119, 65 (1990).
106. E. Duval, *Phys. Rev. B* 46, 5795 (1992).
107. E. Duval, A. Boukenter, and B. Champagnon, *Phys. Rev. Lett* 56, 2052 (1986).
108. B. Champagnon, B. Adrianasolo, A. Ramos, M. Gandais, M. Allais, and J.-B. Bennoit, *J. Appl. Phys.* 73, 2775 (1993).
109. B. Champagnon, B. Adrianasolo, and E. Duval, *J. Chem. Phys.* 94, 5237 (1991).
110. J. Menendez and M. Cardona, *Phys. Rev. B* 31, 3696 (1985).
111. T. Bischof, M. Ivanda, G. Lermann, A. Materny, and W. Kiefer, *J. Raman Spectrosc.* 27, 297 (1996).
112. B. Schreder, T. Schmidt, V. Ptatschek, L. Spanhel, A. Materny, and W. Kiefer, *J. Cryst. Growth* 214, 782 (2000).
113. B. Schreder, T. Schmidt, V. Ptatschek, U. Winkler, A. Materny, E. Umbach, M. Lerch, G. Müller, W. Kiefer, and L. Spanhel, *J. Phys. Chem. B* 104, 1677 (2000).
114. R. G. Alonso, E. K. Suh, A. K. Ramdas, N. Samarth, H. Luo, and J. K. Furdyna, *Phys. Rev. B* 40, 3720 (1989).
115. I. E. Trofimov, M. V. Petrov, F. F. Balakirev, A. E. Milokhin, and V. Kuzḿin, *JETP Lett.* 55, 445 (1992).
116. X. S. Zhao, J. Schroeder, T. G. Bilodeau, and L.-G. Hwa, *Phys. Rev. B* 40, 1257 (1989).
117. O. Brafman and S. S. Mitra, "Proceedings of the 2nd International Conference on Light Scattering in Solids," p. 284. Flammarion Sciences, Paris, 1971.
118. R. J. Briggs and A. K. Ramdas, *Phys. Rev. B* 13, 5518 (1976).
119. J. Schroeder, M. Silvestri, X. Zhao, P. Persans, and L. Hwang, *Mater. Res. Soc. Symp. Proc.* 272, 5888 (1992).
120. A. L. Edwards and H. G. Drickamer, *Phys. Rev.* 122, 1149 (1961).
121. N. B. Owen, P. L. Smith, J. E. Martin, and A. J. Wright, *J. Phys. Chem. Solids* 24, 1519 (1963).
122. C. Solliard, *Solid State Comm.* 51, 947 (1984).
123. A. P. Alivisatos, T. D. Harris, P. J. Carroll, M. L. Steigerwald, and L. E. Brus, *J. Chem. Phys.* 90, 463 (1989).
124. A. P. Alivisatos, T. D. Harris, N. J. Levinos, M. L. Steigerwald, and L. E. Brus, *J. Chem. Phys.* 89, 4001 (1988).
125. J. J. Shiang, A. N. Goldstein, and A. P. Alivisatos, *J. Chem. Phys.* 92, 3232 (1990).
126. J. J. Shiang, I. M. Craig, and A. P. Alivisatos, *Z. Phys. D* 26, 358 (1993).
127. J. F. Scott, R. C. C. Leite, and T. C. Damen, *Phys. Rev.* 188, 1285 (1969).
128. M. Nirmal, C. B. Murray, D. J. Norris, and M. G. Bawendi, *Z. Phys. D* 26, 361 (1993).
129. S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla, *Phys. Rev. B* 35, 8113 (1987).
130. S. Nomura and T. Kobayashi, *Phys. Rev. B* 45, 1305 (1992).
131. V. L. Colvin and A. P. Alivisatos, *J. Chem. Phys.* 97, 730 (1992).
132. A. V. Fedorov, A. V. Baranov, and K. Inoue, *Phys. Rev. B* 56, 7491 (1997).
133. A. l. Efros, A. I. Ekimov, F. Kozlowski, V. Petrova-Koch, H. Schmidbaur, and S. Shumilov, *Solid State Comm.* 78, 853 (1991).
134. S. Hayashi and H. Kanamori, *Phys. Rev. B* 26, 7079 (1982).
135. V. M. Fomin, E. P. Pokatilov, J. T. Devreese, S. N. Klimin, V. N. Gladilin, and S. N. Balaban, *Phys. Status Solidi A* 164, 417 (1997).
136. M. Cardona, "Light Scattering in Solids II," Topics in Applied Physics, Vol. 50. Springer, Berlin, 1983.
137. S. Bauer, H. Berger, P. Link, and W. Gebhardt, *J. Appl. Phys.* 74, 3916 (1993).
138. W. Bala, M. Drozdowski, and M. Kozielski, *Acta Phys. Polon. A* 84, 689 (1993).
139. T. Karasawa, K. Ohkawa, and T. Mitsuyu, *J. Cryst. Growth* 99, 464 (1990).
140. H. Ozaki, D. Suzuki, K. Imai, and K. Kumazaki, *Phys. Status Solidi A* 133, 523 (1992).
141. W. Bala, M. Drozdowski, and M. Kozielski, *Phys. Status Solidi A* 130, K195 (1992).
142. X. Feng and J. P. Hirth, *J. Appl. Phys.* 72, 1586 (1992).
143. J. P. Hirth and X. Feng, *J. Appl. Phys.* 67, 3343 (1990).
144. E. O'Reilly, *Semicond. Sci. Technol.* 4, 121 (1989).
145. D. J. Olego, K. Shazad, J. Petruzello, and D. Cammack, *Phys. Rev. B* 36, 7674 (1987).
146. S. Nakashima, N. Nakakura, H. Fujiyasu, and K. Mochizuki, *Appl. Phys. Lett.* 48, 236 (1986).
147. C. D. Lee, B. K. Kim, J. W. Kim, H. L. Park, C. H. Chung, S. K. Chang, J. I. Lee, and S. K. Noh, *J. Cryst. Growth* 138, 136 (1994).
148. D. J. Olego, K. Shazad, D. Cammack, and H. Cornelissen, *Phys. Rev. B* 38, 5554 (1988).
149. N. Pelekanos, Q. Fu, J. Ding, W. Walecki, A. V. Nurmikko, S. M. Durbin, J. Han, M. Kobayashi, and R. L. Gunshor, *Phys. Rev. B* 41, 9966 (1990).
150. J. Nieto-Jalil and R. Perez-Alvarez, *Phys. Status Solidi A* 164, 699 (1997).
151. M. Kobayashi, M. Konagai, K. Takahashi, and K. Urabe, *J. Appl. Phys.* 61, 1015 (1987).
152. G. C. Osbourn, *J. Appl. Phys.* 53, 1586 (1982).
153. J. W. Matthews and A. E. Blakeslee, *J. Cryst. Growth* 32, 265 (1976).
154. B. Jusserand, P. Voisin, M. Voos, L. L. Chang, E. E. Mendez, and L. Esaki, *Appl. Phys. Lett.* 46, 678 (1985).
155. K. Kumazaki, K. Imai, and A. Odajima, *Phys. Status Solidi A* 119, 177 (1990).

156. S. Nakashima, A. Fujii, K. Mizoguchi, A. Mitsuishi, and K. Yoneda, *Japan. J. Appl. Phys.* 27, 1327 (1988).
157. M. Longo, N. Lovergine, A. M. Mancini, A. Passareo, G. Leo, M. Mazzer, M. Berti, and A. V. Drizo, *Appl. Phys. Lett.* 72, 359 (1998).
158. H. C. Ko and S. Fujita, *Japan. J. Appl. Phys.* 37, 7177 (1998).
159. C. Meyne, U. W. Pohl, J. T. Zettler, and W. Richter, *J. Cryst. Growth* 185, 264 (1998).
160. H. C. Ko, Y. Kawakami, S. Fujita, and S. Fujita, *J. Cryst. Growth* 185, 283 (1998).
161. G. Leo, M. Longo, N. Lovergine, M. Mazzer, A. M. Mancini, M. Berti, and A. V. Drizo, *J. Cryst. Growth* 185, 1332 (1998).
162. A. Krost, W. Richter, D. R. T. Zahn, and O. Brafman, *Semicond. Sci. Technol.* 6, A109 (1991).
163. E. Anastassakis, *J. Appl. Phys.* 68, 4561 (1990).
164. B. Schreder, A. Materny, W. Kiefer, T. Kümmell, G. Bacher, A. Forchel, and G. Landwehr, *J. Appl. Phys.* 88, 764 (2000).
165. B. Schreder, A. Materny, W. Kiefer, T. Kümmell, G. Bacher, A. Forchel, and G. Landwehr, *Solid State Comm.* 114, 435 (2000).
166. B. Schreder, T. Kümmell, G. Bacher, A. Forchel, G. Landwehr, A. Materny, and W. Kiefer, *J. Cryst. Growth* 214, 787 (2000).
167. G. Lermann, T. Bischof, B. Schreder, A. Materny, W. Kiefer, T. Kümmell, G. Bacher, A. Forchel, and G. Landwehr, *Ber. Bunsenges. Phys. Chem.* 101, 1665 (1997).
168. G. Lermann, T. Bischof, A. Materny, W. Kiefer, T. Kümmell, G. Bacher, A. Forchel, and G. Landwehr, *Phys. Rev. B* 56, 7469 (1997).
169. I. H. Tan, R. Mirin, T. Yasuda, E. L. Hu, J. Browers, C. B. Prater, P. K. Hansma, M. Y. He, and A. G. Evans, *J. Vac. Sci. Technol. B* 10, 1971 (1992).
170. M. Notomi, J. Hammersberg, H. Wemau, S. Nojima, H. Sugiura, M. Okamoto, T. Tamamura, and M. Potemski, *Phys. Rev. B* 52, 11147 (1995).
171. Y. M. Niquet, C. Priester, C. Gourgon, and H. Mariette, *Phys. Rev. B* 57, 14850 (1998).
172. G. Lermann, T. Bischof, A. Materny, W. Kiefer, T. Kümmell, G. Bacher, A. Forchel, and G. Landwehr, *J. Appl. Phys.* 81, 1446 (1997).
173. B. Schreder, T. Kümmell, G. Bacher, A. Forchel, G. Landwehr, A. Materny, and W. Kiefer, *J. Cryst. Growth* 214, 792 (2000).
174. B. Schreder, A. Materny, W. Kiefer, G. Bacher, A. Forchel, and G. Landwehr, *J. Raman Spectrosc.* 31, 959 (2000).
175. W. Heiss, G. Prechtl, D. Stifler, H. Sitter, G. Springholz, T. Riemann, F. Bertram, D. Rudloff, J. Christen, G. Bley, U. Neukirch, J. Gutowski, and L. Toth, *Phys. E* 7, 526 (2000).
176. K. Herz, G. Bacher, A. Forchel, H. Straub, G. Brunthaler, W. Faschinger, G. Bauer, and C. Vieu, *Phys. Rev. B* 59, 2888 (1999).
177. M. Tamura, T. Ando, N. Nunoya, S. Tamura, S. Arai, and G. U. Bacher, *Japan. J. Appl. Phys.* 37, 3576 (1998).
178. M. Strassburg, T. Deniozou, A. Hoffmann, R. Heitz, U. W. Pohl, D. Bimberg, D. Litvinov, A. Rosenauer, D. Gerthsen, S. Schwedhelm, K. Lischka, and D. Schikora, *Appl. Phys. Lett.* 76, 685 (2000).
179. A. Kaschner, M. Strassburg, A. Hoffmann, C. Thomsen, M. Bartels, K. Lischka, and D. Schikora, *Appl. Phys. Lett.* 76, 2662 (2000).
180. V. F. Sapega, T. Ruf, M. Cardona, K. Ploog, E. L. Ivchenko, and D. N. Mirlin, *Phys. Rev. B* 45, 4320 (1992).
181. V. F. Sapega, T. Ruf, M. Cardona, K. Ploog, E. L. Ivchenko, and D. N. Mirlin, *Phys. Rev. B* 50, 2510 (1992).
182. J. J. Davies, D. Wolverson, O. Z. Karimov, and I. J. Griffin, *J. Cryst. Growth* 214, 616 (2000).
183. A. A. Sirenko, T. Ruf, N. N. Ledentsov, A. Y. Egorov, P. S. Kopev, V. M. Ustinov, and A. E. Zhukov, *Solid State Comm.* 97, 169 (1998).
184. A. Sirenko, V. Belitsky, T. Ruf, M. Cardona, A. Ekimov, and C. Trallero-Giner, *Phys. Rev. B* 58, 2077 (1998).
185. O. Z. Karimov, D. Wolverson, J. J. Davies, T. Ruf, and L. N. Tenishev, *Phys. Status Solidi B* 215, 373 (1999).
186. T. Ruf, O. Z. Karimov, D. Wolverson, J. J. Davies, A. N. Reznitsky, A. A. Klochikhin, S. Y. Verbin, S. A. Permogorov, and S. V. Ivanov, *Phys. B* 273, 911 (1999).

Raman Spectroscopy in Carbon Nanotubes

M. S. Dresselhaus
Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

A. M. Rao
Clemson University, Clemson, South Carolina, USA

G. Dresselhaus
Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

CONTENTS

1. Introduction
2. Spectra from Isolated Single Wall Carbon Nanotubes
3. Single Wall Carbon Nanotube Bundles
4. Peapods
5. Double Walled Carbon Nanotubes
6. Connection to Multiwall Nanotubes
7. Future Directions
Glossary
References

1. INTRODUCTION

A review of the Raman-active modes in single walled carbon nanotubes is presented both from experimental and theoretical standpoints. A discussion is first given of the Raman spectroscopy of isolated tubes which serve as a building block to understand the Raman spectroscopy of tubes that reside in bundles or in solution, and of nanotubes when subjected to external pressure, high temperature, applied potentials, and interstitial or endohedral doping. Using different nanotube systems, such as bundled tubes, double walled tubes, multiwalled tubes, and peapods, an attempt is made to elucidate the influence of van der Waals forces on the fundamental Raman scattering properties of nanotubes.

Raman spectroscopy has served as a sensitive tool for characterizing new forms of carbon by providing information on structural, vibrational, and electronic properties. The chapter starts with a discussion of Raman spectroscopy of an individual single wall carbon nanotube (SWNT) and how such studies provide direct structural information. The Raman spectra obtained from isolated SWNTs are mainly used here to provide detailed information about each of the features in the Raman spectra and about the resonance Raman process responsible for their observation (Section 2). The spectra thus obtained are used to interpret the Raman spectra observed in SWNTs that interact with each other or with an external environment and especially to illuminate the influence of the van der Waals interaction on the Raman scattering properties of nanotubes. The effect on the Raman spectra of nanotubes coupling to each other as they assemble into bundles of SWNTs (Section 3.1) or in multiwall nanotubes (Section 6) is explicitly considered, as well as the effect of subsequent debundling of the SWNTs into small SWNT bundles and even into individual SWNTs by chemical means (Section 3.3). The effect on the Raman spectra of the nanotubes, when the tubes are subjected to external pressure (Section 3.4), high temperature (Section 3.2), applied potentials (Section 3.3), and interstitial doping (Section 3.3), is also reviewed. The Raman spectra of C_{60} encapsulated single-walled tubes (peapods) (Section 4) and of double-walled nanotubes (Section 5) are given particular attention as model systems for studying the coupling of one nanotube to another and of a nanotube to a molecular endohedral species because of the relative simplicity and theoretical tractability of these systems.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (307–338)

2. SPECTRA FROM ISOLATED SINGLE WALL CARBON NANOTUBES

In this section the Raman spectra from isolated SWNTs are presented in the context of the use of Raman spectroscopy to infer information about the coupling of the SWNT basic building block to form other nanotube-based structures.

In Section 2.1 the basic Raman nanotube spectra are presented at the single nanotube level, while Section 2.2 provides pertinent background for understanding the *unusual* aspects of Raman spectroscopy of SWNTs at the single nanotube level. A review of the use of single nanotube spectroscopy to provide the structural (n, m) characterization of SWNTs is presented in Section 2.3 and the connection between Raman spectra observed at the single nanotube level and in SWNT bundles is given in Section 2.4.

2.1. Single Nanotube Spectroscopy

Figure 1 shows typical Raman spectra from an isolated semiconducting and a metallic nanotube on a Si/SiO_2 substrate. The four most important features seen in Figure 1 (and also seen in the spectra for SWNT bundles) are the radial breathing mode where the carbon atoms are all vibrating in phase in an A symmetry mode in the radial direction of the nanotube, the tangential G-band (derived from the graphitelike in-plane mode), the disorder-induced D-band, and its second-order harmonic (the G'-band) [1]. Table 1 provides a summary of the most important spectral features in Figure 1.

The isolated SWNTs used in these single nanotube Raman experiments were prepared by a chemical vapor deposition method on a Si substrate containing nanometer size iron catalyst particles [3]. Since the silicon substrate is oxidized to form a thin SiO_2 surface coating, no significant charge transfer is expected between the SWNTs and the substrate. Since the nanotubes nucleate and grow from well isolated nanometer size catalyst particles, nanotube bundles are not formed. Figure 2b shows an atomic force microscopy (AFM) image of a sample with a high SWNT concentration

Table 1. Vibrational modes observed for Raman scattering in isolated SWNTs.

Notation	Frequency (cm^{-1})	Symmetry	Type of mode
RBM	$248/d_t$	A	in phase radial displacements
D-band	$\sim$1350	—	defect induced, dispersive
G-band	1550–1605	A, E_1, E_2	graphite-related optical mode[a]
G'-band	$\sim$2700	—	overtone of D-band, highly dispersive

Note: Frequencies for the dispersive D-band and G'-band are given for laser excitation at 2.41 eV (514 nm).

[a] The related graphite mode has E_{2g} symmetry.

(6 ± 3 SWNTs per μm^2) for easy visualization. The inset to Figure 2b shows the wide diameter distribution of the SWNT samples ($\sim 1 < d_t < \sim 3$ nm) typically used in single nanotube Raman studies. However, the actual samples used for single nanotube spectroscopy studies have a very low nanotube density, less than 1 SWNT/μm^2, a factor of 10 more dilute than that shown in Figure 2b.

Resonance Raman spectra in the 100 to $\sim$3000 cm^{-1} spectral range from these individual isolated SWNTs are taken with a readily available Raman microprobe instrument, such as a Renishaw Raman microprobe (1 μm laser spot), using $\sim$1 mW laser power, and typical laser excitation, such as $E_{laser} = 514.5\,nm = 2.41\,eV$. The Raman spectra shown in Figure 1 were taken with a Kaiser micro-Raman spectrograph, operating at a laser excitation energy of 1.58 eV. Relatively high laser powers can be used to probe isolated SWNTs because of their unusually high thermal conductivity [4], their excellent high temperature stability, and their good thermal contact to the substrate.

It is possible to observe the spectrum for one isolated carbon nanotube, only $\sim$1 nm in diameter, when the incident photon energy is close to the energy of a transition

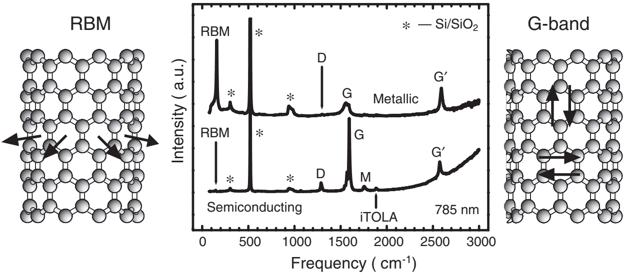
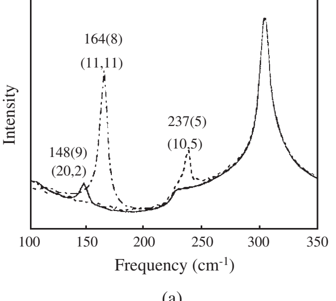


Figure 1. Raman spectra from a metallic (top) and a semiconducting (bottom) SWNT at the single nanotube level using 785 nm (1.58 eV) laser excitation, showing the radial breathing mode (RBM), the D-band, G-band, and G' band features, in addition to weak double resonance features associated with the M-band and the iTOLA feature. Also shown are the mode displacements associated with the RBM (left) and G-band (right) normal mode vibrations. The isolated carbon nanotubes (see Fig. 2b) are sitting on an oxidized silicon substrate, which provides contributions to the Raman spectra denoted by an asterisk, and these Roman features are used for calibration purposes.



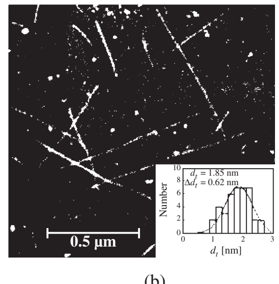


Figure 2. (a) The three Raman spectra (solid, dashed, and dash–dotted curves) come from three different spots on the Si substrate, showing the presence of only one resonant nanotube and one RBM frequency for each of the three laser spots. The RBM frequencies (widths) and the (n, m) assignments for each resonant SWNT are displayed. The 303 cm^{-1} feature comes from the Si substrate and is used for calibration purposes (see Fig. 1) [2]. (b) AFM image of the sample showing isolated single wall nanotubes grown from the vapor phase. The small particles are iron catalyst particles. The inset shows the diameter distribution of this sample ($d_t = 1.85 \pm 0.62$ nm) based on AFM observations of 40 SWNTs. Reprinted with permission from [2], A. Jorio et al., *Phys. Rev. Lett.* 86, 1118 (2001). © 2001, American Physical Society.

between van Hove singularities in the one-dimensional (1D) density of states of the valence and conduction bands, or equivalently a singularity in the joint density of states (JDOS). Under these resonance conditions, the intensity of the Raman signal is strongly increased and becomes comparable to that seen from the silicon substrate (marked by an asterisk in Fig. 1). The enhancement of the resonance process can be seen more clearly in Figure 2a which shows that under favorable resonance Raman conditions, the RBM intensity from only one 1D carbon nanotube can be comparable to that of the silicon spectral features from the quasi-3D Si film substrate (where the Si/C atom ratio within the optical beam is perhaps $\sim 10^6$). In Figure 2a, the resonance of three different isolated SWNTs with the same laser line (1.58 eV or 785 nm), some having their singularity in the joint density of states E_{ii} closer in energy to the laser excitation energy E_{laser}, and some further from E_{laser}. Also determining the Raman intensity is the length of the nanotube, the laser light intensity on the nanotube, and the angle of the optical E field of the laser and the nanotube axis.

Since the energy of these singularities in the JDOS depend strongly on the nanotube diameter, the observation of the Raman spectra using a specific laser excitation energy is highly diameter-selective, as was already pointed out in the first observation of the Raman spectra from single wall carbon nanotubes [5]. The strong dependence on nanotube diameter d_t arises from the proportionality between the nanotube diameter and the number of carbon atoms around the tube circumference [6]. Since the number of allowed wave vectors in 1D relates to the number of carbon atoms around the circumference, the separation between the allowed wave vectors is proportional to $1/d_t$. In addition to the strong diameter selectivity, there is also a weaker selectivity which arises from the chirality of the nanotube and is due to the trigonal warping effect for nanotubes. This trigonal warping effect arises from the corresponding effect in the constant energy surfaces of a 2D graphene sheet shown in Figure 3b. Because of this trigonal warping effect, there is a weak dependence of the electronic energy levels of SWNTs on chiral angle [6], as will be explained below. To gain an understanding of these issues, some background about the geometrical and electronic structure of single wall carbon nanotubes is reviewed.

2.2. Pertinent Background about Carbon Nanotubes

The structure of each nanotube is uniquely described by two integers (n, m), which refer to the number of $\vec{a}_1$ and $\vec{a}_2$ unit vectors of the graphene lattice that are contained in the chiral vector, $\vec{C}_h = n\vec{a}_1 + m\vec{a}_2$, which spans the circumference of each nanotube [6]. From the (n, m) indices, one can calculate the nanotube diameter d_t, the chiral angle θ, the electronic energy bands and the density of electronic states [6, 9], as well as the energies of the interband transitions $E_{ii}(n, m)$ between the van Hove singularities in the valence and conduction band density of states. It is well established [6] that the (n, m) indices are crucial to the nanotube electronic structure. Specifically, SWNTs for which $|n - m| = 3q$ are metallic, and those for which $|n - m| = 3q \pm 1$ are semiconducting, where q is an integer [6, 10]. The Raman spectra

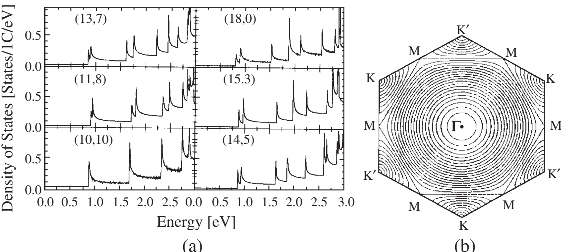


Figure 3. (a) The 1D electronic density of states vs energy (measured from the Fermi level) for several metallic nanotubes of approximately the same diameter, showing the effect of chirality and trigonal warping (see text) on the van Hove singularities in the density of states: (10,10) (armchair), (11,8), (13,7), (14,5), (15,3), and (18,0) (zigzag). Only the densities of states for the conduction π^* band are shown; the mirror image of these plots gives the electronic density of states for the valence π band [7, 8] within the tight binding approximation, assuming the energy overlap integral is $\gamma_0 = 2.9$ eV and the wave function overlap integral vanishes, $s = 0$ [6]. (b) Plot of the 2D equienergy contours of 2D graphite, showing trigonal warping effects in the contours, as we move from the K point in the K–Γ or K–M directions in the 2D graphene Brillouin zone. The equienergy contours are circles near the K point and near the center of the Brillouin zone (Γ). But near the M points on the zone boundary, the contours are straight lines which connect the nearest M points. Reprinted with permission from [7], R. Saito et al., *Phys. Rev. B* 61, 2981 (2000). © 2000, American Physical Society.

of the G-band for metallic and semiconducting nanotubes are qualitatively different from each other, as can be seen in Figure 1 at the single nanotube level, and this difference in lineshape is even more pronounced in SWNT bundles as discussed in Section 3.1. This difference in lineshape can be used to distinguish between metallic and semiconducting nanotubes.

The weak dependence of the electronic energy bands of SWNTs on chiral angle θ relates to the three-fold symmetry of the electronic dispersion relations about the K point of the 2D graphene Brillouin zone (see Fig. 3b), where the graphene valence and conduction bands are degenerate, forming a zero bandgap semiconductor for a 2D graphene sheet (a single basal plane layer of the graphite lattice) [6]. Close to the zone center Γ point, the K point and the K' point (which is related to K by time reversal symmetry), the constant energy contours are circular. But away from the K point, the graphene electronic energy bands exhibit a trigonal warping effect [7, 11–13], as shown in Figure 3b, and this trigonal warping is responsible for the unique energy bands for the conduction and valence bands for each (n, m) SWNT, when the energy bands of the graphene lattice are zone folded to form the energy bands of the SWNTs [7].

The locations of the van Hove singularities for a particular nanotube are found by superimposing the set of *allowed* parallel k-vectors (called cutting lines) on the constant energy contours and looking for the k points where the energy extrema occur. The chirality dependence of the van Hove singularities [7, 8] in the 1D electronic DOS of the conduction band are shown in Figure 3a for several *metallic* (n, m) nanotubes, all having about the same diameter d_t (from 1.31 to 1.43 nm), but having different chiral angles: $\theta =$ 0°, 8.9°, 14.7°, 20.2°, 24.8°, and 30.0° for nanotubes (18,0), (15,3), (14,5), (13,7), (11,8), and (10,10), respectively. Here the energy is measured relative to the Fermi energy, which

is taken to be at $E = 0$. The plots in Figure 3a show that as the chiral angle is varied from the armchair nanotube (10,10) ($\theta = 30°$) to the zigzag nanotube (18,0) ($\theta = 0°$), a splitting of each van Hove singularity, due to the trigonal warping effect, develops in the DOS traces [7, 8], and this splitting increases from zero for $\theta = 30°$ to a maximum for $\theta = 0°$. Because of the dependence of the spacing between cutting lines on $1/d_t$ and because of the trigonal warping effect shown in Figure 3a, every distinct carbon nanotube, denoted by its (n, m) indices, or equivalently by its diameter d_t and chiral angle θ, will have a *unique* set of singularities E_{ii} [2].

This trigonal warping effect consequently causes a spread of the interband energies E_{ii} between the singularities in the JDOS for nanotubes with the same diameter d_t and index i, as shown in Figure 4. Here the integer i labels the singularities in the JDOS, and i increases as the energy relative to the Fermi level E_F increases. As a result of the trigonal warping effect, Figure 4 shows that each nanotube (n, m) has a unique set of interband energies E_{ii} denoting the energy differences between the ith van Hove singularities in the conduction and valence bands. And, conversely, if one interband energy E_{ii} and a nanotube diameter d_t in Figure 4 are specified, then the corresponding unique (n, m) for this nanotube can be identified [2].

2.3. Structural (n, m) Characterization through Raman Spectroscopy

This fact leads to what is special about the resonance Raman effect in *isolated* SWNTs, and this is the structural information that is provided by the single nanotube spectra. Normally the resonance Raman effect just gives spectral information, or information about phonon frequencies, electronic energy levels, and the electron–phonon interaction, but not structural information directly. But in the resonance Raman process for SWNTs, every (n, m) nanotube is in resonance with a unique set of interband energies $E_{ii}(n, m)$,

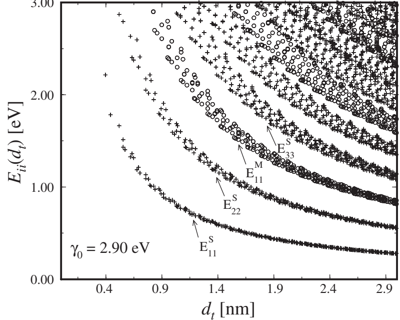


Figure 4. Calculated [14] energy separations E_{ii} between van Hove singularities i in the 1D electronic density of states of the conduction and valence bands for all (n, m) values vs nanotube diameter in the range $0.4 < d_t < 3.0$ nm, using a value for the carbon–carbon energy overlap integral of $\gamma_0 = 2.9$ eV and a nearest neighbor carbon–carbon distance $a_{C-C} = 1.42$ Å [7, 15]. Semiconducting (S) and metallic (M) nanotubes are indicated by crosses and open circles, respectively. The subscript $i = 1$ denotes the index of the lowest energy of a singularity in the joint density of states and the Fermi level for the isolated SWNT is taken at $E = 0$. Reprinted with permission from [14], H. Kataura et al., *Synthetic Metals* 103, 2555 (1999). © 1999, Elsevier Science.

as shown in Figure 4, so that for this 1D system, resonant Raman spectra can also yield structural information through identification of the (n, m) values that correspond to the observed Raman spectra for a given isolated SWNT. Once (n, m) is known from measurements made on the radial breathing mode feature, Raman studies on isolated carbon nanotubes can be carried out to understand in detail the dependence of *all* the features shown in the spectrum of Figure 1 on diameter d_t, chiral angle θ, laser excitation energy E_{laser}, and other pertinent parameters. Furthermore, the spectra observed for these other features in Figure 1 are also sensitive to d_t and θ and can be used to corroborate the (n, m) assignments.

From knowledge of the characteristics of each feature in the Raman spectra at the single nanotube level, the spectrum of SWNT bundles can be elucidated, and by comparison between the spectra measured on SWNT bundles and for isolated SWNTs, the effect of nanotube–nanotube interactions can be deduced. Resonance Raman characterization of the (n, m) indices for one nanotube is important for scientific studies of SWNTs, insofar as many physical properties, such as transport, optical, mechanical, and other properties, that depend sensitively on the (n, m) indices [6], can be studied systematically at the single nanotube level through use of the resonance Raman effect to identify their (n, m) values. The (n, m) characterization provided by resonance Raman spectroscopy [2] is a general, rapid (a few minutes per spectrum), nondestructive technique that operates under ambient temperature ($\sim$300 K) and pressure conditions and uses readily available Raman characterization instrumentation.

Since the frequency of the RBM feature varies as $1/d_t$

$$\omega_{\text{RBM}} = \alpha/d_t \tag{1}$$

the RBM mode provides a convenient tool for the determination of the tube diameter which is one of the parameters needed for the (n, m) characterization (see Fig. 4). For the Si/SiO_2 substrate (see Fig. 2b), the value for α in Eq. (1) is experimentally found to be 248 cm^{-1} nm by measurement of the RBM spectral feature for a large number of isolated SWNTs [2]. Furthermore, ω_{RBM} is observed to be independent of chiral angle θ. *Ab initio* calculations for isolated SWNTs yield a value of $\alpha = 234$ cm^{-1} nm [16], which is within 5% of the experimental value of 248 cm^{-1} nm.

The determination of (n, m) by resonance Raman scattering depends on the determination of E_{ii} and d_t by exploiting the unique relation between E_{ii} and (n, m) shown in Figure 4 and the direct determination of d_t from the radial breathing mode relation $\omega_{\text{RBM}} = 248/d_t$. Although the determination of E_{ii} is most conveniently carried out by measurement of the radial breathing mode feature, other features (such as the G-band, D-band, and G'-band) in the resonance Raman spectrum (see Fig. 1) are sensitive to d_t and can be used to confirm the (n, m) assignment made from analysis of the RBM spectrum. If a tunable laser were available to be tuned to E_{ii} (which could be detected as the E_{laser} value where the maximum intensity in the Raman spectrum occurs), then a simple measurement of ω_{RBM} to yield d_t and to identify the index i in E_{ii} (see Fig. 4) would

be sufficient to determine (n, m). In practice, such a tunable laser system has thus far been applied only once to single nanotube Raman spectroscopy studies and over a small energy range (~0.15 eV) [17], and this is in part due to the fact that only a few laboratories worldwide have available tunable laser systems working over a broad energy range in conjunction with a Raman microprobe system.

Nevertheless, an (n, m) determination can in most cases be made if a nanotube is within the resonant window of a single available laser excitation line, which in practice is satisfied for E_{laser} within $\sim\pm 0.1$ eV of an interband transition for that nanotube. From the measured ω_{RBM} and use of Eq. (1), the nanotube diameter d_t is found, which determines the i index in E_{ii} from Figure 4. In many cases of practical interest, there is only one (n, m) value that has both a diameter close to the measured d_t value, as determined from the ω_{RBM} measurement, and an E_{ii} value within a few meV of E_{laser}.

One method to determine E_{ii} directly requires measurement of both the Stokes and anti-Stokes radial breathing mode spectra [18, 19], where the ratio of the anti-Stokes to Stokes intensities for the radial breathing mode for a given resonant nanotube is used to sensitively (to 10 meV) determine the energy E_{ii} of the resonant van Hove singularity in the joint density of states [18], thereby providing a unique identification of the (n, m) indices. When the anti-Stokes to Stokes intensities are exactly equal to the Boltzmann factor at the measurement temperature for ω_{RBM}, as in Figure 5 of [1], then E_{laser} is in exact resonance with E_{ii} (i.e., $E_{ii} = E_{\text{laser}}$), resulting in very large intensities for all of the features in the Raman spectra. In general, E_{ii} is not exactly equal to E_{laser} ($E_{ii} \neq E_{\text{laser}}$), resulting in a ratio of the intensities of the anti-Stokes to Stokes signals that is not equal to the Boltzmann factor but rather is modified by how well the resonance condition is satisfied. We use this effect to determine the value of E_{ii}.

In Figure 5 we see experimental traces for the Stokes and anti-Stokes spectra for two different isolated SWNTs taken at $E_{\text{laser}} = 1.579$ eV, where in each case the trace for the anti-Stokes spectra is enhanced by the reciprocal of the Boltzmann factor to yield I_{AS}/I_S, which denotes the ratio of the peak intensities of the anti-Stokes and Stokes features in the normalized radial breathing mode spectra, as presented in Figure 5. Although the radial breathing mode frequencies (and consequently the nanotube diameters) for the spectra in Figure 5a and b are almost the same, the ratios of the anti-Stokes to Stokes intensities I_{AS}/I_S are very different. To the right of each pair of experimental traces is a plot of theoretical curves for the expected intensity of the radial breathing mode based on a simple model which assumes constant matrix elements and a joint density of states with the strong frequency dependence that is expected for a 1D system. The profile of the joint density of states for an isolated SWNT was determined by Raman measurements made on the same isolated SWNT, as the laser frequencies were varied by the use of a tunable laser [17]. The curve fitting of the calculated lineshapes and experimental measurements is done by varying the energy of the experimental van Hove singularity E_{ii} iteratively, in order to obtain the experimentally determined I_{AS}/I_S ratio. For example, the two tubes in Figure 5a and b have experimentally determined

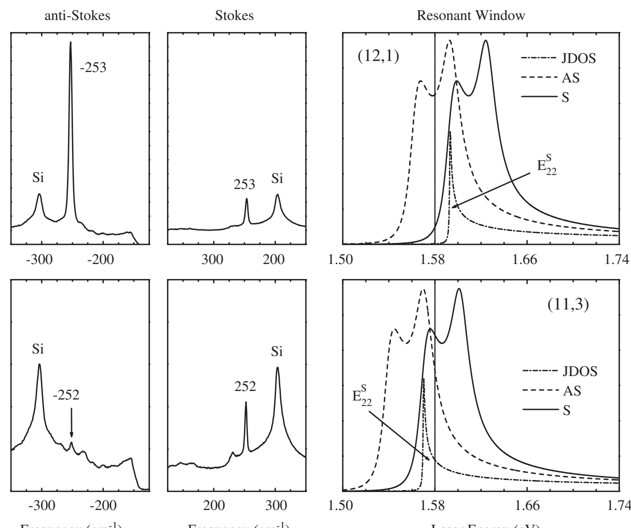


Figure 5. Resonance anti-Stokes and Stokes spectra for an isolated (a) (12,1) and (b) (11,3) SWNT on a Si/SiO_2 substrate using 1.579 eV (785 nm) laser excitation. The peak at 303 cm^{-1} comes from the Si substrate and is used for calibration purposes, while ω_{RBM} for the two tubes is 253 and 252 cm^{-1}, respectively. The anti-Stokes spectra are shown after correcting for the Boltzmann factor $\exp(E_{\text{ph}}/kT)$ to permit easy comparison to be made between the intensities of the Stokes and anti-Stokes spectra. To the right of each pair of experimental spectra is a plot of the calculated anti-Stokes and Stokes spectra predicted for a given profile for the joint density of states (dot–dashed line). Also shown is E_{laser} (solid vertical line) at which the ratio I_{AS}/I_S is evaluated for a selected E^S_{22} value. In the fitting procedure, the energy of the singularity E^S_{22} is varied to obtain the value for the ratio I_{AS}/I_S that is observed experimentally. Reprinted with permission from [18], A. G. Souza Filho et al., *Phys. Rev. B* 63, 24140R (2001). © 2001, American Physical Society.

diameters that differ by only 0.004 nm, but E^S_{22} values that differ by a much more significant amount (33 meV).

In practice, the (n, m) index characterization for isolated nanotubes can be simply carried out by using a single laser excitation energy E_{laser} when the nanotube is in strong resonance with E_{laser}. If the (n, m) values of every nanotube in the sample must be found, then a tunable laser is needed to provide a resonant excitation energy for each nanotube, so that a complete (n, m) Raman characterization can be carried out for each SWNT.

Measurement of the spectral profile of the 1D JDOS for a single wall carbon nanotube near a van Hove singularity gives direct evidence for the sharp singularities in the JDOS, thereby showing how it is possible to get sufficient intensity to observe a Raman spectrum from just one nanotube [17]. The spectral profile of the JDOS is determined by using many laser energies over a narrow energy range, as is provided by a tunable laser, to measure the anti-Stokes to Stokes intensity ratios for an individual SWNT. Evaluating the integrated intensity $I(E_{\text{laser}})$ of the RBM feature as the laser is tuned over the resonant window allows an evaluation of the joint density of states profile $g(E)$. An experimental determination of $g(E)$ shows that the width of the van Hove singularity in the joint density of states has a very small value (~0.5 meV), thus giving rise to the very strong intensity enhancement (~1000) associated with the

sharp 1D van Hove singularities in the joint density of states $g(E)$. Measurements of the Stokes and anti-Stokes spectra over the whole resonance window of a nanotube yield the E_{ii} value more accurately (to 5 meV) than when using a single laser line (~10 meV accuracy). Many physical properties, such as transport, optical, and mechanical properties, are very sensitive to the spectral profile of the electronic density of states, but the profile is difficult to measure directly, because probes, such as scanning tunneling microscopy tips [13], interact too strongly with the nanotube, thereby broadening the van Hove singularities (from less than 1 to more than 10 meV), and consequently also broadening the density of states that is probed by scanning probe methods [13]. Intertube interaction in a SWNT bundle also broadens the linewidth.

2.4. Connecting Single Nanotube and Bundle Raman Spectra

A brief discussion of the various features appearing in the Raman spectra at the single nanotube level is now presented, for later comparison in Section 3.1 with the corresponding results for SWNT bundles. Since most of the measurements on SWNT bundles are for the RBM and G-band, our focus in this section will likewise be on the G-band, since the characteristics of the RBM feature were already discussed.

The G-band in carbon nanotubes, occurring in the 1500–1605 cm^{-1} range (see Fig. 1), is basically derived from the Raman-allowed optical mode E_{2g_2} of graphite by zone folding the 2D graphene Brillouin zone (see Fig. 3b) into the 1D nanotube Brillouin zone, noting that only modes with A, E_1, and E_2 symmetry are Raman active for single wall carbon nanotubes [6, 20]. The G-band is an intrinsic feature of carbon nanotubes that is closely related to vibrations in all sp^2 carbon materials, and the G-band frequency in SWNTs is not dispersive with regard to E_{laser}. This property helps to distinguish features in the harmonics and combination modes that are associated with the nondispersive G-band from those associated with the highly dispersive D-band and G'-band features to be discussed.

The most dramatic features of the G-band spectra are the characteristic differences between the G-band lineshapes for metallic and semiconducting nanotubes, shown in Figure 6. Isolated semiconducting nanotubes (Fig. 6a) characteristically show two dominant Lorentzian features with 6–15 cm^{-1} (full width at half maximum, FWHM) linewidths at room temperature [22], the lower frequency component associated with vibrations along the circumferential direction (ω_G^-), and the higher frequency component (ω_G^+) attributed to vibrations along the direction of the nanotube axis. In contrast, the Raman spectra for metallic nanotubes (Fig. 6b) also have two dominant components with similar origins, but in this case, the upper frequency component ω_G^+ has a Lorentzian lineshape that is almost as narrow as that for the semiconducting nanotubes, but the lower frequency component ω_G^- is a very broad Breit–Wigner–Fano line (see Fig. 6b), with a strong coupling to a continuum, identified with surface plasmons [23]. In general, the coupling of the phonons to surface plasmons is much stronger in SWNT bundles than for isolated SWNTs because the coupling is between the out-of-plane component of the force constant

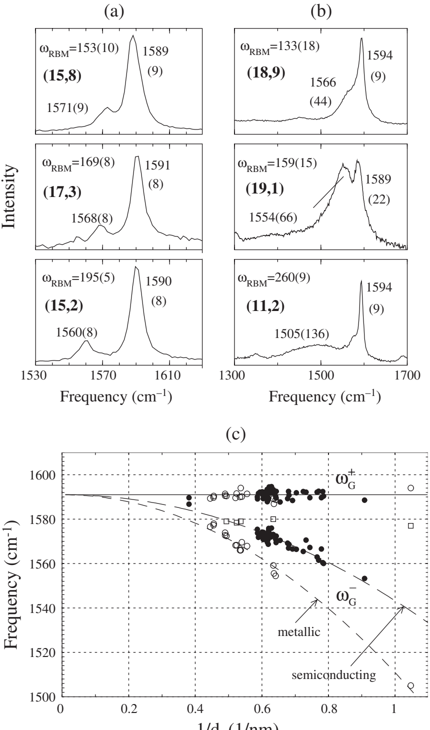


Figure 6. (a,b) Profiles of the G-band spectra for three semiconducting (a) and three metallic (b) nanotubes whose (n, m) indices are determined by analysis of their RBM spectra. The frequencies for ω_G^+ and ω_G^- are indicated in cm^{-1} for each nanotube along with the corresponding linewidths in parentheses. (c) Dependence on reciprocal nanotube diameter $1/d_t$ of ω_G^+ and ω_G^- for both semiconducting nanotubes (solid circles) and metallic nanotubes (open circles). The open squares indicate that a peak near ~1580 cm^{-1} is sometimes observed in the Raman spectra for metallic SWNTs. Reprinted with permission from [21], A. Jorio et al., *Phys. Rev. B* 65, 155412 (2002). © 2002, American Physical Society.

associated with the circumferential ω_G^- mode and the surface plasmon.

Experimental polarization studies on SWNT bundles [24] show that the G-band for a general chiral nanotube consists of six vibrational modes of $A+E_1+E_2$ symmetry along each of the two orthogonal (the axial and circumferential) directions [6], consistent with theoretical predictions. Knowledge of the number of modes is important for the lineshape analysis studies that are used to determine G-band properties. Because of the antenna effect, the strongest coupling of the $\vec{E}$ field of the laser excitation to the nanotube occurs for the optical $\vec{E}$ field along the nanotube axis, and experimentally the most intense modes are found to be the two modes with A symmetry. Therefore, in viewing the Raman spectra of SWNTs with unpolarized light, only two dominant components are generally observed experimentally, one for vibrations in the tangential direction (ω_G^+) and one for vibrations in the circumferential direction (ω_G^-), and this is seen for SWNTs with diameters less than 2 nm, whether they are isolated or in SWNT bundles.

It is the curvature of the nanotube that admixes a small amount of interplanar graphene modes into the circumferential modes of SWNTs, thus lowering their C–C force

constants and mode frequencies, and giving rise to a splitting of the G-band mode frequencies ($\omega_G^+ - \omega_G^-$) into a G^+ axial component and G^- circumferential component, as shown in Figure 6c. Here it is seen that ω_G^+ shows no d_t dependence for either semiconducting or metallic tubes, while ω_G^- has a strong d_t^{-2} dependence that differs according to whether the nanotubes are semiconducting or metallic, and ω_G^- is well fit to the functional form

$$\omega_G^- = \omega_G^+ - \mathscr{C}/d_t^2 \tag{2}$$

where $\mathscr{C} = 45.8$ cm^{-1} nm^2 for semiconducting nanotubes and $\mathscr{C} = 79.5$ cm^{-1} nm^2 for metallic nanotubes. It should be emphasized that this discussion applies to measurements made with unpolarized light, where only the two dominant components in the G-band spectra are considered. At the single nanotube level, no chirality dependence for the G-band ω_G^+ and ω_G^- mode frequencies is found experimentally; nor is $\mathscr{C}$ found to depend on E_{laser}. Therefore it is convenient to use Eq. (2) to obtain an approximate determination of d_t for those SWNTs where the RBM feature is not observed. It should be also mentioned that when polarized light is used, one can distinguish all six components of the G-band expected from theory, the diameter dependence of each component can be found, and some deviation from Eq. (2) is found both experimentally [25] and theoretically [26].

The G-band linewidths for the semiconducting nanotubes at the single nanotube level do not vary much with diameter, with the smallest observed FWHM intensity linewidth values being 6 cm^{-1} [22]. The linewidth for the circumferential ω_G^- component for metallic nanotubes, however, increases very substantially with decreasing d_t, reflecting the increased Breit–Wigner–Fano coupling. These broadening effects in G^- for metallic tubes are more easily observed in SWNT bundles than in isolated SWNTs. Regarding semiconducting SWNTs, the G-band linewidths are not sensitive to chiral angle θ, or to $E_{\text{laser}} - E_{ii}$ [22].

From a theoretical standpoint, zone folding of the graphene sheet to form a nanotube gives rise to Raman-active modes with $A + E_1 + E_2$ symmetry [6] in both the longitudinal (along the tube axis) and transverse (in the circumferential) directions, in addition to a large number of modes with $E_3, E_4, \ldots$ symmetry, which are not Raman-active [6]. Experimental polarization studies on nanotube bundles [24], with an average nanotube diameter $d_t \simeq 1.85$ nm and with $E_{\text{laser}} = 2.41$ eV, have verified the presence of these six modes (three axial and three circumferential modes). These studies have given explicit values for the polarization parameters, namely for the mode frequencies and relative intensities, for each of these six modes for a particular sample of bundled SWNTs [24]. As mentioned above, curvature of the nanotubes is expected to admix some out-of-plane graphite-derived modes with the circumferential G-band mode and thereby to weaken the bonds and lower the force constants in the circumferential direction.

Individual SWNTs are also sensitive to the depolarization or antenna effect, which refers to the tendency of the nanotubes themselves to create a strong local dipole field, so that the optical electric field seen by the nanotube is the vector sum of the applied field and the local field along its axis due to the nanotube dipole antenna. Since the local dipole field is very strong, nanotubes act like polarizers. This antenna effect has been shown by Duesberg et al. to be highly effective for metallic nanotubes or for a very thin nanotube bundle containing only a few nanotubes (see Fig. 7) [27], and they showed that the Raman spectra from a metallic SWNT are strongly suppressed when the light is polarized perpendicular ($\alpha_i = 90°$ in Fig. 7) to the nanotube axis [27].

Polarization measurements at the single nanotube level show that also for isolated semiconducting nanotubes the field component polarized normal to the nanotube axis is completely quenched due to the antenna effect [28]. However, two nanotubes whose axes intersect at some angle each create their own dipole fields, so that the polarization component normal to one of the nanotube axes no longer is completely suppressed, allowing the observation of E_1 and E_2 symmetry modes when $\alpha_i = 90°$ for each of the two tubes. Because of this effect, it is much easier to observe the circumferential modes in SWNT bundles than for isolated SWNTs. The diameter dependence of each of the six G-band modes in the polarized Raman spectra have been studied experimentally, and good agreement was obtained with theoretical predictions [25, 26].

Also observed in the semiconducting and metallic SWNT spectra, both at the single nanotube level and in SWNT bundles, are the D-band and the G'-band features, as seen in Figure 1. The frequencies of the disorder-induced D-band ($1250 < \omega_D < 1450$ cm^{-1}) and of its second-order overtone G'-band ($2500 < \omega_{G'} < 2900$ cm^{-1}) in the Raman spectra of sp^2-bonded carbon materials have been known for many years to exhibit a strongly dispersive behavior as a function of laser excitation energy ($1.0 < E_{\text{laser}} < 4.5$ eV) [29, 30]. The strong linear dispersion of the D-band frequency $\omega_D(E_{\text{laser}})$ in graphite (53 cm^{-1}/eV), and of its second-order overtone

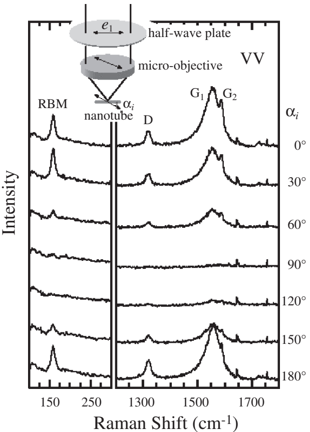


Figure 7. Raman spectra from a thin SWNT bundle where a metallic nanotube is resonant with $E_{\text{laser}} = 1.96$ eV. The spectra were taken with the VV configuration (where the wave vectors for the incident and scattered light are parallel to each other) for various angles α_i between the axis of the bundle and the polarization of the incident laser beam, as depicted in the inset. For $\alpha_i = 0°$ and $\alpha_i = 180°$, the polarization of the incident radiation is parallel to the axis of the SWNTs, as determined from scanning tunneling microscopy (STM) images with an accuracy of ±10°. For $\alpha_i = 90°$, the intensities of all the Raman features (RBM, D-band, G^- and G^+) vanish. Reprinted with permission from [27], G. S. Duesberg et al., *Phys. Rev. Lett.* 85, 5436 (2000). © 2000, American Physical Society.

G' band, has been identified with a double resonance process involving photons, phonons and electronic transitions between linearly dispersive π and π^* states [31–34]. The double resonance model in particular [31, 33, 35] accounts for the observed dispersion of the D-band and the G'-band as a function of photon excitation energy E_{laser} in graphite and in sp^2 carbons. This model was subsequently extended to account for the special properties of the D-band and G'-band of isolated SWNTs and has been successful in accounting for the special features in the D-band and G'-band spectra observed for SWNTs relative to the other sp^2 carbons.

Each van Hove singularity is characterized by an energy E_{ii} and a corresponding electron wave vector k_{ii} where the JDOS (joint density of states) is a maximum. According to double resonance theory, the phonon wave vector at the peak intensity of the D-band and G'-band features occurs at $q_{ii} = -2k_{ii}$. Because of the trigonal warping effect of the electronic structure (see Fig. 8b), the E_{ii}, k_{ii}, and q_{ii} points at resonance are all unique, leading to a nonmonotonic chirality dependent resonance condition for each (n, m) nanotube. The resulting D-band or G'-band frequencies therefore appear to exhibit a significant scatter for a given E_{ii} when ω_D or $\omega_{G'}$ are plotted as a function of tube diameter (see Fig. 8a) [36], because of their nonmonotonic dependence on chiral angle arising from the trigonal warping effect.

Since a given laser energy can excite various (n, m) nanotubes with different diameters and different chiral angles (see Fig. 4) and because the D-band and G'-band are highly dispersive, different D-band and G'-band mode frequencies can be observed at the single nanotube level for different SWNTs using a single laser excitation line. In contrast, for SWNT bundles, a given laser line gives a single D-band or G'-band frequency. This effect results in line broadening in SWNT bundles. The linewidth for the D-band at the single nanotube level can be as small as 6 cm^{-1}, and for the G'-band as small as 11 cm^{-1} under good resonance conditions [22]. The linewidths of the D-band and G'-band in SWNT bundles tend to be significantly larger than for individual SWNTs.

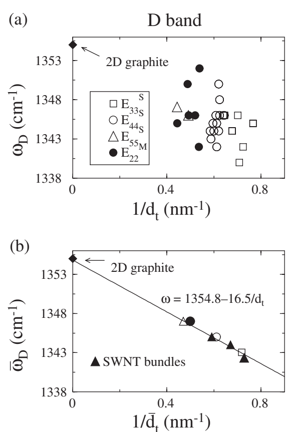


Figure 8. (a) D-band frequencies as a function of reciprocal nanotube diameter for individual SWNTs using $E_{\text{laser}} = 2.41\,\text{eV}$ laser excitation. The data are classified in terms of the E_{ii} interband transition with which the resonancy occurs, including both metallic (M) and semiconducting (S) SWNTs. (b) Plot of $[\bar{\omega}_D(E_{ii})]$, denoting the observed D-band frequencies averaged over all tubes in a given sample that are resonant with a given interband transition E_{ii}, vs $1/\bar{d}_t$, the corresponding average of the reciprocal diameter of the tubes. Data are shown for E_{laser} in resonance with the E_{55}^S, E_{44}^S, E_{33}^S, and E_{22}^M interband transitions in the JDOS. The line is a fit to the data, showing that the D-band frequencies extrapolate (on average) to the graphene (2D graphite) value when $1/\bar{d}_t \to 0$. The solid triangles in (b) denote the D-band frequencies for three different SWNT bundles with different average diameters. Reprinted with permission from [36], A. G. Souza Filho et al., *Phys. Rev. B* 67, 035427 (2003). © 2003, American Physical Society.

In order to compare the SWNT spectra at the single nanotube level with those for SWNT bundles and to gain an understanding of the mechanisms behind the d_t dependence of the D-band and G'-band frequencies, the ω_D and $\omega_{G'}$ peak frequencies are averaged over chiral angles for each E_{ii} and the results are shown in Figure 8b for D-band measurements. For a given interband transition E_{ii}, there is only a small variation in the d_t values for a typical SWNT bundle sample, and this variation is accounted for self-consistently in making the plots in Figure 8. Here the resulting averages of the diameter and D-band frequencies are denoted by $\bar{d}_t(E_{ii})$ and $\bar{\omega}_D(E_{ii})$, respectively, and we plot these pairs of numbers in Figure 8b for $i = 3, 4, 5$ for semiconducting SWNTs and for $i = 2$ for metallic SWNTs, using the same symbols as in Figure 8a [36].

The results of this analysis in Figure 8 give a simple linear dependence of the average $\bar{\omega}_D$ on $1/\bar{d}_t$,

$$\bar{\omega}_D = 1354.8 - 16.5/\bar{d}_t \tag{3}$$

where $1/\bar{d}_t$ is the average $1/d_t$, as shown in Figure 8b. Good agreement in Figure 8b is obtained between the average $\bar{\omega}_D$ results for an ensemble of isolated tubes and the corresponding $\bar{\omega}_D$ results for SWNT bundles measured with the same E_{laser}. The linear downshift of $\bar{\omega}_D$ as a function of $1/\bar{d}_t$, shown in Figure 8b, is attributed to the softening of the spring constants for the vibrations associated with the D-band due to the nanotube curvature. This assertion is based on calculations of the eigenvectors for the D-band [37] which show that the atomic displacements have some components along the nanotube circumference, which soften the modes due to contributions from out-of-plane force constants. Contributions from out-of-plane force constants, which do not occur in graphite, are also responsible for the d_t-dependent downshift of ω_{G^-}, the lower frequency circumferential component of the G-band [38, 39].

However, when the behavior of ω_D vs d_t is investigated at the single nanotube level, within one interband transition, E_{44}^S for example, the D-band frequency ω_D has a tendency to *increase* when the diameter decreases, because E_{ii} is proportional to k_{ii} on the basis of the electronic dispersion relation, and k_{ii} in turn is proportional to $1/d_t$, except for perturbations due to the trigonal warping effect that complicate this simple dependence and give rise to a spread in the data points (see Fig. 8a) [6, 36]. The net result of the formation of E_{ii} subbands is to introduce an opposite dependence of ω_D on d_t when compared to the averaged result shown in Figure 8b [36]. As can be seen in Figure 4, E_{ii} for a given i increases as d_t decreases. Since for a given interband transition E_{ii} the relevant tube diameter range is not so large for

a typical SWNT bundle sample (see Fig. 4), the frequency of the D-band ω_D is mainly determined by the magnitude of the k_{ii} states, and this effect is dominant over the curvature effect throughout the small diameter range over which contributions from resonance within a single E_{ii} subband dominate. When we jump from one van Hove singularity E_{ii} to another singularity E_{jj} ($j \neq i$), the spring constant mechanism becomes dominant in the determination of ω_D. The results shown in Figure 8 also suggest that metallic and semiconducting tubes exhibit the same ω_D diameter-dependent behavior.

A similar analysis carried out on the G'-band spectra at the single nanotube level shows a similar nonmonotonic dependence of $\omega_{G'}$ on chiral angle θ, and a diameter dependence

$$\bar{\omega}_{G'} = 2708 - 35.4/\bar{d}_t \tag{4}$$

Another dispersive feature observed in the Raman spectra is the M-band occurring near 1750 cm^{-1} which has been identified with an in-plane TO + LA combination mode [40] and shows a similar functional d_t dependence for its mode frequency

$$\bar{\omega}_M = \omega_{M,0} - \beta_M/\bar{d}_t \tag{5}$$

where the M-band has two components for which $\omega_{M,0}$ and β_M are, respectively, 1775 cm^{-1} and 18.0 cm^{-1} nm for the M^+ feature and 1754 cm^{-1} and 16.7 cm^{-1} nm for the M^- feature, as determined from single nanotube Raman spectroscopy [40].

One of the unique features of the D-band and the G'-band spectra of SWNT bundles is a step or oscillatory behavior [42–44] superimposed on the linear E_{laser} dependence of both ω_D and $\omega_{G'}$ that is characteristic of graphite and sp^2 carbons. This effect is shown for the G'-band in Figure 9a but also has been observed and studied in the corresponding plot of ω_D vs E_{laser}. The deviation from a linear dispersive behavior is clearly seen in Figure 9b, where the points are obtained after subtraction of the linear solid line fit to the experimental points in Figure 9a. The oscillatory behavior in Figure 9b is identified with quantum effects associated with the strong coupling between the electrons and phonons under resonance conditions. Specifically, the peak near 2.0 eV is identified with the E_{11}^{M} interband transition for metallic nanotubes and the peak near 2.5 eV with the E_{33}^{S} and E_{44}^{S} transitions for semiconducting nanotubes, thus showing the importance of the 1D electronic structure of SWNTs in the physical origin of the D-band and the G'-band spectra. The physical basis for this step or oscillatory behavior was explained by the coupling between electrons and phonons under resonant conditions, consistent with the behavior seen in the plot of the resonance wave vector k_{ii} for SWNTs that are within the resonance window with the various E_{laser} energies (Fig. 9b). Although the oscillations in the D- and G-bands are best observed in SWNT bundles, the mechanism for their origin is best studied in isolated SWNTs. As discussed, the resonant k_{ii} values depend on both the diameter and chirality of the resonant nanotubes, as has been elucidated by studies at the single nanotube level [41]. An independent theory by Kürti et al. [26] that considers the double resonance process in light of the van Hove singularities has been able to explain the anomalous dispersion in SWNT bundles.

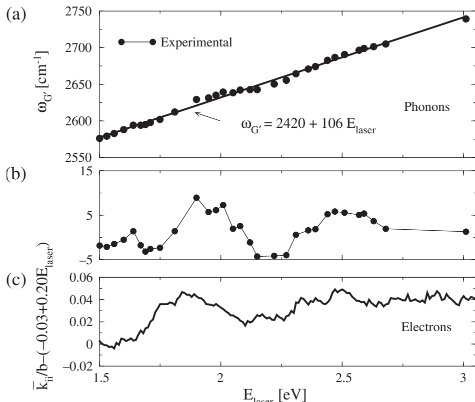


Figure 9. (a) Dependence on the laser excitation energy of $\omega_{G'}$ for the dominant second-order Raman feature for SWNT bundles. In contrast to most sp^2 carbons where $\omega_{G'}$ depends linearly on E_{laser}, the corresponding plot for SWNT bundles shows a superimposed steplike or oscillatory behavior near 2.0 and 2.5 eV for SWNT bundles. (b) The oscillatory component of (a) obtained after subtracting the linear background given by $\omega_{G'} = 2420 + 106E_{\text{laser}}$ in (a). (c) Plot of the E_{laser} dependence of the oscillations in the resonant electron wave vector k_{ii} after normalization to the length of the basis vector b of the reciprocal lattice of 2D graphite and after subtracting the linear background. For each E_{laser} value, the k_{ii} values for each resonant SWNT are calculated, and $\bar{k}_{ii}$ represents an appropriate weighted average of the individual k_{ii} values [41]. The strong correspondence between the E_{laser} dependence of $\omega_{G'}$ for phonons and $\bar{k}_{ii}$ for electrons provides strong support for the identification of each oscillation with a particular interband transition between van Hove singularities in the JDOS. Reprinted with permission from [41], A. G. Souza Filho et al., *Phys. Rev. B* 65, 035404 (2002). © 2002, American Physical Society.

Experimental studies at the single nanotube level have shown that these highly dispersive modes can provide unique information about the electronic structure and the phonon dispersion relations of single wall carbon nanotubes. Some of the new phenomena that have been observed at the single nanotube level will be briefly listed in an effort to stimulate complementary studies in SWNT bundles.

The diameter and chirality dependent D-band and G'-band effects, which are observed at the single nanotube level, are now briefly reviewed. Under special circumstances it has been possible to observe a double resonance process with both the E_{33}^{S} and E_{44}^{S} interband transitions on the same isolated SWNT [45], utilizing the strong dispersion of the G'-band as a function of E_{laser}. Because of the special circumstance under which this two peak G'-band effect can occur, it can be used as an independent confirmation of the (n, m) structural characterization carried out through analysis of the radial breathing mode feature, as discussed in Section 2.2.

Because of the trigonal warping of constant energy surfaces, the D-band and G'-band frequencies produced by cutting lines on opposite sides of the K point in the Brillouin zone will in general be different from each other. If both cutting lines are resonant with the same E_{laser} energy, then it is possible to observe these two contributions, which appear

as a double peak in the Raman spectra of metallic SWNTs. If the frequency difference between the two features is large enough, they can be resolved [45–47].

For the case of metallic SWNTs, the two cutting lines correspond to the same E_{ii}^{M} singularity [46, 47], while for semiconducting SWNTs, the two cutting lines correspond to different E_{ii}^{S} singularities [45, 47]. Unresolved double peak effects for metallic and semiconducting SWNTs are a source of line broadening both at the single nanotube level and for SWNT bundles.

3. SINGLE WALL CARBON NANOTUBE BUNDLES

In this section the Raman spectra for SWNT bundles are presented and discussed in relation to the spectra for isolated SWNTs at the single nanotube level. In Section 3.1, the room temperature baseline Raman spectra for SWNT bundles are presented. The effect of temperature on the RBM, G-band, and D-band mode frequencies for SWNT bundles is discussed in Section 3.2. These temperature-dependent effects are more important for SWNT bundles than for isolated SWNTs because of the poor thermal conductivity between SWNTs within the bundle. In Section 3.3, the effect on the Raman frequencies of the solubilization, debundling, doping, and of applied voltage of SWNTs in SWNT bundles is discussed, while the effect of pressure on the mode frequencies is reviewed in Section 3.4. A comparison is made in Section 3.5 between the perturbations to the Raman frequencies of nanotube bundling and debundling and the effects of the application of pressure, doping, and applied potentials on the Raman spectra of SWNT bundles are also discussed and compared in terms of their effect on the singularities in the joint density of states and on the electrochemical potential.

3.1. Raman Spectra from Single Wall Carbon Nanotube Bundles

Early Raman spectra on SWNTs were taken on samples containing ropes of single wall carbon nanotubes with a narrow diameter distribution in the 1.2–1.4 nm range, and the SWNT bundle samples were prepared by either the electric arc method or by the laser vaporization technique [48, 49]. Raman spectra taken on such a sample for five different laser excitation energies are shown in Figure 10 [5]. Prominent in these spectra are the G-band at ~1590 cm^{-1} and the radial breathing mode at ~186 cm^{-1} [see the trace in Fig. 10 taken for $E_{\text{laser}} = 2.41$ eV (514.5 nm)]. By comparing the various Raman spectra in Figure 10, which were taken at different laser excitation energies E_{laser}, we see large differences in the vibrational frequencies and intensities of the strong A_1 radial breathing mode (see Fig. 11), consistent with a resonance Raman effect involving nanotubes of different diameters for each laser excitation energy.

In Figure 12, we see theoretical calculations showing that for an (n, n) armchair SWNT there are three mode frequencies very close to 1580 cm^{-1}, which have mode symmetries A_{1g}, E_{1g}, and E_{2g} (see Fig. 11), each mode frequency being almost independent of nanotube diameter. A similar behavior is found for the Raman band near 1580 cm^{-1} for zigzag and chiral nanotubes [6, 50]. The atomic displacements associated with the normal modes near ~186 and ~1580 cm^{-1} for a (10,10) nanotube are shown in Figures 1 and 11 [6].

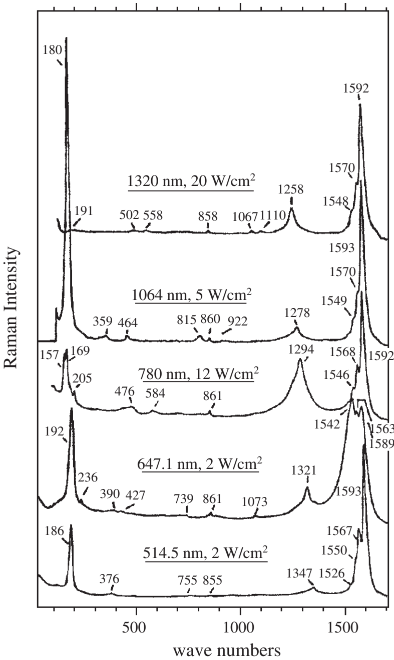


Figure 10. Experimental room temperature Raman spectra for purified single wall carbon nanotubes excited at five different laser excitation wavelengths. The laser wavelength and power density for each spectrum are indicated, as are the vibrational frequencies (in cm^{-1}) [5]. The equivalent incident photon energies for the laser excitation are: 1320 nm → 0.94 eV; 1064 nm → 1.17 eV; 780 nm → 1.58 eV; 647.1 nm → 1.92 eV; 514.5 nm → 2.41 eV. Reprinted with permission from [5], A. M. Rao et al., *Science* 275, 187 (1997). © 1997,

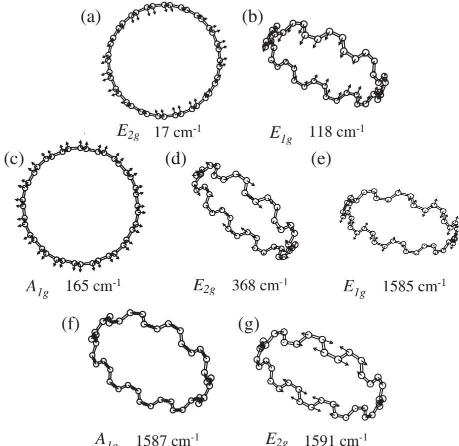


Figure 11. The calculated Raman mode atomic displacements, frequencies, and symmetries for selected normal modes for the (10,10) nanotube. The symmetry and the frequencies for these modes are not strongly dependent on the chirality of the nanotube. In the figure, we show the displacements for only one of the two modes in the doubly degenerate E_{1g} and E_{2g} modes. Reprinted with permission from [6], R. Saito et al., "Physical Properties of Carbon Nanotubes." Imperial College Press, London, 1998. © 1998, Imperial College Press.

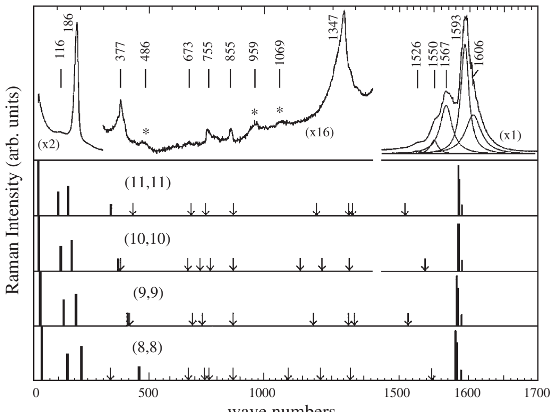


Figure 12. Raman spectra (top) of a rope of single wall carbon nanotubes taken with 514.5 nm laser excitation at ~2 W/cm^2 power level. The features in the spectrum denoted by asterisks are assigned to second-order Raman scattering. The four bottom panels are the calculated Raman spectra (based on a bond polarizability model) for armchair (n, n) nanotubes, $n = 8$ to 11, and the strongest Raman-allowed features are indicated by vertical bars. The arrows in the panels indicate the positions of the remaining weak, Raman-active modes. Reprinted with permission from [5], A. M. Rao et al., *Science* 275, 187 (1997). © 1997,

In contrast to the high frequency band near 1580 cm^{-1}, the feature near ~186 cm^{-1}, which is identified with an A_1 radial breathing mode (ω_{RBM}), is strongly dependent on the nanotube diameter, as shown in Figures 10 and 12. Calculations [51, 52] show high intensities for this radial breathing mode and for the tangential G-band modes (ω_G) near 1580 cm^{-1} (Fig. 12), in agreement with experimental Raman spectra. The other Raman-active modes (see Fig. 12) are predicted to have low Raman cross sections, also in agreement with experiment (Fig. 10). Bond polarizability calculations further predict that the relative intensities of the weaker Raman-active features in the experimental spectra of Figure 10 can be increased by making measurements on carbon nanotubes of lengths (e.g., 100 nm, small compared to an optical wavelength for the laser excitation) [53]. The weak D-band features in the Raman spectrum in Figure 10 at about 1350 cm^{-1} are associated with resonance Raman scattering of phonons near the K point in the 2D Brillouin zone. These identifications are in agreement with experimental and theoretical observations at the single nanotube level, discussed in Section 2.

Historically, these experimental observations of the Raman effect in SWNT bundles [5] also provided the first clear confirmation for the theoretical predictions about the singularities in the 1D electronic density of states of carbon nanotubes through study of both the radial breathing mode features and the G-band mode features. This confirmation of the existence of singularities in the 1D density of states of SWNTs by the resonance Raman effect was soon corroborated by a more direct measurement of the 1D electronic density of states by STM/scanning tunneling spectroscopy [12, 54].

Because of this strong resonance enhancement effect, it was found that only a small concentration of single wall nanotubes in a sample containing other carbon forms can give rise to spectral features showing the characteristic sharp doublet structure in the 1570–1600 cm^{-1} spectral range [55]. The resonance enhancement effect is so strong for small diameter (<2 nm) SWNTs that it is possible to observe the Raman effect from individual SWNTs (see Section 2) and to observe up to fifth order Raman scattering in SWNT bundles (e.g., up to Raman shifts of 6885 cm^{-1} using 488 nm laser excitation) [56].

Since the energies of these sharp features in the 1D electronic density of states are strongly dependent on the nanotube diameter, a change in the laser frequency brings into resonance a different carbon nanotube with a different diameter that satisfies the new resonance condition. This effect is called *diameter selective Raman spectroscopy* and led in 1998 to the first use of Raman spectroscopy to distinguish between metallic and semiconducting nanotubes [57, 58]. The concept of diameter selective spectroscopy follows from Figure 4 where each point on the plot of the energy E_{ii} of the van Hove singularity in the JDOS vs tube diameter d_t denotes a different (n, m) SWNT. For a given diameter distribution in the sample (e.g., 1.37 ± 0.20 nm, for the sample in Fig. 13), we can see in Figure 13 that for E_{laser} below 1.70 and above 2.14 eV, the SWNT bundle contains predominantly semiconducting nanotubes that are resonant with E_{laser}, in agreement with the lineshape shown for these traces in Figure 13, while for E_{laser} in the range 1.83 to 2.14 eV, the G-band shows a very different broad lineshape, characteristic of metallic nanotubes. Here the lower frequency G-band component (designated by G^-) for metallic tubes is described by a Breit–Wigner–Fano lineshape arising from the coupling of the G^- phonons to plasmons through the

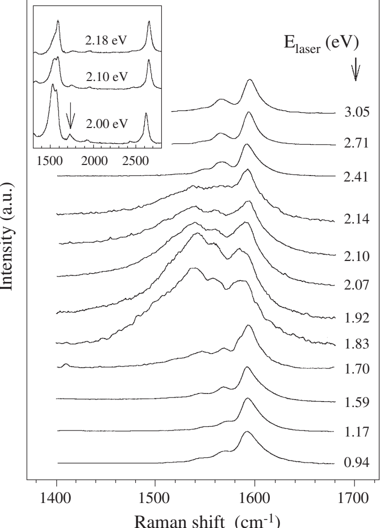


Figure 13. Raman spectra of the tangential modes of carbon nanotubes with diameters in the range $d_t = 1.37 \pm 0.20$ nm, obtained with several different laser excitation energies E_{laser}. The inset shows low resolution Raman spectra between 1300 and 2800 cm^{-1} in the range of laser energies 2.00–2.18 eV where the metallic nanotubes are dominant and the arrow points to a low intensity M-band feature near 1750 cm^{-1} [40] which has its maximum intensity correlated with the maximum intensity of the metallic G-band lineshape. Reprinted with permission from [57], M. A. Pimenta et al., *Phys. Rev. B* 58, R16016 (1998). © 1998, American Physical Society.

tube curvature which allows mixing of out-of-plane phonons with the tangential in-plane phonon modes characteristic of 2D graphite. This coupling is most pronounced for small diameter nanotubes in SWNT bundles relative to isolated SWNTs, since the coupling between phonons in one tube and plasmons in adjacent tubes can readily occur for SWNTs in bundles.

A closely related effect is observed in Figure 14 where Stokes and anti-Stokes spectra for the same SWNT sample with $d_t = 1.49 \pm 0.20$ nm are shown for four different laser excitation energies. Resonance for the scattered photon for the Stokes process occurs when $E_{\text{laser}} - \hbar\omega_G = E_{ii}$, while for the anti-Stokes process, the resonance condition is given by $E_{\text{laser}} + \hbar\omega_G = E_{ii}$. Thus for $E_{\text{laser}} = 2.19$ eV, Figure 4 shows that semiconducting SWNTs in the bundle are resonant for both incident and scattered photons for the anti-Stokes process, but for the Stokes process only a few metallic tubes would be expected to be resonant with the scattered photon. The situation is dramatically different for $E_{\text{laser}} = 1.58$ eV where both the incident and scattered photons are strongly resonant with metallic tubes for both the Stokes and anti-Stokes processes [59]. For $E_{\text{laser}} = 1.92$ eV, the incident photon is not in good resonance for either metallic or semiconducting tubes, while the scattered photon for the Stokes process is in good resonance with metallic tubes and for the anti-Stokes process with semiconducting tubes. For $E_{\text{laser}} = 1.49$ eV, again the incident photons are not in good resonance with the E_{ii} for tubes in the SWNT bundle, but for the scattered photons the resonance in the Stokes process is good with semiconducting tubes, and with metallic tubes in the anti-Stokes process, all of which follows from the plot of E_{ii} vs d_t in Figure 4. Thus the Stokes/anti-Stokes phenomena observed in SWNT bundles can be related to behavior observed at the single SWNT level.

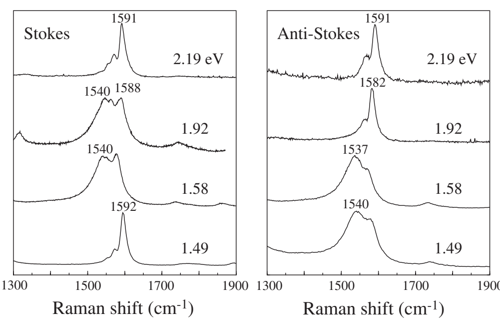


Figure 14. Stokes and anti-Stokes Raman spectra for SWNTs of $d_t = 1.49 \pm 0.20$ nm taken at four different values of E_{laser} to illustrate differences in spectral lineshape between the Stokes and anti-Stokes spectra [59]. For $E_{\text{laser}} = 2.19$ eV, both the Stokes and anti-Stokes processes are in resonance with semiconducting nanotubes: for $E_{\text{laser}} = 1.92$ eV, the Stokes process is in resonance with metallic nanotubes and the anti-Stokes process is in resonance with semiconducting nanotubes. For $E_{\text{laser}} = 1.58$ eV, both processes are in resonance with metallic nanotubes. Finally, for $E_{\text{laser}} = 1.49$ eV, the Stokes process is in resonance with semiconducting nanotubes, and the anti-Stokes process is in resonance with metallic nanotubes. The nanotube-specific resonance Raman process associated with the special characteristics of the 1D density of electronic states is responsible for the observed differences between the Stokes and anti-Stokes lineshapes. Reprinted with permission from [59], S. D. M. Brown et al., *Phys. Rev. B* 61, R5137 (2000). © 2000, American Physical Society.

The G-band lineshape studies and the Stokes/anti-Stokes G-band studies previously described led to the definition of the *resonance window* for metallic tubes within a SWNT bundle as the laser energy range over which the intensity of the G^- component was large enough to be observable. For a SWNT bundle with a diameter distribution of 1.37 ± 0.20 nm, the Stokes resonance window was found to extend from 1.7 to 2.3 eV and the anti-Stokes resonance window was correspondingly downshifted by about 0.2 eV [60]. The early determination of the resonance window from SWNT bundle samples provided a sensitive experimental determination of the energy overlap integral γ_0 for SWNTs [15], which is used for constructing the E_{ii} vs d_t diagram (Fig. 4), within the tight binding approximation, and assuming a zero value for the overlap integral (i.e., $s = 0$). The γ_0 value thus determined yielded (n, m) predictions that were later found to be in good agreement with experimental (n, m) determinations based on RBM mode measurements on isolated SWNTs at the single nanotube level [2].

Studies of the effect of high laser radiation intensity on the Raman spectra of SWNT bundles showed (through observation of the RBM spectra) that the small diameter SWNTs burn off preferentially, thereby affecting the G-band lineshapes irreversibly, depending on whether the smallest diameter SWNTs in the initial SWNT bundle are semiconducting or metallic [61]. High laser intensity also serves to reduce the D-band intensity irreversibly and to enhance the RBM and G-band intensities, indicative of annealing out structural defects in SWNTs [61].

In comparing Raman spectra from SWNT bundles and from individual SWNTs, several differences are noteworthy. (1) The linewidths of the Raman features for isolated tubes are much narrower [22], with FWHM values of the G^+-band and D-band being 6 and 7 cm^{-1}, respectively, for isolated SWNTs, as compared to 15 and 34 cm^{-1} for SWNT bundles [22, 62, 63], reflecting inhomogeneous line broadening that arises from contributions of the various (n, m) nanotubes within the resonant window of the laser. (2) Diameter-dependent effects are observed in the spectra from SWNT bundles, as is clearly seen in the selection of the resonance window in distinguishing between metallic and semiconducting SWNTs (Fig. 13) and in interpreting the Stokes/anti-Stokes spectral lineshapes (Fig. 14). Studies at the single nanotube level are especially useful in addressing diameter-dependent effects in each of the spectral features in the Raman spectrum [1]. (3) Even though chirality-dependent phenomena are observed for many of the features in the Raman spectra at the single nanotube level, the spectra observed for SWNT bundles provides only an average over chiral angles for SWNTs within the resonance window. (4) Certain features in the Raman effect that relate to tube–tube interaction are observed more strongly in SWNT bundles, and these include the reported diameter-independent upshift of the RBM band frequency in SWNT bundles, and the increased strength of the G^- lower frequency component of the G-band in SWNT bundles with a Breit–Wigner–Fano lineshape that arises from the plasmon–phonon coupling [14, 23].

The effect of bundling on the electronic structure of SWNTs has been considered generally from a theoretical standpoint. On the basis of a density functional formalism,

the intertube coupling was calculated within the local density approximation (LDA) for isolated and bundled (8,8), (9,9), (10,10), and (11,11) armchair nanotubes, corresponding to a diameter distribution representative of many experimental samples. The results of this calculation showed that intertube coupling causes a band dispersion of ~0.15–0.20 eV which not only opens up a pseudo-gap at E_F but also broadens the van Hove singularities by the same amount, while the peak positions of the van Hove singularities were found to shift in energy away from E_F [64]. This leads to a net increase of the energy spacing between pairs of van Hove singularities in the electronic JDOS (such as E_{11}^M, etc.) for bundled nanotubes relative to isolated SWNTs [64] (see Table 2). These calculations imply that the force constants for a given nanotube in a SWNT bundle should increase and consequently the mode frequencies would be expected to increase relative to their values in an isolated tube. A general expression is then suggested for the interpretation of radial breathing mode frequency measurements for SWNT bundles, namely $\omega_{\mathrm{RBM}} = \alpha/d_t + \Delta\omega_{\mathrm{RBM}}$. This result implies that the effect of intertube interactions is to upshift the radial breathing mode frequency of a particular SWNT in a bundle relative to that for its ω_{RBM} as an isolated SWNT by $\Delta\omega_{\mathrm{RBM}}$ and perhaps also to modify its coefficient α. The precise value of α and $\Delta\omega_{\mathrm{RBM}}$ would then be expected to depend on the local environment of the tube, the total size of the SWNT bundle, and the location of the SWNT within the SWNT bundle. A detailed procedure was established for the determination of α and $\Delta\omega_{\mathrm{RBM}}$ by Kuzmany et al. [44, 65, 66]. The effect of bundling on ω_G is expected to be small.

3.2. Temperature Dependence

Early studies of Raman-active modes in SWNT bundles reported a softening of the radial breathing mode and tangential G-band frequencies with increasing sample temperature. A few speculations were proposed to explain the origin of these frequency downshifts, attributing them to the stretching of carbon–carbon bonds, defects, and disorder in these materials, as well as to the van der Waals interactions between nanotubes in the bundles [20]. However, no detailed theoretical or experimental basis was provided to support these proposals.

A recent study [67] reexamined the temperature dependence of the RBM, G-band, and D-band, using as-prepared HiPCo SWNTs [68]. Based on experimental data and the results of molecular dynamics simulations, the expected radial breathing mode softening due to three contributing factors was considered: thermal expansion of individual SWNTs in the radial direction, softening of the C–C (intratubular) bonds, and softening of the van der Waals intertubular interactions in SWNT bundles (intertubular interaction). Panels a–d in Figure 15 show a reversible linear shift with temperature of the two RBM frequencies, the G-band frequency, and the D-band frequency observed in the Raman spectrum of as-prepared HiPCo SWNTs. The two RBM peaks, with a peak position of $\omega_{\mathrm{RBM}} \sim 182$ and 264 cm^{-1} at room temperature, correspond to the tube diameters of approximately 1.33 and 0.89 nm, respectively.[1] The temperature dependence in terms of $d\omega_{\mathrm{RBM}}/dT$ was obtained from the slopes of the (linear) plots in Figure 15a and b to be ~ -0.0045 and -0.0090 cm^{-1}/K, respectively. The larger magnitude of $d\omega_{\mathrm{RBM}}/dT$ for ω_{264} relative to ω_{182} suggests that smaller diameter tubes are more susceptible to temperature-induced frequency downshifts. Furthermore, the following experimental values were obtained: $d\omega_{G^+}/dT = -0.0189$ cm^{-1}/K, $d\omega_{G^-}/dT = -0.0238$ cm^{-1}/K, and $d\omega_D/dT = -0.0087$ cm^{-1}/K [67]. The temperature dependences of the G-band and D-band modes (Fig. 15c and d) were used to evaluate the softening of the C–C force constants with temperature [67].

It was found that only a small part of the temperature dependence of ω_{RBM} can be attributed to the very small change in the tube diameter [67]. For nanotubes with diameters $d_t \sim 1.34$ nm, the contribution due to the radial thermal expansion is ~4% over the temperature range $300 < T < 800$ K. Interestingly, this contribution increases to ~10% in the case of nanotubes having $d_t \sim 0.89$ nm, due to the relatively larger curvature of these narrow diameter tubes. The dominant contributions to the RBM mode frequency softening are attributed to temperature-induced softening of the intratubular C–C bond strength and to the SWNT intertubular (van der Waals) interactions [67]. Furthermore, it was found that these factors together contribute to the total change of $\omega_{\mathrm{RBM}}(T)$ with temperature, ~95% and 90%, respectively, for SWNTs having $d_t \sim 1.33$ nm and ~0.89 nm.

In summary, the effect of increased temperature on the Raman spectra of SWNT bundles is to soften all the SWNT phonon modes, some more than others, without a significant effect on the energy of the van Hove singularities.

3.3. Dependence of Debundling, Doping, and Applied External Potential

In this section we consider the effect on the Raman spectra of using chemistry to debundle SWNTs in SWNT bundles, and the spectral modifications resulting from p- and n-type doping and from the application of positive and negative potentials to the SWNT bundles.

In addition to their exotic physical properties, sp^m ($2 < m < 3$) hybridized carbons have been found to exhibit interesting chemical properties [10, 69]. For instance,

Table 2. Calculated energy differences δE_{11} and δE_{22} (in eV) in the electronic DOS for isolated tubes (tube) and weakly interacting tubes (rope) in a bundle (see text).

(n, m)	E'_{11}(tube)	E_{11}(rope)	δE_{11}	E'_{22}(tube)	E_{22}(rope)	δE_{22}
(8,8)	2.03	2.23	0.20	3.69	3.89	0.20
(9,9)	1.83	1.99	0.16	3.39	3.55	0.16
(10,10)	1.64	1.84	0.20	3.12	3.29	0.17
(11,11)	1.49	1.63	0.14	2.87	2.98	0.11

Source: Reprinted with permission from [64], A. M. Rao et al., *Phys. Rev. Lett.* 86, 3895 (2001). © 2001, American Physical Society.

[1] In [67], the relation $\omega_{\mathrm{RBM}} = 223/d_t + 14$ is used, yielding $d_t = 1.33$ and 0.89 nm for the tubes with $\omega_{\mathrm{RBM}} = 182$ and 264 cm^{-1}, respectively, while use of the formula $\omega_{\mathrm{RBM}} = 248/d_t$ for isolated SWNTs yields d_t values of 1.36 and 0.94 nm, respectively.

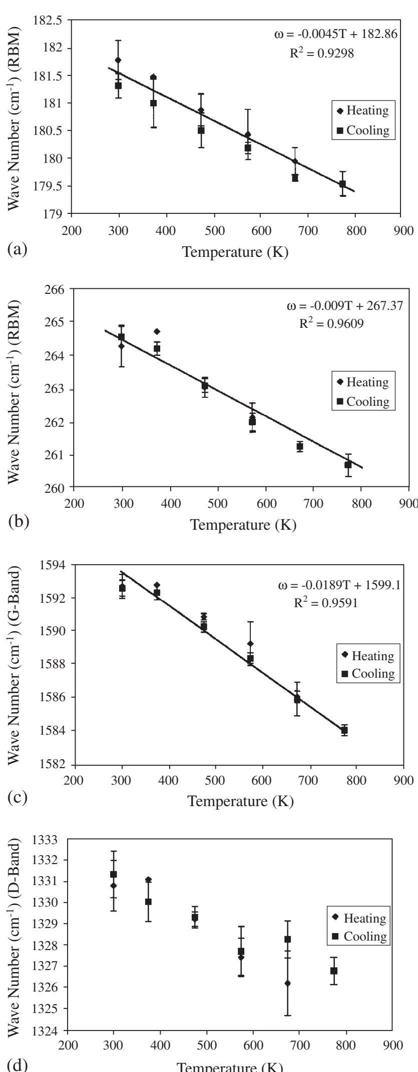


Figure 15. Temperature dependence of Raman frequencies in cm^{-1} of HiPCo SWNTs in SWNT bundles plotted as a function of temperature (K) for the RBM mode at 182 (a) and at 264 cm^{-1} (b) at room temperature, and for the G-band and D-band features in (c) and (d), respectively [67]. Over the temperature range shown in this figure, all modes show a reversible linear shift with temperature upon heating and cooling. Reprinted with permission from [67], N. R. Raravikar et al., *Phys. Rev. B* 66, 235424 (2003). © 2003, American Physical Society.

fullerene chemistry blossomed into a rich subfield, allowing researchers to develop recipes for their purification, chromatographic separation, and functionalization in order to yield fullerenes with tailored chemical properties (see Chap. 10 in [10]). Since carbon nanotubes can be viewed as highly elongated fullerene molecules, it was natural to expect that the chemical properties of carbon nanotubes would be explored soon after the discovery of bulk nanotube synthesis methods for SWNT bundles by the pulsed vaporization [48] and electric arc methods [49]. A major impediment of nanotube chemistry is the insolubility of as-prepared carbon nanotubes in known organic solvents. This chemical inertness stems from the fact that a large fraction of the carbon atoms in a nanotube shell exhibits strong covalent bonding with neighboring atoms, as is the case in a single graphene sheet. Toward the end of 1997, recipes were being developed to chemically purify, debundle, and chop nanotubes into short "fullerene pipes" with a view toward preparing solubilized nanotubes. The following year, a major breakthrough was reported, which revealed that nanotubes attached with long chain octadecylamine molecules become soluble in CS_2 and other organic solvents [70]. AFM images showed that the majority of the bundled SWNTs were separated into isolated tubes and small SWNT bundles (2–4 nm in diameter) during the solubilization process [70]. Henceforth in this chapter, solubilized tubes are called debundled tubes.

Besides literally opening doors to the chemical modification of nanotubes, fullerene pipes and debundled tubes provide an opportunity to investigate the effects of finite-tube length [53] and of debundling on the vibrational and electronic properties of nanotubes [64, 71]. The tube–tube interactions within SWNT bundles are weak, similar to the coupling between adjacent graphene planes in 3D crystalline graphite or the intermolecular coupling found in solid C_{60}. This weak intertube coupling is dominated by the van der Waals (vdW) interaction but contains a nonzero covalent component that has been shown theoretically and experimentally to have a significant influence on the vibrational [16, 72–75] and electronic states for carbon nanotubes [76, 77].

In this section, we focus on Raman spectroscopic studies of debundled SWNTs to probe the effect of vdW interactions on the RBM band and on the G-band frequencies ω_{RBM} and ω_G. Almost all of the early Raman spectra on SWNTs were collected on nanotube bundles that were synthesized by the pulsed laser vaporization or the electric arc methods. These bundles contain on the order of 100 well-aligned SWNTs arranged in an approximately closed-packed triangular lattice. A typical Raman spectrum of the arc-derived SWNT bundles obtained using the 1064 nm excitation wavelength (see Fig. 16) exhibits two prominent features at $\omega_{\mathrm{RBM}} \sim 160$ cm^{-1} (radial band) and $\omega_{G^+} = 1591$ cm^{-1} (tangential band). The frequencies of the peaks in the RBM bands have been used extensively to determine the SWNT diameter distribution (see Section 2.3). The line-shape of the G band has been used to determine the semiconducting/metallic nature of SWNTs (see Section 3.1).

In Figure 16, the Raman spectrum of debundled tubes in CS_2 (solid trace) is compared to that obtained for bundled SWNTs (dotted trace), the latter in powder form. The debundled and bundled SWNTs referred to in Figure 16 stem from the same as-prepared arc-derived material. Three G-band modes near 1590 cm^{-1}, previously identified with A_{1g}, E_{1g}, and E_{2g} symmetries (see Section 3.1), are expected for achiral tubes [6] with frequencies that are nearly independent of the tube diameter [20], while for chiral tubes, six G-band modes are Raman-active with $2A_1$, $2E_1$, and $2E_2$ symmetries [20]. In contrast to the high frequency G-band at 1590 cm^{-1}, the low frequency RBM band centered at $\sim$160 cm^{-1} is identified with a band of A_1 symmetry radial breathing modes representing contributions from the various SWNTs within the resonance window of the excitation laser, each contributing at a frequency in accordance with its diameter $\omega_{\mathrm{RBM}} = \alpha/d_t + \Delta\omega_{\mathrm{RBM}}$ [6]. Consistent with the calculated mode frequencies and intensities, other weak

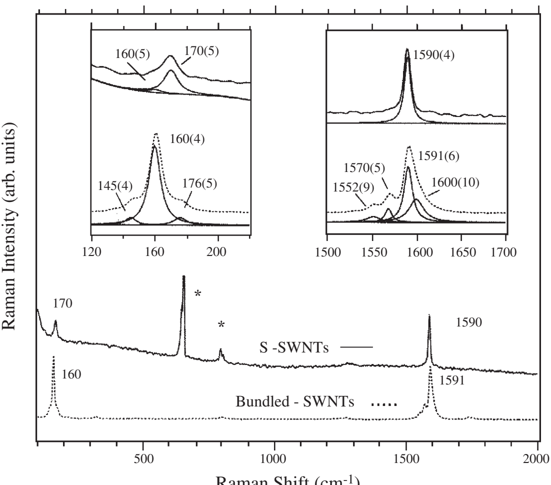


Figure 16. Room temperature Raman spectrum of arc-synthesized bundled SWNTs (dotted curve) and of solubilized SWNTs (solid curve) denoted by S-SWNTs in CS_2. The laser excitation wavelength is 1064 nm. The peaks denoted by asterisks are attributed to CS_2. The *G*-band peak intensities are set to a constant value so that the intensities for the RBM bands for the two spectra can be compared. The left and right insets show, respectively, the deconvolution of the RBM and *G*-band features in the bundled SWNTs and S-SWNTs. The vertical and horizontal axes for the two insets correspond, respectively, to the Raman intensity and frequency. The numbers within parentheses refer to the full width at half maximum intensity of the various constituents of the deconvolved bands. Reprinted with permission from [64], A. M. Rao et al., *Phys. Rev. Lett.* 86, 3895 (2001). © 2001, American Physical Society.

Raman-active features have been observed in the intermediate phonon frequency region between the RBM and *G*-band features [53].

At least two striking differences in the Raman spectra of bundled versus debundled tubes stand out in the spectra of Figure 16. The first difference is the narrower linewidths for the RBM and *G*-bands of the debundled SWNTs in CS_2 relative to those observed in bundled SWNTs, albeit the linewidth of the each Lorentzian component is nearly the same for bundled and debundled tubes. Likewise the linewidths for the G^+-band components for isolated tubes are also considerably smaller than for SWNT bundles, as is also the case for the RBM band, because of the large decrease in inhomogeneous line broadening for isolated SWNTs. In fact, the linewidths for the RBM and *G*-band features for debundled SWNTs can be as narrow as for isolated SWNTs [22]. This is also shown by the Lorentzian lineshape analysis of the *G*-band (right inset of Fig. 16) which reveals at least four narrow components for the bundled SWNTs and a single narrow mode for debundled SWNTs in CS_2. The loss of the G^--band in the debundled spectrum is apparent and can be understood from studies at the single nanotube level which show the suppression of the G^--band in isolated semiconducting tubes due to the antenna effect [1]. Similarly, from the left hand inset in Figure 16 for the RBM band, a reduced number of Lorentzians is observed in the spectrum of the S-SWNT sample when compared to that of the bundled SWNTs, indicating a much narrower tube diameter distribution in the debundled sample.

The second difference revealed in Figure 16 is the upshift of ω_{RBM} for bundles with respect to the corresponding value in isolated tubes due to the space restrictions imposed by the presence of neighboring tubes in a SWNT bundle. The tube–tube interactions within a SWNT bundle have been predicted to cause a ~6–20 cm^{-1} upshift in ω_{RBM}, depending on the detailed considerations of the various theoretical calculations [74, 75, 77]. Figure 16 shows that the RBM band in debundled tubes is upshifted by ~10 cm^{-1} relative to the RBM band in bundled tubes, in apparent contradiction with lattice dynamics predictions of a stiffening of the force constants in a bundled cluster of tubes relative to isolated tubes because the atomic displacements in a bundled tube are somewhat hindered by the radial atomic displacements of neighboring tubes.

In Figure 17, the RBM band (panel a) and the *G*-band (panel b) spectra taken at 1064, 647.1, and 514.5 nm excitations are compared for laser-synthesized bundled SWNTs and solid laser-synthesized debundled tubes (i.e., SWNTs that are first solubilized along their full lengths by organic solvents which are then evaporated). The RBM bands in these solid debundled tubes are consistently upshifted relative to those of SWNTs that are in bundles, as can be seen in all the spectra depicted in Figure 17. On the other hand, the narrowing of the *G*-band observed in the solubilized debundled tubes in CS_2 (see right-hand inset in Fig. 16) is absent in the tangential *G*-band in solid debundled tubes (Fig. 17). Nevertheless, there is essentially no change in the *G*-band peak frequencies as a result of debundling within the resolution of the Raman experiments.

At least two important conclusions can be drawn from the Raman data presented in Figure 17:

(i) Both semiconducting and metallic tubes exhibit about a 10 cm^{-1} upshift in ω_{RBM} when the bundled tubes are solubilized. As Pimenta et al. [57, 58] have argued, the 647.1 nm excitation couples predominantly to metallic tubes in a sample with a mean SWNT diameter of ~1.4 nm.

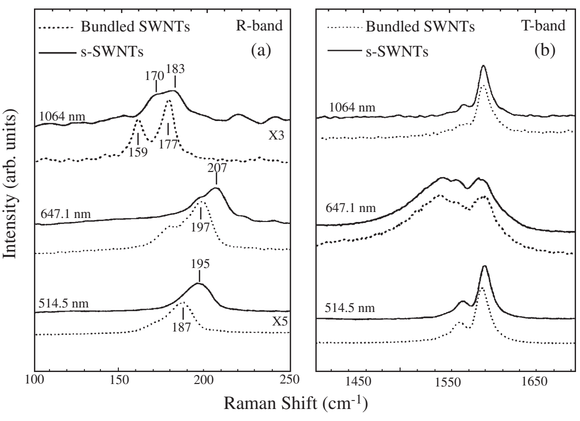


Figure 17. Comparison of the radial breathing mode band (labeled R-band) (panel a) and the tangential *G*-band (labeled T-band) (panel b) of laser-synthesized bundled SWNTs and subsequent solid debundled tubes based on these bundled SWNTs for three different laser excitation wavelengths. Reprinted with permission from [64], A. M. Rao et al., *Phys. Rev. Lett.* 86, 3895 (2001). © 2001, American Physical Society.

(ii) The relatively broader tangential G-band in solid S-SWNTs compared to that of debundled tubes in CS_2 suggests that the organic solvent–nanotube interactions have a smaller influence on the tangential mode lifetime than the tube–tube interactions in solid debundled tubes.

Duesberg et al. [27] have also reported an upshift in ω_{RBM} in their Raman spectrum collected on SWNT bundles of different thicknesses. In Figure 18, the 5 nm thick SWNT bundle (#3 bundle in Fig. 18) showed $\omega_{\mathrm{RBM}} \sim 188$ cm^{-1} while a 1.5 nm thick bundle (#1 bundle in Fig. 18) showed $\omega_{\mathrm{RBM}} \sim 193$ cm^{-1}. These observations are consistent with those discussed in Figures 16 and 17.

We now consider two other perturbations to SWNT bundles which modify their Raman spectra. First we consider the effect of an applied potential, and then we consider the case of doping the SWNT bundles with donors and acceptors. Both of these experiments provide insights into how external perturbations modify the observed Raman spectra.

Electrochemical studies of the resonance Raman spectra of SWNT bundles show that the electrochemical potential, Fermi level, and charge carrier density of SWNTs can be controlled by an externally applied potential, thereby radically changing the Raman spectra as initial or final states become unavailable for the resonance Raman effect [78]. Figure 19a shows the change in the spectra at $E_{\mathrm{laser}} = 1.96$ eV for SWNT bundles ($d_t = 1.25 \pm 0.20$ nm) cast on a Pt substrate immersed in an H_2SO_4 (0.5 M) aqueous solution as the external voltage V is varied from 0 to +1.3 V and then back to 0 V. While the dependence of ω_D and ω_G on V was studied, it was found that the dependence of $\omega_{G'}$ on V could be explained quantitatively in terms of the lowering of the Fermi level with increasing V, thereby emptying states from van Hove singularities in the valence band and, in the case of metallic tubes, also from the constant nonzero density of states between the lowest energy pair of van Hove singularities E_{11}^M, as shown in Figure 19b. The good agreement

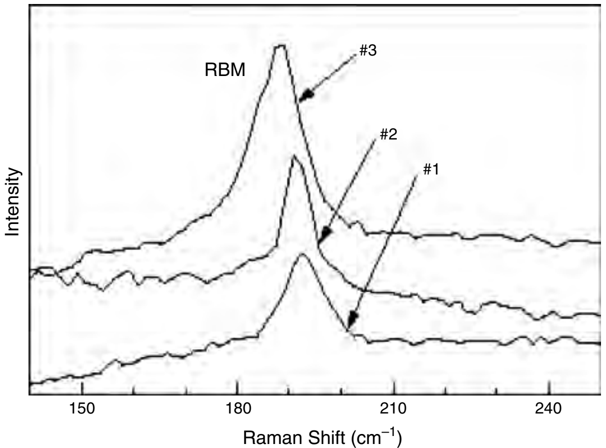


Figure 18. Raman spectra of the radial breathing mode using $E_{\mathrm{laser}} = 1.96$ eV for three different thicknesses of SWNT bundles, with #2 and #3 having a thickness of 5 nm, with #1 having a thickness of 1.5 nm. The results show a dependence of ω_{RBM} and its intensity on the thickness of the SWNT bundle. Reprinted with permission from [27], G. S. Duesberg et al., *Phys. Rev. Lett.* 85, 5436 (2000). © 2000, American Physical Society.

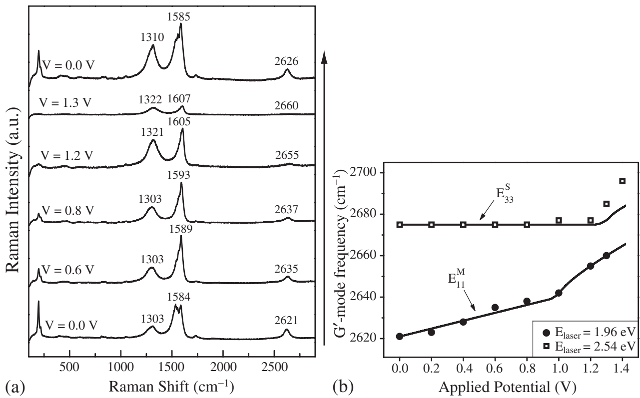


Figure 19. (a) *In-situ* Raman spectra of a SWNT film cast on a platinum surface in an H_2SO_4 0.5 M aqueous solution. The spectra are obtained at the indicated positive applied potentials and $E_{\mathrm{laser}} = 1.96$ eV. The frequencies for the dominant G-band, D-band, and G'-band features are indicated in cm^{-1}. (b) The points denote G'-band frequencies for the same SWNT film as in (a), obtained at the indicated applied potentials for $E_{\mathrm{laser}} = 1.96$ and 2.54 eV. The solid lines in (b) are theoretical curves for the G'-band calculated from charge transfer in SWNT density of states. Reprinted with permission from [78], P. Corio et al., *Chem. Phys. Lett.* 370, 675 (2003).

between experiment and theory regarding the dependence of $\omega_{G'}$ on V and the observation of a different behavior for semiconducting and metallic tubes indicates that the applied voltage V has only a small effect on the energy of the van Hove singularities, but the application of V mainly affects the occupation of valence and conduction band states [78].

Closely connected to these electrochemical experiments are doping experiments where either electrons or holes are added to the nanotubes by the addition of either donor or acceptor dopant species. For the case of saturation-doped SWNT bundles with halogen acceptors (for example, Br_2), the ω_{RBM} and ω_{G^+} frequencies upshift, respectively, by ~74 and ~24 cm^{-1}, relative to the corresponding frequencies in the undoped SWNT bundles. On the other hand, doping with alkali metals like K or Rb leads to a softening (or downshift) of ω_{RBM} and ω_{G^+}, accompanied by dramatic changes in the lineshape for the tangential G-band. For example, the G-band in saturation K-doped or Rb-doped bundles, ω_{G^+}, downshifts by ~35 cm^{-1} and exhibits a characteristic Breit–Wigner–Fano lineshape [80]. The radial breathing mode band in saturation K-doped or Rb-doped bundles is not evident in the Raman spectrum. The RBM band may have shifted to very low frequencies or it may have broadened to such an extent that it cannot be detected above the background signal. These shifts provide evidence for charge transfer between the dopants and SWNT bundles, indicating an ionic character for the doped samples. Doping with halogens transfers electrons from the π states in the SWNTs to the halogens, creating hole carriers in the SWNTs (p-type). Likewise, alkali-metal dopants transfer electrons to the π^* states, making the nanotubes n-type. The authors explained their results within a rigid band model, assuming no modification to the E_{ii} values for their doped SWNTs.

We now offer some explanation for the information provided by these Raman studies on the effect of these perturbations on the nanotubes. With regard to interpreting

the debundling effects discussed previously and depicted in Figures 16, 17, and 18, we note that the shifts in ω_{RBM} or in ω_G observed in doped nanotube bundles have been related to the charge transfer between the nanotubes and the alkali metal (downshift) or halogen (upshift) dopants [79]. In this case, the van Hove singularities do not change much in energy, but the Fermi level (or the filling of electron and hole states) is greatly affected, resulting in large shifts in G-band properties and in the suppression of interband transitions. The effect of positively and negatively applied potentials is similar to that of chemical doping insofar as the values of E_{ii} do not change much, but large shifts in Fermi level (the filling and emptying of electron states) are observed. In contrast, the debundling of SWNTs mainly affects the tube–tube coupling and sensitively affects ω_{RBM} but not ω_G. Clearly, for the spectra depicted in Figure 16, the 10 cm^{-1} upshift in ω_{RBM} upon forming debundled from bundled SWNTs cannot be attributed to charge transfer, since ω_G for debundled tubes exhibits an insignificant frequency shift from the observed frequency in bundled SWNTs. Furthermore, the 10 cm^{-1} upshift in Figures 16 and 17 is not due to the presence of long chain molecules (which are attached to the tube walls or to the tube ends during the debundling process) [70], since the effect of debundling on ω_{RBM} should result in a downshift in frequency, since the attachment of molecules to the SWNT would be expected to weaken the carbon–carbon force constants involved in the normal mode vibrations of the SWNTs. The anomalous upshift of ~10 cm^{-1} has therefore been attributed to a decrease in the energy spacing of the van Hove singularities in debundled tubes relative to the spacings in bundled tubes, thereby allowing the same laser excitation to excite SWNTs with smaller diameters in the debundled SWNTs relative to the bundled SWNTs [64].

3.4. Pressure-Dependent Studies

The study of Raman-active modes in SWNTs as a function of external pressure has been useful in probing both the influence of vdW (van der Waals) interactions on the vibrational modes in carbon nanotubes [80–82] and the pressure-induced structural transitions in SWNTs [80, 82]. Pressure-induced changes in the electrical [83] and optical [84] properties of SWNT bundles have also been reported. In the high-pressure experiments reported to date, SWNT material synthesized either by the electric arc [49] or pulsed laser vaporization [48] methods has been used.

Generalized tight binding molecular dynamic (GTBMD) simulations predicted different pressure dependences for ω_{RBM} in bundled and isolated tubes [80]. These GTBMD calculations showed that the pressure dependence for ω_{RBM} is sensitive to the coupling between the tubes in the SWNT bundles. In the GTBMD scheme, a Lennard–Jones type potential was used to describe the vdW forces between SWNTs in a bundle, using Lennard–Jones parameters similar to those needed to simulate the c-axis bonding in bulk graphite. In the calculations, the hydrostatic pressure, P, was introduced through a radial force $F_P = P \times A$, where A is a cross-sectional area perpendicular to the tube axis [80]. Three scenarios (I, II, and III) by which the external applied pressure can be transmitted to a SWNT bundle were considered (Fig. 20). In model I, the pressure transmitting liquid resides external to the SWNT bundle and pressure is transmitted only to the outer tubes, while the tubes in the interior of the bundle are coupled to the external force through vdW interactions. Model II describes a situation in which the vdW interactions were neglected and the pressure transmitting liquid penetrates the bundle via the interstitial channels to exert uniform pressure on individual tubes within the bundle. This model also applies to isolated tubes, where the pressure transmitting liquid can freely access the entire perimeter of each tube and the vdW interactions are thereby strongly reduced. Finally, model III is identical to model II except for the inclusion of the vdW interactions in the GTBMD calculations.

Large SWNT bundles (containing more than 100 tubes) produced by pulsed laser vaporization of graphite were debundled using chemical methods described in [70]. This procedure resulted in some isolated tubes, but the majority of the tubes were in small bundles containing three to seven tubes, as seen in the height profile analysis of AFM images. A schematic representation of the debundled SWNT sample containing a smaller number of tubes is shown in Figure 21, where the number within the tube represents the number of neighboring tubes. For example, an isolated tube has zero neighboring tubes, and within a bundle of three tubes, each tube has two neighboring tubes. Also the peripheral tubes in a bundle of seven tubes have three neighboring tubes, while the central tube has six neighboring tubes, as shown in the figure.

In Figure 22, the pressure-induced frequency shifts of the RBM (panel a) and the tangential G-band (panel b) frequencies for debundled SWNTs (squares) are compared to those for bundled SWNTs (triangles). In this figure, the difference between the observed frequency at any given pressure and the atmospheric pressure frequency, that is, $[\omega(P) - \omega(0)]$, is plotted along the vertical axis for convenience of comparison [85]. This choice for the ordinate enables us to eliminate the shifts in ω_{RBM} between the bundled and debundled SWNTs, which was discussed in Figures 16 and 17. In Figure 22, the two dominant G-band modes (G^+ and G^-) observed in both bundled and debundled SWNTs are labeled as T_2 and T_1, corresponding to G^+ and G^-, respectively. The solid lines represent fits to the experimental data (linear in panel a; quadratic in panel b in Fig. 22) while the dashed curves represent theoretical pressure dependences for ω_{RBM} and ω_G from models I–III described previously. Interestingly, the experimentally observed pressure dependence for both ω_{RBM} and

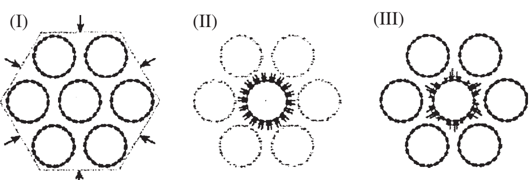


Figure 20. Three scenarios (I, II, and III) by which the external applied pressure can be transmitted to a SWNT bundle (see text). Reprinted with permission from [85], U. Schlecht et al., *J. Nanosci. Nanotechnol.* 3, 139 (2003). American Scientific Publishers.

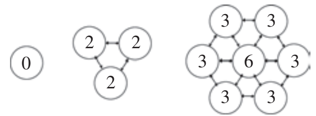


Figure 21. A schematic representation of the number of neighbors in debundled SWNT samples containing a small number of tubes, namely one, three, and seven tubes (see text). Reprinted with permission from [85], U. Schlecht et al., *J. Nanosci. Nanotechnol.* 3, 139 (2003). American Scientific Publishers.

ω_G in debundled SWNTs is nearly the same as in bundled SWNTs.

It should be noted that the GTBMD calculations (left panel in Fig. 22) predict different pressure dependences for ω_{RBM} in bundled (model I) and isolated tubes (models II and III). For the pressure dependence of ω_{RBM} in debundled SWNTs, Figure 22 shows an experimentally determined linear shift of $\sim$8.4 cm^{-1}/GPa, which is comparable to the values of 9.7 [81, 86] and 10.1 cm^{-1}/GPa [82] reported, respectively, for $d\omega_{\mathrm{RBM}}/dP$ in arc-derived and laser-derived bundled tubes. Thus, the experimentally determined pressure coefficient for ω_{RBM} in debundled SWNTs is significantly higher than the pressure dependence of $\sim$1.3 cm^{-1}/GPa for ω_{RBM} predicted by model II for isolated tubes [85]. This would imply that the magnitude of the vdW interactions on ω_{RBM}, as reflected in the pressure dependence of the Raman-active radial breathing mode, is nearly the same whether a SWNT resides within a large bundle (with 100 neighboring tubes) or inside a small bundle containing three to seven tubes, indicating the greater importance of near-neighbor intertube interactions. The Raman data described in Figures 16 and 22 collectively suggest that the debundling into small diameter bundles causes a 10 cm^{-1} upshift in ω_{RBM} but does not lead to a change in its dependence to external applied pressure. To unravel these subtle differences, further high pressure Raman studies on a homogeneous sample of isolated SWNTs prepared either by improved chemical methods or by chemical vapor deposition are required.

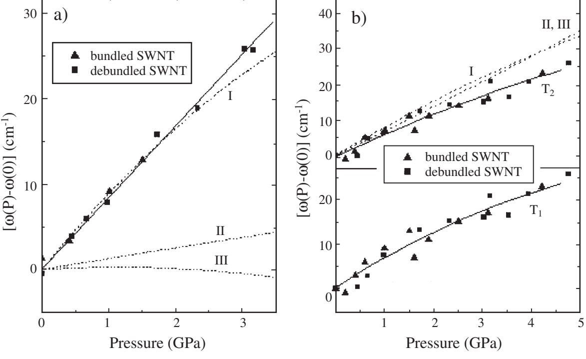


Figure 22. Pressure dependence of Raman modes in solubilized SWNTs and bundled SWNTs. The pressure-induced frequency shifts of the ω_{RBM} (panel a) and tangential G-band frequencies ω_{G^+} and ω_{G^-} (labeled T_2 and T_1, respectively) (panel b) for debundled SWNTs (squares) to those for bundled SWNTs (triangles) are compared. The dashed curves represent the calculated pressure dependence for ω_{RBM} and ω_G from models I, II, and III. Reprinted with permission from [85], U. Schlecht et al., *J. Nanosci. Nanotechnol.* 3, 139 (2003). American Scientific Publishers.

Application of external pressure on a SWNT bundle can be expected to modify the electron density of states resulting in the "band tuning" phenomenon. Since Raman scattering from nanotubes is a resonant process, one might expect the laser excitation E_{laser} to resonate with different diameter tubes in the bundle as the external pressure is ramped from 0 to 5 GPa (Fig. 22). In panel (a) of Figure 22, the radial band cannot be discerned at external pressures above $\sim$3 GPa, while the tangential band persists even at 5 GPa. The loss of intensity of the radial band is most likely due to faceting (hexagonal distortion) of the nanotubes and a constraint on the radial atomic displacements of the carbon atoms by adjacent SWNTs, rather than due to a decrease in the resonance cross-section.

A direct measurement of the pressure dependence of the E_{11}^S and E_{22}^S transition energies is provided by the high pressure optical absorption studies of SWNT films by Kazaoui et al. [84], who experimentally observed that the optical absorption spectra of SWNTs changed reversibly due to external pressure, as shown in Figure 23. The features labeled S1 and S2 in Figure 23 were attributed to allowed electronic transitions E_{11}^S and E_{22}^S between the first and second pair of van Hove singularities in semiconducting SWNTs [84]. By monitoring the peak energy for S1, the authors found that the energy gap downshifts by only $\Delta E_{11}^S \sim$ 0.04 eV over the pressure range of 0–4 GPa (see inset to Fig. 23). This small shift in ΔE_{11}^S would predict a downshift of only 2 cm^{-1} in ω_{RBM} over the pressure range 0–4 GPa, which is much smaller than the experimental observation shown in Figure 22. This value of ΔE_{11}^S due to applied pressure is also relatively small compared to the energy shift of $\Delta E_{11}^M \sim$ 0.15–0.20 eV predicted from the Table 2 calculations [64] for the effect of bundling (intertube interactions) on E_{11}^M for metallic SWNTs (see Section 3.1). The measured ΔE_{11}^S due to applied pressure of 4 GPa is insufficient in magnitude to bring different diameter tubes into resonance with a given laser line E_{laser}.

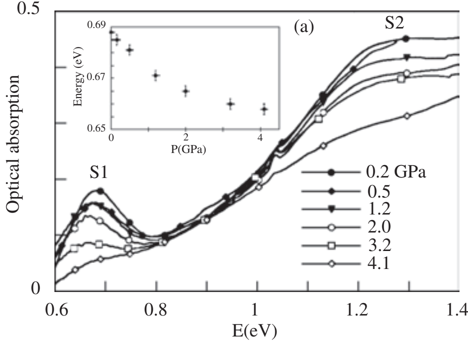


Figure 23. Room temperature optical absorption spectra of a SWNT film on KBr recorded by optical transmission measurements through the diamond anvil cell [84]. The insert shows the pressure dependence of ΔE_{11}^S as obtained from the measurements on the S1 absorption peak. Reprinted with permission from [84], S. Kazaoui et al., *Phys. Rev. B* 62, 1643 (2000). © 2000, American Physical Society.

3.5. Effects of External Perturbations on Raman Spectra

The effects of various perturbations to the Raman spectra have been reviewed in Sections 3.1–3.4, including the effects of bundling (Section 3.1), temperature (Section 3.2), debundling, doping, and applied voltage (Section 3.3), and pressure (Section 3.4). In this section, all of these effects are brought together and are discussed in terms of their effect on: (i) the Raman features of a given SWNT that is assumed to always be in resonance with a tunable laser (whose E_{laser} can be varied to satisfy the resonance condition) and (ii) the Raman measurements that are actually carried out.

First we consider the case of bundling the SWNTs which occurs in a self-assembly process during synthesis when the SWNT bundles are formed. Calculations show that the effect of bundling is to introduce an attractive intertube interaction, which for typical tube bundles (containing approximately 100 SWNTs in an ordered hexagonal arrangement) leads to: (i) an increase in the separation between the van Hove singularities and an upshift in E_{ii} by 0.15–0.20 eV and (ii) an increase in ω_{RBM} by $\sim$10 cm^{-1} [64]. From a physical standpoint, the effect of bundling and debundling mostly affects out-of-plane displacements, so that one would expect bundling and debundling to mainly affect ω_{RBM}, as is observed experimentally. When the experimental spectra on SWNT bundles are taken with the same laser line, then the ω_{RBM} of SWNT bundles would appear *downshifted* by $\sim$10 cm^{-1} relative to the isolated SWNTs, since different tubes of larger diameter would now be brought into resonance. In actuality, this experiment is not carried out, because the starting material is already in bundles.

The experiment that is actually carried out in the laboratory is the inverse experiment of debundling (see Section 3.3), which, from the theoretical argument given previously, has the effect of reducing the separation between the van Hove singularities, bringing into resonance smaller diameter SWNTs, thereby giving rise to a higher ω_{RBM}, in agreement with experimental observations.

These results also suggest that the coupling of the excitation laser energy (E_{laser}) is sensitive to the environment in which the nanotube resides, and thus future studies aimed at correlating nanotube diameters with ω_{RBM} should pay attention to the size of the bundles. Also, a detailed theoretical study of the dependence of E_{ii}, ω_{RBM}, and ω_G on bundle size is needed, including possible differences between semiconducting and metallic tubes. The techniques developed to fit ω_{RBM} to d_t through the relation $\omega_{RBM} = \alpha/d_t + \Delta\omega_{RBM}$ from measured diameter distributions using four or more laser lines [87] would be very useful in carrying out such systematic studies.

Another mechanism by which tubes move in and out of resonance with the excitation energy can be found in the electrochemical studies of SWNT bundles [78]. These experiments show that the electrochemical potential, Fermi level, and charge carrier density of SWNTs can be controlled by an externally applied potential, thereby radically changing the Raman spectra as initial or final states become unavailable for the resonance Raman effect as the applied potential is varied. The applied voltage V is found to affect the energy of the van Hove singularities E_{ii} only mildly, but the application of V mainly affects the emptying of states in the valence band and the occupation of states in the conduction band [78]. More detailed studies at the single nanotube level are needed to clearly establish the shifts in E_{ii} induced by application of a voltage V to isolated metallic and semiconducting SWNTs.

In closely related optical absorption experiments by Kazaoui et al. [88], essentially no shifts in the E_{ii} peak frequencies are observed upon donor doping with Cs and acceptor doping with Br_2, although an attenuation in the peak intensities is observed, consistent with the emptying and filling of states through acceptor and donor doping. Shifts in ω_{RBM} and ω_G with doping have been observed and have been correlated with the optical absorption data.

Other experiments by Kazaoui et al. [84] show that the optical absorption peak energy depends only weakly on pressure, although its intensity decreases strongly as the pressure is increased. This result implies that essentially the same SWNTs remain in resonance with E_{laser} as the pressure is increased. Because of this weak dependence of E_{ii} on pressure, the observed upshifts in ω_{RBM} and ω_G with pressure must be connected to increases in both the out-of-plane (RBM) and the in-plane (G-band) force constants of the *same nanotubes* with increasing pressure. The much larger compressibility of the tube bundles in the radial direction compared to the tangential direction results in a much larger percentage increase in ω_{RBM} as compared to ω_G with pressure, as is seen experimentally. Correspondingly, increasing the temperature [67] weakens the force constants for both the RBM and the G-band, without significantly shifting E_{ii}, so a given laser line remains in resonance with essentially the same SWNTs upon heating the sample.

Future experimental work is needed to provide a more systematic experimental framework which yields a more reliable model to account for the changes in the Raman spectra induced by the various perturbations to SWNTs discussed in Section 3, thereby allowing Raman spectroscopy to provide a sensitive and informative characterization tool for measuring these perturbations.

4. PEAPODS

Peapods refer to single wall carbon nanotubes containing C_{60} (or other fullerene) molecules within their cores. Early samples of peapods were discovered accidentally in high resolution transmission electron microscopy (HRTEM) studies of acid-purified laser-derived SWNT bundles [89–92]. Detailed HRTEM studies revealed the presence of structural defects induced in the tube walls due to acid treatment, which presumably served as entry ports for trace amounts of C_{60} that were coproduced during the nanotube synthesis. Soon after this discovery, an efficient route to prepare peapods became obvious, since nanotube researchers had already used mild oxidation conditions to open nanotube ends and fill the inner core of MWNTs with metals, such as Pb, by exploiting the capillary properties of nanotubes [93, 94]. It is thus not surprising to find that oxidized as-prepared SWNTs with open tube ends when exposed to C_{60} vapors could provide an efficient route for the synthesis of peapods based on vapor-phase transport and the capillary filling of fullerenes [95]. Filling fractions as high as 60% have now been demonstrated in the preparation of

peapods. Some authors have also prepared peapods using synthesis techniques similar to those used to prepare endohedral fullerenes. A controlled insertion of fullerenes inside SWNTs has been reported using plasma irradiation methods where various atoms (Li, Na, K) or molecules (C_{60}) can be encapsulated [96]. HRTEM studies of irradiated tubes showed that controlled amounts of defects could be created in the SWNT sidewall through ion irradiation, and that these defects could serve as additional entry ports (besides the open tube ends) for irradiated C_{60} into the nanotubes. Typical transmission electron microscope images of peapods are shown in Figure 24. A string of evenly spaced encapsulated C_{60} molecules, denoted by $(C_{60})_n$@SWNT, is clearly observed in panel A of Figure 24.

Raman scattering experiments on SWNT bundles [71, 75, 97] at ambient pressure (Figs. 16 and 17) and solubilized SWNTs [5, 85, 98–100] at high pressures (~5 GPa) inside a diamond anvil cell (see Fig. 22) revealed the effect of van der Waals interactions on the Raman-active modes in SWNT bundles [71, 97]. By comparing the experimentally observed radial breathing mode and tangential *G*-band frequencies in the high pressure Raman studies of SWNT bundles to those predicted (using a generalized tight binding model calculation) in the presence or absence of van der Waals interactions, Venkateswaran et al. [75] reported a negligible shift in the tangential *G*-band frequency, but a ~14 cm^{-1} upshift in the radial band frequency due to tube–tube interactions [71, 97]. Likewise, the Raman investigations of peapods could in principle probe the effect of the encapsulation of fullerene molecules on the vibrational and electronic properties of the host SWNTs.

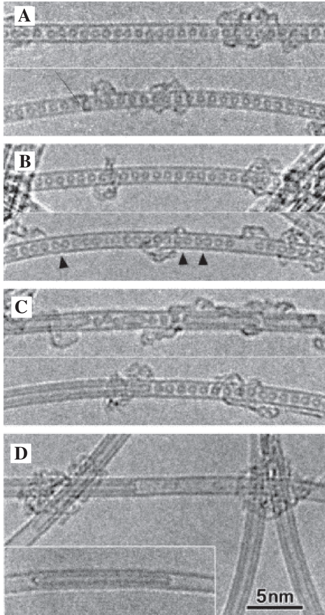


Figure 24. Structure of peapods taken by HRTEM [95]. (A) As prepared peapods. (B) The formation of dimers at 800 °C. (C) The formation of DWNT segments as the C_{60} molecules coalesce at 1000 °C. (D) The complete coalescence of C_{60} molecules to form DWNTs at 1200 °C. The inset in (D) shows inner tubes with terminating end caps. Reprinted with permission from [95], S. Bandow et al., *Chem. Phys. Lett.* 337, 48 (2001). © 2001, Elsevier Science.

A detailed Raman study using 10 different laser excitation energies (over the range 1000–413 nm or 1.24–3.0 eV) of empty and fullerene filled SWNTs has been reported by Pfeiffer et al. [101]. Laser-derived SWNTs were used in their preparation of peapods with approximately 60% C_{60} occupancy. Figure 25 shows a collection of Raman spectra obtained at 80 K on empty and C_{60} filled SWNTs. Interestingly, the RBM band in the peapod bundles is qualitatively similar to that observed in empty SWNT bundles and is found to be downshifted by only ~2 cm^{-1} relative to that of empty SWNT bundles. The authors attribute this small frequency downshift to a slightly increased peapod diameter or to a change in the intertube interactions [101]. The tangential *G*-bands of the empty and C_{60} filled SWNTs are also qualitatively similar without a noticeable shift between the two Raman spectra. Similar to the response of empty SWNTs, the peapods exhibit diameter selective resonance Raman scattering [5] (as shown in Fig. 25) and exhibit a strong dependence of the tangential band lineshape on excitation energy (narrow tangential band for the blue–green excitations and Fano-distorted lineshape for red laser excitation) [101]. The Raman spectra for empty SWNTs when *subtracted* from the corresponding Raman spectra for peapods reveal interesting Raman features in the vicinity of the pentagonal pinch mode frequency of C_{60} (Fig. 26) [101]. Two weak modes at 1466 and 1474 cm^{-1} are found with an intensity that is 20 times lower than the intensity of the *G*-mode of SWNTs. The relative intensities of these weak modes change as a function of the excitation energy (Fig. 26). Since each peak in the doublet observed in Figure 26 lies on either side of the expected 1469 cm^{-1} peak for C_{60}, it was concluded that encapsulated C_{60} seems to be strongly affected by the encapsulation, perhaps related to the orientation of the pentagons and hexagons of C_{60} with respect to the closest contact of the C_{60} molecule with respect to the SWNT hexagons, though a detailed interpretation for the origin of two weak peaks rather than one is presently lacking.

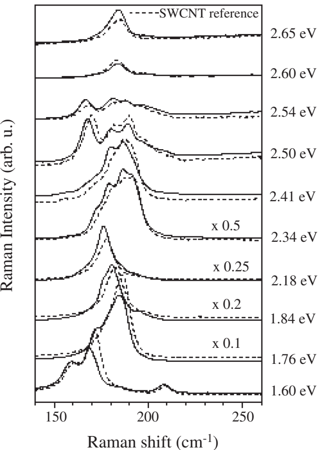


Figure 25. Raman spectra taken in the RBM region at 80 K on an ensemble of peapods for various E_{laser} excitation energies. The dashed lines denote the spectra from similar SWNT bundles prior to the introduction of fullerenes into the SWNTs. Reprinted with permission from [101], R. Pfeiffer et al., *Diamond Relat. Mater.* 11, 957 (2002). © 2002, Elsevier Science.

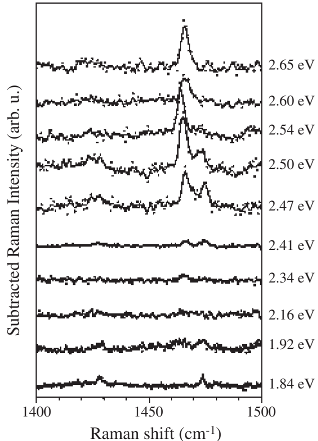


Figure 26. Raman spectra obtained from peapod bundles in the vicinity of the $A_2(g)$ pentagonal pinch mode of encapsulated C_{60}, plotted after *subtracting* the Raman signal of empty SWNTs from the Raman signal of the corresponding peapod sample. Several laser excitation energies were used in the experiment and the data were collected at 80 K. Reprinted with permission from [101], R. Pfeiffer et al., *Diamond Relat. Mater.* 11, 957 (2002). © 2002, Elsevier Science.

In contrast with observations reported by Pfeiffer et al. [101], a comparative Raman study (using 514.5 nm excitation) was done with a purified SWNT bundle reference sample and a peapod sample prepared using the irradiation method [96]. The results of this study showed the presence of the C_{60} A_{1g} Raman peak in the peapod spectrum at 1469 cm^{-1} [96]. From the positions of the peaks in the RBM band, the authors estimated a 1.2–1.3 nm tube diameter, in contrast to the diameters determined from TEM (1.2–1.8 nm). This discrepancy in diameter evaluation is largely due to the fact that nanotubes exhibit resonant Raman scattering, and several laser excitation energies should be used to probe the full range of tube diameters in the peapod sample. Other notable features reported in their work [96] included a low intensity for the 1353 cm^{-1} D-band feature, signifying that little damage is caused to SWNTs by the low level of alkali metal plasma irradiation flux [(1–5) × 10^9 cm^{-3}] used in their experiment. Regarding the C_{60} modes, the low frequency H_g mode, the radial breathing C_{60} mode $A_g(1)$, and the pentagonal pinch $A_g(2)$ mode appear, respectively, at 273, 497, and 1469 cm^{-1}, which are very close to the frequencies for free C_{60} molecules, suggesting a weak interaction between the encapsulated C_{60} and the SWNTs.

Electrochemistry offers more favorable conditions for both reductive and oxidative doping than chemical doping methods. A thin film electrode was prepared by evaporating an ethanolic slurry of peapods on Pt or on an ITO (indium tin oxide) transparent electrode to simultaneously monitor the electronic and vibrational properties of peapods [102]. In this work, the peapods on ITO served as a working electrode for *in-situ* visible near-infrared spectroelectrochemistry spectra taken on peapods as the population of the valence-band electronic states was tuned electrochemically in a solution of 0.2 M $LiClO_4$ in acetonitrile. The electrochemical cell was assembled inside a glove box with a glass optical window for *in-situ* Raman characterization. The intensities of the RBM and G-band modes for the peapod SWNT bundles were observed to decrease as a result of cathodic (panels a) or anodic (panels b) doping as shown in Figure 27, similar to the behavior observed for empty SWNT bundles. The lower frequency RBM component (165–175 cm^{-1}) is less affected by the positive and negative potentials as compared to the higher frequency RBM component (180–190 cm^{-1}), revealing a diameter dependence of the electronic transition energies (i.e., the so-called E_{33}^S transition, between van Hove singularities, E_3^v and E_3^c in the valence and conduction band density of states), whereby wide diameter tubes are depleted of/filled with carriers before those of narrower diameter tubes, as the potential is changed toward more positive/negative values. In analogy to empty tubes, the tangential G-band significantly and monotonically upshifts with increasing anodic potential (panel b) and is less sensitive to cathodic charging (panel a). In contrast to empty SWNT bundles [79], no detectable shifts of the RBM peaks is observed upon progressive cathodic or anodic charging of the peapod sample (Fig. 27). Clearly, the intensities of the RBM peaks decrease as a result of cathodic or anodic charging. These doping-induced behav-

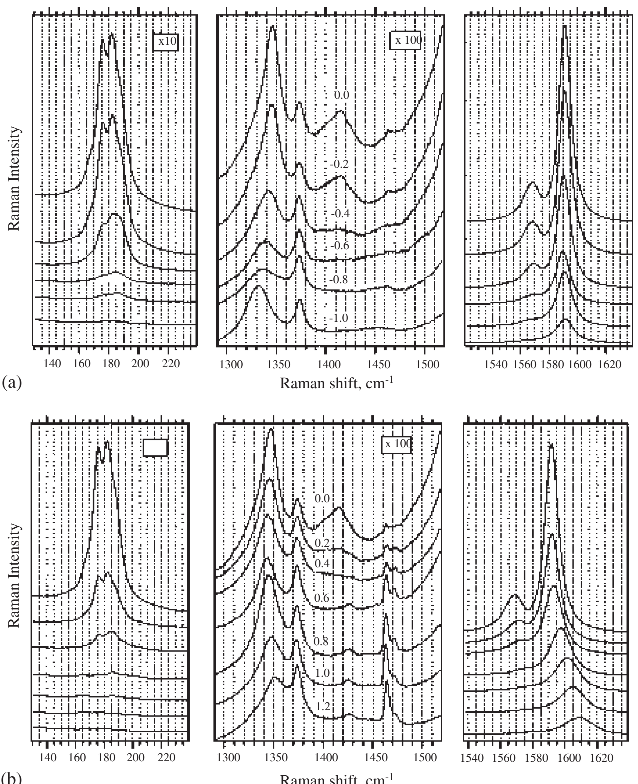


Figure 27. Raman spectra of peapods on Pt electrode (excited at 2.41 eV) in 0.2 M $LiClO_4$ + acetonitrile. (a) Spectra taken for various electrode potentials (vs Ag-pseudo-reference electrode) varying from 0 to –1.0 V (cathodic potentials). The spectra are offset for clarity, but the intensity scale is identical for all spectra in the respective window. As indicated, the intensities are multiplied by a factor of 10 and 100 for the first and second charts, respectively. The peak at 1375 cm^{-1} belongs to the δC–H vibration in acetonitrile. (b) The corresponding spectra taken for various electrode potentials varying from 0 to 1.2 V (anodic potentials). Reprinted with permission from [102], L. Kavan et al., *Chem. Phys. Lett.* 361, 79 (2002). © 2002, Elsevier Science.

iors of the RBM have been attributed to a sharp tube diameter distribution in the peapod sample, and a quenching of the resonance Raman cross-section due to bleaching of the transition between van Hove singularities [102].

Doping-induced shifts in the *D*-band and *G*-band in empty SWNTs are well understood [79, 88], based on charge-transfer-induced charge in the high-frequency interlayer mode (E_{2g_2}) in acceptor and donor graphite intercalation compounds (for a review see [69]). Softening (stiffening) of the E_{2g_2}-derived modes in empty SWNT bundles occurs with increasing charge transfer into (from) the π^* (π) bands. The cathodic downshift per electron and anodic upshifts per hole equals 140 and 460 cm^{-1}, respectively, in graphite. While chemical doping of empty SWNTs exhibits similar doping-induced trends, the effect of cathode charging of peapods [panel (a) in Fig. 27] shows exceptions which are unresolved to date. While the *D*-band softens by ~15 cm^{-1}, the G^--band essentially remains unchanged in frequency, exhibiting a gradual loss in peak intensity with cathodic charging. The zero frequency shift in the G^--band is an unexpected result which needs to be addressed in further electrochemical doping experiments of peapods. The loss in peak intensity is consistent with a loss of Raman cross-section for the G^--band in empty SWNTs [79]. The anodic charging of peapods also shows similar inconsistencies. For example, the G^--band upshifts by ~24 cm^{-1} with a concomitant loss in peak intensity as expected. However, the *D*-band frequency remains essentially unshifted in frequency [panel (b) in Fig. 27].

Like Pfeiffer et al. [101], the strongest Raman line of C_{60} in peapods is observed at 1465 cm^{-1} with a satellite line at 1474 cm^{-1}. A trace of the $H_g(7)$ line at 1430 cm^{-1} is also visible in the central panels in Figure 27. The overall effect of cathodic doping leads to decreased Raman peak intensities, and the C_{60} Raman features eventually drop below detectable limits. On the other hand, anodic doping leads to an increase in the Raman intensity of the $A_g(2)$ mode at 1465 cm^{-1} with a gradual drop in intensity of the satellite peak at 1474 cm^{-1}. This is an interesting observation which may be explained on the basis of the Raman spectra of doped C_{60} prepared by chemical methods.

Several experiments have shown that solid C_{60} can be reduced by alkali metal doping to form A_xC_{60} compounds, where $x = 1, 3, 4$, and 6 [10]. The pentagonal pinch mode frequency observed in solid C_{60} at 1469 cm^{-1} is found to downshift in frequency by approximately 6 cm^{-1} per electron which is transferred from the alkali metal dopant to the host C_{60} solid. In addition, the intensity of pentagonal pinch mode decreases dramatically in A_xC_{60} relative to that in undoped pristine C_{60} [10]. The disappearance of the pentagonal pinch mode in the case of cathode charging of peapods [panel (a) in Fig. 27] may be due to partial doping of the "peas" inside the peapods. Alternatively, the enhancement of the pinch mode during anodic charging of peapods [panel (b) in Fig. 27] may be due to two factors: (i) the Raman peak intensities for the SWNT host goes out of resonance and (ii) C_{60} is well known to resist oxidation. Since no frequency shift of the 1465 cm^{-1} was observed, the authors concluded that there was no transfer of holes to the C_{60} and no dimerization of C_{60} [102]. If dimerization of C_{60} had occurred, it would have resulted in a ~10 cm^{-1} downshift of the $A_g(2)$ mode of C_{60} [103].

Synthesis techniques have now improved to a point where researchers are able to prepare doped peapods, such as Sm–metallofullerene peapods, $(Sm@C_{82})_n$@SWNTs [104–106]. While *in-situ* HRTEM and electron energy loss spectroscopy (EELS) spectroscopic studies have been done, no reports on Raman scattering from this interesting class of peapods have yet been reported. Since EELS studies reveal that the encapsulated Sm ion assumes a +2 charge state in $(Sm@C_{82})_n$@SWNTs, exciting Raman properties can be expected for such metallofullerene peapod samples.

5. DOUBLE WALLED CARBON NANOTUBES

The double wall carbon nanotubes (DWNTs) are a special case of MWNTs containing only two coaxial nanotubes. DWNTs therefore provide a unique opportunity to examine simultaneously the Raman scattering from the very small diameter interior tubes along with Raman scattering from the exterior tubes.

DWNTs were first reported in the products obtained in an electric arc discharge with catalyst mixtures of Ni, Co, Fe, and S in an Ar/H_2 (1:1 volume ratio) atmosphere at 350 Torr [107]. Others have prepared DWNTs from a thermal decomposition of C_2H_2 over ferrocene/sulfur catalysts [108], or of methane over ferrocene/thipohene [109] in the temperature range near ~1000 °C. Several Raman studies have been reported on DWNTs [95, 107–109], but the most comprehensive study thus far was reported by Bandow et al. [95] who made a joint systematic HRTEM and Raman evaluation of the transformation of peapods into DWNTs when the peapod sample was annealed in vacuum at 1200 °C and followed up this work with a more detailed evaluation of the Raman spectra using several different laser excitation energies [110].

Returning to Figure 24, the peapods prepared using purified laser-derived SWNTs (panel A) were systematically annealed in vacuum (~10^{-6} Torr) and characterized by HRTEM and Raman spectroscopy. Surprisingly, the C_{60} molecules showed no change in structural or Raman features until the sample temperature reached 800 °C. At this elevated temperature, the C_{60} molecules do not leave the tube but instead come closer to one another to form C_{60} dimers (indicated by arrows in panel B), which provides the first step toward coalescing several C_{60} molecules into elongated encapsulated tubes. This experiment is a good example to show how the shell of the SWNT serves as a nanoscale reaction chamber. At 1000 °C (panel C) the peapods are observed to transform into DWNTs, and at 1200 °C (panel D), most of the C_{60} molecules have coalesced to form DWNTs for the whole length of the tube. At each step during the transformation, Raman spectra were gathered using 514.5 nm excitation wavelength (Fig. 28). The weakly observable A_g symmetry RBM and pentagonal pinch modes of molecular C_{60} are found to be downshifted by ~10 cm^{-1} relative to their corresponding mode frequencies in molecular C_{60} solids. Several Raman studies on polymerized C_{60} have reported a downshift of the pentagonal pinch

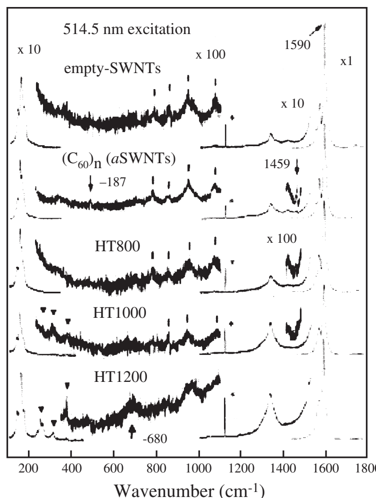


Figure 28. Raman spectra in the radial breathing mode region up to 400 cm^{-1}, in the intermediate frequency range $400 < \omega < 1100$ cm^{-1}, in the range of the *D*-band and of the pentagonal pinch modes of C_{60} and finally in the tangential *G*-band region of SWNTs and DWNTs. Reprinted with permission from [95], S. Bandow et al., *Chem. Phys. Lett.* 337, 48 (2001). © 2001, Elsevier Science.

mode by ~10 cm^{-1} relative to that for isolated C_{60} molecules [79, 103, 111, 112]. In particular, solid C_{60} was found to undergo photoinduced polymerization in the absence of oxygen when excited with laser power densities of greater than 50 W/cm^2, exhibiting a characteristic $A_g(2)$ pinch mode at 1460 cm^{-1} [10, 103]. Thus, the feature at 1459 cm^{-1} in Figure 28 may be due to photoinduced polymerization of encapsulated C_{60} caused by the high laser power used in the experiment. Unfortunately, the authors did not state the laser power density used in their paper.

In the Raman spectrum labeled HT800 we see that heat treatment and the formation of dimers strongly suppress the pentagonal pinch mode, and the RBM is hardly discernible above the background signal (Fig. 28). Dramatic changes in the HT1200 spectrum are seen in which new modes appear in the vicinity of the RBM and in the 400–1100 cm^{-1} region. The appearance of these new modes between 200 and 400 cm^{-1} correlates with the DWNT formation, as seen in the HRTEM images in panel D of Figure 24, and signals the formation of the inner tube inside the DWNT, since the RBM frequency varies as $1/d_t$ where d_t is the tube diameter [see Eq. (1)]. These changes in the Raman activity in the intermediate frequency region (400–1100 cm^{-1}) are consistent with a nonresonant bond polarization model calculation applied to short length SWNTs [53]. The intensity of the *G*-band is also observed to increase in the HT1200 sample, while the *D*-band intensity also increases as the heat treatment temperature increases, perhaps due to the formation of short segments of inner tubes.

The new Raman features which appear in the RBM region as the DWNTs are formed are of particular interest. Again a ~2 cm^{-1} softening for the radial band is observed in the peapod sample [95]. However, unlike Pfeiffer et al. [101], Bandow et al. [95] do not attribute the mode softening due to an expansion of tube diameter in the peapod sample, but instead they attribute the mode downshift to a small charge transfer to the SWNT shell from the encapsulated C_{60} molecules. Also, in the second spectrum from the top in Figure 28, the *G*-band is found to upshift.

From a detailed analysis in which the van der Waals distance of 0.34 nm between the shells of the inner and outer tubes was taken into account, (n, m) indices were elicited from the positions of the new peaks in the RBM region which suggested that the inner tube diameters ranged from 0.93 to 0.59 nm [95]. On average, the diameter difference between the inner and outer tubes was 0.71 ± 0.05 nm, consistent with HRTEM studies.

In another independent experiment, three samples were prepared from the same batch of as-prepared SWNTs for a detailed comparative Raman study: (1) the baseline as-prepared SWNT bundle sample, (2) C_{60} doped SWNT bundles (peapods), and (3) DWNT bundles. Raman characterization was performed using four laser excitation wavelengths, 488.0, 514.5, 647.1, and 1064 nm (2.54, 2.41, 1.92, and 1.16 eV) as shown in Figures 29 and 30 [110], for the RBM bands, and for the *G*-band region, respectively, for these three types of samples. Clearly, diameter-selective Raman scattering from these samples is observed. Collectively, the RBM data can be broadly classified into two regions: (1) In the low frequency region below 250 cm^{-1}, the three samples exhibit nearly identical RBM features with any given laser excitation, suggesting that these features correspond to the exterior tube ($d_t \sim 1.3$–1.6 nm) properties. The encapsulated C_{60}, or the presence of the interior

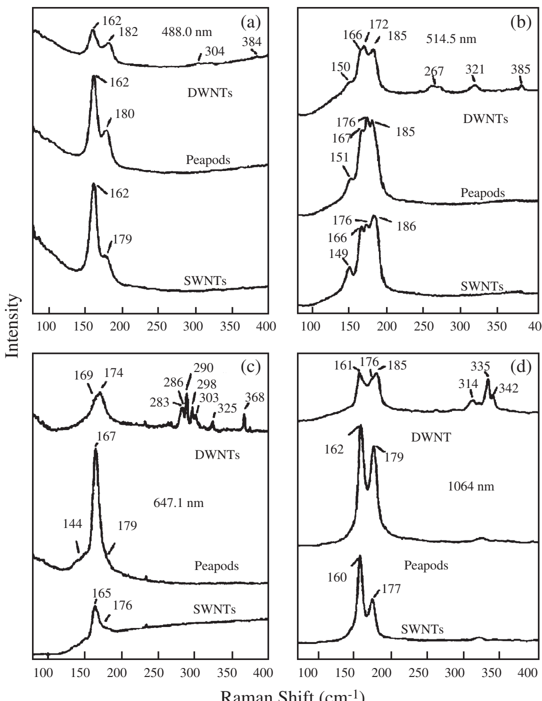


Figure 29. radial breathing mode Raman spectra for SWNTs, DWNTs, and C_{60}-SWNT peapods using different excitation energies of 488.0 nm (2.54 eV) (a), 514.5 nm (2.41 eV) (b), 647.1 nm (1.92 eV) (c), and 1064 nm (1.17 eV) (d). The spectra were fitted to an appropriate sum of Lorentzians and the obtained peak positions are marked on the respective spectra. Reprinted with permission from [110], S. Bandow et al., *Phys. Rev. B* 66, 075416 (2002). © 2002, American Physical Society.

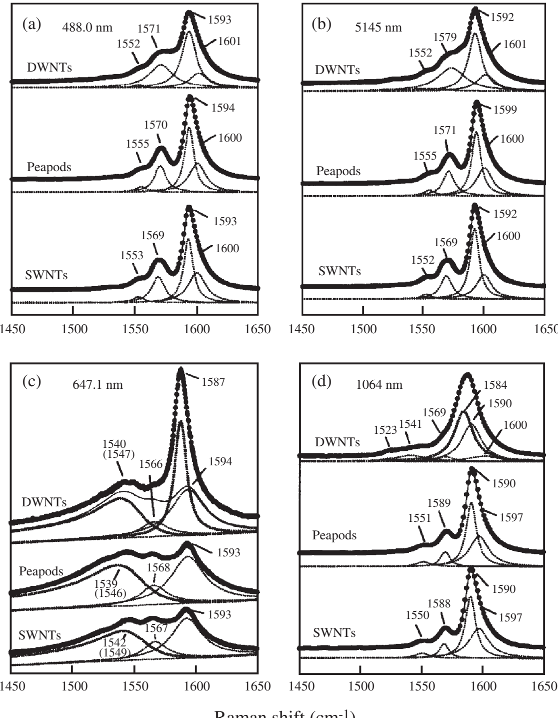


Figure 30. Raman scattering spectra for the tangential *G*-band region for three kinds of bundled nanotubes. The excitation wavelengths are indicated in each panel (see also Fig. 29). Lorentzian components associated with the primary semiconducting tubes are indicated by the thin dotted lines and those with the secondary tubes in DWNTs are indicated by the thick dotted lines in (c) and (d) (see text). The asymmetric Breit–Wigner–Fano (BWF) component is also represented by the thick dotted lines in (c) and (d), where the coupling parameter $(1/q)$ between the discrete phonon mode and the continuum modes, determining the asymmetry of the lineshape, was fixed at -0.35 (from [79]). The numerical values in the parentheses are the renormalized phonon frequencies of the BWF lines (see text). The thin solid line in (c) is a sum of thin dotted Lorentzian components and the BWF component, which is similar to the spectra taken for SWNTs and peapods. Reprinted with permission from [110], S. Bandow et al., *Phys. Rev. B* 66, 075416 (2002). © 2002, American Physical Society.

tube, appears to have a negligible influence on the RBM mode frequencies in this region where the peak positions for the RBM features in the SWNT and DWNT samples are almost identical. (2) For modes with frequencies above 250 cm^{-1}, the DWNT bundles exhibit new peaks that are absent in the SWNT or peapod samples. From the experimentally observed RBM frequencies, the diameters of interior tubes were determined to be between 0.6 and 0.9 nm [110]. The difference between the mean exterior and interior tube diameter is found to be ~0.7 nm, close to twice the stacking distance between graphene sheets in graphite, suggesting that the interior tube is tightly nested inside the exterior tube.

The Raman scattering from the *G*-band carbon atom vibrations shown in Figure 30 is normally detected at ~1593 cm^{-1} for typical SWNT bundles [5]. The diameter dependence of the tangential band is expected on theoretical grounds to be very weak for sp^2 bonded SWNTs [6]. Interestingly, the tangential *G*-band in the DWNT sample when excited with red excitation (panels c and d of Fig. 30) is found to be downshifted by ~7 cm^{-1} relative to its position in the host tubes at 1593 cm^{-1} [110]. This can be understood by looking at the E_{ii} vs $1/d_t$ plot shown in Figure 31. It is clear from Figure 31 that the inner tubes resonate with 647 and 1064 nm excitations, and not with 488 or 514 nm excitations (only one open circle intersects the 514 nm line in Fig. 31) and consequently, the *G*-bands of DWNTs in the 647 and 1064 nm spectra are considerably downshifted from 1592 cm^{-1} (panels c and d in Fig. 30). From a detailed lineshape analysis of the RBM and tangential bands of the DWNT sample, the frequency of the downshifted tangential modes has been identified with interior tubes and associated with wall curvature of the interior nanotube [110]. No features due to the fullerene $A_g(2)$ pentagonal pinch modes within the peapod samples were reported in these studies of the *G*-band spectral region [110].

In summary, based on published research to date, the effects of van der Waals interactions in peapods seem subtler than those in SWNTs bundles. Further experimental evidence to elucidate the van der Waals interaction in peapods is highly desirable. For example, future research should investigate whether the origin of the doublet structure in the peapod Raman spectra (described in Fig. 26) may be due to subtle differences in nanotube interactions with C_{60} molecules frozen in different orientations.

6. CONNECTION TO MULTIWALL NANOTUBES

In this section, we discuss the Raman spectra of multiwall carbon nanotubes (MWNTs) with emphasis on Raman modes that have been observed below ~300 cm^{-1}. Raman-active modes in this spectral region correspond to radial breathing modes that are typically observed in SWNTs and DWNTs. In MWNTs, the inner-most tube can have a very

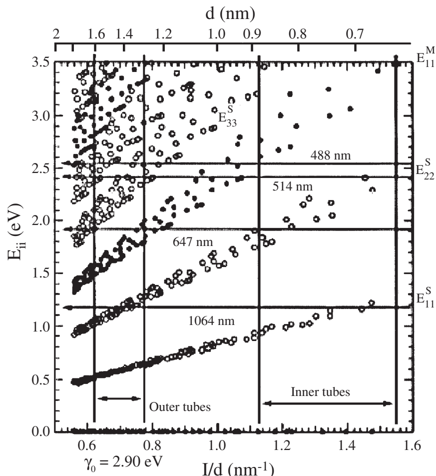


Figure 31. Plot of E_{ii} vs $1/d_t$, which is a replot of the points in Figure 4 where E_{ii} is plotted vs d_t. The solid and open circles correspond respectively to metallic and semiconducting tubes. The primary (outer) and secondary (inner) tube diameters existing in the samples occupy the indicated regions between the thick vertical lines. Reprinted with permission from [110], S. Bandow et al., *Phys. Rev. B* 66, 075416 (2002). © 2002, American Physical Society.

small diameter, in fact as small as the smallest diameter that has yet been found for any SWNTs. This implies that the many shells of the MWNT are able to stabilize very small diameter nanotubes at the core of a MWNT. Figure 32 shows the Raman spectrum using 514.5 nm excitation of high purity MWNTs deposited as a film on a quartz substrate. The MWNTs in this figure are prepared by a thermal decomposition of ferrocene–xylene mixtures at 700 °C [113]. The dominant outer- and inner-tube diameters ranged between 25–30 and 2–7 nm, respectively. Besides the strong D- and G-bands at 1355 and 1582 cm^{-1}, and the second-order bands at 2454, 2704, and 2942 cm^{-1}, weak features can be discerned on either side of the D–G-bands (Fig. 32). The peak marked by an asterisk corresponds to a plasma line frequency, while peaks at 165, 295, and 452 cm^{-1} can be assigned to weak radial breathing modes of inner tubes with approximate diameters of 1.5, 0.84, and 0.55 nm, respectively, with d_t determined from the relation $\omega_{\text{RBM}} \sim 248/d_t$, or alternatively the higher frequency modes can be identified with the second and third harmonics of the first-order ω_{RBM} mode. Other examples of reported radial breathing mode features in MWNTs can be found in the literature [14, 114]. Weak modes at ∼390 and 490 cm^{-1} were observed in the Raman spectrum of purified arc-derived MWNTs by Kataura et al. [14] and they identified these modes with the radial breathing modes of tubes that define the core of the MWNTs (core diameter of ∼0.65–0.5 nm). Chandrabhas et al. [114] have also observed weak modes at 49, 58, 470, and 700 cm^{-1} in their arc-produced MWNTs, as well as stronger features at 1353 and 1583 cm^{-1}, and second-order features at 2455, 2709, and 3250 cm^{-1}. In Figure 32, peaks at 2454 and 2942 cm^{-1} can be associated, respectively, with a K-point overtone ($2 \times 1220 = 2440$ cm^{-1}) or with ($1355 + 1095 = 2450$ cm^{-1}) and ($1355 + 1582 = 2937$ cm^{-1}) combination modes, while the 2704 cm^{-1} band is identified as the second-order ($2 \times 1355 = 2710$ cm^{-1}) G'-band.

In a more recent report, micro-Raman data excited with 514.5 nm excitation for highly oriented pyrolytic graphite (HOPG), SWNTs, and three samples of MWNTs were compared to identify the tube diameter distribution in the nanotube samples (Fig. 33) [115]. The MWNT samples were prepared from a hydrogen arc discharge and HRTEM showed that the outer tube diameters ranged from ∼2 to 28 nm (dominant outer tube diameter ∼10 nm) [115]. In contrast to the Raman spectrum of HOPG, the nanotube spectra show peaks in the RBM region, with some peaks having FWHM linewidths as narrow as 3 cm^{-1} and frequencies as high as 570 cm^{-1}. The ∼392, 387, and 279 cm^{-1} peaks were observed in the spectrum of all three MWNT samples. The presence of the 524 and 570 cm^{-1} peaks suggest inner tube diameters on the order of ∼0.45 nm, consistent with their HRTEM studies [115]. Polarized Raman scattering studies were also performed on a single aligned bundle of MWNTs to confirm that the Raman features observed in the low frequency region all correspond to the RBM of MWNTs.

The most detailed experimental polarization results for Raman spectra presently available are for high purity aligned arrays of MWNTs [116]. These aligned MWNTs were prepared on silica substrates from the thermal decomposition of a ferrocene–xylene mixture and had an average nanotube diameter $d_t \sim 25$ nm, with the MWNTs aligned normal to bare silica substrates [113]. The nanotube diameters in this MWNT sample are probably large enough so that the resonance Raman scattering effect is substantially weaker than is typical for SWNTs ($1 < d_t < 2$ nm). The thickness of the aligned nanotube bundle was intentionally kept below 1 μm so that the focused laser excitation beam (beam diameter $d \sim 1$ μm) overfilled the aligned MWNTs during the polarized Raman scattering measurements. This study measured polarization effects associated with ($\|, \|$) and ($\|, \perp$) scattering for the G-band and the D-band, as well as providing confirmation for a predicted angular dependence of the scattered intensity as a function of the measured angle θ_m between the polarization direction and the nanotube axis [116].

The polarized spectra in Figure 34 for four polarization geometries, labelled (XY), (YX), (YY), and (XX), are shown for the Raman intensities of the D-band and the tangential G-band for these MWNTs in the backscattering

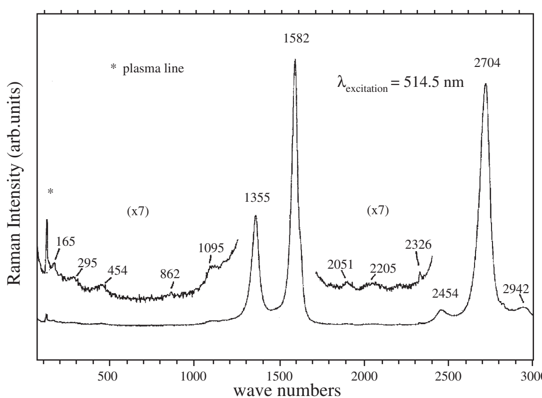


Figure 32. The Raman spectrum for a high purity multiwall nanotube sample for laser excitations $E_{\text{laser}} = 2.41$ eV (514.5 nm) (see text). Reprinted with permission from [113], R. Andrews et al., *Chem. Phys. Lett.* 303, 467 (1999). © 1999, Elsevier Science.

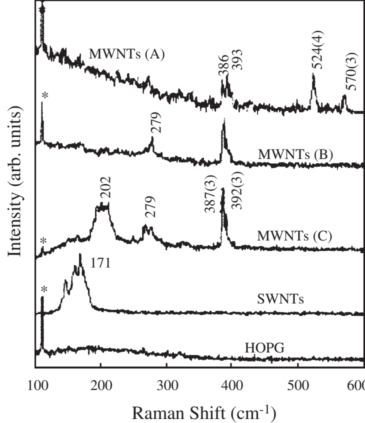


Figure 33. Comparison of the micro-Raman spectra taken at $E_{\text{laser}} = 2.41$ eV (514.5 nm) in the low frequency range for HOPG, SWNTs, and three samples of MWNTs. Reprinted with permission from [115], X, Zhao et al., *Chem. Phys. Lett.* 361, 169 (2002). © 2002, Elsevier Science.

configuration, with the greatest intensities observed for the (XX) polarization geometry. The geometry and notation used to describe these polarization studies on MWNTs are shown in the inset to Figure 34, where x, y, z refer to the coordinates of the MWNTs and X, Y, Z refer to the laboratory frame of the light. In this notation, a backscattering Raman experiment is described by $Z(XY)\overline{Z}$ to indicate light incident along the Z axis with the electric vector in the X direction for the incident light beam, while the Poynting vector for the scattered light is along the $-Z$ direction, with the electric vector for the scattered light beam in the Y direction. For the polarization results on the MWNTs shown in Figures 34 and 35, the backscattering geometry was used with light normal to the nanotube axis, so that the abbreviated notation (XY), giving the directions of the incident and scattered electric field polarizations, is used to label the polarization geometries shown in these figures.

A Lorentzian lineshape analysis for the MWNTs spectra in Figure 34 shows four peaks at 1354, 1576, 1583, and 1624 cm^{-1} in almost all the polarization geometries, but with very different polarization-dependent relative intensities [116]. From Figure 34, the experimental polarized integrated intensity ratios for the tangential band (at 1584 cm^{-1} for the XX spectrum) are I_{XX}/I_{XX}:I_{YY}/I_{XX}:I_{XY}/I_{XX}:I_{YX}/I_{XX} = 1.00:0.29:0.19:0.39 for this MWNT sample. These polarization results can in part be explained by the nanotube geometry which gives rise to slightly different force constants along the nanotube axis relative to the circumferential direction, where the nanotube curvature reduces the force constant. This anisotropy in the force constant accounts physically for the frequency difference between the components at 1575 cm^{-1} (attributed to vibrations in the circumferential direction) and at 1584 cm^{-1} (attributed to vibrations along the nanotube axis) for the (XX) polarization. Similar frequency differences are observed for the other polarization geometries, and all the upshifted and downshifted modes in Figure 34 average to 1580.8 ± 0.8 cm^{-1}, in good agreement with the E_{2g_2} graphite mode frequency. Because of the small number of allowed k vectors in the circumferential direction, we expect the frequency difference between the lowest and middle frequency components to increase with decreasing nanotube diameter [117]. For example, SWNTs with diameters of d_t ~1.4 nm show these two spectral features [5] to be at 1567 and 1593 cm^{-1} (see Section 3.1), and these two mode frequencies also average to ~1580 cm^{-1}. The feature near 1620 cm^{-1} which is quite pronounced in the spectra for MWNTs, and not present in SWNTs, is associated with the maximum in the graphene 2D phonon density of states [118].

No theory has yet been developed to describe polarization effects for MWNTs. Therefore the experimental polarization results for the G-band in Figures 34 and 35 have been interpreted in terms of the theory developed for (10,10) SWNTs [52] as a first approximation. The theory of polarization effects in single wall nanotubes was first used to distinguish between the contributions to the G-band from modes with A_{1g}, E_{1g}, and E_{2g} symmetry [52], which according to bond polarization model calculations [52, 53] for a (10,10) SWNT predict Raman-active frequencies at 1587, 1585, and 1591 cm^{-1}, respectively, all occurring within a range of 6 cm^{-1}, which is much smaller than the observed linewidth of the tangential band for Raman scattering, even for semiconducting nanotubes [52, 119].

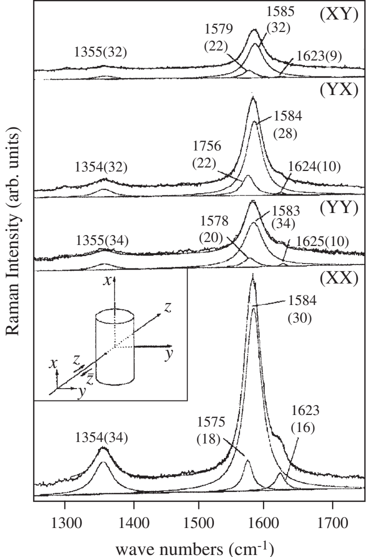


Figure 34. Polarized Raman spectra for MWNTs taken at 514.5 nm (2.41 eV) for four scattering geometries. The peak frequencies for the various features are in cm^{-1} and the FWHM linewidths in cm^{-1} are given in parenthesis. The inset defines the polarization geometries for the aligned nanotubes that lie parallel to the x-axis but there is no preferred angular alignment of the nanotubes in the plane normal to the nanotube axis. Reprinted with permission from [116], A. M. Rao et al., *Phys. Rev. Lett.* 84, 1820 (2000). © 2000, American Physical Society.

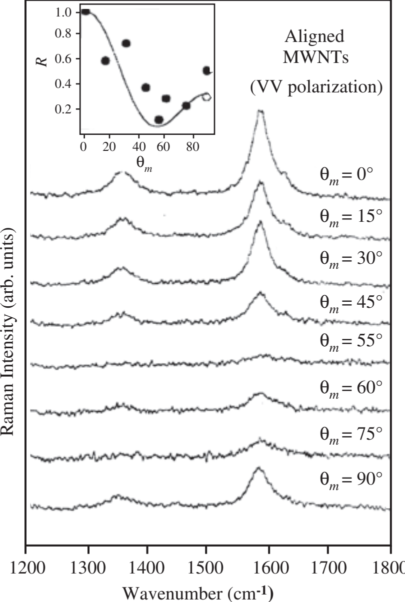


Figure 35. Dependence of the measured Raman intensity for VV scattering as a function of the angle θ_m between the polarization direction and the nanotube axis (see inset to Fig. 34). The inset to this figure shows the dependence of the experimental G-band intensity ratio $R = I_{VV}(\theta_m)/I_{VV}(0)$ (open and filled circles) as a function of θ_m in comparison to the theory for SWNTs [53, 116]. Reprinted with permission from [116], A. M. Rao et al., *Phys. Rev. Lett.* 84, 1820 (2000). © 2000, American Physical Society.

The experimental intensities for the (XX) and (YY) geometries in Figure 34 and in the inset to Figure 34 correspond, respectively, to $\theta_m = 0$ and $\theta_m = 90°$ for the VV polarization configuration, while the (XY) and (YX) geometries, respectively, correspond to $\theta_m = 0$ and $\theta_m = 90°$ for the VH configuration, where V and H denote vertical and horizontal, respectively. Calculations show that the A_{1g} mode is dominant for the tangential band in the (XX) configuration, while in the (YY) configuration, both the A_{1g} and E_{2g} modes should contribute to the scattering intensity. The intensity ratios predicted for the A_{1g} mode at $\theta_m = 0°$, the A_{1g} mode at $\theta_m = 90°$, and the E_{2g} mode at $\theta_m = 90°$ are, respectively, 2.91, 0.72, and 0.33. Thus the theoretical value for the intensity ratio for I_{XX}/I_{YY} is 2.91/1.05 = 1.00/0.36, which is to be compared to the experimental values of $I_{XX}/I_{YY} = 1.00/0.29$. This observation would be consistent with the peak at 1584 cm^{-1} containing the unresolved E_{2g} and A_{1g} modes of the tangential band.

Based on theoretical considerations, the intensity of the tangential A_{1g} mode as a function of θ_m is expected to exhibit a minimum at $\theta_m = \cos^{-1}(1/\sqrt{3}) = 54.7°$ [52]. Figure 35 shows a collection of polarized Raman spectra that were obtained on aligned MWNTs in the VV configuration as a function of θ_m. Clearly the experimental polarized tangential band intensity ratio $R = I_{VV}(\theta_m)/I_{VV}(0)$ exhibits a minimum near $\theta_m = 55°$, in good agreement with the theoretical curve for the intensity of the VV signal vs θ_m shown in the inset to Figure 35. It should be noted that the intensity ratio R denoted by the open circle in the inset to Figure 35 was obtained from the XX and YY data depicted in Figure 34. The data set in Figure 35 also suggests that the G-band intensity is dominated by the intensity of the A_{1g} symmetry mode at 1584 cm^{-1}. It is interesting that the experimental D-band intensity is also a minimum near $\theta_m = 55°$. No theory is yet available for the angular dependence of the D-band intensity.

Returning to Figure 34 we note that the optical absorption spectra of a single wall carbon nanotube, calculated for polarized light by Ajiki and Ando [120, 121], show that the selection rules for optical absorption between the valence and conduction π subbands with subband index n are $\Delta n = 0$ and $\Delta n = 1$ for polarization parallel and perpendicular to the nanotube axis, respectively. The optical absorption for the polarization perpendicular to the nanotube axis is predicted to be suppressed almost completely when the depolarization effect is taken into account. Thus the optical absorption occurs only when the polarization is parallel to the nanotube axis for intersubband transitions, with $\Delta n = 0$ for single wall carbon nanotubes. The depolarization effect should, however, be relaxed in MWNTs, especially as the diameter increases. On the basis of the depolarization effect, we expect more effective excitation of resonant Raman scattering for the (XX) scattering geometry, and for metallic nanotubes, for which the nanotube acts as a radiation pipe for the laser excitation. The polarized VV Raman spectra for aligned MWNTs in Figure 34 appear to be consistent with the optical anisotropy present in the MWNTs. The discrepancy between the experimental polarized Raman intensity ratios (regarding Fig. 34) and theoretical predictions may be due to a depolarization (antenna) effect operative also on the scattered light which tends to enhance the experimental $YX{:}XY$ intensity ratio relative to the calculations.

As the diameter of the MWNTs increases to over 100 nm, the spectra in Figure 32 show features associated with the D-band, the G-band, and the G'-band of sp^2 carbons, and the interlayer separation between adjacent shells in the MWNTs approaches that of 2D turbostratic graphite (0.344 nm), which is characteristic of nanofibers. For yet larger diameter MWNTs, the interlayer separation drops below 0.344 nm so that the interlayer C–C correlations become important, and in this regime the sp^2 carbon is called a vapor grown carbon fiber (VGCF), often refereed to as an “Endo fiber” [122, 123]. In the diameter range above 100 nm, the G-band feature in VGCFs becomes more and more similar to that for the 3D graphite spectra, while the D-band continues to provide information about the disorder in the sample. The G'-band, which is a single Lorentzian line in the Raman spectrum of carbon nanofibers, becomes a doublet feature as interplanar correlations become established. When a VGCF or an carbon nanofiber is broken open, a thin MWNT is typically observed [124].

7. FUTURE DIRECTIONS

In this section we propose future directions for carbon nanotube research, and especially for the use of Raman spectroscopy as a characterization tool for carbon nanotubes. As a first step, synthesis and separation techniques for SWNTs need to advance further. Since many of the SWNT properties depend critically on the (n, m) indices of the SWNT and, especially, on whether the SWNT is semiconducting or metallic, the demands on the control of the carbon nanotube growth process are severe. An approach based on developing a synthesis process for growing only metallic SWNTs or only semiconducting SWNTs seems like a fruitful strategy, but until such a time that a synthesis process of this kind is developed, it will be necessary to focus on SWNT separation processes that overwhelmingly favor either metallic or semiconducting SWNTs. In this connection, Raman spectroscopy should be useful to sensitively measure the efficacy of a given separation method regarding the percentage of metallic and semiconducting SWNTs with $d_t < 2$ nm, as well as the diameter distribution of the SWNTs in the sample. Research to extend present Raman characterization techniques to do such a sample characterization would be highly desirable. Because of the lower sensitivity of Raman spectroscopy to the geometrical structure of carbon nanotubes for tube diameters greater than $\sim$2 nm, we should expect Raman spectroscopy to play a decreasing role in the characterization of large diameter SWNTs and MWNTs.

Detailed combined experimental and theoretical studies are needed to comprehend how E_{ii} and E_F change as a function of external perturbations. Here again investigations should start at the single nanotube level, as discussed, and then should be extended to investigate SWNT bundles as a function of the number of SWNTs in a bundle, to DWNTs, and finally to the inner-most layer of a MWNT. Below a critical diameter d_{critical} it may not be possible to understand the properties of DWNTs or MWNTs because of the large curvature effects of very small diameter SWNTs. At large diameters (>2.5 nm), the sensitivity of Raman scattering

decreases with regard to distinguishing between the detailed structural properties of nanotubes relative to those of general sp^2 carbons.

The effect of intertube interactions to upshift the radial breathing mode frequency relative to that for isolated SWNTs and the use of the expression $\omega_{RBM} = \alpha/d_t + \Delta\omega_{RBM}$ requires detailed systematic study following the pioneering work of Kuzmany et al. regarding the structural characterization of SWNT bundles [44, 65, 66]. More systematic studies are needed of the dependence of α and $\Delta\omega_{RBM}$, the local environment of the tube, the total size of the SWNT bundle, and the location of the SWNT within the SWNT bundle. Experimental observations described in Figures 16 and 22 collectively suggest that the debundling into small diameter bundles causes a 10 cm^{-1} upshift in ω_{RBM} but does not lead to a change in the dependence of ω_{RBM} on external applied pressure. Further high pressure Raman studies on a homogeneous sample of isolated SWNTs prepared either by improved chemical methods or by chemical vapor deposition are needed. These pressure-dependent results also suggest that the coupling of the excitation laser energy (E_{laser}) is sensitive to the environment in which the nanotube resides, and thus future studies aimed at correlating nanotube diameters with ω_{RBM} should pay attention to the size of the bundles, and to the procedures used to functionalize SWNTs.

Future doping experiments should investigate, in the context outlined previously for systematic joint experimental and theoretical studies, the effects of donor and acceptor doping and variation of applied positive and negative voltages to SWNTs, starting with isolated SWNTs and extending the studies to SWNT bundles, including studies of debundling effects. With regard to doping studies, systematic studies of boron and nitrogen doping are of particular interest, since these dopants behave in part as substitutional dopants. Regarding studies, such as investigation of the effect of applied potentials, particular attention should be paid to the use of the G-band to characterize the charge transfer, since the G-band feature in metallic SWNTs is known to be strongly affected by the coupling of phonons to plasmons through the electron–phonon interaction, and the intensity of this interaction is strongly affected by the number of SWNTs in the SWNT bundles. Therefore the use of the G-band feature in the Raman spectra of sp^2 carbons needs to be reexamined theoretically and experimentally. Other Raman features that are also experimentally observed to be sensitive to charge transfer, such as the G'-band, should be further investigated theoretically and experimentally as a sensitive probe of charge transfer in SWNTs.

Raman spectroscopic studies on isolated peapod or DWNT samples could provide a wealth of information about inter- and intratube force constants. The origin of the two weak modes at 1466 and 1474 cm^{-1} in the vicinity of the pentagonal pinch mode frequency of C_{60} (Fig. 26) [101] may be due to subtle differences in the interaction of a pentagonal vs hexagonal face of a C_{60} molecule with the interior surface of a nanotube. Such detailed experimental and theoretical studies of peapods and DWNTs should initially be carried out at the single nanotube level, if possible, to illuminate the physical mechanisms involved. This should be followed by systematic studies on bundles of peapods and DWNTs, including systematic variation of the various perturbations discussed in this chapter, including external pressure, high pressure, applied potentials, endohedral doping, chemical donor and acceptor doping, etc.

GLOSSARY

Anisotropy Pertaining to a material whose electrical or optical properties vary with the direction of propagation of a traveling wave or with different polarizations of a traveling wave.

Armchair nanotube Armchair nanotubes correspond to the roll up of the nanotube so that the indices (n, m) are equal $(n = m)$ and the chiral angle is 30°.

Band tuning The variation of parameters such as chemical composition to vary the electronic band gap of a semiconductor.

Boltzmann factor The factor $\exp(-E/k_BT)$ that compares the energy of a system E to its thermal energy k_BT and gives the probability of occupying the state with energy E.

Breit–Wigner–Fano lineshape Observed Raman lineshape when phonon modes couple to a Raman-active electronic continuum.

Brillouin zone A primitive cell in the reciprocal lattice.

Bundled tubes Single wall nanotubes arranged in a bundle containing between about 10 and 100 nanotubes and when ordered the tubes are arranged in a triangular lattice.

C_{60} or fullerene A C_{60} molecule consists of 60 carbon atoms bonded to each other and occupying corner sites in a regular truncated icosahedron.

Carbon Nanotubes A graphene sheet (2D graphite) rolled into a seamless tube.

Chemical vapor deposition (CVD) A chemical reaction which transforms gaseous molecules, called a precursor, into a solid material, in the form of thin film, powder, or individual nanotubes on the surface of a substrate.

Chiral angle The angle between the chiral vector $(na_1 + ma_2)$ and a_1.

Chiral vector C_h Chiral vector can be expressed by the real space unit vectors the unit cell basis vectors of the graphene lattice $\mathbf{a}_1$ and $\mathbf{a}_2$ of the hexagonal lattice by $C_h = na_1 + m\mathbf{a}_2 \equiv (n, m)$, where n, m are integers, $0 \leq |m| \leq n$.

Chirality The values of n and m determine the chirality, or the angle that the nanotube axis makes with the a_1 axis of the graphene lattice of the nanotube.

Chromatography The chemical method of separating constituents dissolved in one phase (usually mobile) through its equilibration with a second phase (usually stationary).

Density of states (DOS) The spectrum of the number of energy levels per eV usually plotted as a function of energy.

Depolarization effect The change in polarization of the incident optical electric field induced by the polarization of the nanotube.

Electric arc method A method of producing nanotubes through the arc-vaporization of two carbon rods placed end to end with a high electric field between the electrodes to create a plasma of carbon atoms.

Electrochemical potential Amount of energy (work) required to add an amount n of charge-bearing species into

a solution when a potential is applied, and in the limit of zero applied potential reduces to the Fermi level.

Endohedral doping A doping method in which an ion (or ions) is (are) encaged inside the carbon framework of a fullerene molecule or a nanotube.

Fermi energy The energy E_F below which all electron states are filled and above which all are empty at zero temperature.

Hybridization Mixing of different orbitals on the same atom, such as the mixing of a π and σ orbital on a carbon atom.

Interband energies Refers to the difference in energy between a state in the valence band and a state in the conduction band to which a transition is made.

Lorentzian lineshape The functional form of the frequency dependence of a line in the optical absorption spectra and includes its spectral width of the optical absorption line.

Multiwall tubes Coaxial arrangement of single wall carbon nanotubes.

Phonon frequency The frequency of vibration of a quantized lattice vibration in a solid.

Photoinduced polymerization Bonding between isolated neighboring molecules induced by light.

Photons A discrete packet, i.e., a quantum, of electromagnetic energy.

Plasmons The quantized ensemble or gas of charged carriers such as electrons.

Polarization Refers to a vector denoting the sum of dipole moments per unit volume or a vector denoting the direction of the optical electric field.

Radial breathing mode Refers to the vibrational displacements whereby every atom in a carbon nanotube vibrates in phase in the radial direction.

Raman active modes Those vibrational modes of a system that are excited by an incident photon in an inelastic scattering process.

Raman spectroscopy Refers to the response of a system to an inelastic scattering of a photon in a system where the scattered photon that is either upshifted or down shifted in frequency from the incident photon by the natural vibrational frequency of the system.

Resonance Refers to the strong response of a system to an external excitation at a particular frequency of excitation.

Single-walled carbon nanotube (SWNT) The rolling of a single layer of a graphite lattice to form a seamless tube one atom in thickness.

Stokes and Anti-Stokes Most of the Raman scattered photons are shifted to longer wavelengths (Stokes shift) corresponding to emission of a phonon, but a small portion are shifted to shorter wavelengths (anti-Stokes shift) corresponding to the absorption of a phonon.

Surface plasmons Collective excitation of the electrons of a solid or liquid at the surface of the material.

Symmetry mode Refers to the symmetry operations (rotations, reflections, inversion, etc.) that are preserved in the atomic displacements of the atoms as they execute the vibrational mode.

Tangential band A small range of frequencies over which the nearest neighbor carbon atoms vibrate.

Trigonal warping The departure in the symmetry of the constant energy contours away from circles, where the distortion or warping of the constant energy contours shows a three-fold axis of symmetry.

Van der Waals force A weak non-covalent attractive force acting between neutral atoms or molecules.

Van Hove singularities The divergence of the density of states of a one dimensional system such as a carbon nanotube showing a density of states feature with an energy dependence $(E - E_0)^{-1/2}$.

Zigzag nanotube Zigzag nanotubes are formed when either n or m in the nanotube indices (n, m) is zero and the chiral angle is 0°.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Professors Marcos Pimenta, Ado Jorio, Dr. Antonio G. Souza Filho, Riichiro Saito, Peter Eklund, U. Venkateswaran, Hans Kuzmany and many others for many discussions about the research reviewed in this review article. AMR would like to thank Mr. Saurabh Chopra for assistance in the preparation of the manuscript. The authors gratefully acknowledge support under NSF Grants DMR 01-16042 (MSD and GD) and grants through NSF funded ERC EEC-9731680; 0132573 & 0244290 (AMR).

REFERENCES

1. M. S. Dresselhaus, G. Dresselhaus, A. Jorio, A. G. Souza Filho, and R. Saito, *Carbon* 40, 2043 (2002).
2. A. Jorio, R. Saito, J. H. Hafner, C. M. Lieber, M. Hunter, T. McClure, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 86, 1118 (2001).
3. J. H. Hafner, C. L. Cheung, T. H. Oosterkamp, and C. M. Lieber, *J. Phys. Chem. B* 105, 743 (2001).
4. S. Berber, Y.-K. Kwon, and D. Tománek, *Phys. Rev. Lett.* 84, 4613 (2000).
5. A. M. Rao, E. Richter, S. Bandow, B. Chase, P. C. Eklund, K. W. Williams, S. Fang, K. R. Subbaswamy, M. Menon, A. Thess, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *Science* 275, 187 (1997).
6. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, "Physical Properties of Carbon Nanotubes." Imperial College Press, London, 1998.
7. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 61, 2981 (2000).
8. R. Saito and H. Kataura, in "Carbon Nanotubes: Synthesis, Structure, Properties and Applications" (M. S Dresselhaus, G. Dresselhaus, and P. Avouris, Eds.), Springer Series in Topics in Applied Physics, Vol. 80, pp. 213–246. Springer-Verlag, Berlin, 2001.
9. M. S. Dresselhaus, G. Dresselhaus, and Ph. Avouris, "Carbon Nanotubes: Synthesis, Structure, Properties and Applications," Vol. 80 of Springer Series in Topics in Appl. Phys. Springer-Verlag, Berlin, 2001.
10. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, New York/San Diego, 1996.
11. R. Saito, A. Jorio, J. H. Hafner, C. M. Lieber, M. Hunter, T. McClure, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 64, 085312 (2001).

12. T. W. Odom, J. L. Huang, P. Kim, and C. M. Lieber, *Nature* (*London*) 391, 62 (1998).
13. T. Odom, J. H. Hafner, and C. Lieber, in "Carbon Nanotubes: Synthesis, Structure, Properties and Applications" (M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds.), Springer Series in Topics in Applied Physics, Vol. 80, pp. 173–212. Springer-Verlag, Berlin, 2001.
14. H. Kataura, Y. Kumazawa, Y. Maniwa, I. Umezu, S. Suzuki, Y. Ohtsuka, and Y. Achiba, *Synthetic Metals* 103, 2555 (1999).
15. G. Dresselhaus, M. A. Pimenta, R. Saito, J.-C. Charlier, S. D. M. Brown, P. Corio, A. Marucci, and M. S. Dresselhaus, in "Science and Applications of Nanotubes" (D. Tománek and R. J. Enbody, Eds.), pp. 275–295. Kluwer Academic, New York, 2000; Proceedings of the International Workshop on the Science and Applications of Nanotubes, Michigan State University, East Lansing, MI, 24–27 July 1999.
16. J. Kürti, H. Kuzmany, B. Burger, M. Hulman, A. G. Rinzler, and R. E. Smalley, in "Proc. of the XII Int. Winter School on Electronic Properties of Novel Materials: Molecular Nanostructures" (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), pp. 101–105, Amer. Inst. of Physics, New York, 1998.
17. A. Jorio, A. G. Souza Filho, G. Dresselhaus, M. S. Dresselhaus, R. Saito, J. H. Hafner, C. M. Lieber, F. M. Matinaga, M. S. S. Dantas, and M. A. Pimenta, *Phys. Rev. B* 63, 245416 (2001).
18. A. G. Souza Filho, A. Jorio, J. H. Hafner, C. M. Lieber, R. Saito, M. A. Pimenta, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 63, 241404R (2001).
19. Z. Yu and L. E. Brus, *J. Phys. Chem. B* 105, 6831 (2001).
20. M. S. Dresselhaus and P. C. Eklund, *Adv. Phys.* 49, 705 (2000).
21. A. Jorio, A. G. Souza Filho, G. Dresselhaus, M. S. Dresselhaus, A. K. Swan, M. S. Ünlü, B. Goldberg, M. A. Pimenta, J. H. Hafner, C. M. Lieber, and R. Saito, *Phys. Rev. B* 65, 155412 (2002).
22. A. Jorio, C. Fantini, M. S. S. Dantas, M. A. Pimenta, A. G. Souza Filho, Ge. G. Samsonidze, V. W. Brar, G. Dresselhaus, M. S. Dresselhaus, A. K. Swan, M. S. Ünlü, B. B. Goldberg, and R. Saito, *Phys. Rev. B* 66, 115411 (2002).
23. S. D. M. Brown, A. Jorio, P. Corio, M. S. Dresselhaus, G. Dresselhaus, R. Saito, and K. Kneipp, *Phys. Rev. B* 63, 155414 (2001).
24. A. Jorio, G. Dresselhaus, M. S. Dresselhaus, M. Souza, M. S. S. Dantas, M. A. Pimenta, A. M. Rao, R. Saito, C. Liu, and H. M. Cheng, *Phys. Rev. Lett.* 85, 2617 (2000).
25. A. Jorio, M. A. Pimenta, A. G. Souza Filho, Ge. G. Samsonidze, A. K. Swan, M. S. Ünlü, B. B. Goldberg, R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 20, 107403 (2003).
26. J. Kürti, V. Zólyomi, A. Grüneis, and H. Kuzmany, *Phys. Rev. B* 65, 165433 (2002).
27. G. S. Duesberg, I. Loa, M. Burghard, K. Syassen, and S. Roth, *Phys. Rev. Lett.* 85, 5436 (2000).
28. A. Jorio, A. G. Souza Filho, V. W. Brar, A. K. Swan, M. S. Ünlü, B. B. Goldberg, A. Righi, J. H. Hafner, C. M. Lieber, R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 65, 121402(R) (2002).
29. F. Tuinstra and J. L. Koenig, *J. Phys. Chem.* 53, 1126 (1970).
30. R. P. Vidano, D. B. Fishbach, L. J. Willis, and T. M. Loehr, *Solid State Commun.* 39, 341 (1981).
31. A. V. Baranov, A. N. Bekhterev, Y. S. Bobovich, and V. I. Petrov, *Opt. Spectrosk.* 62, 1036 (1987) [*Opt. Spectrosc.* 62, 612 (1987)].
32. M. J. Matthews, M. A. Pimenta, G. Dresselhaus, M. S. Dresselhaus, and M. Endo, *Phys. Rev. B* 59, R6585 (1999).
33. C. Thomsen and S. Reich, *Phys. Rev. Lett.* 85, 5214 (2000).
34. I. Pócsik, M. Hundhausen, M. Koós, and L. Ley, *J. Non-Cryst. Solids* 227–230, 1083 (1998).
35. R. Saito, A. Jorio, A. G. Souza Filho, G. Dresselhaus, M. S. Dresselhaus, and M. A. Pimenta, *Phys. Rev. Lett.* 88, 027401 (2002).
36. A. G. Souza Filho, A. Jorio, Ge. G. Samsonidze, G. Dresselhaus, M. A. Pimenta, M. S. Dresselhaus, A. K. Swan, M. S. Ünlü, B. B. Goldberg, and R. Saito, *Phys. Rev. B* 67, 035427 (2003).
37. A. Grüneis, R. Saito, Ge. G. Samsonidze, T. Kimura, M. A. Pimenta, A. Jorio, A. G. Souza Filho, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 67, 165402 (2003).
38. R. Saito, A. Jorio, J. H. Hafner, C. M. Lieber, M. Hunter, T. McClure, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 64, 085312 (2001).
39. O. Dubay, G. Kresse, and H. Kuzmany, *Phys. Rev. Lett.* 88, 235506 (2002).
40. V. W. Brar, Ge. G. Samsonidze, G. Dresselhaus, M. S. Dresselhaus, R. Saito, A. K. Swan, M. S. Ünlü, B. B. Goldberg, A. G. Souza Filho, and A. Jorio, *Phys. Rev. B* 66, 155418 (2002).
41. A. G. Souza Filho, A. Jorio, G. Dresselhaus, M. S. Dresselhaus, R. Saito, A. K. Swan, M. S. Ünlü, B. B. Goldberg, J. H. Hafner, C. M. Lieber, and M. A. Pimenta, *Phys. Rev. B* 65, 035404 (2002).
42. M. A. Pimenta, E. B. Hanlon, A. Marucci, P. Corio, S. D. M. Brown, S. A. Empedocles, M. G. Bawendi, G. Dresselhaus, and M. S. Dresselhaus, *Brazilian J. Phys.* 30, 423 (2000).
43. S. D. M. Brown, A. Jorio, M. S. Dresselhaus, and G. Dresselhaus, *Phys. Rev. B* 64, 073403 (2001).
44. A. Grüneis, M. Hulman, Ch. Kramberger, H. Peterlik, H. Kuzmany, H. Kataura, and Y. Achiba, in "Proc. of the XV Int. Winter School on Electronic Properties of Novel Materials (IWEPNM'01)" (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), AIP Conference Proceedings, p. 319. American Institute of Physics, Woodbury, NY, 2001.
45. A. G. Souza Filho, A. Jorio, Anna K. Swan, M. S. Ünlü, B. B. Goldberg, R. Saito, J. H. Hafner, C. M. Lieber, M. A. Pimenta, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 65, 085417 (2002).
46. A. G. Souza Filho, A. Jorio, Ge. G. Samsonidze, G. Dresselhaus, M. S. Dresselhaus, A. K. Swan, M. S. Ünlü, B. B. Goldberg, R. Saito, J. H. Hafner, C. M. Lieber, and M. A. Pimenta, *Chem. Phys. Lett.* 354, 62 (2002).
47. Ge. G. Samsonidze, R. Saito, A. Jorio, A. G. Souza Filho, A. Grüneis, M. A. Pimenta, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 90, 027403 (2003).
48. A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tománek, J. E. Fischer, and R. E. Smalley, *Science* 273, 483 (1996).
49. C. Journet, W. K. Maser, P. Bernier, A. Loiseau, M. Lamy de la Chapelle, S. Lefrant, P. Deniard, R. Lee, and J. E. Fischer, *Nature* (*London*) 388, 756 (1997).
50. R. A. Jishi, L. Venkataraman, M. S. Dresselhaus, and G. Dresselhaus, *Chem. Phys. Lett.* 209, 77 (1993).
51. E. Richter and K. R. Subbaswamy, *Phys. Rev. Lett.* 79, 2738 (1997).
52. R. Saito, T. Takeya, T. Kimura, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 57, 4145 (1998).
53. R. Saito, T. Takeya, T. Kimura, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 59, 2388 (1999).
54. J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, *Nature* (*London*) 391, 59 (1998).
55. J. M. Holden, P. Zhou, X.-X. Bi, P. C. Eklund, S. Bandow, R. A. Jishi, K. Das Chowdhury, G. Dresselhaus, and M. S. Dresselhaus, *Chem. Phys. Lett.* 220, 186 (1994).
56. P.-H. Tan, Y. Tang, Y.-M. Deng, F. Li, Y. L. Wei, and H. M. Cheng, *Appl. Phys. Lett.* 75, 1524 (1999).
57. M. A. Pimenta, A. Marucci, S. Empedocles, M. Bawendi, E. B. Hanlon, A. M. Rao, P. C. Eklund, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 58, R16016 (1998).
58. M. A. Pimenta, A. Marucci, S. D. M. Brown, M. J. Matthews, A. M. Rao, P. C. Eklund, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *J. Mater. Res.* 13, 2396 (1998).
59. S. D. M. Brown, P. Corio, A. Marucci, M. S. Dresselhaus, M. A. Pimenta, and K. Kneipp, *Phys. Rev. B* 61, R5137 (2000).
60. M. S. Dresselhaus, M. A. Pimenta, K. Kneipp, S. D. M. Brown, P. Corio, A. Marucci, and G. Dresselhaus, in "Science and

Applications of Nanotubes" (D. Tománek and R. J. Enbody, Eds.), pp. 253–274. Kluwer Academic, New York, 2000; Proceedings of the International Workshop on the Science and Applications of Nanotubes, Michigan State University, East Lansing, MI, 24–27, July 1999.
61. P. Corio, P. S. Santos, M. A. Pimenta, and M. S. Dresselhaus, *Chem. Phys. Lett.* 360, 557 (2002).
62. M. N. Iliev, A. P. Litvinchuk, S. Arepalli, P. Nikolaev, and C. D. Scott, *Chem. Phys. Lett.* 316, 217 (2000).
63. H. Wilhelm, M. Lelausian, E. McRae, and B. Humbert, *J. Appl. Phys.* 84, 6552 (1998).
64. A. M. Rao, J. Chen, E. Ricther, U. Schlecht, P. C. Eklund, R. C. Haddon, U. D. Venkatesvaran, Y. K. Kwon, and D. Tománek, *Phys. Rev. Lett.* 86, 3895 (2001).
65. M. Milnera, J. Kürti, M. Hulman, and H. Kuzmany, *Phys. Rev. Lett.* 84, 1324 (2000).
66. H. Kuzmany, W. Plank, M. Hulman, Ch. Kramberger, A. Grüneis, Th. Pichler, H. Peterlik, H. Kataura, and Y. Achiba, *Euro. Phys. J. B* 22, 307 (2001).
67. N. R. Raravikar, P. Keblinski, A. M. Rao, M. S. Dresselhaus, L. S. Schadler, and P. M. Ajayan, *Phys. Rev. B* 66, 235424 (2003).
68. O. Gorelik, P. Nikolaev, W. Holmes, R. Devivar, V. Hadjiev, S. Arepalli, and L. Yowell, submitted for publication.
69. M. S. Dresselhaus and G. Dresselhaus, *Adv. Phys.* 30, 139 (1981); 50, 1 (2002).
70. J. Chen, M. A. Hamon, H. Hu, Y. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon, Science 282, 95 (1998).
71. U. D. Venkateswaran, E. A. Brandsen, U. Schlecht, A. M. Rao, E. Richter, I. Loa, K. Syassen, and P. C. Eklund, *Phys. Status Solidi B* 223, 225 (2001).
72. D. Kahn and J. P. Lu, *Phys. Rev. B* 60, 6535 (1999).
73. P. Delaney, H. J. Choi, J. Ihm, S. G. Louie, and M. L. Cohen, *Nature* (*London*) 391, 466 (1998).
74. L. Henrard, E. Hernández, P. Bernier, and A. Rubio, *Phys. Rev. B* 60, R8521 (1999).
75. U. D. Venkateswaran, A. M. Rao, E. Richter, M. Menon, A. Rinzler, R. E. Smalley, and P. C. Eklund, *Phys. Rev. B* 59, 10928 (1999).
76. Y. K. Kwon, S. Saito, and D. Tománek, *Phys. Rev. B* 58, R13314 (1998).
77. L. Alvarez, A. Righi, T. Guillard, S. Rols, E. Anglaret, D. Laplaze, and J.-L. Sauvajol, *Chem. Phys. Lett.* 316, 186 (2000).
78. P. Corio, P. S. Santos, V. W. Brar, Ge. G. Samsonidze, S. G. Chou, and M. S. Dresselhaus, *Chem. Phys. Lett.* 370, 675 (2003).
79. A. M. Rao, P. C. Eklund, S. Bandow, A. Thess, and R. E. Smalley, *Nature* (*London*) 388, 257 (1997).
80. A. K. Sood, P. V. Teresdesai, D. V. S. Muthu, R. Sen, A. Govindaraj, and C. N. R. Rao, *Phys. Status Solidi B* 215, 393 (1999).
81. C. Thomsen, S. Reich, A. R. Goni, H. Jantoliak, P. M. Rafailov, I. Loa, K. Syassen, C. Journet, and P. Bernier, *Phys. Status Solidi B* 215, 435 (1999).
82. M. J. Peters, L. E. McNeil, J. P. Lu, and D. Kahn, *Phys. Rev. B* 61, 5939 (2000).
83. R. Gaál, J.-P. Salvetat, and L. Forró, *Phys. Rev. B* 61, 7320 (2000).
84. S. Kazaoui, N. Minami, Yamawaki H, K. Aoki, H. Kataura, and Y. Achiba, *Phys. Rev. B* 62, 1643 (2000).
85. Ulrich Schlecht, U. D. Venkateswaran, E. Richter, J. Chen, R. C. Haddon, P. C. Eklund, and A. M. Rao, *J. Nanosci. Nanotechnol.* 3, 139 (2003).
86. S. Reich, H. Jantoljak, and C. Thomsen, *Phys. Rev. B* 61, R13389 (2000).
87. M. Hulman, W. Plank, and H. Kuzmany, *Phys. Rev. B* 63, R081406 (2001).
88. S. Kazaoui, N. Minami, R. Jacquemin, H. Kataura, and Y. Achiba, *Phys. Rev. B* 60, 13339 (1999).
89. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Nature* 396, 323324 (1998).
90. D. E. Luzzi and B. W. Smith, *Carbon* 38, 1751 (2000).
91. J. Vavro, M. C. Llaguno, B. C. Satishkumar, D. E. Luzzi, and J. E. Fischer, *Appl. Phys. Lett.* 80, 1450 (2002).
92. D. J. Hornbaker, S.-J. Kahng, S. Misra, B. W. Smith, A. T. Johnson, E. J. Mele, D. E. Luzzi, and A. Yozdani, *Science* 295, 828 (2002).
93. P. M. Ajayan and S. Iijima, *Nature* (*London*) 361, 333 (1993).
94. P. M. Ajayan, T. W. Ebbesen, T. Ichihashi, S. Iijima, K. Tanigaki, and H. Hiura, *Nature* (*London*) 362, 522 (1993).
95. S. Bandow, M. Takizawa, K. Hirahara, M. Yudasaka, and S. Iijima, *Chem. Phys. Lett.* 337, 48 (2001).
96. G.-W. Jeong, T. Hirata, R. Hatakeyama, K. Tohji, and K. Motomiya, *Carbon* 40, 2247 (2002).
97. U. D. Venkateswaran, E. A. Brandsen, M. E. Katakowski, A. Harutyunyan, G. Chen, A. L. Loper, and P. C. Eklund, *Phys. Rev. B* 65, 054102 (2002).
98. J. E. Riggs, D. B. Walker, D. L. Carroll, and Y.-P. Sun, *J. Phys. Chem. B* 104, 7071 (2000).
99. S. Bandow, S. Asaka, Y. Saito, A. M. Rao, L. Grigorian, E. Richter, and P. C. Eklund, *Phys. Rev. Lett.* 80, 3779 (1998).
100. Y. Chen, R. C. Haddon, S. L. Fang, A. M. Rao, P. C. Eklund, W. H. Lee, E. C. Dickey, E. A. Grulke, J. C. Pendergass, A. Chavan, B. E. Haley, and R. E. Smalley, *J. Mater. Res.* 13, 2423 (1998).
101. R. Pfeiffer, H. Kuzmany, W. Plank, T. Pichler, H. Kataura, and Y. Achiba, *Diamond Relat. Mater.* 11, 957 (2002).
102. L. Kavan, L. Dunsch, and H. Kataura, *Chem. Phys. Lett.* 361, 79, (2002).
103. A. M. Rao, P. Zhou, K.-A. Wang, G. T. Hager, J. M. Holden, Ying Wang, W. T. Lee, Xiang-Xin Bi, P. C. Eklund, D. S. Cornett, M. A. Duncan, and I. J. Amster, *Science* 259, 955 (1993).
104. T. Okazaki, K. Suenaga, K. Hirahara, S. Bandow, S. Iijima, and H. Shinohara, *Physica B* 323, 97 (2002).
105. K. Suenaga, T. Tence, C. Mory, C. Colliex, H. Kato, T. Okazaki, H. Shinohara, K. Hirahara, S. Bandow, and S Iijima, *Science* 290, 2280 (2000).
106. K. Suenaga, K. Hirahara, S. Bandow, S. Iijima, T. Okazaki, H. Kato, and H. Shinohara, *AIP Conf. Proc.* 591, 256 (2001).
107. J. L. Hutchison, N. A. Kiselevi, E. P. Krinichnaya, A. V. Krestinin, R. O. Loutfy, A. P. Morawsky, V. E. Muradyan, E. D. Obraztsova, J. Sloan, S. V. Terekhov, and D. N. Zakharov, *Carbon* 39, 761 (2001).
108. L. Ci, Z. Rao, Z. Zhou, D. Tang, X. Yan, Y. Liang, D. Liu, H. Yuan, W. Zhou, G. Wang, W. Liu, and S. Xie, *Chem. Phys. Lett.* 359, 63 (2002).
109. W. Ren, F. Li, J. Chen, S. Bai, and H.-M. Cheng, *Chem. Phys. Lett.* 359, 196 (2002).
110. S. Bandow, G. Chen, G. U. Sumanasekera, R. Gupta, M. Yudasaka, S. Iijima, and P. C. Eklund, *Phys. Rev. B* 66, 075416 (2002).
111. S. Pekker, A. Jànossy, L. Mihaly, O. Chauver, M. Carrard, and L. Forró, *Science* 265, 1077 (1994).
112. Y. Iwasa, T. Arima, R. M. Fleming, T. Siegrist, O. Zhou, R. C. Haddon, L. J. Rothberg, K. B. Lyons, H. L. Carter, Jr., A. F. Hebard, R. Tycko, G. Dabbagh, J. J. Krajewski, G. A. Thomas, and T. Yagi, *Science* 264, 1570 (1994).
113. R. Andrews, D. Jacques, A. M. Rao, F. Derbyshire, D. Qian, X. Fan, E. C. Dickey, and J. Chen, *Chem. Phys. Lett.* 303, 467 (1999).
114. N. Chandrabhas, A. K. Sood, D. Sundararaman, S. Raju, V. S. Raghunathan, G. V. N. Rao, V. S. Satry, T. S. Radhakrishnan, Y. Hariharan, A. Bharathi, and C. S. Sundar, *PRAMANA-J. Phys.* 42, 375 (1994).
115. X. Zhao, Y. Ando, L.-C. Qin, H. Kataura, Y. Maniwa, and R. Saito, *Chem. Phys. Lett.* 361, 169 (2002).
116. A. M. Rao, A. Jorio, M. A. Pimenta, M. S. S. Dantas, R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 84, 1820 (2000); see also *Phys. Rev. Lett.* 85, 3545 (2000).

117. A. Kasuya, Y. Sasaki, Y. Saito, K. Tohji, and Y. Nishina, *Phys. Rev. Lett.* 78, 4434 (1997).
118. M. S. Dresselhaus, G. Dresselhaus, K. Sugihara, I. L. Spain, and H. A. Goldberg: "Graphite Fibers and Filaments," Vol. 5 of Springer Series in Materials Science. Springer-Verlag, Berlin, 1988.
119. H. D. Sun, Z. K. Tang, J. Chen, and G. Li, *Solid State Commun.* 109, 365 (1999).
120. H. Ajiki and T. Ando, *Physica B* 201, 349 (1994).
121. H. Ajiki and T. Ando, *Jpn. J. Appl. Phys. Suppl.* 34-1, 107 (1995).
122. M. Endo, K. Nishimura, Y. A. Kim, K. Hakamada, T. Matushita, M. S. Dresselhaus, and G. Dresselhaus, *J. Mater. Res.* 14, 4474 (1999).
123. M. Endo, Y. A. Kim, T. Takeda, S. H. Hong, T. Matusita, T. Hayashi, and M. S. Dresselhaus, *Carbon* 39, 2003 (2000).
124. M. S. Dresselhaus and M. Endo, in "Carbon Nanotubes: Synthesis, Structure, Properties and Applications" (M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds.), pp. 11–28. Topics in Applied Physics, Vol. 80: Relation of Carbon Nanotubes to Other Carbon Materials. Springer-Verlag, Berlin/Heidelberg, 2001.

www.aspbs.com/enn

Resists for Nanolithography

P. Argitis

NCSR Demokritos, Athens, Greece

CONTENTS

1. Introduction
2. Resist Fundamentals
3. Resist Design Strategies
4. Resists for Sub-100 nm Patterning
5. Polymeric Materials in Unconventional Lithographic Processes
6. Summary
Glossary
References

1. INTRODUCTION

Reliable fabrication of nanostructures represents one of the most essential steps that are encountered in a variety of applications in the expanding field of nanotechnology. In the specific area of materials encountered in photonic and electronic applications these nanostructures must be also formed on the basis of a predetermined pattern. A number of different approaches have been followed so far for the formation of such patterned nanostructures, including both polymeric resist based methods and novel unconventional methods. A few years ago progress in standard resist based lithography has been reviewed [1] in a special issue of *Chemical Reviews* devoted to nanostructures whereas the different "unconventional" approaches for nanostructure formation have been also reviewed in the same issue [2].

Resist based lithography is the predominant methodology for pattern formation in the general area of electronic and optical device fabrication and the one almost exclusively used by the semiconductor industry, allowing the ever-increasing number of individual elements in a chip so far. Nevertheless, as the discussion on the resolution limits of resists has already started [3, 4] some experts view that other approaches [5] will help in keeping Moore's law [6], which states that the number of transistors integrated on silicon chip doubles every 18 months. Currently, the resist technology has allowed semiconductor devices of critical dimensions down to 130 nm to have already entered production, whereas intense research and development activity is underway for the development of reliable industrial technologies suitable for 90 and 65 nm critical dimension devices [3, 7, 8]. According to the International Technology Roadmap for Semiconductors industrial nanofabrication processes will gradually reach 32 nm gate transistors within the next 10 years, and most probably this progress will be based on 193 nm, 157 nm, and extreme ultraviolet (EUV; 13 nm) lithography. Electron beam and ion beam lithographies are also being extensively studied as alternative nanofabrication methods.

The aforementioned technical goals can be accomplished through the development of resist materials suitable for the aforementioned wavelengths that combine a number of improved performance characteristics allowing subwavelength resolution and effective transfer of the pattern to the substrates of interest [7]. During the last decade the research activities related to lithography and in particular to resist materials have been greatly expanded. The progress in the field has been mostly recorded in the proceedings of the main relevant conferences, which include the annual SPIE Microlithography Conference, the Electron Ion Photon Beam and Nanofabrication (EIPBN) conference, Photopolymer Conferences of the Society for Plastic Engineers (Mid-Hudson), and certain American Chemical Society Symposia in USA, Photopolymer and Micro-Process Conferences in Japan, and the Micro- and Nano-Engineering Conference (MNE) in Europe. Nevertheless, the number of resist related papers published in broader scope journals has been also greatly increased during this last decade.

In this chapter we will discuss aspects of the progress on resist based nanolithography during recent years and in general nanolithography based on the use of photosensitive polymeric films. In particular we will focus on the photosensitive polymeric materials issues. We will discuss mainly chemical approaches but also physicochemical issues and technological challenges.

2. RESIST FUNDAMENTALS

2.1. Resist Based Lithographic Schemes

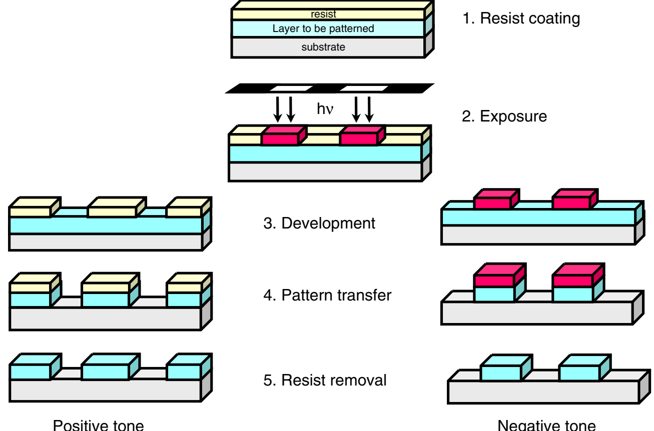
The basic principles of lithography have been discussed in a number of excellent articles or books in the past [3, 9–13]. Here, we present only some basic concepts necessary for understanding the latest developments in this field.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

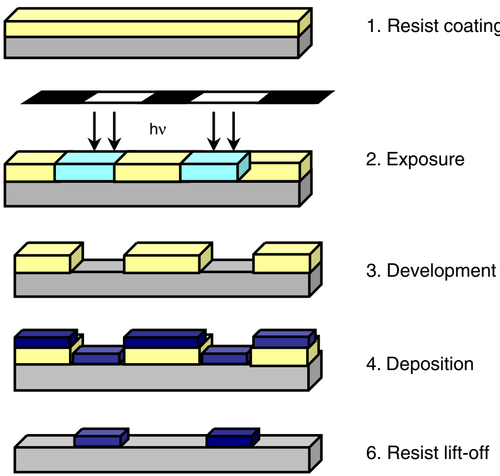
The most common lithographic process is shown in Scheme 1. There, a photosensitive polymeric material is coated over the layer that needs to be patterned and it is exposed through a mask. Radiation causes suitable chemical reactions in the exposed areas that change the polymeric material solubility. In the case where the exposed areas can be selectively removed by a suitable solvent we have positive tone lithography. In the opposite case, where the unexposed areas can be selectively removed, we have negative tone lithography. In a next step, the photosensitive polymeric film acts as a protective mask for the transfer of the pattern to the underlayer using suitable wet chemistry or reactive plasmas. In other words the polymeric film at this stage resists the reagent used for the etching of the bottom layer. The name *photoresist* has come from the main functions of the polymeric material according to this scheme: *photo*sensitivity and *resist*ance to the reagent used for etching of the underlayer. Since the radiation used can be also particles (electrons or ions) instead of photons the more generic term *resist* has been mostly used to characterize these materials.

In addition to the basic scheme presented several variations have been also proposed. The lithography represented by Scheme 1 is also called subtractive lithography since it is based on the deposition of the layer to be patterned and then on the removal (subtraction) of the nonprotected area by the patterned resist areas. An alternative lithographic process is presented in Scheme 2. According to this scheme the layer to be patterned is deposited after the patterning of the photosensitive polymeric material. The polymeric material is finally removed along with the material, which has been deposited on top of it. For this reason this process, which was introduced more than 30 years ago [14], is better known as the "lift-off" process. In this process the photosensitive polymeric material is not actually used to resist the action of a chemical reagent in the sense described in the first scheme, but to protect certain substrate areas during the deposition of the film to be patterned. Nevertheless, since the same or similar photosensitive polymeric materials are used in both schemes the term resist is commonly used in this case as well. The lift-off scheme could be considered a process of "prosthetic" lithography.

Photosensitive polymeric materials are also used in new processes for micro- and nanostructure fabrication proposed



Scheme 1. Typical resist process in semiconductor manufacturing.

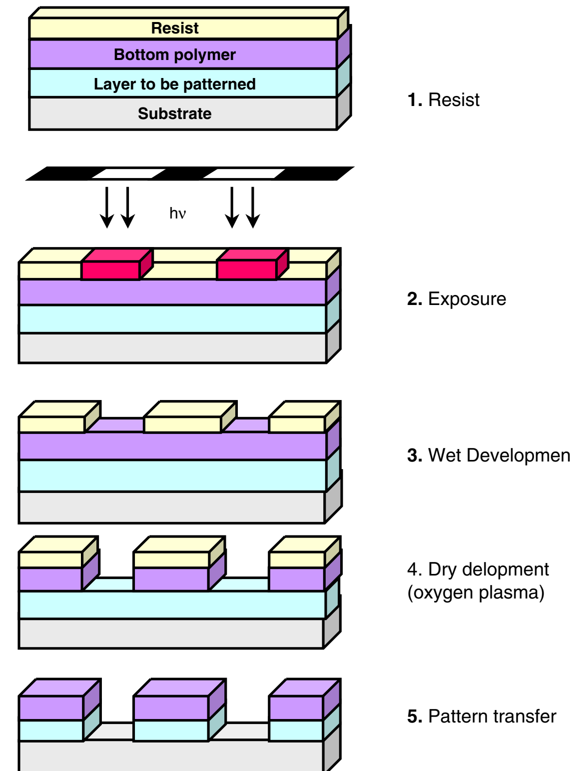


Scheme 2. Principle of lift-off process.

during the last decade, the most interesting probably being nanoimprint lithography [15]. Other radiation-guided processes have also been introduced especially for the patterning of biomolecule layers and the fabrication of microarrays. Some of these processes will be discussed further at the end of this chapter.

On the other hand more complicated resist based processes than the one presented in Scheme 1 have been also introduced in order to facilitate the fabrication of high resolution structures. In the double layer process shown in Scheme 3 two polymer layers are coated over the layer that needs to be patterned. The first, or bottom, polymer layer does not need to be photosensitive but must provide adequate etch resistance to the plasmas used for the patterning of the underlayer. The second, or top, polymer layer must be photosensitive and capable of providing high resolution patterning through wet development, but it is not necessary to provide high etch resistance to the plasmas used for the etching of the underlayer. It must, though, contain an element that does not form volatile oxides in order to provide enough etch resistance to the oxygen plasmas that can be used for the patterning of the first polymer layer. Si containing polymers have been mostly used in the top layer [16–22] but polymeric materials containing other elements that do not form volatile oxides, like W and Ge, have been used as well [23, 24]. In the previously described scheme the top resist layer is usually thinner than the bottom layer, since this bottom layer will actually act as the layer that will provide the etch resistance for the pattern transfer to the substrate of interest. The top layer needs just to resist during the etching of the bottom polymer layer with oxygen plasma, the so-called dry development step. The presence of elements that form nonvolatile oxides in the upper layer usually helps to easily achieve this goal. Thus, often the top layer can be substantially thinner than the resist layers used in the standard, single layer resist processes of Scheme 1. This fact facilitates lithography and often double layer processes are used in order to obtain higher resolution and higher aspect ratio features.

Other processes where a thin layer near the top surface of the resist needs to be imaged have been developed as well. These processes are generally known as top surface



Scheme 3. Double layer resist process.

imaging processes and they are used also in many cases to facilitate lithography and obtain higher resolution. The best known examples of this type are provided by the silylation processes [25–27]. There, light induces chemical reactions, usually cross-linking, in the exposed areas and this way the penetration and reaction of a silylating agent in the resist film are controlled. This silylation agent can be brought either in gas [25] or in liquid [26, 27] phase, giving rise to dry or wet processes correspondingly. The areas where the silylating agent is inserted (i.e., the unexposed areas in the case of systems that cross-link upon exposure) acquire a high etch resistant to oxygen plasmas and thus can form structures during a dry development, or oxygen plasma, step in an analogous manner to the case of double layer processes. Processes that combine double layer and silylation lithography have also been proposed and developed [28–31].

Related to the double layer are also the triple layer processes [11]. In this case between the two polymer layers there is also one additional layer, usually a thin inorganic layer. Thus, after the wet development of the top-imaging layer the pattern is transferred to the thin inorganic layer and then this layer provides the necessary etch resistance to the oxygen plasma used for the etching of the bottom layer. In the case of triple layer processes the top polymer layer does not need to contain an element giving nonvolatile oxides. Thus every resist can be used in such a process and thus many possibilities for the development of ultrahigh resolution processes exist.

2.2. Basic Resist Performance Requirements

According to the lithographic schemes presented resists must possess a number of properties to meet performance requirements needed in high resolution lithography. Here we will discuss briefly the basic requirements that need to be fulfilled.

First, the resists should form uniform films by spin-coating. This practically has led to the adoption of linear polymers as the main resist components. Nevertheless, an increasing research effort has started lately targeting the use of small molecules as main components of resist formulations. Mostly oligomers or polymers of different architecture like dendrimers and cyclic polymers have been used so far as the main resist components [32–36]. This issue is under intense investigation currently toward the design of ultrahigh resolution resists for e-beam and next generation lithography (see also the following relevant sections), that is, resists targeting to sub-50 nm resolution, where polymer MW, radius of gyration, and in general resist component molecular dimensions are thought to play a vital role on the quality of the patterns.

Inorganic resists have also been proposed for ultrahigh resolution patterning [37, 38]. These resists could not be deposited by spin-coating but instead they were deposited by physical vapor deposition methods. Attempts to design spin-coatable inorganic resists have also been reported [39]. Inorganic resists will be discussed along with molecular resists in the section on e-beam lithography.

Resist absorbance at the exposure wavelength is probably the most widely discussed issue in recent years. This is due to the fact that as resolution demands push the technology to shorter and shorter optical wavelengths the selection of polymers, and in general resist components, with suitable absorbance characteristics becomes more challenging. The demand for optimizing absorbance has been the driving force for designing new resists during the last 20 years. In general, resist films must have a moderate absorbance at the wavelength of exposure to allow proper imaging of the film in all its depth. This practically means that we need to have a limited number of chromophores inside the resist. Nevertheless, at short UV wavelengths most organic groups start to absorb. Especially in the 157 nm spectral region the problem is very severe, since C–H and C–C bonds also absorb. On the other hand in EUV (13 nm), X-ray, and charged particle lithography (e-beam, ion beam) the resist absorbance is defined by the atomic composition and material density.

Photochemistry and dissolution change chemistry play a vital role in resist design as well. After exposure, a latent image is formed inside the resist film defined by the distribution of chromophores (or absorbers) inside this material. Depending on the photochemistry that follows and on possible subsequent reactions that lead to the resist solubility change this latent image is probably modified. In the most common today case, that is, in the case of the so-called "chemically amplified resists" [40–44], which are based on photochemical generation of acid that next catalyzes reactions leading to solubility change, acid diffusion determines this latent image modification. The optimization of acid diffusion length defines to a great extent the high resolution performance of a specific resist process.

Development plays also a very important and, often, critical role in resist performance. During this step the latent image is transformed to a real relief structure based on the developer selectivity and the development mechanism. The most common problem during development is the swelling of the resist areas that will not be removed, that is, unexposed areas of positive resists or exposed areas of negative resists. This problem is very often encountered in negative resists based on cross-linking mechanisms but is not limited to these systems. In recent years, swelling is less common due to the adoption of aqueous base developable systems, where development mechanisms minimize swelling, as will be discussed in the next section.

Pattern transfer is the final goal of the resist process in most cases. In microelectronics during recent years the pattern transfer is almost exclusively performed by using reactive plasmas for the etching of the substrate. The chemical composition of resist defines its resistance to these plasmas. Empirical rules based on experimental data and mechanistic considerations can be used to predict the general performance of different materials [45–49]. In general, carbon rings and especially aromatic rings are desirable for improving the etch resistance of the material, due to their low reactivity toward decomposition and formation of gaseous products. On the other hand the presence of reactive groups and in general heteroatoms such as oxygen reduces the etch resistance of the material. In recent years as the lateral structure dimensions shrink, thin photoresist films, lately below 0.4 micrometers, are used to keep aspect ratios lower than about 3 and thus allow the fabrication of dense structures. Consequently, the improvement of the film resistance during the plasma etching step appears critical for effective pattern transfer especially at the sub-100 nm regime. Increased plasma resistance would allow us to obtain high resolution structures with improved aspect ratio.

3. RESIST DESIGN STRATEGIES

3.1. General Considerations

The design of resist materials that could fulfill the requirements discussed represents in most cases a very challenging problem. The resists should encompass suitable components to control the desired properties, components that should also be mutually compatible and if possible should work in a synergistic way toward optimum performance.

A polymer is almost always the basic component of the resist for a number of reasons. The most important include the necessity for film formation properties by spin-coating, as discussed, and the capabilities for dissolution change induced by small extent chemical changes. Indeed, the selection of a suitable strategy for photochemically induced property change, usually solubility change, in a resist film plays a predominant role in the design of a resist material. The main routes followed so far are discussed briefly in the following.

(1) The simpler case is the one where the light induces polymer cross-linking. Thus, the exposed areas become insoluble to the solvents of the unexposed regions.
(2) The second case is referred to as radiation induced polymer main chain scission. In this case the exposed areas become more soluble than the unexposed.
(3) Most interesting is the third case where the light induces chemical changes in characteristic functional groups resulting in hydrophilicity changes that differentiate solubility. Most common is the case where hydrophobic groups change to hydrophilic and thus the solubility of the polymeric film to aqueous developers increases.

The driving force of the research in the area of resists during the last 30 years is the need for the decrease of the electronic device dimensions. This decrease has proceeded according to the aforementioned Moore's law during these years and has resulted in an increase of the electronic device number in memory and microprocessor chips from 10^3 in 1970 to more than 10^9 today. The progress in lithography, including resists, played a critical role in the realization of this tremendous technology evolution. The main direction followed during the last 15 years in lithography technology development was the reduction of exposure wavelength first from 436 to 365 and then to 248 and to 193 nm respectively. Currently there is intense ongoing research for the development of 157 and 13 nm lithographic technologies.

Each time a new exposure wavelength was considered for introduction in lithography new resists had to be developed with absorbance and photochemical properties suitable for exposure at this wavelength of choice. This need determined the evolution of chemistry approaches in the resist field. In the following we will discuss the basic resist approaches that are currently considered for high resolution lithography.

Although this chapter targets mostly the systems used for nanolithography, that is, systems with sub-100 nm resolution capabilities, characteristic systems with resolution in the 0.1–1 μm range will be discussed as well since a lot of new systems proposed for sub-100 nm lithography have evolved from these resist materials.

3.2. Novolac Based Resists with Diazonaphthoquinone Dissolution Inhibitors

The resist system of choice for 436 nm lithography was the one based on novolac polymer matrices and diazonaphthoquinone (DNQ) photosensitizers. Variations of this resist approach were also used in 365 nm lithography. The novolac DNQ system finds up to now very wide use and it can still be considered as the standard resist material [50]. The solubility changing photochemistry in this system is presented in Scheme 4. The starting phenolic polymer has a small dissolution rate in aqueous base. Nevertheless, upon addition of the photosensitizer the solubility drops dramatically and increases again after the photochemical conversion of DNQ to indene carboxylic acid to a rate higher than the one of the starting novolac polymer, as shown in Scheme 5. Thus, the diazonaphthoquinone sensitizer acts in this system as strong **dissolution inhibitor** and its photochemistry offers the desired solubility change. The reasons for this effect and the physical chemistry of dissolution in **aqueous base** were studied in detail by a number of researchers during the

Novolac Diazonaphthoquinone

Diazonaphthoquinone photochemistry

Scheme 4. Solubility change photochemistry in novolac resists.

last decade. An explanation proposed for the strong dissolution inhibition effect of DNQ in phenolic matrices attributes this action to strong hydrogen bonding of the diazo group to the phenolic OH, which gives rise to a chain of hydrogen bonds involving phenolic hydroxyl groups. On the other hand the physical chemistry of phenolic polymer dissolution in aqueous base has been extensively studied. According to the accepted picture the dissolution mechanism proceeds through ionization of phenolic groups by the base, which enhances the polymer dissolution in the aqueous developer. Thus, the developer diffusion into the polymer matrix is the slowest step followed by fast polymer chain removal. This way swelling effects are practically eliminated. This dissolution behavior of phenolic resists contributed at most to their adoption for high resolution lithography as opposed to previously used negative systems based on cross-linking mechanisms, which suffered strongly from swelling problems. The development mechanisms of phenolic polymers have been extensively studied during recent years, mostly by Reiser et al. [51] and Flanagin et al. [52–56].

3.3. Chemically Amplified Resists

High sensitivity was always one of the desirable properties one resist should have in order to be useful for practical applications, especially under exposure with the extremely high cost, and high operational cost, laser source based exposure systems of optical lithography used in the last decade in semiconductor manufacturing. Enhancement of sensitivity was also always of high demand in the area of electron beam lithography to reduce the long times necessary for the writing of complicated circuits. High sensitivity resists had been actually designed for use in electron beam lithography that were also sensitive in deep UV, based on reactions resulting to polymer backbone scission, like polysulfones [57], but these resists had poor performance in pattern transfer through plasma processing. The strategy that came to revolutionize resist design by offering very high sensitivity and at the same time a lot of capabilities to control other resist properties was the so-called "**chemical amplification**" introduced in the early 1980s [58–62]. This strategy was gradually adopted in most resists designed since then and it is currently the basis for 248 and 193 nm resists used in the semiconductor industry [40–44, 63]. Moreover most of the resists proposed for 157 nm, EUV, and next generation lithography are also based on chemical amplification, whereas chemical amplification based resists are also used in electron beam, ion beam, and X-ray lithography.

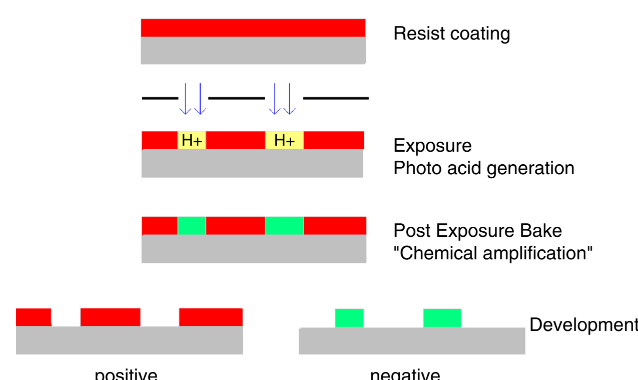
Dissolution rate

Scheme 5. Dissolution rate changes in novolac systems.

The chemical amplification strategy is based on the photogeneration of a catalyst during exposure, which in the next step, typically at elevated temperatures, catalyzes the reactions that result in solubility change. The catalyst produced is, in almost all cases, a Brønsted acid, which then catalyzes, or in some cases only initiates, the desired chemical reactions. Exposure is usually followed by a postexposure bake step, where these reactions take place. Typical reactions that take place in chemically amplified resists are shown in Scheme 6. The photosensitizers used in these resist systems, which are also called photoacid generators, can come from different organic compound classes [64–71]. The most commonly used are the sulfonium or iodonium salts [64, 65]. These types of compounds have been proposed for use in systems where acid catalyzed chemistries are desired a few years before the introduction of chemically amplified resists and they are very effective photoacid generators (PAGs). On the other hand pure organic photoacid generators have also been proposed to improve compatibility with other resist components [66]. Lately a quite intense activity is reported in this area and different new PAGs are evaluated in order to improve among others resist transparency, uniformity, line edge roughness, and outgassing performance of new materials designed for 193 nm, 157 nm, and next generation lithographies [67–71].

Acid catalyzed or acid initiated chemistries can be used in numerous different ways in order to bring about desired solubility changes. Extensive reviews have already been presented covering numerous such systems [40–44]. Here, we discuss briefly just a few representative cases. Since positive systems continue to dominate resist design the most common reactions that are encountered in the commercially available resists today are reactions of acid catalyzed transformation of hydrophobic to hydrophilic groups in order to render them soluble in aqueous base. In Scheme 6 we present three representative such acid catalyzed cases based on (a) deprotection of *t*-BOC groups to phenols, (b) acid catalyzed thermolysis of *t*-butyl acrylate esters, and (c) acid catalyzed deprotection of ketals. These chemistries were first

Scheme 6. Typical reactions encountered in chemically amplified resists.



Scheme 7. Lithography with chemically amplified photoresists.

introduced for 248 nm resists but are currently used also in numerous systems designed for 193 and 157 nm lithography. Acid catalyzed main chain cleavage represents another interesting case [73]. On the other hand one of the first systems proposed for 248 nm lithography was a negative resist based on acid catalyzed cross-linking of phenolic polymer by melamine through *trans*-etherification reactions [74]. Similar formulations were used in e-beam lithography and X-ray lithography (commercially available resist SAL 601). Another negative resist system is based on acid initiated cationic polymerization of epoxy polymers. Quite a few such systems were reported as will be further discussed in the e-beam resist section. Some negative systems have also been used for ultrahigh resolution lithography by e-beam exposure and they will be also discussed in the relevant section.

The process needed when using chemically amplified resists is more complicated than the conventional lithography process presented in Scheme 1, since it requires an extra step (i.e., the postexposure bake step) as shown in Scheme 7. This step plays actually a dominant role in the final lithographic performance since it controls the kinetics of the reactions resulting in solubility change. First, the photogenerated acid has to diffuse in order to catalyze the desired reactions and the diffusion length has to be controlled in order not to cause undesirable pattern alterations. Second, the temperature and the length of the step also have to be optimized according to a complicated diffusion–reaction kinetics encountered in most polymeric resist systems. In addition to diffusion and reaction processes taking place during the postexposure bake step similar phenomena that can strongly influence the lithographic performance are also encountered at the time intervals between the different processing steps. Most important are the ones between exposure and postexposure bake, depending on the *Tg* of the specific polymeric material, the *Ea* of the possible reactions, and the presence of potential contaminants, mostly basic compounds, in the atmosphere. These issues have been discussed in a large number of publications based on both experimental and simulation studies. These publications have appeared mostly in proceedings of lithography symposia and conferences like SPIE Microlithography, EIPBN, Micro-process, and MNE. Lately there is intense activity that involves the use of physicochemical methods in order to monitor the chemical reactions and acid diffusion processes taking place in the chemically amplified resist systems [75–78].

These processing issues will be covered here only briefly and to the extent they directly influence the resist design strategies. Thus, we have to mention that a family of resists, the so-called ESCAPs (environmentally stable chemically amplified positive resists), have been introduced to address environmental contamination [79, 80]. According to this approach the adoption of tert-butyl acrylate esters instead of *t*-BOC protected phenols allows the postapply bake to take place at higher temperatures, without thermal stability problems, and this way the free volume inside the film is reduced, minimizing also contaminant uptake from the environment. On the other hand base additives at small quantities are also introduced inside the resist film to control acid diffusion phenomena [81]. These basic additives can also be photosensitive in order to render to acids by the light at the exposed areas but to function as bases in the unexposed ones, where they need to scavenge the diffused acid [82, 83].

Finally, it should be mentioned that the basic processing issues encountered in the chemically amplified resists were presented a couple of years ago by Hinsberg and co-workers [84] and Ito [85].

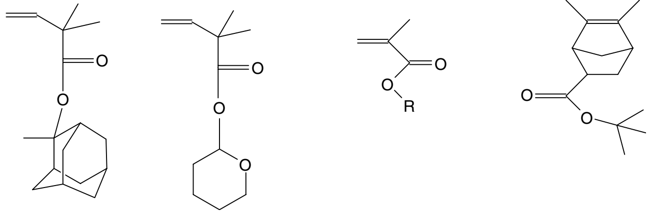
4. RESISTS FOR SUB-100 nm PATTERNING

4.1. The 193 nm Resists

Resists suitable for exposure at 193 nm have been introduced in production in recent years for sub-130 nm lithography. These resists are also believed to be the main choice for the production of 90 nm critical dimension circuits during the next few years and probably their use could be further extended. Reviews covering the efforts toward 193 nm resist design have appeared in the scientific literature in recent years [44, 86]. On the other hand new materials are still being proposed and the resolution limits to be reached are still under investigation. Here, we will discuss the main topics related to the resist materials considered and developed.

The main challenge that had to be addressed in the design of resists suitable for 193 nm lithography was the unacceptably high absorbance of aromatic polymers used up to that point as main resist components. Aromatic polymers offered high resistance to the plasmas used in pattern transfer and thus served as a suitable platform to build useful functionalities in previous wavelengths. Nevertheless, at 193 nm these polymers could be imaged only in top-surface imaging processes [88–91].

On the other hand methacrylate homopolymers and in particular poly(methyl methacrylate) (PMMA) were first used as single layer resists in 193 nm to demonstrate resolution capabilities offered by this wavelength [92, 93]. Acid catalyzed resist systems based on acrylate copolymers were then proposed to improve sensitivity [94] and imaging properties. These polymers, though, had very poor etch resistance and thus new components had to be introduced to bring the necessary improvements. The strategy proposed first by Kaimoto et al. [95] and Endo et al. [96], and adopted by most researchers in the field, was the incorporation of cycloaliphatic groups in the polymer. Empirical etch resistance prediction models and experimental studies proved the suitability of this type of component. Thus, one approach for 193 nm resist polymer design was based on the use of (meth)acrylate copolymers [97–107] synthesized from monomers such as the ones shown in Scheme 8. There, in addition to the cycloaliphatic moiety containing monomer, other monomers bear an acid cleavable group necessary for the imaging through a chemical amplification mechanism, possibly a carboxylic acid group for easier change to a base soluble material and probably a secondary hydrophilic group for adhesion and probably compatibility reasons. The exact chemical composition in each case is mostly unknown and it is the result of fine structure tuning toward optimum performance. Resist compositions in 193 nm involve also the selection of a suitable photoacid generator, possibly a base component, and even other additives to control the overall performance. In this (meth)acrylate platform described belong some of the commercial resists now available for 193 nm. The (meth)acrylate backbone offers many possibilities for copolymer optimization and thus high performance systems have been obtained. Nevertheless, in most cases the etch resistance is still considered as rather moderate due to the presence of a high number of ester groups which show high reactivity in the plasma environment. For this reason the cycloaliphatic component(s) content is increased to the maximum level allowed by the need to fulfill other resist properties as well. In a characteristic case reported by NEC researchers all the copolymer components contain cycloaliphatic moieties properly functionalized [106]. Photoacid structure effects in similar systems have also been studied [107].



Scheme 8. Typical monomers that are used in copolymers for 193 nm lithography.

As an alternative to the (meth)acrylate backbone systems, polyolefines have been also introduced and adopted by many researchers in the field [108–114]. Examples of such systems provide norbornene *alt* copolymers with maleic anhydride. A typical monomer used in the synthesis of the copolymers is also shown in Scheme 8. The resists based on such copolymers exhibit higher etch resistance compared to the (meth)acrylates, but their composition is most difficult to be optimized. For instance they very often present adhesion problems, unacceptably high *Tg*, and also self-life problems. Nevertheless, high quality, high resolution patterns with such type of resists have been demonstrated. Lately novel polymers that belong to this family have also been proposed. A characteristic example of a novel material based on cyclopolymers has also been reported [115].

Other polycarbocycle compounds have been also examined as possible components of 193 nm resists. In particular cholic acid derivatives have been mostly evaluated as dissolution inhibitors and relevant resist formulations have been presented [114]. Furthermore, the possibility of including a percentage of an aromatic component in 193 nm resists has been also considered. As indicated by Toshiba researchers [116, 117], the UV spectrum of polyaromatics is redshifted in the 193 nm region compared to the simple aromatic rings resulting to an increased transparency in this region. Thus, 193 nm resist formulations containing a certain extent of naphtahalene components have been presented, but the transparency was not still adequate. On the other hand anthracene compounds [118] have been examined as possible etch resistance additives and recently new anthracene derivatives have been presented and evaluated for this application [119].

Silicon containing polymers were also considered as the basis of bilayer resist formulations. Si in this case was incorporated in a pendant group [120]. High quality patterns have been demonstrated with resists of this type. Nevertheless, concerns for outgassing and contamination of lenses with difficult to remove Si species have limited the use of such resists.

Outgassing is referred to as the undesirable formation of gaseous products during illumination of the resist [121]. This is due to photochemical processes that result in direct bond breaking. The phenomenon was first discussed in the case

of 193 nm lithography, since at this wavelength the photon energy is high enough to induce breaking of certain organic molecule bonds. These direct photochemical processes cannot obscure the main solubility changing chemistry in chemically amplified systems but they mostly raise concerns about possible contamination of the lenses with gaseous products. It has been reported that carbon species can be removed from the lenses using relatively easy procedures based on treatment with oxygen. Nevertheless, things are much worse with Si containing species, which can form silicon oxides and thus eventually destroy the lenses.

Lately, the etch resistance enhancement with the addition of Si nanoparticles in the resist formulation has also been proposed [122]. Most interesting seems to be the case of incorporating silsesquioxane groups in 193 nm resist formulations by using polymerizable compounds bearing such groups to synthesize methacrylate copolymers [122, 123]. These approaches will be also discussed in the next sections.

Development mechanistic studies related to 193 nm resists have been also reported [124]. On the other hand it has also been pointed out that diluted developers could present some advantages for process optimization in these nonaromatic resist systems [125–127], but the semiconductor industry is very reluctant to adopt changes from the standard base concentration (0.26 N) used today.

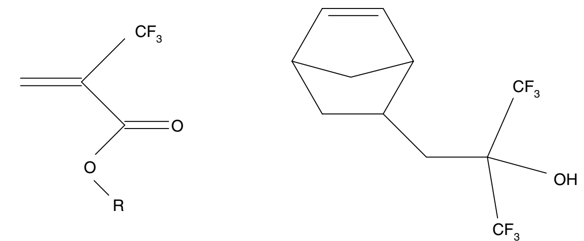
At the end of this section on resist materials for 193 nm lithography efforts for negative resists will be briefly discussed. The demand for negative tone resists is very limited during recent years despite the fact that resists of this tone are considered to present advantages for certain types of patterns. New chemistries have been reported for negative resists, mostly acid catalyzed esterifications and *trans*-esterifications. Moderate resolution results, compared with the positive systems, have been reported [128–131].

4.2. The 157 nm Resists

Lithography at 157 nm is currently the most serious candidate to succeed the 193 nm lithography in the production of integrated circuits. It is believed that it will be the technology of choice for devices having critical dimensions of 65 nm, which according to the Semiconductor Industry Association roadmap will probably enter in production during the next few years [7, 8]. Thus, at the moment a huge research effort worldwide has been directed for the timely development of the relevant components that are necessary. The progress in this area is being reviewed relatively often mainly in the SPIE, Photopolymer, Microprocess, and International Sematech Conferences, helping give timely information to the research community.

The development of photoresists with acceptable lithographic performance at this wavelength represents one of the major challenges. The most critical issue for 157 nm resist development is the accomplishment of adequate transparency, since traditional resist platforms are too opaque to allow imaging in sufficiently thick films. Indeed even C–C bonds and C–H bonds absorb considerably strongly at this wavelength and thus almost all organic materials are too opaque to be used as main components of single layer resists. This difficulty was well recognized a few years ago [132–137], along with the severe outgassing problems anticipated and first investigated [138–140] or the different photochemical pathways that are possible at this wavelength [141, 142], when 157 nm lithography appeared as a strong candidate for entering the semiconductor industry in coming years. The absorbance of typical organic polymers is mostly in the range of 5–10 μm^{-1}. Although aromatic polymers exhibit generally high absorbance values aliphatic polymers also absorb unacceptably high. Only two main classes of polymeric materials have reduced absorbance at this wavelength: first, polymers containing a high percentage of C–F bonds and, second, polymers containing a high percentage of Si–O bonds. Nevertheless, existing polymers bearing high percentages of either type of bonds are far from fulfilling criteria to provide resist materials with desired performance, such as presence of imageable functional groups for positive lithography, aqueous base development, high resolution patterning, etch resistance, etc. Thus, a huge research effort is in progress in recent years aiming at the synthesis of novel polymers that could be the basis of candidate 157 nm resists materials.

Partial fluorination for absorbance reduction of polymers that were used in longer wavelengths was the main strategy followed [143–153]. Since models for predicting absorbance of different materials at the 157 nm had, at least in the beginning, a rather limited prediction capability a lot of new polymers had to be synthesized and evaluated experimentally. A few groups appeared to be extremely useful in this absorbance reduction effort. Through this effort great progress in absorbance reduction has been achieved so far. Characteristic groups that were proved very effective in reducing the total resist absorbance are shown in Scheme 9. Polymers with absorbance in the order of 1 μm^{-1}, or even slightly below, have been recently reported. Thus, lithographic performance has been substantially improved as the necessary film thickness for single layer resists in this case is 200–250 nm. Nevertheless, a number of problems still have to be solved. The materials must additionally satisfy numerous other requirements, such as adhesion, high resistance to etching by reactive plasmas, controlled dissolution behavior, and aqueous base development. Therefore, the design of materials suitable for 157 nm resist applications presents a great challenge. For instance etch resistance of heavily fluorinated resist materials is low. Incorporation of fluorine can dramatically affect the physicochemical characteristics of the resist, making aqueous base dissolution more difficult and reducing adhesion. The presence of a big number of fluorinated groups causes problems in film homogeneity, since these groups tend to phase separate from the rest for the resist components. Surface composition differentiation



Scheme 9. Monomers that are used for increased transparency in copolymers for 157 nm lithography.

phenomena are also expected. Thus, irregular development in such systems is common.

Siloxanes and silsesquioxanes offer an alternative route for designing resist materials with optimized performance at 157 nm. Thus research is also taking place in this area in order to develop appropriate materials with high resolution positive imaging capabilities [154, 155]. Siloxanes with small vinyl content have been used in the past as solvent developable negative resists in deep UV and recently negative behavior has also been shown in 193 nm. On the other hand even simple siloxanes, like poly(dimethyl siloxane) (PDMS), behave as negative resists under ion beam and electron beam exposure. The same behavior has been observed in 157 nm [155]. Thus, novel polymers have to be designed in order to formulate positive tone resists, with aqueous base development capability. A few years ago synthesis of polysiloxanes with acid-sensitive side groups was reported. The main drawback of these materials appears to be their low glass transition temperature, and thus novel designs are under investigation for 157 nm lithography. Details of this research effort have not been reported, although very good lithographic results of sub-100 nm resolution have been shown in lithography symposia in the last year. Silsesquioxanes of the ladder-type structure have also been extensively examined as negative materials, mostly for e-beam applications, in recent years. Recently a fluorocarbinol-substituted polymer with acid-labile ester functionalities has been suggested as a promising positive tone material for 157 nm applications [154]. Finally, silsesquioxanes of the cage-type structures, belonging to the same family with the ones first studied for 193 nm lithography, have been recently investigated for 157 nm lithography [155]. Partial fluorination can also be applied in these systems for absorbance optimization.

Outgassing, first discussed in the section devoted to the 193 nm resists, is a much more important issue at 157 nm. The high photon energy (7.6 eV) at this wavelength is enough to directly break a great number of the molecular bonds encountered in organic molecules. As discussed in the 193 nm resist section outgassing from carbon based resists is not considered so a big problem since it is relatively easy to avoid deposition of the gaseous products on the lenses or even to clean it afterward. In the contrary potential damage of lenses is mostly a big concern when Si containing polymers are used, since in this case Si species can form nonremovable silicon oxides on the lens surface. Specific protocols for the evaluation of deposits on the lenses have been proposed. It has been shown that outgassing of Si containing species from polymers bearing Si in the side chain is unacceptable. On the other hand in siloxane and silsesquioxane polymers only traces of Si containing species, if any, have been detected [140]. The fact that the SiO bond is almost transparent at 157 nm and the binding of Si to two different chain atoms in siloxanes, or to three different chain atoms in silsesquioxanes, reduce substantially the outgassing risk.

4.3. E-Beam Resists

Electron beam lithography is the technology providing the highest resolution structures and, thus, it is the predominant tool used in nanolithography. On the other hand often electron beam lithography could also serve as a benchmark to evaluate the ultimate resolution capabilities of the different resist systems.

The best known high resolution e-beam resist is poly(methyl methacrylate) (PMMA), which actually is one of the simpler resist systems, since it is a one component resist. Although this resist was introduced more than 30 years ago [14, 156], it is still intensively studied. The interest in this resist has been partly renewed after Chen and Ahmed demonstrated sub-10 nm lithographic features [157] in the early 1990s, and since then it can be considered the main vehicle used in nanolithography investigations.

The radiation chemistry involved in the imaging of this resist has been extensively studied and is discussed in most of the standard lithography sources, for instance in [11]. General considerations on the behavior of polymeric materials under exposure in high energy radiation are also available [158, 159]. PMMA is a typical example of a polymer where chain scission occurs under high energy exposure, whereas cross-linking is predominant in other polymer types, for instance in polystyrene. Chain scission also takes place under exposure of PMMA to X-rays [161] but also to deep UV irradiation [162]. In any case, due to this chain scission imaging mechanism, the ultimate resolution is not determined by the size of the starting macromolecule but by the size of the produced chain fragments after exposure.

On the other hand PMMA is known to have very low sensitivity and poor etch resistance to the plasmas used for pattern transfer. For this reason over the last 30 years many efforts were undertaken in order to modify its structure with the goal to improve its performance in these aspects. Copolymers of methyl (methacrylalate) with other acrylate monomers were mostly proposed by many researchers and substantial improvement in sensitivity was obtained with quite a few of them, for instance in copolymers with methacrylic acid [163, 164]. Nevertheless, the PMMA sensitivity remained generally low (doses close to or higher than 100 μ C/cm^2 for 20 keV exposure are needed in most of the cases). Thus, PMMA was not preferred for use in mask fabrication where other systems based on poly(alkyl sulfones), like PBS, were developed [165–168] in the mid 1970s. More recently, though, certain acrylate polymers, like poly(fluoroalkyl methacrylates) [169] or poly(methyl chloroacrylates) [170, 171] proved to give improved sensitivities, whereas copolymers bearing phenyl moieties [172] improved etch resistance. Thus some commercial products based on these chemistries are now available. Especially interesting is the case of the resist ZEP520, which is a copolymer of methyl *a*-chloro acrylate and *a*-methyl styrene and provides good sensitivity, good resolution capabilities, and improved etch resistance.

In recent years many efforts have been devoted in improving the characteristics of PMMA and also other resists of similar imaging chemistry, like ZEP520, by using special additives, usually for etch resistance enhancement. These additives include fullerene derivatives, silicon and metal nanoparticles, and anthracene derivatives. Different fullerene additives, bearing solubility promotion moieties, have been used in ZEP520 by Ishi and co-workers [173–176]. On the other hand Gonsalves and co-workers use silicon oxide and metal nanoparticles in ZEP520 [122, 177, 178]. The incorporation of these nanoparticles in addition to the

etch resistance enhancement has been reported to improve energy deposition profiles, thus enhancing resolution. The activity in this area has been also recently reviewed by these authors and co-workers [122]. In a related approach PMMA has been used as host of tungsten polyoxometallates that are investigated as possible components of lithographic materials or of molecular electronic systems [179].

On the other hand intense effort is devoted to understanding and improving development characteristics of PMMA. PMMA is characterized by very well controlled dissolution behavior, without swelling, despite the fact that it is developed traditionally in organic solvent mixtures [180–182]. First, ultrasonic agitation proved very effective in nanolithography applications [183, 184]. Recently also some new developers have been tried, such as pure isopropyl alcohol and water/alcohol mixtures with very good results from the point of view of resolution [185–187]. Some of these processes developed aim at the use of PMMA structures in nanoimprint lithography.

Development studies have been reported also lately in order to improve high resolution performance of ZEP520 resist. These studies have led to a new processing scheme (SAGEX) where the resist is first cross-linked in order to improve development uniformity [188, 189].

In addition to PMMA and in general to acrylate systems that undergo chain scission by high energy radiation, other resist systems especially popular for e-beam nanolithography recently are negative systems based on small molecules. One such system represents calixarenes, which can be viewed as cyclic oligomer analogs of phenolic resists, of $\sim$1 nm dimensions. These materials were first proposed by Fujita and co-workers in 1996 [190], but currently quite a few groups work with them. Although they are difficult to dissolve, optimized processes can provide acceptable films by spin coating [190–192]. Upon exposure to e-beam radiation calixarenes give negative tone imaging [190, 193, 194, 196–206]. Organic solvents are used as developers. Different derivatives have been reported in order to improve solubility and sensitivity, which currently is higher than 0.7 mC/cm^2. Despite these problems the calixarenes are used in different nanolithography applications since they can provide sub-10 nm resolution. On the other hand Fujita and co-workers [195] have shown that low MW polystyrene (MW < 1000) can also give comparable resolution but at similar or higher doses.

Similar to calixarenes, a low MW molecular system represents the hydrogen silsesquioxanes, which also give negative tone imaging upon e-beam exposure. These systems were proposed by Namatsu and co-workers [207, 208] and since then a few other groups have also worked with these due to the high resolution capabilities they provide. They work also as negative resists, forming acceptable films by spin-coating. According to the mechanism proposed by Namatsu et al. [208], radiation breaks Si–H bonds and the system cross-links by the mediation of water molecules, which render the formed radicals to silanol groups. The starting material is provided by Dow Corning, but stability (shelf-life) problems have to be taken into account [213]. On the other hand simpler polymers like poly(dimethyl siloxane) are known to cross-link under exposure to e-beam radiation at relatively low doses. Recently sub-100 nm structures with such systems have been presented, and their pattern transfer behavior when used as double layer resists has been investigated [214, 215].

Certain modified fullerenes have been also used as molecular negative e-beam imaging materials [216–218]. Fullerenes are known also to react photochemically [219] to form larger molecules, and a similar mechanism is proposed for their e-beam reactivity as well [220]. Vapor deposition is mostly used for their deposition, whereas the doses needed are very high (on the order of 25 mC/cm^2).

Ion beam sputtering is used for the film deposition in the case of inorganic resist systems [37, 38, 221]. These are mostly different metal halides like $LiF(AlF_3)$ or silica. Almost isolated lines (60 nm periodicity) have been reported with these materials [221]. The doses in these cases are very high (few C/cm^2). Ion beam exposure can be used with these systems for improved sensitivity [222].

A very interesting work targeting the development of both negative and positive molecular electron beam resists has been presented by Shirota and co-workers in a number of communications [223–228]. The work of the group has been reviewed in a recent paper in *Proceedings of SPIE* [36]. Considerations of different chemistry routes, including chemical amplification mechanisms, have been reported. On the other hand other molecular systems like dendrimers and hyperbranched polymers have been reported as mentioned in a previous section [32–34], where chemically amplified systems have been presented as well. In all these cases the interest comes from the use of alternatives to the standard linear polymers, which form the basis in current resists. Thus systems based on well-defined molecular structures that could give benefits regarding line edge and surface roughness are envisioned. On the other hand issues like sensitivity, compatibility with current resist process requirements (e.g., aqueous base development, spin-coating, shelf-life stability, pattern transfer properties) have to be taken into account. In many cases such resist systems can find applications in optical lithography or in EUV, but currently the e-beam is mostly used to benchmark resolution capabilities.

On the other hand extensive use of chemically amplified resists that are the same or very close in composition to counterparts designed for optical lithography has taken place in electron beam lithography during the last decade. These resists offer the very significant advantage of high sensitivity, which is very important in electron beam lithography due to the inherent low throughput of the method. In most cases the resolution regime targeted is close to 100 nm, but lately a lot of effort has been devoted to the evaluation of the resolution limits of such systems. Both positive and negative systems have been evaluated. In the case of negative systems a lot of work has been done with systems based both on epoxides [229–233] and on melamine as cross-linker [234–238]. In the first case both solvent developable and aqueous base developable systems have been widely explored while in the second case the aqueous base development is standard. Commercial melamine based products are very well known such as SAL 601 (Shipley) and AZPN (Clariant). Quite a few reports on the use of these systems have been reported. Top surface imaging silylation processes have also been used that can reverse the tone of negative systems to positive [239, 240]. On the other hand standard deep UV positive resists, like UV III of Shipley, have also been evaluated

[238, 241]. In many cases the reported best resolution results are close to 50 nm. Currently a lot of effort has been devoted to the understanding of the factors that may limit the resolution of chemically amplified resists, in relation to the work also started for EUV lithography. Especially popular negative commercial resist systems in these applications lately are Sumitomo NEB 22 and NEB 31 [242–244]. Recently [244] some more information on the chemical composition of these resists was disclosed, along with studies on the influence of resist components in lithographic performance.

4.4. Resists for EUV and Other Next Generation Lithographic Technologies

With the term next generation lithography we mean nowadays mostly technologies competing to succeed optical lithography (including 157 nm) in production of integrated circuits. In this group we include extreme UV lithography (13 nm), along with projection particle technologies including electron and ion beams, low energy e-beam lithography, and X-ray lithography. Currently EUV [245] is considered as the strongest candidate and it will be discussed first.

The resist transparency at EUV is controlled by the atomic composition and density of the material and not by the molecular bonding, as was the case in the optical lithography regimes. Several theoretical and experimental investigations on the absorbance of polymeric films in the EUV region have been reported [246–250] . Since it is considered that EUV resists will be mostly used for patterning features of sub-50 nm lateral dimensions [7, 8], and taking into account that the aspect ratio targeted is ∼3, or slightly smaller, the higher resist thickness expected to be used at this wavelength is less than 150 nm. At this thickness the transparency of polymeric systems conventionally used in microlithography, such as the aromatic polymer based 248 nm resists, seems adequate. Thus, single layer resist processes are mostly investigated in recent years, although trilayer or top surface imaging schemes had first been considered. Nevertheless, new challenges have to be overcome in the design of EUV resists. These challenges actually come from the very high resolution targets imposed.

First, film thickness reduction was recognized by many researchers as an important issue that can influence the pattern quality. Resist film uniformity and adhesion studies have been reported [251–254]. Lately an increasing activity focuses also on Tg changes in thin polymeric films. These Tg changes have implications on the optimization of the thermal processing parameters but, more important, indicate strong interfacial phenomena that are expected to strongly affect critical performance aspects such as development rates and etching. Second, the molecular size of the candidate resist components has been recognized as a parameter that can influence the ultimate resist performance, especially line edge roughness. Thus, first reports of so-called molecular resists have appeared as discussed in the e-beam section. Third, dissolution differentiation mechanisms also need further consideration. For instance, doubts are raised about the suitability of catalytic mechanisms as the ones encountered in chemically amplified resists for sub-50 nm lithography. On the other hand high resist sensitivity is highly desirable; a target of 5 mJ/cm^2 has been imposed, to compensate the low fluences of most of the exposure sources available so far. Finally outgassing is also a concern, although first reports are considered rather encouraging [255, 256].

First lithographic results, both open field exposures but also high resolution results, have been reported in a few cases [257–264], showing good potential for the technology. Especially encouraging is the behavior of ESCAP-type 248 nm resists in this wavelength and the sensitivity target of the 5 mJ/cm^2 seems to show that it can be reached [263].

The requirements and challenges related to resists that can be used in other next generation lithographic technologies are similar to the ones presented for EUV. So far resists developed for optical or e-beam lithography have been mainly used in these applications [265–272]. Especially in the case of X-ray lithography there is considerable experience, since this technology has been investigated for more than two decades. Different chemically amplified resists, especially resists designed for deep UV (248 nm), have been mostly used, in addition to the conventional PMMA-type resists. A comparison of resist performance at EUV and X-ray, showing in general similar behavior despite the big difference in absorbance, has also been reported recently [260].

Similar is the situation in the case of e-beam projection lithography, where very high sensitivity is an important demand. Sensitivity issues are also mostly considered in the case of ion beam lithography. In the case of low energy beam lithography, which mostly refers to energies lower than ∼2 keV, the small penetration depth of electrons inside the resist imposes an extra demand for addressing thin film resist issues. Scanning tunneling microscope based lithography could also be considered as an extreme case of low energy e-beam lithography, where conventional e-beam resists have also been used.

5. POLYMERIC MATERIALS IN UNCONVENTIONAL LITHOGRAPHIC PROCESSES

As discussed in the first section of this chapter, in recent years a number of unconventional patterning schemes have appeared in the scientific literature aiming mostly at the fabrication of nanostructures. Polymeric materials issues involved in these schemes will be discussed here.

We will start with one very recent development related to the use of two photon processes in lithography [273–277]. These processes were well recognized as capable of providing ways of overcoming diffraction limits in optical lithography. Nevertheless, they were impractical due to the high intensities that were required in order for multiphoton processes to take place. Lately there is a lot of activity for both intensity enhancement and for improved response of the resist material. In the last case, which is more within the scope of this chapter, lately new photoacid generators have been reported [275, 276] with faster response versus multiphoton excitation compared to the standard ones. A new route has been probably opened in the resist design approaches based on these developments.

One other area of intense activity during recent years is related to self-assembling processes in thin films. Such processes encountered in polymers of certain architectures have

drawn the attention of the scientific community as possible novel routes for the formation of ordered nanostructures with applications in the fabrication of electronic and photonic devices [277–279]. Different types of simple ordered structures are being reported. Among them special interests present cases where self-assembling phenomena are mixed with radiation guided processes resulting in the formation of novel-type structures [279]. There has been also some effort for using materials, such as block copolymers, with a self-assembling tendency in standard lithographic schemes aiming at the improvement of lithographic properties, but with moderate results so far [277, 278]. Nevertheless, this direction could probably draw more interest in the near future since in a lot of cases (e.g., in resists based on fluoropolymers) self-assembling phenomena could probably be controlled for the enhancement of lithographic performance.

Resists are also used in numerous novel applications aiming at microstructure fabrication for use in specific microsystems, or MEMs. In most cases these structures have dimensions in the microdomain [280, 281] and, thus, they are not considered here where the emphasis is on nanolithography. An exception is probably the case of biomolecule microarrays. In this case thin biomolecule layers need to be patterned for use in chemical or biochemical analytical applications including DNA and protein analysis. Mechanical delivery methods like microspotting and ink-jetting or direct light guided *in-situ* synthesis [282] are the predominant methodologies used. Photoresists have been used in limited cases so far in such applications. First positive resists like PMMA were used to pattern monolayers by lift-off processes where biomolecules are physisorbed or chemisorbed [283]. In another approach, photoresists served after patterning as layers for selective physi/chemisorption of certain biomolecules [284]. A double layer scheme, based on a negative epoxy-based photoresist, has also been proposed for oligonucleotide synthesis [285]. Lately, chemically amplified positive photoresists processable under biocompatible conditions were proposed for patterning biomolecules by lift-off in successive lithographic steps [286, 287].

6. SUMMARY

Photoresist technology has been enormously improved during the last decade. The demand for fabricating structures of dimensions decreasing from sub-quarter micrometer to sub-100 nm regimes has led to the introduction of a plethora of new polymeric materials. Most of these materials are based on the chemical amplification principle introduced during the 1980s. Positive tone materials, developable in aqueous base, are almost exclusively used in the semiconductor industry and are predominant in other applications.

Although the design of photoresists suitable for exposure at decreasing optical wavelengths (248, 193, 157 nm) was the main driving force for new resist materials development, novel issues have started lately to draw considerable attention. Thus, designs of photoresists suitable for sub-50 nm have to take account of physicochemical issues like ultrathin film homogeneity, surface phenomena, microphase separation, and molecular size. Novel photoresist chemistries would also probably be needed in the near future.

Finally novel areas for photoresist applications emerge in the area of microsystems and in the broader field of nanotechnology including bionanotechnology.

GLOSSARY

Chemically amplified resists Resist materials in which the radiation causes the formation of a chemical species inside the resist film, which species further catalyses or initiates a series of chemical events that result to solubility change. Thus, in these materials the radiation action is "enhanced" by chemistry. In most chemically amplified resist systems radiation generates Brønsted acids that act further as catalysts or cationic polymerization initiators.

Dissolution inhibitor A chemical compound or a functional group, which reduces the solubility of a polymeric film. For instance, diazonaphthoquinone sensitizers act as dissolution inhibitors in novolac photoresists, whereas tert-butyl ester groups can be considered as dissolution inhibitors in certain positive chemically amplified resists. The dissolution inhibitors usually undergo radiation induced chemical changes that finally result to enhanced resist solubility as shown in Scheme 5.

Lithography Processing technology for patterning films and surfaces. When electromagnetic radiation is used to guide the patterning process, as in the case of processes shown in Schemes 1, 2, and 3, the term *photolithography* is also used. When other forms of radiation are used, or if emphasis on the specific spectral range is desired, the technology is named accordingly, (e.g., *e-beam lithography*, *ion beam lithography*, *UV lithography*, *X-ray lithography*, *193 nm lithography* etc.).

Positive/negative tone process When radiation exposure results to the removal of the exposed area the patterning process is characterized as positive, as shown in Scheme 1. On the other hand when the unexposed area is removed the patterning process is characterized as negative.

Resist Radiation sensitive material, usually having an organic polymer as the basic component, used for the patterning of films or surfaces. Typical resist-based patterning processes are shown in Schemes 1, 2 and 3 of this chapter. When the resist is sensitive to electromagnetic radiation the term *photoresist* is also used. Depending on the radiation type the terms *e-beam resist, X-ray resist, UV resist, 193 m resist* etc. are also used.

REFERENCES

1. G. M. Wallraff and W. D. Hinsberg, *Chem. Rev.* 99, 1801 (1999).
2. Y. Xia, J. A. Rogers, K. E. Paul, and G. M. Whitesides, *Chem. Rev.* 99, 1823 (1999).
3. T. Ito and S. Okazaki, *Nature* 406, 1027 (2000).
4. E. Reichmanis and O. Nalamasu, *Science* 297, 349 (2002).
5. R. F. Pease, *Nature* 417, 802 (2002).
6. G. E. Moore, *Electronics* 38, 19 April (1965).
7. K. Ronse, *Microelectron. Eng.* 67–68, 300 (2003).
8. International Technology Roadmap for Semiconductors, http://public.itrs.net.

9. H. Steppan, G. Buhr, and H. Vollmann, Angew. *Chem. Int. Ed. Engl.* 21, 455 (1982).
10. "Introduction to Microlithography" (L. F. Thompson, C. G. Willson, and M. J. Bowden, Eds.), 2nd ed. American Chemical Society, Washington, DC, 1993.
11. W. M. Moreau, "Semiconductor Lithography." Plenum, New York, 1988.
12. A. Reiser, "Photoreactive Polymers." Wiley, New York, 1989.
13. S. Nonogaki, T. Ueno, and T. Ito, "Microlithography Fundamentals in Semiconductor Devices and Fabrication Technology." Dekker, New York/Basel, 1998.
14. M. Hatzakis, *J. Electrochem. Soc.* 116, 1033 (1969).
15. S. Y. Chou, P. R. Krauss, W. Zhang, L. Guo, and L. Zhuang, *J. Vac. Sci. Technol. B* 2897 (1997).
16. Q. Lin, A. Katnani, T. Brunner, C. De Wan, C. Fairchok, D. LaTulipe, J. Simons, K. Petrillo, K. Babich, D. Seeger, M. Angelopoulos, R. Sooriyakumaran, G. Wallraff, and D. Hofer, *Proc. SPIE* 3333, 278 (1998).
17. J. Shaw, E. Babich, M. Hatzakis, and J. Paraszczak, *Solid State Technol.* 83 (1987).
18. M. Hatzakis, J. Shaw, E. Babich, and J. Paraszczak, *J. Vac. Sci. Technol. B* 6, 2224 (1988).
19. M. Ishikawa and K. Nate, *ACS Symp. Ser.* 360, 209 (1988).
20. M. A. Hartney, D. W. Hess, and D. S. Soane, *J. Vac. Sci. Technol. B* 7, 1 (1989).
21. J. M. DeSimone, G. A. York, J. E. Mc Grath, A. S. Gozdz, and M. J. Bowden, *Macromolecules* 24, 5330 (1991).
22. I. Zharov, J. Michl, M. H. Sherwood, R. Sooriyakamaran, C. E. Larson, R. A. DiPietro, G. Breyta, and G. M. Wallraff, *Chem. Mater.* 14, 656 (2002).
23. J. C. Carls, P. Argitis, and A. Heller, *J. Electrochem. Soc.* 139, 786 (1992).
24. S.-J. Kim, B.-S. Park, and H. Lee, *Proc. SPIE* 1672, 477 (1992).
25. F. Coopmans and B. Roland, *Proc. SPIE* 633, 126 (1986).
26. J. M. Shaw, M. Hatzakis, E. D. Babich, J. R. Paraszczak, D. F. Witman, and K. I. Stewart, *J. Vac. Sci. Technol. B* 7, 1709 (1989).
27. E. Gogolides, D. Tzevelekis, S. Grigoropoulos, E. Tegou, and M. Hatzakis, *J. Vac. Sci. Technol. B* 14, 3332 (1996).
28. M. Sebald, H. Ahne, R.Leuschner, and R. Sezi, *Polymers Adv. Technol.* 5, 41 (1994).
29. K. Elian, W.-D. Domke, E. Günther, and M. Irmscher, *Microelectron. Eng.* 45, 319 (1999).
30. I. Satou, H. Watanabe, M. Endo, and H. Morimoto, *Jpn. J. Appl. Phys.* 39, 442 (2000).
31. J. Horgreaves, *Microelectron. Eng.* 45, 351 (1999).
32. M. Ueda, D. Takahashi, T. Nakayama, and O. Haba, *Chem. Mater.* 10, 2230 (1998).
33. O. Haba, K. Haga, M. Ueda, O. Morikawa, and H. Konishi, *Chem. Mater.* 11, 427 (1999).
34. D. C. Tully, A. R. Trimble, and J. M. J. Fréchet, *Adv. Mater.* 12, 1118 (2000).
35. D. Pasini, E. Low, and J. M. J. Fréchet, *Adv. Mater.* 12, 347 (2000).
36. T. Kadota, M. Yoshiiwa, H. Kageyama, F. Wakaya, K. Gamo, and Y. Shirota, *Proc. SPIE* 4345, 891 (2001).
37. X. Pan, D. R. Allee, A. N. Broers, Y. S. Tang, and W. Wilkinson, *Appl. Phys. Lett.* 59, 3157 (1991).
38. M. Isaakson and A. Marray, *J. Vac. Sci. Technol. B* 19, 111 (1981).
39. T. Kudo, A. Ishikawa, H. Okamoto, K. Miyauchi, F. Murai, K. Mochiji, and U. Umezaki, *J. Electrochem. Soc.* 134, 2607 (1987).
40. C. G. Wilson, H. Ito, J. M. J. Fréchet, T. G. Tessier, and F. M. Houlihan, *J. Electrochem. Soc.* 133, 181 (1986).
41. E. Reichmanis, F. M. Houlihan, O. Nalamasu, and T. X. Neenan, *Chem. Mater.* 3, 394 (1991).
42. H. Ito, *Jpn. J. Appl. Phys.* 31, 4273 (1992).
43. S. A. MacDonald, C. Grant Willson, and J. M. J. Fréchet, *Acc. Chem. Res.* 27, 151 (1994).
44. H. Ito, *Proc. SPIE* 3678, 2 (1999).
45. Y. Ohnishi, M. Mizuko, H. Gokan, and S. Fujiwara, *J. Vac. Sci. Technol.* 19, 1141 (1981).
46. H. Gokan, S. Esho, and Y. Ohnishi, *J. Electrochem. Soc.* 130, 143 (1983).
47. L. A. Pederson, *J. Electrochem. Soc.* 129, 205 (1982).
48. Tepermeister and H. H. Sawin, *J. Vac. Sci. Technol. A* 10, 3149 (1992).
49. R. R Kunz, S. C. Palmateer, A. R. Forte, R. D. Allen, G. M. Wallraff, R. A. DiPietro, and D. C. Hofer, *Proc. SPIE* 2724, 365 (1996).
50. For an overview of these resist systems, see R. Dammel, in "Diazonaphthoquinone-Based Resist" (D. Shea, Ed.), p. 70. SPIE Press, Bellingham, WA, 1993.
51. A. Reiser, H.-Y Shih, T.-F. Yeh, and J.-P. Huang, *Angew. Chem. Int. Ed. Engl.* 35, 2428 (1995).
52. P. C. Tsiartas, L. W. Flanagin, C. L. Henderson, W. D. Hinsberg, I. C. Sanchez, R. T. Bonnecaze, and C. G. Wilson, *Macromolecules* 30, 4656 (1997).
53. C. L. Mc Adams, L. W. Flanagin, C. L. Henderson, A. R. Pawlovski, P. C. Tsiartas, and C. G. Wilson, *Proc. SPIE* 3333, 1171 (1998).
54. L. W. Flanagin, V. K. Singh, and C. G. Wilson, *J. Vac. Sci. Technol. B* 17, 1371 (1999).
55. L. W. Flanagin, V. K. Singh, and C. G. Wilson, *J. Polym. Sci. B* 37, 2103 (1999).
56. L. W. Flanagin, C. L. Adams, W. D. Hinsberg, I. C. Sanchez, and C. G. Wilson, *Macromolecules* 32, 5337 (1999).
57. M. J. Bowden and E. A. Chandross, *J. Electrochem. Soc.* 122, 1370 (1975).
58. H. Ito and C. G. Wilson, in "SPE Regional Technical Conference in Photopolymers," Ellenville, NY, 1982.
59. H. Ito, C. G. Wilson, and J. M. J. Fréchet, in "Digest of Technical Papers of 1982 Symposium on VLSI Technology," 1982, p. 86.
60. J. M. J. Fréchet, H. Ito, and C. G. Wilson, *Proc. Microcircuit Eng.* 260 (1982).
61. C. G. Wilson, H. Ito, J. M. J. Fréchet, and F. M. Houlihan, in "Proceedings of IUPAC 28th Symposium on Macromolecules," Amherst, MA, 1982, p. 448.
62. H. Ito and C.G. Wilson, in "Polymers in Electronics" (T. Davidson, Ed.), ACS Symposium Series 242, p. 11. American Chemical Society, Washington, DC, 1984.
63. D. Seeger, *Solid State Technol.* 115 (1997).
64. J. V. Crivello and J. H. W. Lam, *J. Org. Chem.* 43. 3055 (1978).
65. J. Dektar and N. P. J. Hacker, *J. Am. Chem. Soc.* 112, 6004 (1990).
66. F. M. Houlihan, T. X. Neenan, E. Reichmanis, J. M. Kometani, and T. Chin, *Chem. Mater.* 3, 462 (1991).
67. F. Ortica, J. C. Scaiano, G. Pohlers, J. F. Cameron, and A. Zampini, *Chem. Mater.* 12, 414 (2000).
68. G. Pohlers, Y. Suzuki, N. Chan, and J. F. Cameron, *Proc. SPIE* 4690, 178 (2002).
69. D. K. Lee, X. Ma, W. Lamanna, and G. Pawlowski, *Proc. SPIE* 4690, 169 (2002).
70. Y.-S. Kim, Y. Kim, S.-H. Lee, Y.-G. Yim, D.-B. Kim, and J. Kim, *Proc. SPIE* 4690, 829 (2002).
71. H. Wu and K. E. Gonsalves, *Adv. Mater.* 13, 670 (2001).
72. F. M. Houlihan, O. Nalamasu, and E. Reichmanis, *J. Fluorine Chem.*, in press.
73. M. Hiro and J. M. J. Fréchet, *ACS Symp. Ser.* 620, 381 (1996).
74. J. W. Thackeray, G. W. Orsula, E. K. Pavelchek, D. Canistro, L. E. Bogan, A. K. Berry, and K. A. Graziano, *Proc. SPIE* 1086, 34 (1989).
75. A. Sekiguchi, Y. Miyake, and M. Isono, *Jpn. J. Appl. Phys.* 39, 1392 (2000).
76. A. R. Pawloski and P. F. Nealy, *Chem. Mater.* 13 4154 (2001).
77. G. D. Fenke, R. D. Grober, G. Pohlers, K. Moore, and J. F. Cameron, *Anal. Chem.* 73, 3472 (2001).

78. E. K. Lin, C. L. Soles, D. L. Goldfarb, B. C. Trinque, S. B. Burns, R. D. Jones, J. L. Lenhart, M. Angelopoulos, C. Grant Willson, S. K. Satija, and W.-L. Wu, *Science* 297, 372 (2002).
79. H. Ito, W. P. England, R. Sooriyakumaran, N. J. Clecak, G. Breyta, W. D. Hinsberg, H. Lee, and D. Y. Yoon, *J. Photopolym. Sci. Technol.* 6, 547 (1993).
80. H. Ito, G. Breyta, D. Hofer, R. Sooriyakumaran, K. Petrillio, and D. Seeger, *J. Photopolym. Sci. Technol.* 7, 433 (1994).
81. K. Asakawa, T. Ushiroguchi, and N. Nakase, *Proc. SPIE* 2438, 563 (1995).
82. F. M. Houlihan, J. M. Kometani, A. G. Timko, R. S. Hutton, R. A. Cirelli, E. Reichmanis, O. Nalamasu, A. H. Gabor, A. N. Medina, J. J. Biafore, and S. G. Slater, *Proc. SPIE* 3333, 73 (1998).
83. F. M. Houlihan, J. M. Kometani, A. G. Timko, R. S. Hutton, R. A. Cirelli, E. Reichmanis, O. Nalamasu, A. H. Gabor, A. N. Medina, J. J. Biafore, and S. G. Slater, *J. Photopolym. Sci. Technol.* 11, 419 (1998).
84. W. D. Hinsberg, F. A. Houle, M. I. Sanchez, and G. M. Wallraff, *IBM J. Res. Dev.* 45, 667 (2001).
85. H. Ito, *IBM J. Res. Dev.* 45, 683 (2001).
86. E. Reichmanis, O. Nalamasu, and F. M. Houlihan, *Acc. Chem. Res.* 32, 659 (1999).
87. M. A. Hartney, R. R. Kunz, D. J. Ehrlich, and D. C. Shaver, *Proc. SPIE* 1262, 119 (1990).
88. M. A. Hartney, D. W. Johnson, and A. C. Spencer, *Proc. SPIE* 1466, 238 (1991).
89. M. Rothschild, R. B. Goodman, M. A. Hartney, M. W. Horn, R. R. Kunz, J. H. C. Sedlacek, and D. C. Shaver, *J. Vac. Sci. Technol. B* 10, 2989 (1992).
90. M. A. Hartney and J. W. Thackeray, *Proc. SPIE* 1672, 486 (1992).
91. M. A. Hartney, R. R. Kunz, L. M. Eriksen, and D. C. LaTlipe, *Proc. SPIE* 1925, 270 (1993).
92. M. Sasago, Y.Tani, M. Endo, and N. Nomura, *Proc. SPIE* 1264, 466 (1990).
93. Y. Ozaki, Y. Kawai, and A. Yoshikawa, *Jpn. J. Appl. Phys.* 29, 2553 (1990).
94. R. D. Allen, G. M. Wallraf, W. D. Hinsberg, W. E. Conley, and R. R. Kunz, *Solid State Technol.* 11, 53 (1993).
95. Y. Kaimoto, K. Nozaki, S. Takechi, and N. Abe, *Proc. SPIE* 1672, 66 (1992).
96. M. Endo, K. Hashimoto, K. Yanashita, A. Katsuyama, T. Matsuo, Y. Tani, M. Sasago, and N. Nomura, *IEDM Tech. Dig.* 45 (1992).
97. R. R. Kunz, R. D. Allen, W. D. Hinsberg, and G. M. Wallraff, *Proc. SPIE* 1925, 167 (1993).
98. R. D. Allen, G. M. Wallraff, R. A. Di Pietro, D. C. Hofer, and R. R. Kunz, *Proc. SPIE* 2437, 474 (1995).
99. M. Takahashi, S. Takechi, K. Nozaki, Y. Kaimoto, and N. Abe, *J. Photopolym. Sci. Technol.* 7, 31 (1994).
100. K. Nakano, K. Maeda, S. Iwasa, J. Yano, Y. Ogura, and E. Hasegawa, *Proc. SPIE* 2195, 194 (1994).
101. A. Gabor, L. Pruette, and C. Ober, *Chem. Mater.* 8, 2282 (1996).
102. K. Maeda, K. Nakano, S. Iwasa, and E. Hasegawa, *Proc. SPIE* 3049, 55 (1997).
103. K. Nozaki and E. Yano, *J. Photopolym. Sci. Technol.* 10, 545 (1997).
104. K. Nakano, K. Maeda, S. Iwasa, and E. Hasegawa, *J. Photopolym. Sci. Technol.* 10, 561 (1997).
105. M. Padmanaban, J. B. Bae, W. K. Kim, T. Kudo, M. D. Rahmen, and R. R. Dammel, *J. Photopolym. Sci. Technol.* 13, 607 (2000).
106. K. Maeda, K. Nakano, S. Iwasa, and E. Hasegawa, *Jpn. J. Appl. Phys.* 40, 7162 (2001).
107. H. Yoshino, T. Itani, M. Takimoto, and H. Tanabe, *Jpn. J. Appl. Phys.* 38, 7099 (1999).
108. F. M. Houlihan, T. Wallow, A. Timko, E. Neria, R. Hutton, R. Cirelli, J. M. Kometani, O. Nalamasu, and E. Reichmanis, *J. Photopolym. Sci. Technol.* 10, 511 (1997).
109. F. M. Houlihan, T. I. Wallow, O. Nalamasu, and E. Reichmanis, *Macromolecules* 30, 6517 (1997).
110. J. B. Kim, H. J. Yun, Y. G. Kwon, and B. W. Lee, *J. Photopolym. Sci. Technol.* 13, 629 (2000).
111. U. Okoroanyanwu, T. Shimokawa, J. Byers, and C. G. Willson, *Chem. Mater.* 10, 3319 (1998).
112. U. Okoroanyanwu, J. Byers, T. Shimokawa, and C. G. Willson, *Chem. Mater.* 10, 3328 (1998).
113. U. Okoroanyanwu, J. D. Byers, T. Cao, S. E. Webber, and C. G. Willson, *ACS Symp. Ser.* 706, 174 (1998).
114. O. Nalamasu, F. M. Houlihan, R. A. Cirelli, G. P. Watson, and E. Reichmanis, *Solid State Technol.* 5, 29 (1999).
115. J. M. Klopp, D. Pasini, J. D. Byers, C. Grant Willson, and J. M. J. Fréchet, *Chem. Mater.* 13, 4147 (2001).
116. M. Nakase, T. Naito, K. Asakawa, A. Hongu, N. Shida, and T. Ushirogouchi, *Proc. SPIE* 2438, 445 (1995).
117. T. Ushirogouchi, T. Naito, K. Asakawa, N. Shida, M. Nakase, and T. Nada, *ACS Symp. Ser.* 614, 239 (1995).
118. P. Argitis, M. A. Vasilopoulou, E. Gogolides, E. Tegou, M. Hatzakis, Z. Kollia, and A. C. Cefalas, *Microelectron. Eng.* 41/42, 355 (1998).
119. E. Gogolides, P. Argitis, E. Couladouros, V. Vidali, M. Vasilopoulou, G. Cordoyannis, C. D. Diakoumakos, and A. Tserepi, *J. Vac. Sci. Technol. B* 21 (2003).
120. B. W. Smith, D. A. Mixon, A. E. Novembre, and S. Butt, *Proc. SPIE* 2438, 504 (1995).
121. R. R. Kunz and D. K. Downs, *J. Vac. Sci. Technol. B* 17, 3330 (1999).
122. K. E. Gonsalves, L. Merhari, H. Wu, and Y. Hu, *Adv. Mater.* 13, 703 (2001).
123. H. Wu, Y. Hu, K. E. Gonsalves, and M. J. Yakaman, *J. Vac. Sci. Technol. B* 19, 851 (2001).
124. A. Yamaguchi, M. Takahashi, S. Kishimura, N. Matsuzawa, T. Ohfuji, T. Tanaka, S. Tagawa, and M. Sasago, *Jpn. J. Appl. Phys.* 38, 4033 (1999).
125. R. D. Allen, C. E. Larson, H. D. Treuong, P. J. Brock, and H. Ito, *J. Photopolym. Sci. Technol.* 13, 595 (2000).
126. K. Tanaka, H. Iwaki, Y. yamada, Y. Kiba, S. Kamei, and K. Goto, *Proc. SPIE* 4690, 557 (2002).
127. C. D. Diakoumakos, A. Douvas, I. Raptis, S. Kakabakos, D. Dimotikali, G. Terzoudi, and P. Argitis, *Microelectron. Eng.* 61-2 819 (2002).
128. J. Lee, T. Aoai, S. Kondo, N. Miyakawa, S. Takahara, and T. Yamaoka, *J. Polym. Sci. A Polym. Chem.* 40, 1858 (2002).
129. S. Isawa, K. Nakano, K. Maeda, and E. Hasegawa, *Proc. SPIE* 3333, 417 (1998).
130. S. C. Fu and K. H. Hsieh, *J. Polym. Sci. Chem.* 38, 954 (2000).
131. C. D. Diakoumakos, I. Raptis, A. Tserepi, and P. Argitis, *Polymer* 43, 1103 (2002).
132. A. K. Bates, M. Rotschild, T. M. Bloomstein, T. H. Fedynyshyn, R. R. Kunz, V. Liberman, and M. Switkes, *IBM J. Res. Dev.* 45, 605 (2001).
133. T. M. Bloomstein, M. W. Horn, M. Rothschild, R. R. Kunz, S. T. Palmacci, and R. B. Goodman, *J. Vac. Sci. Technol. B* 15, 2112 (1997).
134. T. M. Bloomstein, M. Rothschild, R. R. Kunz, D. E. Hardy, R. B. Goodman, and S. T. Palmacci, *J. Vac. Sci. Technol. B* 16, 3154 (1998).
135. M. Rothschild, T. M. Bloomstein, J. E. Curtin, D. K. Downs, T. H. Fedynyshyn, D. E. Hardy, V. Liberman, J. H. C. Sedlacek, R. S. Uttaro, A. K. Bates, and C. Van Peski, *J. Vac. Sci. Technol. B* 17, 3262 (1999).
136. R. R. Kunz, T. M. Bloomstein, D. E. Hardy, R. B. Goodman, D. K. Downs, and J. E. Curtin, *J. Vac. Sci. Technol. B* 17, 3267 (1999).
137. S. Kishimura, A. Katsuyama, M. Sasago, M. Shirai, and M. Tsunooka, *Jpn. J. Appl. Phys.* 38, 7103 (1999).

138. A. C. Cefalas, E. Sarantopoulou, P. Argitis, and E. Gogolides, *Appl. Phys. A* 69S, 929 (1999).
139. A. C. Cefalas, E. Sarantopoulou, E. Gogolides, and P. Argitis, *Microelectron. Eng.* 53, 123 (2000).
140. S. Hien, S. Angood, D. Asworth, S. Basset, T. Bloomstein, K. Dean, R. R. Kunz, D. Miller, S. Patel, and G. Rich, *Proc. SPIE* 4345, 439 (2001).
141. T. H. Fedynyshyn, R. R. Kunz, S. P. Doran, R. B. Goodman, M. L. Lind, and J. E. Curtin, *Proc. SPIE* 3999, 335 (2000).
142. T. H. Fedynyshyn, R. R. Kunz, R. F. Sinta, R. B. Goodman, and S. P. Doran, *J. Vac. Sci. Technol. B* 18, 3332 (2000).
143. T. H. Fedynyshyn, R. R. Kunz, R. F. Sinta, M. Sworin, W. A. Mowers, R. B. Goodman, and S. P. Doran, *Proc. SPIE* 4345, 2396 (2001).
144. R. R. Dammel, R. Sakamuri, A. R. Romano, R. Vicari, C. Hacker, W. Conley, and D. A. Miller, *Proc. SPIE* 4345, 350 (2001).
145. A. K. Bates, M. Rothschild, T. M. Bloomstein, T. H. Fedynyshyn, R. R. Kunz, V. Liberman, and M. Switkes, *IBM J. Res. Dev.* 45, 605 (2001).
146. Y. C. Bae, K. Douki, T. Yu, J. Dai, D. Schmaljohann, H. Koerner, and C. K. Ober, *Chem. Mater.* 14, 1306 (2002).
147. H. Ito, G. Wallraff, P. Brock, N. Fender, H. Troang, G. Breyta, D. Miller, M. Sherwood, and R. Allen, *Proc. SPIE* 4345, 273 (2001).
148. H. Ito, G. M. Wallraff, N. Fender, P. J. Brock, W. D. Hinsberg, A. Mahorowala, C. E. Larson, H. D. Troong, G. Breyta, and R. D. Allen, *J. Vac. Sci. Technol. B* 19, 2678 (2001).
149. H. V. Tran, R. J. Hung, T. Chiba, S. Yamada, T. Mrozek, Y.-T. Hsieh, C. R. Chambers, B. P. Osborn, B. C. Trinque, M. J. Pinnow, D. Sanders, E. Connor, R. Grubbs, W. Conley, S. A. McDonald, and C. Grant Willson, *J. Photopolym. Sci. Technol.* 14, 669 (2001).
150. B. C. Trinque, T. Chiba, R. J. Hung, C. R. Chambers, M. J. Pinnow, B. P. Osburn, H. V. Tran, J. Wunderlich, Y.-T. Hsieh, B. H. Thomas, G. Shafer, D. D. DeesMarteau, W. Conley, and C. Grant Willson, *J. Vac. Sci. Technol. B* 20, 531 (2002).
151. H. V. Tran, R. J. Hung, T. Chiba, S. Yamada, T. Mrozek, Y.-T. Hsieh, C. R. Chambers, B. P. Osborn, B. C. Trinque, M. J. Pinnow, S. A. McDonald, C. Grant Willson, D. P. Sanders, E. F. Connor, R. H. Grubbs, and W. Conley, *Macromolecules* 35, 6539 (2002).
152. M. Toriumi, N. Shida, T. Yamazaki, H. Watanabe, S. Ishikawa, and T. Itani, *Microelectron. Eng.* 61–62, 717 (2002).
153. T. Yamazaki and T. Itani, *Jpn. J. Appl. Phys.* 41, 4065 (2002).
154. R. Sooriyakumaran, D. Fenzer-Alexander, N. Fender, G. M. Wallraff, and R. D. Allen, *Proc. SPIE* 4345, 319 (2001).
155. V. Bellas, E. Tegou, I. Raptis, E. Gogolides, P. Argitis, H. Iatrou, N. Hatjichristidis, E. Sarantopoulou, and C. E. Cefalas, *J. Vac. Sci. Technol. B* 20, 2902 (2002).
156. I. Haller, M. Hatzakis, and R. Shrinivasan, *IBM J. Res. Dev.* 12, 251 (1968).
157. W. Chen and H. Ahmed, *Appl. Phys. Lett.* 62, 1499 (1993).
158. D. Evans and M. A. Crook, *MRS Bull.* April 1997, p. 36.
159. K. Sugita, *Progr. Organic Coatings* 31, 87 (1997).
160. T. Ichikawa, H. Ueda, and H. Koizumi, *Chem. Phys. Lett.* 13, 363 (2002).
161. D. L. Spears and H. I. Smith, *Solid State Technol.* 15, 21 (1972).
162. B. J. Lin, *J. Vac. Sci. Technol.* 12, 1317 (1975).
163. I. Haller, R. Feder, M. Hatzakis, and E. Spiller, *J. Electrochem. Soc.* 126, 154 (1979).
164. W. Moreau, D. Meritt, W. Moyer, M. Hatzakis, D. Johnson, and L. Pedersen, *J. Vac. Sci. Technol. B* 16, 1989 (1979).
165. L. F. Thompson and M. J. Bowden, *J. Electrochem. Soc.* 120, 1722 (1973).
166. M. J. Bowden and L. F. Thompson, *J. Appl. Polym. Sci.* 17, 3211 (1973).
167. M. J. Bowden and L. F. Thompson, *J. Appl. Polym., Sci., Polym. Symp.* 23, 99 (1974).
168. M. J. Bowden, L. F. Thompson, and J. P. Ballantyne, *J. Vac. Sci. Technol.* 12, 1294 (1975).
169. M. Kakuchi, S. Sugawara, K. Murase, and K. Matsuyama, *J. Electrochem. Soc.* 124, 1648 (1977).
170. T. Tada, *J. Electrochem. Soc.* 128, 1791 (1981).
171. T. Tada, *J. Electrochem. Soc.* 130, 912 (1983).
172. K. Harada, O. Kogure, and K. Murase, *IEEE Trans. Electron Dev.* ED-29, 518 (1982).
173. T. Ishi, H. Nozawa, and T. Tamamura, *Appl. Phys. Lett.* 70, 110 (1997).
174. T. Shibata, T. Ishi, H. Nozawa, and T. Tamamura, *Jpn. J. Appl. Phys.* 36, 7642 (1997).
175. T. Ishi, T. Tamamura, and K. Shigehara, *Jpn. J. Appl. Phys.* 39, L1068 (2000).
176. T. Ishi, Y. Murata, and K. Shigehara, *Jpn. J. Appl. Phys.* 40, L478 (2001).
177. Y. Hu, H. Wu, K. Gonsalves, and L. Merhari, *Microelectron. Eng.* 56, 289 (2001).
178. L. Merhari, K. E. Gonsalves, Y. Hu, W. He, W.-S. Huang, M. Angelopoulos, W. H. Bruenger, C. Dzionk, and M. Torkler, *Microelectron. Eng.* 63, 391 (2002).
179. N. Glezos, P. Argitis, D. Velessiotis, P. Koutsolelos, C. D. Diakoumakos, A. Tserepi, and K. Beltsios, in "Proceedings MRS Fall Meeting," Boston, MA, 2001.
180. J. S. Greeneich, *J. Electrochem. Soc.* 122, 970 (1975).
181. A. Uhl, J. Bendig, J. Leistner, U. Langhold, L. Bauch, and M. Böttcher, *J. Vac. Sci. Technol. B* 16, 2968 (1988).
182. I. Raptis, D. Velessiotis, M. Vasilopoulou, and P. Argitis, *Microelectron. Eng.* 53, 489 (2000).
183. S. Yasin, D. G. Hasko, and H. Ahmed, *J. Vac. Sci. Technol. B* 17, 3390 (1999).
184. S. Yasin, A. Mumtaz, D. G. Hasko, and H. Ahmed, *Microelectron. Eng.* 53, 471 (2000).
185. O. Dial, C. C. Cheng, and A. Scherer, *J. Vac. Sci. Technol. B* 3887 (1998).
186. C. Vieu, F. Carcenac, A. Pépin, Y. Chen, M. Mejias, A. Lebid, L. Manin-Ferlazzo, L. Couraud, and H. Launois, *Appl. Surf. Sci.* 164, 111 (2000).
187. E. Lavallée, J. Beauvais, and J. Breens, *J. Vac. Sci. Technol. B* 16, 1255 (1998).
188. T. Yamaguchi, H. Namatsu, M. Nagese, K. Yamazaki, and K. Kurihara, *Jpn. J. Appl. Phys.* 38, 7114 (1999).
189. T. Yamaguchi and H. Namatsu, *Jpn. J. Appl. Phys.* 41, 4217 (2002).
190. J. Fujita, Y. Ohnishi, Y. Ochiai, and S. Matsui, *Appl. Phys. Lett.* 68, 1297 (1996).
191. N. Mita, U.S. Patent, 5, 143, 784, 1992.
192. N. Wamme and Y. Ohnishi, *Proc. Am. Chem. Soc., PMSE* 67, 451 (1992).
193. Y. Ohnishi, J. Fujita, Y. Ochiai, and S. Matsui, *Microelectron. Eng.* 35, 117 (1997).
194. J. Fujita, Y. Ohnishi, S. Mamako, Y. Ochiai, E. Nomura, T. Sakamoto, and S. Mitsui, *Jpn. J. Appl. Phys.* 36, 7769 (1997).
195. S. Mamako, J. Fujita, K. Tanigaki, Y. Ochiai, and E. Nomura, *Jpn. J. Appl. Phys.* 37, 6785 (1998).
196. M. Ishida, K. Kobayashi, J. Fujita, Y. Ochiai, H. Yamamoto, and S. Touno, *Jpn. J. Appl. Phys.* 41, 4228 (2002).
197. M. Ishida, J. Fujita, T. Ogura, Y. Ochiai, E. Ohshima, and J. Monoda, *Jpn. J. Appl. Phys.* 42, 3913 (2003).
198. Y. Ochiai, S. Manako, J. Fujita, and E. Nomura, *J. Vac. Sci. Technol. B* 17, 933 (1999).
199. F. E. Prins, J. Pfeiffer, S. Raidle, D. P. Kern, and V. Shurig, *Microelectron. Eng.* 41/42, 359 (1998).
200. H. Sailer, A. Ruderisch, D. P. Kern, and V. Schurig, *J. Vac. Sci. Technol. B* 20, 2958 (2002).
201. A. Ruderisch, H. Sailer, V. Schurig, and D. P. Kern, *Microelectron. Eng.* 67–68, 292 (2003).

202. J. Kadzierski, E. Anderson, and J. Bokor, *J. Vac. Sci. Technol. B* 16, 3428 (2000).
203. P. Jedrasik, M. Hanson, and B. Nilsson, *Microelectron. Eng.* 53, 497 (2000).
204. A. Tilke, M. Vogel, F. Simmel, A. Kriele, R. H. Blick, H. Lorenz, D. A. Wharam, and J. P. Kotthaus, *J. Vac. Sci. Technol. B* 17, 1594 (1999).
205. Y. Miyamoto, A. Kokubo, T. Hattori, and H. Hongo, *J. Vac. Sci. Technol. B* 16, 3894 (1998).
206. S. Yasin, D. G. Hasko, and F. Carecenac, *J. Vac. Sci. Technol. B* 19, 311 (2001).
207. H. Namatsu, T. Yamaguchi, M. Nagase, K. Yamasaki, and K. Kurihara, *Microelectron Eng.* 41–42, 331 (1998).
208. H, Namatsu, Y. Takahashi, K. Yamazaki, T. Yamaguchi, M. Nagase, and K. Kurihara, *J. Vac. Sci. Technol. B* 16, 69 (1998).
209. F. C. M. J. M. van Delft, J. P. Weterings, A. K. van Langen-Suurling, and J. Romijn, *J. Vac. Sci. Technol. B* 18, 3419 (2000).
210. B. E. Maile, W. Henschel, H. Kurz, B. Rienks, R. Polman, and P. Kaars, *Jpn. J. Appl. Phys.* 39, 6836 (2000).
211. L. Mollard, G. Gunge, S. Tedesco, B. Dal'zotto, and J. Foucher, *Microelectron. Eng.* 61–62, 755 (2002).
212. H. Namatsu, *J. Vac. Sci. Technol. B* 19, 2709 (2001).
213. F. C. M. J. M. van Delft, *J. Vac. Sci. Technol. B* 20, 2932 (2002).
214. A. Tserepi, E. S. Valamontes, E. Tegou, I. Raptis, and E. Gogolides, *Microelectron. Eng.* 57–58, 547 (2001).
215. A. Tserepi, G. Cordoyiannis, G. P. Patsis, V. Constantoudis, E. Gogolides, E. S. Valamontes, D. Eon, M. C. Peignon, Ch. Cardinaud, and G. Turban, *J. Vac. Sci. Technol. B*, in press.
216. T. Tada and T. Kanayama, *Jpn. J. Appl. Phys.* 35, L63 (1996).
217. A. P. G. Robinson, R. E. Palmer, T. Tada, T. Kanayama, and J. A. Preece, *Appl. Phys. Lett.* 72, 1302 (1998).
218. A. Matsutani, F. Koyama, and K. Iga, *Jpn. J. Appl. Phys.* 337, 4211 (1998).
219. A. M. Rao, P. Zhou, K.-A. Wang, G. T. Hager, J. M. Holden, Y. Wang, W.-T. Lee, X.-Y. Bi, P. C. Eklund, D. S. Cornett, M. A. Duncan, and I. J. Amster, *Science* 259, 955 (1993).
220. T. Hara, J. One, H. Tanaka, Y. Li, and K. Takeuchi, *Jpn. J. Appl. Phys.* 40, L478 (2001).
221. J. Fujita, H. Watanabe, Y. Ochiai, S. Manako, J. S. Tsai, and S. Matsui, *Appl. Phys. Lett.* 66, 3065 (1995).
222. J. Melngailis, A. A. Mondelli, I. L. Berry, and R. Mohondro, *J. Vac. Sci. Technol. B* 16, 927 (1998).
223. Y. Shirota, *J. Mater. Chem.* 10, 1 (2000).
224. M. Yoshiwa, H. Kageyama, F. Wakaya, K. Gamo, and Y. Shirota, *J. Photopolym. Sci. Technol.* 9, 57 (1996).
225. M. Yoshiiwa, H. Kageyama, Y. Shirota, F. Wakaya, K. Gamo, and M. Takai, *Appl. Phys. Lett.* 69, 2605 (1996).
226. T. Kadota, H. Kageyama, F. Wakaya, K. Gamo, and Y. Shirota, *J. Photopolym. Sci. Technol.* 11, 147 (1998).
227. T. Kadota, H. Kageyama, F. Wakaya, K. Gamo, and Y. Shirota, *J. Photopolym. Sci. Technol.* 12, 375 (1999).
228. T. Kadota, H. Kageyama, F. Wakaya, K. Gamo, and Y. Shirota, *J. Photopolym. Sci. Technol.* 13, 203 (2000).
229. K. J. Stewart, M. Hatzakis, J. M. Shaw, and D. E. Seeger, *J. Vac. Sci. Technol. B* 7, 1734 (1989).
230. P. Argitis, I. Raptis, C. J. Aidinis, N. Glezos, M. Bacciochi, J. Everett, and M. Hatzakis, *J. Vac. Sci. Technol. B* 13, 3030 (1995).
231. P. Argitis, N. Glezos, M. Vasilopoulou, I. Raptis, M. Hatzakis, J. Everett, G. Meneghini, A. Palumbo, M. Ardito, P. Hudek, and I. Kostic, *Microelectron. Eng.* 53, 453 (2000).
232. N. Glezos, P. Argitis, D. Velessiotis, I. Raptis, M. Hatzakis, P. Hudek, and I. Kostic, *J. Vac. Sci. Technol.* 18, 3431 (2000).
233. G. P. Patsis, A. Tserepi, I. Raptis, N. Glezos, E. Gogolides, and E. S. Valamontes, *J. Vac. Sci. Technol. B* 18, 3292 (2000).
234. E. A. Dobisz, C. R. K. Marrian, R. E. Salvino, M. A. Ancona, F. K. Perkins, and N. H. Turner, *J. Vac. Sci. Technol. B* 11, 2733 (1993).
235. E. A. Dobisz and C. R. K. Marrian, *J. Vac. Sci. Technol. B* 9, 3024 (1991).
236. E. A. Dobisz, T. N. Fedynyshyn, D. Ma, L. M. Shirley, and R. Baas, *J. Vac. Sci. Technol. B* 16, 3773 (1998).
237. I. Raptis, M. Gentilli, E. Difabrizio, R. Maggiora, M. Baciocchi, and L. Mastrogiacomo, *Microelectron. Eng.* 27, 417 (1995).
238. Z. Cui and P. Prewett, *Microelectron. Eng.* 46, 255 (1999).
239. E. Tegou, E. Gogolides, P. Argitis, and M. Hatzakis, *Microelectron. Eng.* 41/42, 335 (1998).
240. E. Tegou, E. Gogolides, P. Argitis, I. Raptis, M. Hatzakis, G. Meneghini, and Z. Cui, *Jpn. J. Appl. Phys.* 37, 77 (1998).
241. L. Grella, M. Gentilli, E. Di Fabrizio, M. Baciocchi, L. Mastrogiacomo, and R. Maggiora, *J. Vac. Sci. Technol. B* 15, 2596 (1997).
242. S. Tedesco, T. Mourier, B. dal'Zotto, A. Mc Dougal, S. Blanc-Coquant, Y. Quéré, P. J. Paniez, and B. Mortini, *J. Vac. Sci. Technol. B* 16, 3676 (1998).
243. L. Pain, C. Higgins, B. Scarfoglière, S. Tedesco, B. dal'Zotto, C. Courgon, M. Ribeiro, T. Katsumoto, M. Suetsugu, and R. Hanawa, *J. Vac. Sci. Technol. B* 18, 3388 (2000).
244. J. Saint-Pol, S. Landis, C. Courgon, S. Tedesco, R. Hanawa, M. Suetsuga, M. Akita, and S. Yamamoto, *Microelectron. Eng.* 67–68, 274 (2003).
245. R. H. Stulen and D. W. Sweeney, *IEEE J. Quantum Electron.* 35, 694 (1999).
246. N. Matswuzawa, H. Oizumi, S. Mori, S. Irie, E. Yano, S. Okazaki, and A. Ishitani, *Microelectron. Eng.* 53, 671 (2000).
247. N. N. Matsuzawa, H. Oizumi, S. Mori, S. Irie, E. Yano, S. Okazaki, and A. Ishitani, *Microelectron. Eng.* 53, 671 (2000).
248. N. N. Matsuzawa, H. Oizumi, S. Mori, S. Irie, S. Shirayone, E. Yano, S. Okazaki, A. Ishitani, and D. A. Dixon, *Jpn. J. Appl. Phys.* 38, 7109 (1999).
249. S. Irie, M. Endo, M. Sasago, N. Kandaka, H. Kando, and K. Murakami, *Jpn. J. Appl. Phys.* 41, 4027 (2002).
250. S. Irie, M. Endo, M. Sasago, N. Kandaka, H. Kando, and K. Murakami, *Jpn. J. Appl. Phys.* 41, 5864 (2002).
251. R. L. Jones, C. L. Soles, F. W. Starr, E. K. Lin, J. L. Lenhart, W.-L. Wu, D. L. Goldfarb, and M. Angelopoulos, *Proc. SPIE* 4690, 342 (2002).
252. X. Chen, L. M. Tolbert, D. W. Hess, and C. Henderson, *Macromolecules* 34, 4104 (2001).
253. C. D. Diakoumakos and I. Raptis, *Polymer* 44, 251 (2003).
254. I. Raptis, D. Niakoula, et al., *Microelectron Eng.,* in press.
255. M. C. Chauhan and P. F. Nealy, *J. Vac. Sci. Technol. B* 18, 3402 (2000).
256. P. M. Dentiger, *J. Vac. Sci. Technol. B* 18, 3364 (2000).
257. B. M. Martens, B. van Zwan, P. W. H. de Jager, M. Leenders, H. G. C. Werij, J. P. H. Benschop, and A. J. J. van Dijsseldonk, *Microelectron Eng.* 53, 659 (2000).
258. S. A. Bulgakova, A. Ya. Lopatin, V. I. Lunchin, L. M. Maranova, S. A. Molodnjakov, and N. N. Salaghchenko, *Nucl. Instr. Methods A* 448, 487 (2000).
259. F. Cerrina, S. Bollepalli, M. Khan, H. Solak, W. Li, and D. He, *Microelectron Eng.* 53, 13 (2000).
260. D. He, H. Solak, W. Li, and F. Cerrina, *J. Vac. Sci. Technol. B* 17, 3379 (1999).
261. H. H. Solak, D. He, W, Li, and F. Cerrina, *J. Vac. Sci. Technol. B* 17, 3052 (1999).
262. R. L. Brainard, C. Henderson, J. Cobb, V. Rao, J. F. Mackevich, U. Okoroanyanwu, S. Gunn, J. Chambers, and S. Connoly, *J. Vac. Sci. Technol. B* 17, 3384 (1999).
263. P. M. Dentiger, L. L. Hunter, D. J. O'Connel, S. Cunn, D. Goods, T. H. Fedynyshyn, R. B. Goodmann, and D. K. Astolfi, *J. Vac. Sci. Technol. B* 2962 (2002).
264. M. Azam Ali, K. E. Gonsalves, V. Golovkina, and F. Cerrina, *Microelectron Eng.* 65, 454 (2003).
265. H. I. Smith, D. L. Spears, and S. E. Bernacki, *J. Vac. Sci. Technol. B* 10, 913 (1973).

266. P. V. Lenzo, and E. G. Spencer, *Appl. Phys. Lett.* 24, 289 (1974).
267. L. F. Thomson, E. D. Feit, M. J. Bowden, P. V. Lenzo, and E. G. Spencer, *J. Electrochem. Soc.* 121, 1500 (1974).
268. N. Ueno, T. Mitsuhata, K. Sugita, and K. Tanaka, "Polymers in Microlithography," ACS Symp. Ser., Ch. 26, p. 425. *Am. Chem. Soc.*, Washington, DC, 1989.
269. M. C. K. Tinone, K. Tanaka, and N. Ueno, *J. Vac. Sci. Technol. A* 13, 1885 (1995).
270. M. C. K. Tinone, K. Tanaka, J. Maruyama, N. Ueno, M. Imamura, and N. Matsubayashi, *J. Chem. Phys.* 100, 5988 (1994).
271. N. Rau, F. Stratton, C. Fields, T. Ogawa, A. Neureuther, R. Kubena, and G. Willson, *J. Vac. Sci. Technol. B* 16, 3784 (1998).
272. M. Peuker, M. H. Lim, H. I. Smith, R. Morton, A. K. van Langen-Suurling, J. Romijn, E. W. J. M. van der Drift, and F. C. M. J. M. van Delft, *Microelectron. Eng.* 61–62, 803 (2002).
273. X. Yin, N. Fang, X. Zhang, I. B. Martini, and B. J. Schwartz, *Appl. Phys. Lett.* 81, 3663 (2002).
274. K. D. Beldfield et al., *J. Phys. Org. Chem.* 13, 837 (2000).
275. B. H. Cumpston et al., *Nature* 398, 51 (1999).
276. W. Zhou, S. M. Kuelber, K. L. Braun, T. Yu, J. Kevin Cammak, C. K. Ober, J. W. Perry, and S. R. Mardev, *Science* 296, 1106 (2002).
277. T. Yu, C. K. Ober, S. M. Kuelber, W. Zhou, S. R. Marder, and J. W. Perry, *Adv. Mater.* 15, 517 (2003).
278. A. H. Gabor, and C. K. Ober, in "Microelectronics Technology," ACS Symp. Ser. 614, p. 281. American Chemical Society, Washington, DC, 1995.
279. J. Y. Cheng, C. A. Ross, E. L. Thomas, H. I. Smith, and G. J. Vansco, *Appl. Phys. Lett.* 81, 3657 (2002).
280. K. Lee, N. LaBianca, S. Rishton, and S. Zohlgharnain, *J. Vac. Sci. Technol. B* 13, 3012 (1995).
281. I. Raptis, M. Chatzichristidi, C. D. Diakoumakos, A. Douvas, D. Niakoula, and P. Argitis, *J. Photopolym. Sci. Technol.* 14, 445 (2001).
282. S. P. A. Fodor, R. Leighton, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas, *Science* 251, 767 (1991).
283. D. Kleinfeld, K. H. Kahler, and P. E. Hockberger, *J. Neurosci.* 8, 4098 (1988).
284. D. V. Nicolau, T. Taguchi, H. Tanigawa, and S. Yoshikawa, *Langmuir* 15, 3845 (1999).
285. G. McGall, J. Labadie, P. Brock, G. Wallraf, T. Nguyen, and W. Hinsberg, *Proc. Natl. Acad. Sci. USA* 93, 13555 (1996).
286. A. Douvas, P. Argitis, C. D. Diakoumakos, K. Misiakos, D. Dimotikali, and S. E. Kakabacos, *J. Vac. Sci. Technol. B.* 19, 2820 (2001).
287. A. Douvas, P. Argitis, K. Misiakos, D. Dimotikali, P. S. Petrou, and S. Kakabacos, *Biosensors Biolectron.* 17, 269 (2002).

Resonant Tunneling Devices

A. F. M. Anwar
University of Connecticut, Storrs, Connecticut, USA

Mirza M. Jahan
Intel Corporation, Chandler, Arizona, USA

CONTENTS

1. Introduction
2. Resonant Tunneling Mechanism
3. Methods of Analysis
4. Carrier Dynamics (Tunneling Time)
5. DBRTS Capacitance
6. Shot Noise
7. Conclusion
Glossary
References

1. INTRODUCTION

Resonant tunneling structures (RTSs) or double barrier RTS (DBRTSs) have attracted a lot of attention because of their possible application in microwave and millimeter wave technologies as well as their application in multivalued logic systems. A resonant tunneling structure in it simplest form consisting of the low bandgap material, GaAs, sandwiched between two wide bandgap semiconductors, AlGaAs, was first investigated by Esaki and Tsu in the early 1970s [1]. Interest has been building up since Sollner et al. [2] reported in 1983 a negative differential resistance (NDR) with a large peak to valley ratio in current. This breakthrough was possible due to the improvements in epitaxial growth techniques.

The fast switching time and NDR in RTS can be utilized in many ways in realizing novel device structures. RTS has been incorporated in the emitter and base of bipolar transistors to obtain unique performance [3–11]. Liu et al. reports a multiple negative-differential-resistance InGaP/GaAs RT bipolar transistor where double and quaternary NDR phenomena are observed at room temperature [12]. RTS has been incorporated in FET structures to obtain high-speed low power electronics [13–15]. RT bipolar/unipolar transistors have been successfully demonstrated for applications in high-density memories using multivalued logic [16–20] and high speed analog-to-digital (A/D) converters with potential applications that include ultrahigh speed A/D converters [21]. RTS has been widely implemented in different combinations in FET and bipolar transistors to realize different digital logic and adder circuits [22–27]. Gonzalez et al. [28] reports the implementation of a multivalued logic signed digital full adder using a standard 0.6 μm CMOS process. The circuit is based on a two peak NDR device obtained by combining a MOS and an NDR device. Recently, Lind et al. [29] reported the fabrication of a resonant tunneling permeable base transistor where a room temperature peak-to-valley ratio (PVR) in current of 8 was achieved. HEMTs have also been successfully incorporated with RTS to realize the different analog and digital applications [30–33]. RTSs in different forms have been utilized in analog and mixed signals circuitry to obtain high frequency oscillators, frequency multipliers, and current mirrors [34–38]. The high switching speed of RTD has enabled its use as a local oscillator in timing extraction in optical communication systems [39]. RTS has also been used in optical modulators and millimeter and submillimeter wave applications [40, 41].

The fabrication of RTD and related devices is not only confined to GaAs based material systems. Kikuchi et al. [42] reports the fabrication of AlN/GaN/AlN resonant tunneling structure where the room temperature PVR of 3 was measured. See and Paul [43] have reported a PVR of 2.43 at a peak current density of 282 kA/cm^2 using a 5 $\times$ 5 μm^2 Si/SiGe sample. Structures using ZnO/Zn$_{0.8}$Mg$_{0.2}$O, CdF$_2$/CaF$_2$, InAlAsSb/InGaAs, amorphous SiO$_2$/Si/SiO$_2$, amorphous SiC/Si, and InGaAs/AlAs/InGaAsP double barrier structures have also been reported [44–50].

Transport in RTS has been widely investigated as evident from the extensive literature. Nonequilibrium quantum transport, three-dimensional transport, and quantum transport using hydrodynamic formulation has been reported [51–55]. The Wigner function has been widely used to model open systems like resonant tunneling structures [56–60]. Jensen and Ganguly [57] report the computation of current–voltage characteristics and incorporation of scattering using the Wigner function and compare that to the standard

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (357–370)

technique using the transmission coefficient approach. The incorporation of scattering in the Wigner function formulation has been reported by Mains and Haddad [58] and Jensen and Bout [59]. The transient response of resonant tunneling structures has been reported by Frensley [60] and Pinaud [61].

Shot microwave, and flicker noise has been widely investigated and reported [61–71]. Davies et al. [65] report the computation of shot noise using a classical description based upon the rate equation for sequential tunneling. Suppression of shot noise in asymmetric DBRTS has been reported by Liu et al. [66]. Enhancement of shot noise when the DBRTS is biased in the negative differential resistance region is reported by Iannaconne et al. [67]. Sun and Milburn [68] report the computation of current noise in DBRTS and the treatment includes both coherent and incoherent quantum tunneling processes. Blanter and Buttiker [69] report the transition from sub-Poissonian to super-Poissonian shot noise in DBRTS that takes place near the instability threshold. Ng and Surya [70] and Ouacha and Willander [71] report the theory and measurement of flicker noise in DBRTS. Resonant tunneling in quantum cascade lasers has been reported by Sirtori et al. [72] and the physics of resonant tunneling in quantum dots has been reported by Luscombe et al. [73].

2. RESONANT TUNNELING MECHANISM

Resonant tunneling structures have an interesting property of transmitting electrons through a quantum well (QW) with high transmission probability if the electrons are incident at energies equal to one of the virtual energy levels of the QW. Electrons with energies other than the virtual energy levels have a very low probability of transmission or in other words they are not allowed to pass. Figure 1, shows the conduction band diagram of a typical DBRTS. The barriers are composed of a higher bandgap semiconductor (e.g., AlGaAs), and the QW is formed in the lower bandgap semiconductor, namely GaAs. The discussion is restricted to heterostructures with type-I transitions (i.e., ΔE_C and ΔE_V, both being positive).

The operation of a RTS has been explained in the light of a resonant Fabry–Perot effect [74]. In the absence of scattering mechanisms, which can destroy the phase coherence of an electron wave, the amplitude of the resonant modes builds up in the QW. Concomitantly, the electron wave leaks out in both directions and in doing so cancels the reflected waves and enhances the transmitted waves. On the other hand, if the time evolution of the electron wave function in the well is long compared to the scattering time, the collisions randomize the phase of the wave function. In this case, the electrons tunnel through both the barriers sequentially and the RTS has a lower quantum mechanical transmission peak.

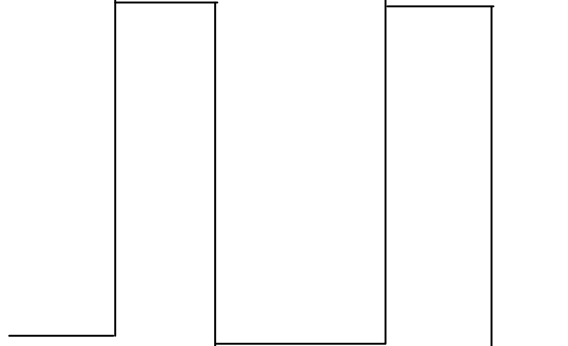


Figure 1. Conduction band profile of a typical DBRTS.

3. METHODS OF ANALYSIS

Matrix methods using the continuity of the wave function and its first derivative, at any heterointerface, have been used quite extensively to determine the incident energy dependence of transmission probability. Ando and Itoh [75] approximated the arbitrary potential well by multistep functions and then used a matrix method to determine the transmission coefficient. The position dependence of electron effective mass, m_n^*, and the permittivity were also approximated by multistep functions. Using the envelope wave function under effective mass approximation the amplitudes of the incident and reflected wave functions at any segment n, $\varphi_n(x) = A_n e^{jk_n x} + B_n e^{-jk_n x}$ with $k_n = \sqrt{2m_n^*(E - U_n)/\hbar^2}$ where E is the incident electron energy and U_n is the potential, were related to that at the reference segment by the following matrix equation:

$$\begin{pmatrix} A_j \\ B_j \end{pmatrix} = \prod_{l=0}^{j-1} M_l \begin{pmatrix} A_0 \\ B_0 \end{pmatrix}$$

The matrix M_l is generated by invoking the continuity of the wave function $\varphi(x)$ and its first derivative by properly accounting for the effective mass,

$$\frac{1}{m_n^*} \frac{d\varphi_n}{dx}$$

at the interface n. The transmission probability at any energy $D(E)$ is given as

$$D(E) = \frac{m_0^*}{m_{N+1}^*} \frac{k_0}{k_{N+1}} |A_{N+1}|^2$$

where $n = N + 1$ and $A_0 = 1$ are assumed.

Though the matrix method is straightforward it becomes slow with the increase in heterointerfaces and prone to round-off errors. Lui and Fukuma [76] approximated the arbitrary potential well by piecewise linear functions and then used a numerical method to calculate the transmission coefficient. In the treatment presented here we will use the method of logarithmic derivates to determine the transmission probability and other parameters required to investigate the system

3.1. Logarithmic Derivatives

Using logarithmic derivatives [77], defined as the ratio of the derivative of the wave function $\Psi(x)$ and the wave function $\Psi'(x)$

$$\Xi(x, E_x) = \frac{2\hbar}{m^*} \frac{\Psi'(x)}{\Psi(x)}$$

the one electron Schrödinger equation within the effective mass approximation can be written as [78]

$$\frac{d\Xi(x, E_x)}{dx} = -j\left[\frac{m^*}{2\hbar}\Xi(x, E_x)^2 + \frac{4}{\hbar}(V_{\text{TOTAL}} - E_x)\right] \quad (1)$$

where m^* is the electronic effective mass, $\hbar$ is the modified Planck's constant, and $V_{\text{TOTAL}} = \Delta E_C$ is the conduction band offset. In general, $V_{\text{TOTAL}} = \Delta E_C + V_a(x) + V_{\text{sc}}(x)$, where V_{sc} is due to the accumulation of the space charge inside the quantum well and V_a is due to the applied bias voltage and needs to be evaluated to calculate the current–voltage characteristics. Equation (1) can be solved numerically and for a piecewise constant potential profile, the recurrence relation may be written as

$$\Xi_n(x, E_x) = \sqrt{\frac{2m^*A}{8}}\tanh\left[-j\sqrt{\frac{2m^*A}{8}}\Delta x + \arctan h\left(\Xi_{n+1}\sqrt{\frac{m^*}{8A}}\right)\right] \quad (2)$$

where $A = -[\Delta E_C - E_x]$, $\Delta x = x_{n+1} - x_n$.

The transmission coefficient is calculated as

$$D(E_x) = \sqrt{1 - \rho(E_x)^2} \quad (3)$$

where

$$\rho(E_x) = \frac{\Xi_n(x, E_x) - \Xi_0}{\Xi_n(x, E_x) + \Xi_0} \quad (4)$$

is the reflection coefficient and

$$\Xi_0 = \sqrt{\frac{2m^*}{\hbar^2}(E_x - \Delta E_C)} \quad (5)$$

In Figure 2, the transmission probability of a simple DBRTS is plotted as a function of the incident electron energy. The structure as shown in the inset is composed of $Al_{0.3}Ga_{0.7}As$ as barriers and GaAs as QW. The QW is 30 Å wide with 30 Å barriers. The conduction band offset is assumed to be 0.25 eV and the effective masses in the barrier and QW are $0.09m_0$ and $0.067m_0$, respectively. Here m_0 is the free electron mass. The peaks in the transmission probabilities correspond to the eigenstates and the widths are related to the lifetime of the carriers of the QW. The dips in the transmission probability for energies greater than the band offset are due to the reflections from the edges of the QW.

The barrier height dependence of transmission probability is investigated in Figure 3. Conduction band offsets of 0.1, 0.2, and 0.3 eV are considered. The QWs with band offsets of 0.1 and 0.2 eV have two eigenstates which move toward higher energy with increasing ΔE_C. The QW with $\Delta E_C = 0.3$ eV has an additional eigenstate close to 0.3 eV. Figure 4 shows the nature of the transmission probability curves with varying QW width. QW widths of 30, 50, and 100 Å with $\Delta E_C = 0.3$ eV are considered. With increasing QW widths more eigenstates are accommodated within the QW. Figure 5 shows the nature of the transmission probability curves in the presence of scattering. The simulated results are with localized elastic and inelastic (phase breaking) scattering at the center of the QW. With elastic scattering at the center of the QW the odd-resonant energies shift toward higher energies. On the other hand, with phase

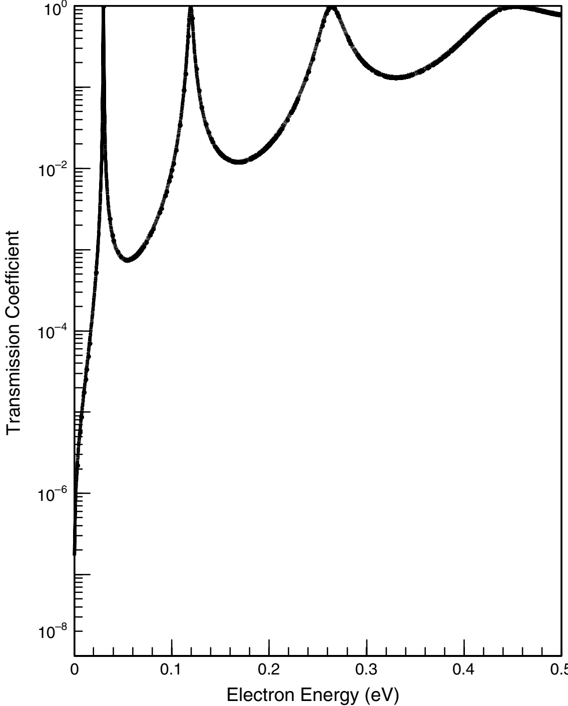


Figure 2. Transmission coefficient of DBRTS.

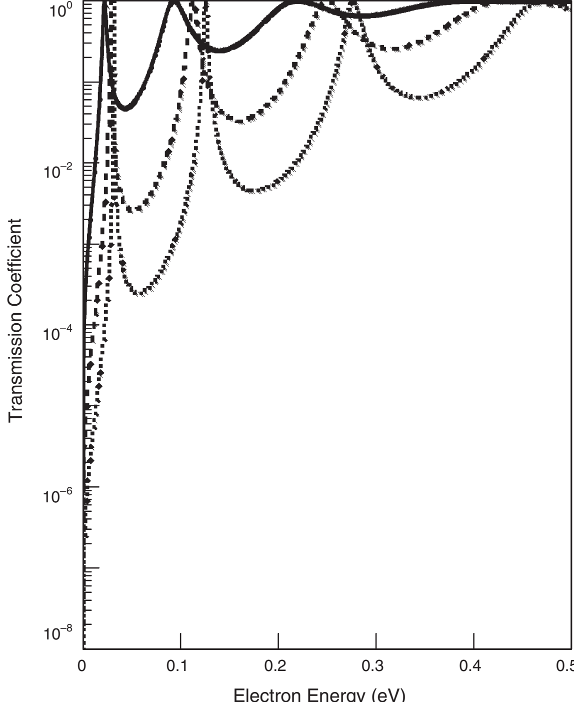


Figure 3. Transmission coefficient of DBRTS with conduction band offset as a parameter. Solid, dashed, and dotted lines represent conduction band offsets of 0.1, 0.2, and 0.3 eV, respectively.

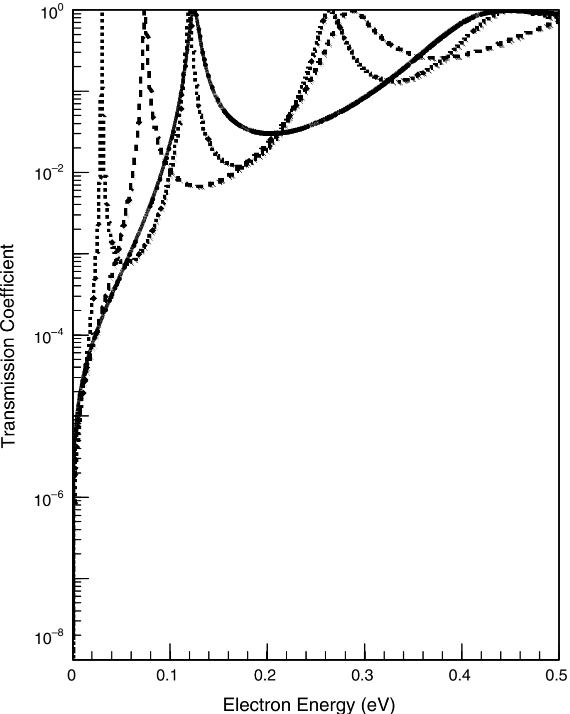


Figure 4. Transmission coefficient of DBRTS with varying QW widths. Solid, dashed, and dotted lines represent 30, 50, and 100 Å, respectively.

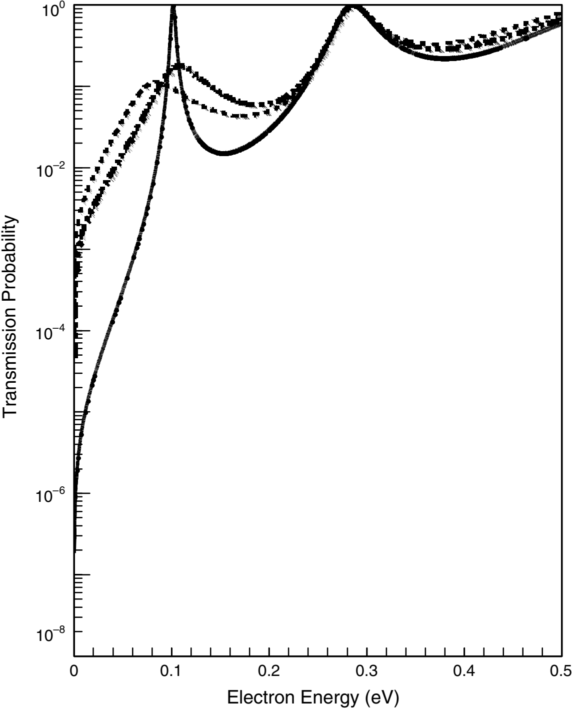


Figure 5. Transmission coefficient of DBRTS in the presence of scattering. Solid and dashed lines represent transmission with elastic and phase breaking scattering at the center of the QW. Dotted line represents the effect of both elastic and phase breaking scattering.

breaking scattering at the center of the QW the transmission coefficient decreases.

3.2. Carrier Group Velocity and Density of States

Current density in its simplest form, $J = qnv$, is expressed as a product of the carrier concentration, n, and the corresponding drift velocity, v. The determination of the carrier concentration requires the density of states (DOS) and the carrier statistics. The formulation of the local DOS as well as the drift velocity will be presented in terms of the logarithmic derivatives.

3.2.1. Group Velocity

The definition of logarithmic derivative can be recasted in the form [79]

$$\begin{aligned}\frac{\Psi'(x)}{\Psi(x)} &= \alpha(x) + j\beta(x)\\ &= \frac{m^*}{2\hbar}[-\mathrm{Im}(\Xi(x, E_x)) + j\,\mathrm{Re}\Xi(x, E_x))]\end{aligned}$$

which upon integration from $x = 0$ to any x

$$\Psi(x) = \Psi(0)e^{\int_0^x \alpha(x)\,dx}e^{j\int_0^x \beta(x)\,dx} \tag{6}$$

where $\Psi(x)$ is the wave function at point x and $\Psi(0)$ is the wave function of the incident particle at $x = 0$. Depending upon the energy of the incident particle the composite wave, Eq. (6), in the well and in the barriers may build up, giving rise to resonant tunneling, or simply tunnel through the device structures as in ordinary tunneling. At resonance, the logarithmic derivative of an electron outside a symmetrical structure is purely real and has a constant value. The wave function described by (6), in this case, reduces to positive traveling plane waves outside the barrier.

The average probability current density at any x, of the quantum mechanical system

$$\begin{aligned}S(x) &= \frac{1}{2}\mathrm{Re}\left(\frac{2\hbar}{jm^*(x)}\Psi'\Psi\right)\\ &= \frac{1}{2}\,|\Psi(0)|^2\,\mathrm{Re}(\Xi(x, E_x))e^{2\int_0^x \alpha(x)\,dx}\end{aligned} \tag{7}$$

when equated to the particle current as obtained from conservation principle

$$S(x) = v_g(x)\,|\Psi(x)|^2 \tag{8}$$

yields the group velocity which is expressed as

$$v_g(x) = \frac{1}{2}\mathrm{Re}(\Xi(x, E_x)) \tag{9}$$

The group velocities of the electron at the first and second resonant energies are plotted in Figure 6. A 100 Å QW with conduction band offset of 0.25 eV is considered. It should be noticed that the velocity peaks at the same location where the electron wave function is at its minimum.

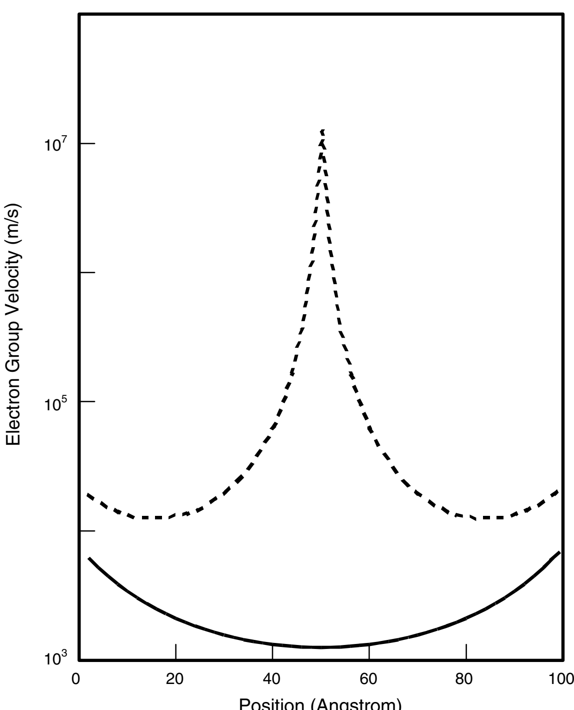


Figure 6. Electron group velocity for the first (solid line) and second (dashed line) eigenstates.

3.2.2. Density of States

For a given DBRTS, the one-dimensional density of states is given as [80, 81]

$$N(x, E_x) = \frac{2}{\pi} \mathrm{Im}[G^R(x, x, E_x)] \tag{10}$$

where $G^R(x, x, E_x)$ is the retarded Green's function and is obtained as the solution of the following Schrödinger equation:

$$\left[E_x - H_0 + \frac{j\hbar}{2\tau_\varphi(x, E_x)}\right] G^R(x, x, E_x) = \delta(x - x_1) \tag{11}$$

The Hamiltonian

$$H_0 = -\frac{\hbar^2}{2m^*}\left(\frac{d^2}{dx^2} + \Delta E_C(x)\right)$$

is the unperturbed part of the total Hamiltonian and τ_ϕ is the phase breaking time. The calculation of DOS necessitates the computation of only the diagonal elements of the retarded Green's function and the off-diagonal elements are less significant. The diagonal elements of the Green's function can be computed by using the logarithmic derivatives $\Xi(x, E_x)$and are expressed as

$$N(x, E_x) = \mathrm{Im} \frac{-\frac{j4}{\hbar}}{\Xi^+(x, E_x) - \Xi^-(x, E_x)} \tag{12}$$

where

$$\Xi(x, x_1, E_x) = \frac{2\hbar}{jm^*} \frac{\frac{\partial G^R(x, x_1, E_x)}{\partial x}}{G^R(x, x_1, E_x)} \tag{13}$$

The global DOS $N(E_x)$ is obtained by integrating Eq. (12) along the QW,

$$N(E_x) = \int_0^{L_W} N(x, E_x)dx \tag{14}$$

where L_W is the length of the QW.

3.3. Self-Consistent *I–V* Calculation

The determination of the current–voltage characteristic that takes into account the potential due to the space charge developed within the QW demands that the Schrödinger and Poisson equations are solved self-consistently. Kluksdahl et al. [82] report self-consistently calculated current–voltage characteristics using the Wigner function to model quantum transport. In this section we use the method of logarithmic derivatives to self-consistently calculate the current–voltage characteristics.

The total current flowing from the emitter to the collector is given in terms of the currents from the emitter to the collector J^+ and from the collector to the emitter J^-, respectively [80, 83],

$$\begin{aligned} J^+ dE_x &= J_{E\to C}(E_x)dE_x \\ &= \frac{qm^*kTD^+(E_x)}{2\pi^2\hbar^3} \frac{e^{qV_{\text{Applied}}/kT}}{e^{qV_{\text{Applied}}/kT} - 1} \\ &\quad \cdot \ln\left[\frac{1 - e^{(E_F - E_x)/kT}}{1 + e^{(E_F - E_x - qV_{\text{Applied}})/kT}}\right] dE_x \end{aligned} \tag{15a}$$

$$\begin{aligned} J^- dE_x &= J_{C\to E}(E_x)dE_x \\ &= \frac{qm^*kTD^-(E_x)}{2\pi^2\hbar^3} \frac{1}{e^{qV_{\text{Applied}}/kT} - 1} \\ &\quad \cdot \ln\left[\frac{1 - e^{(E_F - E_x)/kT}}{1 + e^{(E_F - E_x - qV_{\text{Applied}})/kT}}\right] dE_x \end{aligned} \tag{15b}$$

where V_{Applied} is the applied bias. So the total current is given by

$$J = \int_0^\infty [J^+(E_x) - J^-(E_x)]\, dE_x \tag{16}$$

The expression for J^+ and J^- assumes a Fermi–Dirac distribution with E_F as the Fermi level. D^+ (D^-) is the transmission probability looking from the emitter toward the collector (collector toward emitter).

The contribution of space charge will change the potential profile in the quantum well as dictated by Poisson's equation. The charge concentration in the quantum well is written as

$$n^\pm(x) = \int_0^\infty n^\pm(x, E_x)\, dE_x \tag{17}$$

where

$$n^\pm(x, E_x) = \frac{J^\pm(E_x)}{qv_g^\pm(x, E_x)} \tag{18}$$

where the superscript + (−) indicates the contribution due to the current J^+ (J^-). The total charge is obtained as

$$n(x) = \int_0^\infty dE_x \sum_\pm n^\pm(x, E_x) \tag{19}$$

The updated potential profile is obtained by solving Poisson's equation and is given by

$$\phi(x) = \phi(0) - \frac{q}{\varepsilon} \int dx' \int dx''\, n(x'') \tag{20}$$

where ε is the dielectric constant and $V_{sc}(x) = -q\phi(x)$.

The self-consistently obtained current–voltage characteristic is shown in Figure 7. The structure considered is similar to that shown in Figure 1 with 30 Å barriers and 50 Å QWs. The doping concentrations of the contact GaAs layers are 2×10^{18} cm^{-3}. The Fermi level E_F is assumed to be 0.03 eV above the bottom of the conduction band at the emitter contact. The total applied voltage is the sum of the voltage drops across the DBRTS, the voltage drop across the depletion region at the collector contact, V_d, and the drop across the accumulation region at the emitter contacts, V_a. V_d and V_a are determined by solving the equations [75]

$$V_d = \frac{qn_s^2}{2\varepsilon_{\text{collector}} N_D}$$
$$e^{qV_a/kT} - \frac{qV_a}{kT} - 1 = \frac{q^2 n_s^2}{2\varepsilon_0 kTN_D} \tag{21}$$

where N_D is the doping concentration at the contact layers. The current peaks at an applied bias of 0.13 V and corresponds to the first virtual state of the DBRTS. On the same plot, current obtained when the voltage was scanned from 0.6 to 0 V, is shown. The two graphs are different in the region showing negative differential resistance and this is attributed to the presence of space charge within the QW. The observed bistability is a strong function of the width of the QW.

In Figure 8, current–voltage characteristics are plotted in the presence of elastic and phase breaking scattering. Comparing Figures 7 and 8, it is to be noticed that in the presence of elastic scattering the peak current shifts to higher bias that is attributed to the shift in the first resonant energy as depicted in Figure 5. In the presence of phase breaking scattering a drastic reduction in the peak current magnitude is observed and is in accordance with Figure 5.

Figure 9a and b shows the self-consistently obtained group velocity as a function of the incident electron energy and position inside the QW for positive and negative voltage sweeps, respectively [78]. Figure 10a and b shows the self-consistently obtained effective density of states as a function of incident electron energy and position [80]. Figure 10a is obtained by assuming $\tau_\varphi = \infty$ or, in other words, scattering was not considered. Figure 10b is obtained by self-consistently calculating τ_φ.

4. CARRIER DYNAMICS (TUNNELING TIME)

DBRTS, biased in the NDR region, is widely investigated for possible applications requiring very high frequency oscillations. The determination of the high frequency of oscillation is based upon the evaluation of the lifetime of the resonant states. The study of traversal or tunneling time of

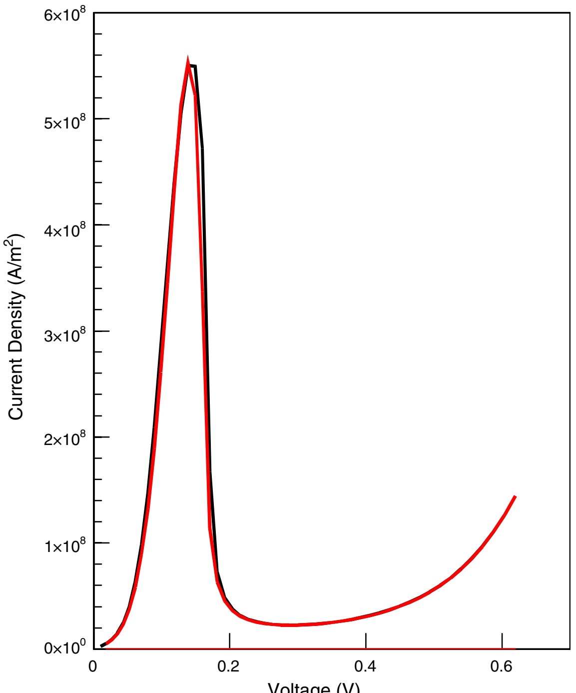


Figure 7. Self-consistently obtained current–voltage characteristics.

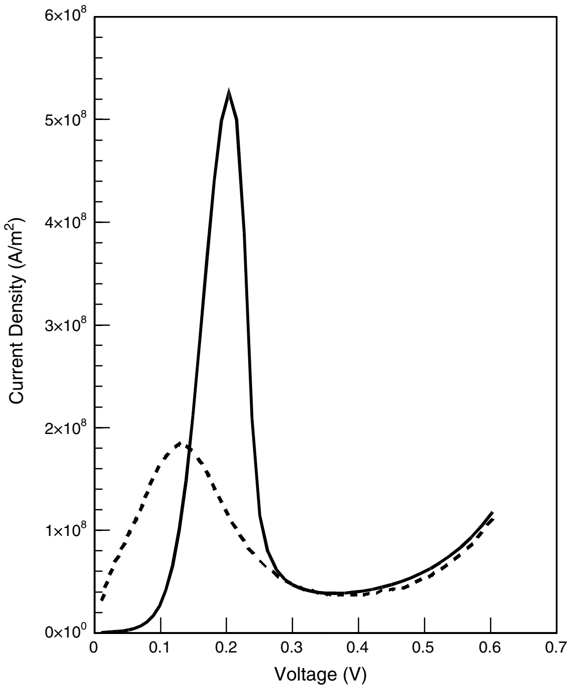


Figure 8. Self-consistently obtained current–voltage characteristics of a DBRTS in the presence of scattering. Solid and dashed lines represent elastic and phase breaking scattering, respectively.

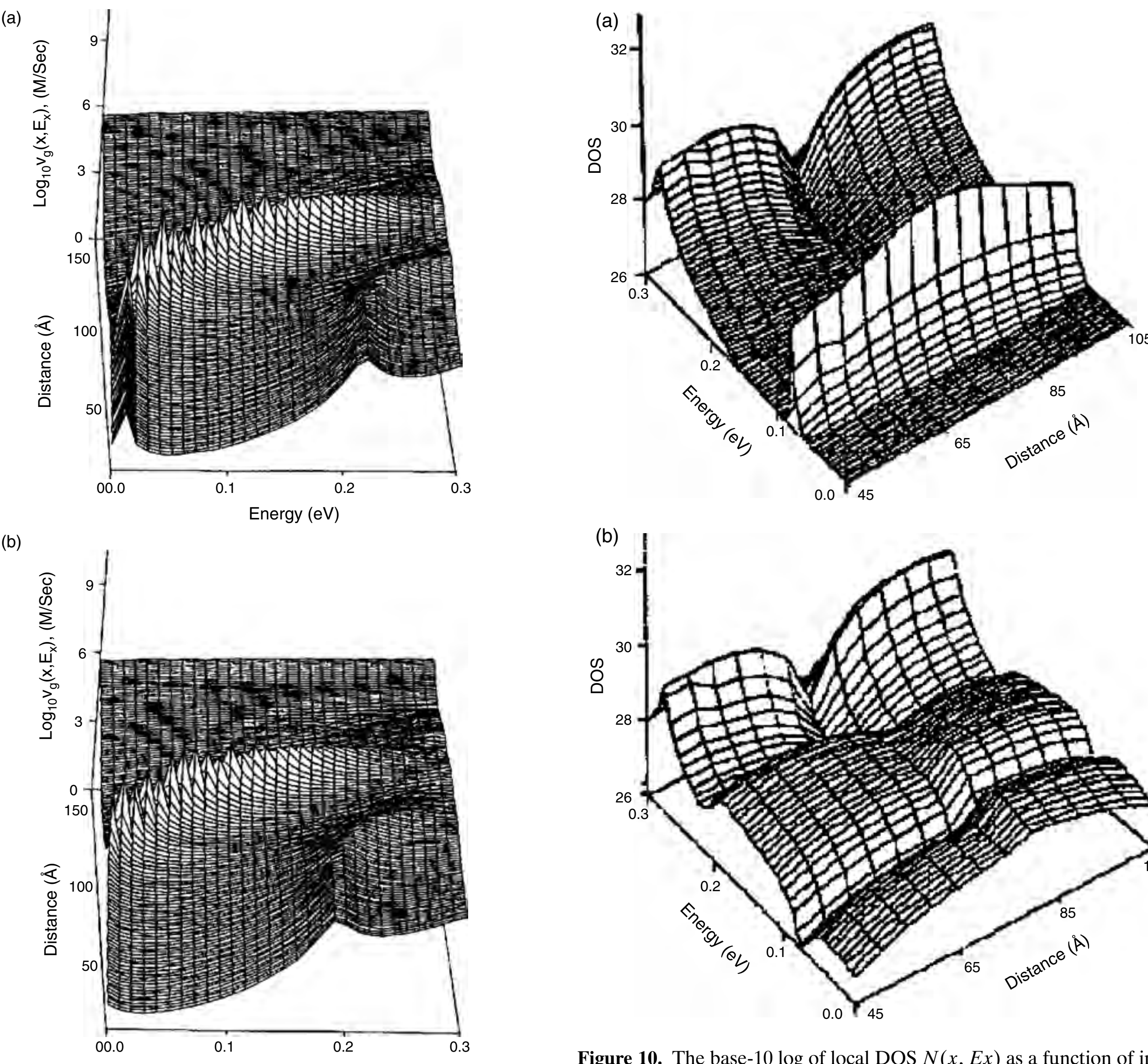


Figure 9. Group velocity is plotted as a function of incident electron energy and position at 0.1 V (positive sweep). Reprinted with permission from [78], A. F. M. Anwar and M. M. Jahan, *IEEE J. Quantum Electron.* 31, 3 (1995). © 1995, IEEE.

Figure 10. The base-10 log of local DOS $N(x, Ex)$ as a function of incident electron energy and position inside the QW. The plot to the left is with $\tau_\varphi = 0$ and the one on the right is obtained by self-consistently calculated τ_φ. Reprinted with permission from [80], M. M. Jahan, Self-consistent Study of the Effect of Transverse Magnetic Field on the Performance of the Double Barrier Resonant Tunneling Structure, M.S. Thesis, University of Connecticut, Storrs, CT, 1993.

an electron through a resonant tunneling structure, in general, has attracted much attention. The early work on tunneling time using the phase method was by Wigner [84]. Phase time can be defined locally as it is an asymptotic quantity [85]. The concept of dwell time was proposed by Smith [86] and was used by Leavens and Aers [87]. The speed of operation of NDR devices has been investigated by several authors, namely Ricco and Azbel [88], Luryi [89], and Harada and Kuroda [90]. Luryi [89] estimated the speed of operation by calculating the *RC* delay of a single barrier. Harada and Kuroda [90] studied the time evolution of an electron wave packet localized in a QW. Ricco and Azbel [88] proposed that the average time spent by an electron in the device (dwell time) should provide a good measure of the time constant required for the system to approach its final configuration. The concept of a Larmor clock to calculate particle lifetime was introduced by Baz [91] which later was extended by Rybachenko [92] and Buttiker [93]. Büttiker and Landauer [94] calculated tunneling time using the WKB approximation by considering tunneling through a time-modulated barrier. Tunneling time through a single barrier in the presence of a transverse magnetic field has been addressed by Guéret et al. [95] These calculations can be extended to study the effect of the different structure parameters, such as barrier width or height, on tunneling time. The readers will find the review article by Landauer and Martin [96] very informative in dealing with topics on tunneling time.

Tunneling time for a DBQWS of length L can be written as [78, 79, 83, 96–99]

$$\tau^{\pm}(E_x) = \int_0^L \frac{dx}{v_g^{\pm}(x, E_x)} \tag{22}$$

where the $\pm$ sign implies that the electron is moving from the emitter (collector) to the collector (emitter) electrode. It has been argued by Anwar et al. [79] that the traversal time thus defined is the sum of the time required for the electron to build up inside the QW and the corresponding decay time. In the event the buildup and decay times are equal, the traversal time τ_T equals twice the electron lifetime. Table 1 provides a comparison of the calculated traversal time and lifetime obtained from numerical calculation [99].

The dependence of traversal time on barrier thickness for a given QW width is depicted in Figure 11 [99]. For symmetric barriers, traversal time increases exponentially with barrier thickness. To investigate the sensitivity of traversal time on barrier widths an asymmetric structure is considered as shown in Figure 12 [99]. The QW width is fixed at 100 Å and traversal times for the two lowest resonant energy levels are shown. It is observed that the traversal time is rather insensitive to the variation of the width of the left (emitter) barrier, whereas it is increases exponentially with the increase of the right (collector) barrier width.

The self-consistently calculated traversal time (the time it takes for the carrier to traverse the DBRTS) is plotted as a function of the applied bias in Figure 13. The self-consistent calculation takes into account the space charge formed within the QW. A double barrier resonant tunneling structure consisting of a 100 Å wide quantum well is used while the barriers are 30 Å wide. Traversal time initially decreases with increasing bias and the rate slows down followed by a change in slope around 0.17 V and can be correlated with the device in the negative differential resistance region as seen from the I–V characteristics. The decrease beyond 0.2 V is correlated to the positive differential resistance region of the I–V characteristics.

The effect of barrier width on the traversal time has been investigated by varying both the emitter and the collector barrier widths. The traversal times for the symmetric and the asymmetric cases are then compared. It is observed that the traversal time for a given incident energy is a function of the degree of the asymmetry of the structure.

Table 1. Comparison of the traversal time, τ_T, with the lifetime, τ, obtained from simulation [90], for a symmetrical well.

V_0 (mV)	L_w (Å)	d (Å)	Lifetime, $\tau^{[90]}$ (ps)	Traversal time, τ (ps)
400	20	20	0.022	0.04549
		40	0.37	0.751
		60	6.1	12.31
	50	20	0.23	0.458
		40	8.9	17.93
		60	360.0	716.0
	100	20	2.1	4.23
		40	110.0	223.88
		60	5900.0	11769.0

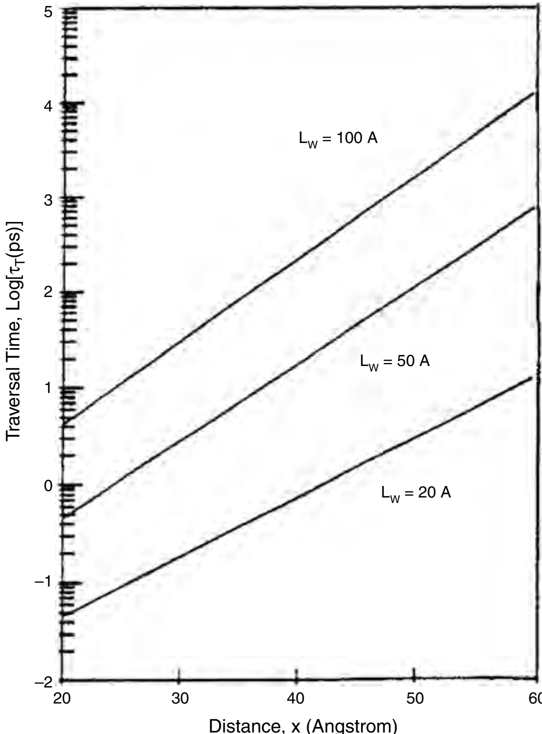


Figure 11. Traversal time of a symmetric DBRTS with QW width as a parameter. The barriers are 80 Å wide. Reprinted with permission from [99], A. F. M. Anwar, On Modelling of Quantum Well Devices, Ph.D. Dissertation, Clarkson University, Potsdam, NY, 1988.

In Figure 14, the quantity τ_V^+/τ_V^- is plotted as a function of energy and bias voltage for structures II (70–50–50 Å) and III (50–50–70 Å) respectively. A flatter surface means less bistability. It is observed from the figures that as the collector barrier is made wider the surface of the plot becomes flatter as compared to the case if emitter barrier is made wider. Thus, the wider emitter structure shows greater bistability than the wider collector structure. The observed bistability can be explained in terms of the space charge developed inside the quantum well which is a strong function of the asymmetry of DBQWS. Bistability in tunneling time depends on the difference in the amount of space charge between two sweeps for a particular bias voltage. As is evident from Figure 15, the space charges for structure III for both positive and negative sweeps are of the same order of magnitude, whereas, for structure II, they differ by an order of magnitude. This can be explained in terms of two current components $J^+$and J^-. For all structures, J^+ is of the same order of magnitude for both sweeps. J^- for structures I and III is the same for positive sweep. However, J^- for structure II is a few orders of magnitude higher in the potential range of 0 to 0.03 V and then follows the rest. In the negative sweep, the J^- characteristics are the same in the voltage range beyond 0.04 V. In the range 0 to 0.04 V, J^- for structure III is twice that of structure I (50–50–50 Å), whereas, the structure II current is a few orders of magnitudes higher. Though J^- is negligible compared to J^+, its behavior controls the extent of the bistable region in traversal time. We believe that a comparatively larger J^- in structure II is responsible for the difference in space charge

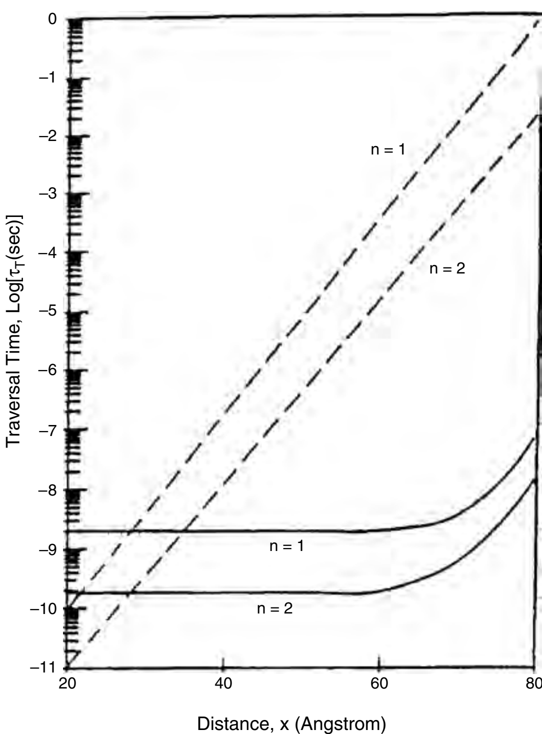


Figure 12. Traversal time for the first two eigenstates of a 100 Å wide asymmetric QW. Solid lines represent traversal time with varying d_L with $d_R = 30$ Å. Dashed lines represent traversal time with varying d_R with $d_L = 30$ Å. Reprinted with permission from [99], A. F. M. Anwar, On Modelling of Quantum Well Devices, Ph.D. Dissertation, Clarkson University, Potsdam, NY, 1988.

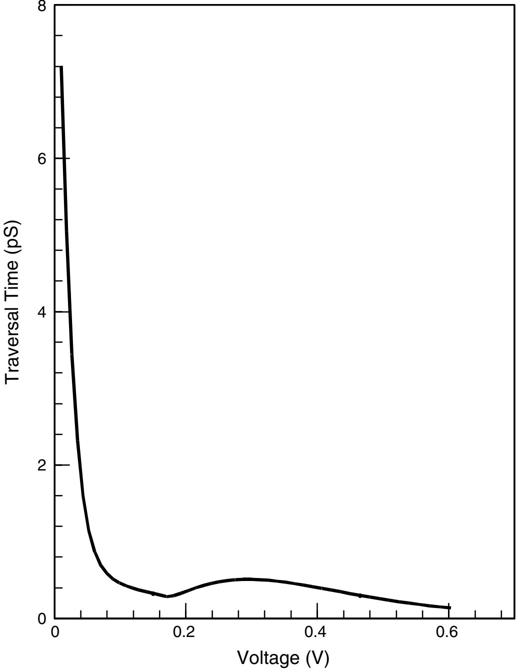
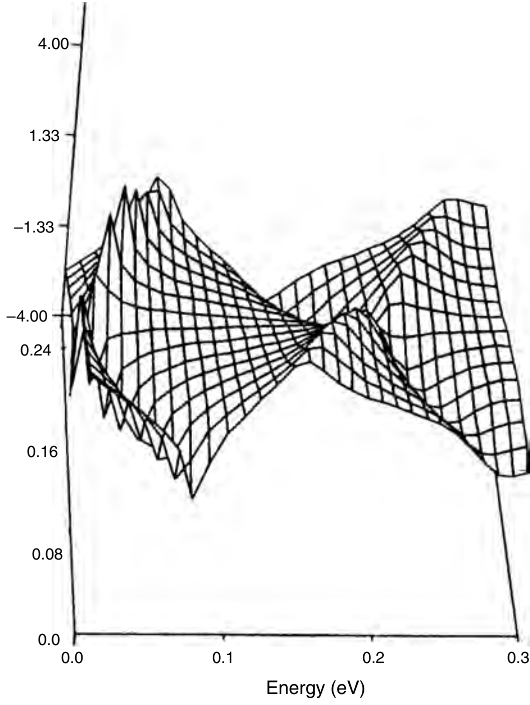


Figure 13. Self-consistently calculated traversal time as a function of applied bias.



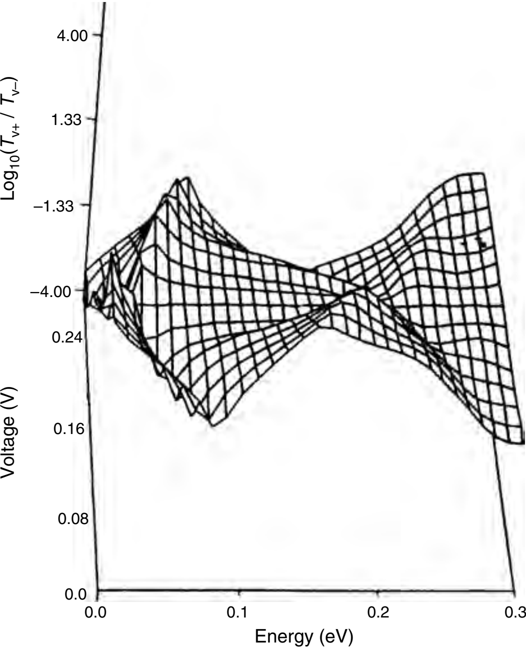


Figure 14. The ratio of τ_{v+} and τ_{v-} are plotted for structure II and structure III. Reprinted with permission from [78], A. F. M. Anwar and M. M. Jahan, *IEEE J. Quantum Electron*. 31, 3 (1995). © 1995, IEEE.

and may explain the large bistability in this structure. Also to be noted is that the space charge for structure III is much greater than that for structure II. This can be explained by noting that structure III has a higher transmission than structure II and the group velocity is of the same order of magnitude. In this connection, it is instructive to study the behavior of effective group velocity that is calculated by evaluating the quantity $J/[qn(x)]$.

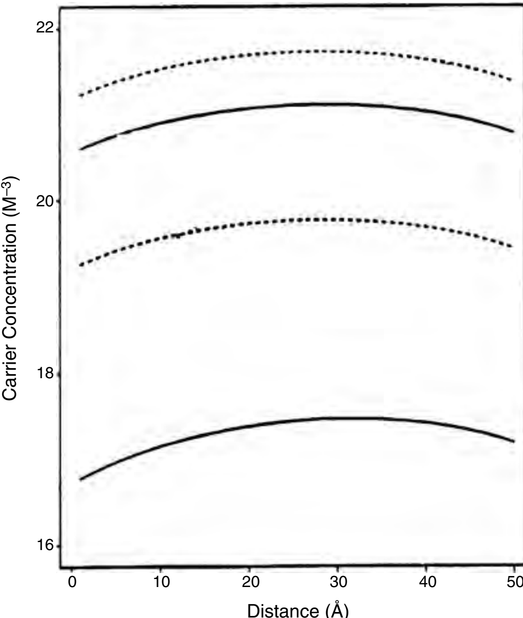


Figure 15. Space charge is plotted as a function of position. Solid and dashed lines represent positive and negative voltage sweep, respectively. The upper (lower) two curves are for structure III (structure II). Reprinted with permission from [78], A. F. M. Anwar and M. M. Jahan, *IEEE J. Quantum Electron.* 31, 3 (1995). © 1995, IEEE.

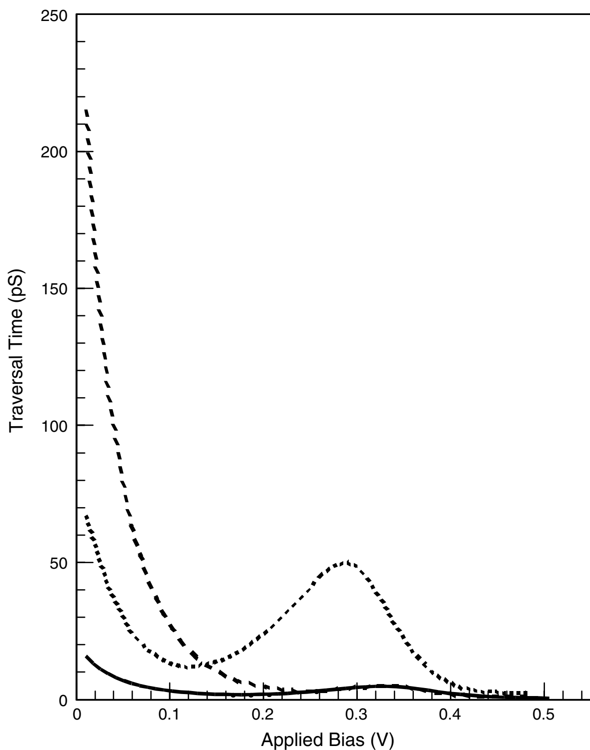


Figure 16. Traversal time as a function of applied bias. Structure I (solid line), structure II (dashed line), and structure III (dotted line).

The issue of group velocity can further be studied as a function of incident energy and position along the structure. Figure 9 shows the results for structure II for both positive and negative sweep of bias voltage. It is observed that the two sweep group velocities differ significantly in the neighborhood of 0.04 eV. As mentioned earlier, this behavior in the low energy region makes the two sweep space charge differ significantly. However, for structures I and II, the difference in behavior in the low energy region is less pronounced and hence the space charge for two sweeps also differs less.

Figure 16 shows the plot of effective tunneling time as a function of bias voltage for all three structures. Here again it is observed that a wider emitter structure requires a longer time than a wider collector structure.

5. DBRTS CAPACITANCE

The promise of ultrahigh speed operation of a DBRTS has stimulated a host of research devoted to the understanding of the factors that affect the frequency of operation of the device [100–104]. Capacitance and the associated conductance of the device are particularly important in this regard. The capacitance of a DBRTS consists of three principal components: (a) accumulation region capacitance C_a, (b) depletion region capacitance C_d, and (3) capacitance due to stored charge in the quantum well C_w. This warrants the inclusion of the capacitance C_w due to stored charge in the quantum well [102].

Brown et al. [100] calculated the capacitance by approximating DBRTS as a parallel plate capacitor with a spacing that is a sum of the widths of the depletion region, the accumulation region, and the structure itself. The neglect of space charge in the quantum well was not justified and was later pointed out by Hu and Stapleton [102]. Consequently, Hu and Stapleton provide a scheme [102, 103] to include the effects of space charge inside the quantum well in the calculation of the capacitance of DBRTS. They assumed a horizontal bar to replace the actual conduction band profile and used the Airy function to solve the trapezoidal barrier. As the discussion in the last paragraph suggests, the assumption of a horizontal conduction band in the quantum well might not be a valid approximation. Fernando and Frensley [104] solved the periodic time-dependent Schrödinger equation under a weak ac field to determine the high frequency characteristics of DBRTS.

Figure 17 shows the equivalent circuit of DBRTS. C_a and C_d are in series and C_w is in parallel to the combination. Also in parallel is the conductance of DBRTS, G. C_w is calculated as

$$C_w = A\frac{dQ}{dV}\frac{dV}{dV_{\text{Applied}}} \tag{23}$$

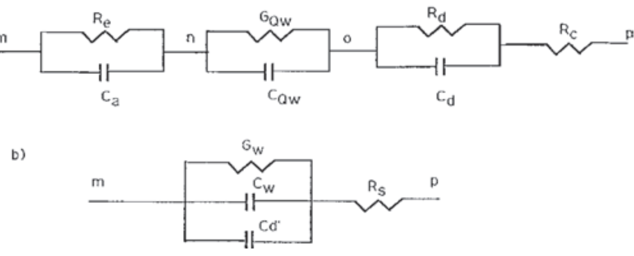


Figure 17. Small signal equivalent circuit of a DBRTS.

where Q is the total space charge in the quantum well due to traversing electrons, A is the cross-sectional area of the device, V is the voltage across the DBRTS, and V_{Applied} is the total applied bias. The calculation of $Q = -q\int_0^{L_w} n(x)\,dx$ requires the computation of the current–voltage characteristics of DBRTS.

The total capacitance is plotted as a function of the applied bias in Figure 18. The stored charge in the QW redistributes upon the application of an applied bias due to the perturbation of the eigenstates.

The frequency, f, at which a DBRTS oscillates is obtained from the requirement that the real part of the impedance be negative:

$$f = \frac{1}{2\pi C}\left(\frac{G}{R_s} - G^2\right) \tag{24}$$

For a 11–45–11 Å GaAs/AlAs DBRTS an oscillation frequency of 430 GHz is obtained which is in agreement with the experimental observation of 420 GHz by Brown et al. [100].

6. SHOT NOISE

In this section shot noise is calculated in a DBRTS by taking the space charge accumulated inside the quantum well into account. The calculation is self-consistent in nature and is obtained by simultaneously solving the Schrödinger and Poisson equations. The calculation shows the suppression of shot noise in the positive differential resistance (PDR) region and an enhancement in the NDR of the DBRTS. This behavior is explained in terms of the auto- and cross-power density spectrum of current in the presence of space charge in the quantum well. The calculated shot noise factor shows a suppression in the PDR region ($\gamma = 0.31$) and an enhancement in the NDR region ($\gamma = 9$) as observed experimentally.

Shot noise [80, 105–108] is observed in almost all electronic devices and is due to the fluctuations in device current. The fluctuations are attributed to the discreteness and randomness of the carrier flowing through the device. Therefore, any process that will affect the random nature of the carriers or eliminate the discreteness of charge will give rise to shot noise that deviates from the full shot noise. Although it is not possible to eliminate the discrete nature of the charge, however, it is possible to affect the randomness associated with the emission of the carriers in the device. This is the phenomenon that occurs in a DBRTS where the stored space charge in the quantum well alters the randomness of the injected carriers and may result either in the suppression or enhancement of shot noise [102]. It is customary to represent the deviation from the full shot noise behavior by shot noise factor γ. $\gamma > I$ for shot noise enhancement and $\gamma < I$ for shot noise suppression.

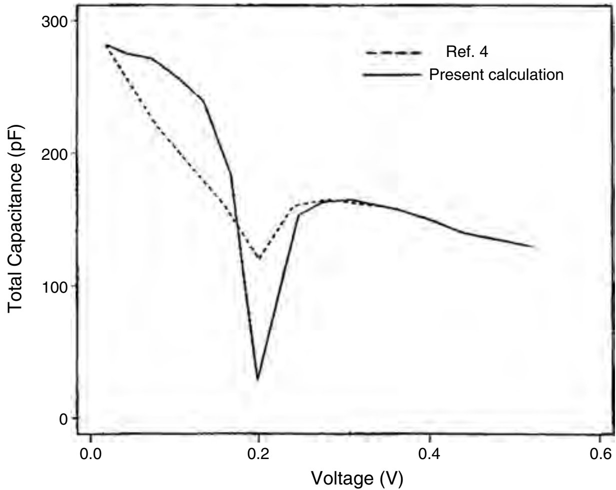


Figure 18. Total capacitance of a DBRTS.

Li et al. [106] reported the measurement of shot noise suppression of DBRTS below 10 kHz. Ran and Barnes [107] showed that shot noise may be significantly lowered by using quantum wells. They assumed distributions for both the velocity and the density of carriers incident on the quantum well structure and then used the results of quantum mechanics for describing the transport through the well. Brown [105] has presented a theoretical analysis of the suppression and enhancement of shot noise. In his analysis, the space charge is modeled as delta function at the center of the quantum well. However, the calculation was not self-consistent in nature and an exact space charge distribution was not taken into account.

The treatment presented here takes into account an exact space charge distribution and is used to compute the shot noise factor self-consistently. As an electron traverses through a DBRTS, it spends a finite amount of time inside the structure and thus the space charge gets accumulated in the quantum well. The magnitude of the space charge contributed by a traversing electron is a function of the energy level. The contribution of space charge due to an electron is also affected by the presence of other electrons. This gives rise to two different types of power density spectrums: the auto-power density spectrum (APDS) that is related to the state itself and the cross-power density spectrum (CPDS) that depends on the cross-correlation among different current carrying states. To make a self-consistent calculation of shot noise, these correlation terms need to be computed by solving the Schrödinger and Poisson equations simultaneously as has been described in previous sections.

The self-consistently calculated potential energy $V_{\text{SC}}(x)$ may also be expressed in the form [80]

$$V_{\text{SC}}(x) = \sum_{i=1}^{N} V_{\text{SC}}(x, E_x^i) \tag{25}$$

where $V_{\text{SC}}(x, E_x^i)$ denotes the space charge induced potential due to the flow of current at E_x^i. By considering $V_{\text{SC}}(x, E_x^i)$ as a perturbation potential, the shift in the eigenenergy levels and the modified $D(Ex)$ can readily be evaluated. The shift in the energy levels produces an effective perturbation in the output current that may be written as

$$(\Delta J)^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\Delta J(E_x^i, E_x^i)\right]^2 \tag{26}$$

where $\Delta J(E_x^i, E_x^i)$ is the current density at E_x^i due to perturbation of E_x^j. $(\Delta J)^2$ contains all the mean square information about APDS and CPDS and is used to calculate the shot noise factor of the structure. Shot noise power spectrum S is obtained as [80]

$$S = \frac{(\Delta J)^2}{\delta f} \tag{27}$$

where δf is the measurement bandwidth. Shot noise factor γ is expressed as

$$\gamma = \frac{S}{2qJ} \tag{28}$$

where J is the total current density, and q is the electronic charge, 1.6×10^{-19} Coulomb.

In Figure 19, the shot noise factor is plotted as a function of bias voltage. The plot shows a suppression of shot noise factor ($\gamma < I$) in the PDR region and an enhancement ($\gamma > I$) in the NDR region. The departure of γ from the full shot noise ($\gamma = I$) behavior is the consequence of the space charge developed inside the quantum well. The stored charge affects shot noise in essentially two different manners in the PDR and NDR regions. Figure 20 shows the space charge that is stored in the quantum well as a function of bias voltage. The space charge affects the I–V characteristics on two fronts: (a) First, a voltage drop is associated with the space charge which effectively reduces the net voltage across the DBRTS. Consequently, the current gets peaked at a voltage which is little higher than the case if no space charge were considered. (b) Second, space charge introduces a perturbation which affects the energy levels as well as transmission coefficients of the quantum well. In general, the effect of the perturbation is to introduce a positive shift into the transmission peaks along the energy scale. The latter phenomenon may enhance or diminish the current depending on the bias of the DBRTS. For example, if the device is biased in the PDR region, the positive shift in energy will reduce the current. On the other hand, current will increase if the device is biased in the NDR region. It is the global effect of these two mechanisms that brings about a net decrease of current in the NDR region and a net increase in the PDR region. The similar ratiocination also applies to the noise current. In Figure 19, experimental data reported by Brown [105] are also shown. The agreement of the calculated shot noise factor and the experimentally obtained one is good.

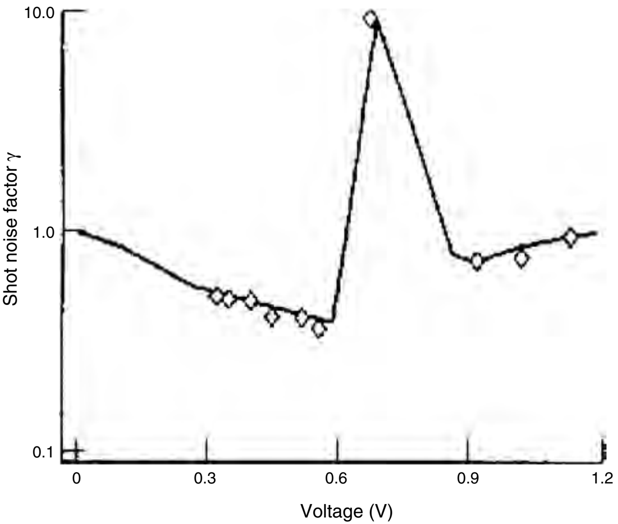


Figure 19. Shot noise as a function of applied bias. Measured data (diamond) and calculated results (solid line). Reprinted with permission from [80], M. M. Jahan, Self-consistent Study of the Effect of Transverse Magnetic Field on the Performance of the Double Barrier Resonant Tunneling Structure, M.S. Thesis, University of Connecticut, Storrs, CT, 1993.

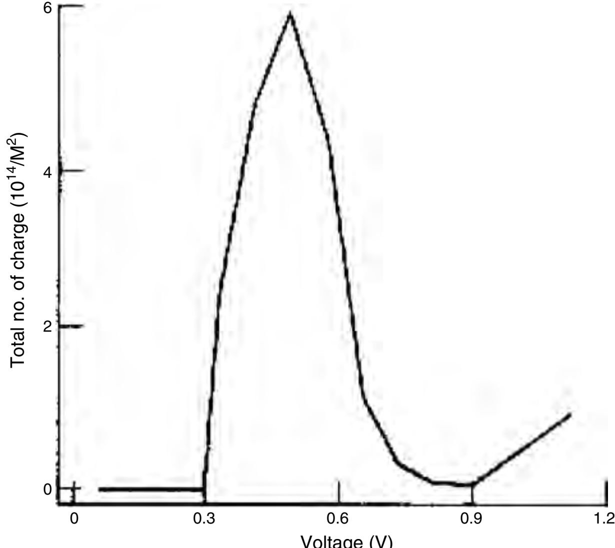


Figure 20. Space charge as a function of applied bias. Reprinted with permission from [80], M. M. Jahan, Self-consistent Study of the Effect of Transverse Magnetic Field on the Performance of the Double Barrier Resonant Tunneling Structure, M.S. Thesis, University of Connecticut, Storrs, CT, 1993.

7. CONCLUSION

In recent years DBRTS has provided the means to realize transport in reduced dimension. This has necessitated the formulation of carrier transport that incorporates the effect of scattering in reduced dimension. Moreover, DBRTS has facilitated the realization of devices with unique performance.

GLOSSARY

FET Transistor in which output drain current is controlled by the voltage applied to the gate. In a MOSFET an oxide insulator is present in between the metal and semiconductor.

Negative differential resistance (NDR) Where the current decreases with increasing voltage observed in current-voltage characteristics.

PVR The ratio of the peak (maximum) current to that of the valley (minimum) current.

Quantum well (QW) A quantum well is like a well where a particle can be trapped. Layers made by using different

semiconductor materials may result in the trapping of carriers (electrons, holes) in a particular layer. The trapped carriers and confined and are restricted to make certain movements or move in certain directions. The presence of QWs has resulted in modern electronic devices such as high electron mobility transistors (HEMTs), heterojunction bipolar transistors (HBTs) and photonic devices such as QWIPs.

Schottky barrier height Potential barrier in between the metal and semiconductor in metal semiconductor contact.

REFERENCES

1. L. Esaki and R. Tsu, IBM Internal Res. Rep. RC2418, 1969.
2. T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, *Appl. Phys. Lett.* 43, 588 (1983).
3. S. Thomas III, C. H. Fields, M. Sokolich, K. Kiziloglu, and D. Chow, in "International Conference on Indium Phosphide and Related Materials," pp. 286–289, 2000.
4. C. H. Lin, K. Yang, A. F. Gonzalez, J. R. East, P. Mazumder, and G. I. Haddad, in "International Conference on Indium Phosphide and Related Materials," pp. 419–422, 1999.
5. C. H. Lin, K. Yang, M. Bhattacharya, X. Wang, X. Zhang, J. R. East, P. Mazumder, and G. I. Haddad, in "International Conference on Indium Phosphide and Related Materials," 1998, pp. 419–422.
6. P. Velling, M. Agethen, W. Prost, G. Janssen, R. M. Bertenburg, and F. J. Tegude, in "Device Research Conference," 2001, pp. 89–90.
7. W. C. Liu, S. Y. Cheng, J. H. Tsai, P. H. Lin, J. Y. Chen, and W. C. Wang, *IEEE Electron Device Lett*. 18, 515 (1997).
8. F. Cappasso and R. A. Kheil, *J. Appl. Phys.* 58, 1366, (1987).
9. Y. F. Yang, W. L. Wang, and E. S. Yang, *Electron. Lett.* 30, 90 (1994).
10. W. Liu, A. C. Seabaugh, T. S. Henderson, A. Yuksel, E. A. Beam III, and S. Fan, *IEEE Trans. Electron Devices* 40, 1384 (1993).
11. C. E. Chang, P. M. Asbeck, K. C. Wang, and E. R. Brown, *IEEE Trans. Electron Devices* 40, 685 (1993).
12. W. C. Liu, H. J. Pan, W. C. Wang, S. C. Feng, K. W. Lin, K. H. Yu, and L. W. Laih, *IEEE Trans. Electron Devices* 48, 1054 (2001).
13. W. Prost, U. Auer, J. Degenhardt, A. Brennemann, C. Pacha, K. F. Goser, and F. J. Tegude, in "IEEE International Conference On Indium Phosphide and Related Materials," 2001.
14. R. H. Mathews, J. P. Sage, T. C. Sollner, S. D. Calawa, C. L. Chen, L. J. Mahoney, P. A. Maki, and K. M. Molvar, *Proc IEEE* 87, 596 (1999).
15. J. Robertson, T. Ytterdal, W. C. B. Peatman, R. S. Tsai, E. R. Brown, and M. Shur, *IEEE Trans. Electron Devices* 44, 1033 (1997).
16. T. Waho, K. J. Chen, and M. Yamamoto, *IEEE Electron Device Lett.* 17, 223 (1996).
17. J. P. A. Van-der-Wagt, H. Tang, T. P. E. Broekaert, A. C. Seabaugh, and Y. C. Kao, *IEEE Trans. Electron Devices* 46, 55 (1999).
18. M. Takatsu, T. Adachihara, T. Mori, Y. Awano, and N. Yokoyama, *IEDM Tech. Dig. Int.* 219 (1999).
19. C. H. Mikkelson, A. C. Seabaugh, E. A. Beam III, J. H. Luscombe, and G. A. Frazier, *IEEE Trans. Electron Devices* 41, 132 (1994).
20. Z. X. Yan and M. J. Deen, *IEEE J. Solid-State Circuits* 27, 1198 (1992).
21. T. Waho, K. Hattori, and Y. Takamatsu, in "Proceedings of the 31st IEEE International Symposium on Multiple-Valued Logic," 2001, pp. 94–99.
22. J. Stock, J. Malindretos, K. M. Indlekofer, M. Pottgens, A. Forster, and H. Luth, *IEEE Trans. Electron Devices* 48, 1028 (2001).
23. C. Pacha, U. Auer, C. Burwick, P. Glosekotter, A. Brennemann, W. Prost, F. J. Tegude, and K. F. Goser, *IEEE Trans. Very Large Scale Integration Syst.* 8, 558 (2000).
24. P. Mazumder, *IEEE Trans. Analog Digital Signal Process.* 47, 1080 (2000).
25. W. Williamson III, S. B. Enquist, D. H. Chow, H. L. Dunlap, S. Subramaniam, L. Peiming, G. H. Bernstein, and B. K. Gilbert, *IEEE J. Solid State Circuits* 32, 222 (1997).
26. K. J. Chen, T. Waho, K. Maezawa, and M. Yamamoto, *IEEE Electron Device Lett.* 17, 309 (1996).
27. K. Imamura, M. Takatsu, T. Mori, T. Adachihara, H. Ohnishi, S. Muto, and N. Yokoyama, *IEEE Trans. Electron Devices* 39, 2707 (1992).
28. A. F. Gonzalez, M. Bhattacharya, S. Kulkarni, and P. Mazumder, *IEEE J. Solid State Circuits* 36, 924 (2001).
29. E. Lind, L. E. Wernersson, I. Pietzonka, and W. Seifert, *IEEE Trans. Electron Devices* 49, 1066 (2002).
30. H. Fukuyama, K. Maezawa, M. Yamamoto, H. Okazaki, and M. Muraguchi, *IEEE Trans. Electron Devices* 46, 281 (1999).
31. B. Zhang, K. J. Chen, R. Gang, and R. M. M. Chen, in "Proceedings of the IEEE International Symposium on Circuits and Systems," 1, 178–181, 1999.
32. T. Waho, K. J. Chen, and M. Yamamoto, *IEEE J. Solid State Circuits* 33, 268 (1998).
33. K. J. Chen and M. Yamamoto, *IEEE Electron Device Lett*. 17, 235 (1996).
34. I. I. Abramov and A. V. Korolev, in "International Conference on Microwave and Telecommunication Technology," 2001, pp. 445–448.
35. U. Auer, W. Prost, G. Janssen, M. Agethen, R. Reuter, and F. J. Tegude, *IEEE J. Selected Topics Quantum Electron*. 2, 650 (1996).
36. T. P. E., B. Brar, F. Morris, A. C. Seabaugh, and G. Frazier, in "Proceedings of the Ninth Great Lakes Symposium on VLSI," 1999, pp. 123–126.
37. R. J. Pflueger, *IEEE Trans. Circuits Syst. I* 41, 744 (1994).
38. R. J. Pflueger, *IEEE Trans. Circuits Syst. I* 41, 740 (1994).
39. M. Kahn, J. Lasri, G. Eisenstein, D. Ritter, and M. Orenstein, in "International Topical Meeting on Microwave Photonics," 2001, pp. 207–210.
40. P. Holmstrom, L. Thylen, and U. Ukenberg, *IEEE J. Quantum Electron*. 37, 224 (2001).
41. R. Bouregba, O. Vanbesien, P. Mounaix, D. Lippens, L. Palmateer, J. C. Pernot, G. Beaudin, P. Encrenaz, E. Bockenhoff, J. Nagle, P. Bois, F. Chevoir, and B. Vinter, *IEEE Trans. Microwave Theory Techniques* 41, 2025 (1993).
42. A. Kikuchi, R. Bannai, and K. Kishino, in "60th Device Research Conference," 2002, pp. 157–158.
43. P. See and D. J. Paul, *IEEE Electron Device Lett.* 22, 582 (2001).
44. P. See, D. J. Paul, B. Hollander, S. Mantl, I. V. Zozoulenko, and K. F. Berggren, *IEEE Electron Device Lett*. 22, 182 (2001).
45. A. A. Iliadis, A. Inumqudi, S. Choopun, R. D. Vispute, and T. Venkatesan, in "Semiconductor Device Research Symposium," 2001, pp. 450–453.
46. M. Watanabe, N. Sakamaki, and T. Ishikawa, "IEEE International Conference On Indium Phosphide and Related Materials," 2001, pp. 244–247.
47. Y. K. Su, J. R. Chang, Y. T. Lu, C. L. Lin, and K. M. Wu, *IEEE Trans. Electron Devices* 47, 895 (2000).
48. R. Lake, B. Brar, G. D. Wilk, A. Seabaugh, and G. Klimeck, "IEEE International Symposium on Compound Semiconductors," 1998, pp. 617–620.
49. Y. K. Fang, K. H. Chen, C. R. Liu, J. D. Hwang, K. S. Wu, and W. R. Liou, *IEEE J Quantum Electron*. 30, 2293 (1994).
50. W. L. Chen, G. O. Munns, J. R. East, and G. I. Haddad, *IEEE Trans. Electron Devices* 41, 155 (1994).
51. B. Gelmont, D. Woolard, W. Zhang, and T. Globus, in "Semiconductor Device Research Symposium," 2001, pp. 256–259.
52. R. Lake and S. Datta, *Phys. Rev. B* 45, 6670 (1992).
53. D. Z. Y. Ting, M. Gu, J. Cao, X. Chi, and J. N. Schulman, in "IEEE International Symposium on Compound Semiconductors," 2000, pp. 185–190.

54. S. E. Hosseini and R. Faez, in "Proceedings of the 12th International Conference on Microelectronics," 2000, pp. 191–194.
55. Y. Hu and S. P. Stapleton, *IEEE J. Quantum Electron.* 29, 327 (1993).
56. W. R. Frensley, *Phys. Rev. B* 36, 1570 (1987).
57. K. L. Jensen and A. K. Ganguly, *J. Appl. Phys.* 73, 4409 (1993).
58. R. K. Mains and G. I. Haddad, *J. Appl. Phys.* 64, 5041 (1988).
59. K. L. Jensen and F. A. Bout, *J. Appl. Phys.* 67, 7602 (1990).
60. W. R. Frensley, *Phys. Rev. Lett.* 57, 2853 (1986).
61. O. Pinaud, *J. Appl. Phys.* 92, 1987 (2002).
62. Y. Ando, W. Contrata, K. Maruhashi, and H. Miyamoto, in "IEEE International Symposium on Compound Semiconductors," 1997, pp. 609–612.
63. C. Surya, N. Sze-Him, E. R. Brown, and P. A. Maki, *IEEE Trans. Electron Devices* 41, 2016 (1994).
64. K. M. Hung and G. Y. Wu, "Electron Devices and Materials Symposium," 1994.
65. J. H. Davies, P. Hydgaard, S. Hershfield, and J. W. Wilkins, *Phys. Rev. B* 46, 9620 (1992).
66. H. C. Liu, J. Li, G. C. Aers, C. R. Leavens, M. Buchanan, and Z. R. Wasilewski, *Phys. Rev. B* 51, 5116 (1995).
67. G. Iannaccone, G. Lombardi, M. Macucci, and B. Pellegrini, *Phys. Rev. Lett.* 80, 1054 (1998).
68. H. B. Sun and G. J. Milburn, *Phys. Rev. B* 59, 10748 (1999).
69. Y. M. Blanter and M. Buttiker, *Phys. Rev. B* 59, 10217 (1999).
70. S. H. Ng and C. Surya, *J. Appl. Phys.* 73, 7504 (1993).
71. A. Ouacha and M. Willander, *J. Appl. Phys.* 77, 6026 (1995).
72. C. Sirtori, F. Capasso, J. Faist, A. L. Hutchinson, D. L. Sivco, and A. Y. Cho, *IEEE J. Quantum Electron.* 34, 1722 (1998).
73. J. H. Luscombe, J. N. Randall, and A. M. Bouchard, *Proc. IEEE* 79, 1117 (1991).
74. F. Capasso, S. Sen, and A. Y. Cho, *Proc. SPIE* 792, 10 (1987).
75. Y. Ando and T. Itoh, *J. Appl. Phys.* 61, 1497 (1987).
76. W. W. Liu and M. Fukuma, *J. Appl. Phys.* 60, 1555 (1986).
77. A. N. Khondker, M. R. Khan, and A. F. M. Anwar, *J. Appl. Phys.* 63, 5191 (1988).
78. A. F. M. Anwar and M. M. Jahan, *IEEE J. Quantum Electron.* 31, 3 (1995).
79. A. F. M. Anwar, A. N. Khondker, and M. R. Khan, *J. Appl. Phys.* 65, 2761 (1989).
80. M. M. Jahan, Self-consistent Study of the Effect of Transverse Magnetic Field on the Performance of the Double Barrier Resonant Tunneling structure," M.S. Thesis, University of Connecticut, Storrs, CT, 1993.
81. A. F. M. Anwar and M. M. Jahan, *Phys. Rev. B* 50, 10864 (1994).
82. N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry, *Phys. Rev. B* 39, 7720 (1989).
83. A. F. M. Anwar and M. M. Jahan, *Phys. Rev. B* 49, 17440 (1994).
84. E. P. Wigner, *Phys. Rev.* 98, 145 (1955).
85. E. H. Hauge and J. A. Støvneng, *Rev. Mod. Phys.* 61, 917 (1989).
86. F. T. Smith, *Phys. Rev.* 118, 349 (1960).
87. C. R. Leavens and G. C. Aers, *Phys. Rev. B* 39, 237 (1991).
88. B. Ricco and M. Y. Azbel, *Phys. Rev. B* 29, 1970 (1984).
89. S. Luryi, *Appl. Phys. Lett.* 47, 490 (1985).
90. N. Harada and S. Kuroda, *Jpn. J. Appl. Phys.* 25, L871 (1986).
91. A. I. Baz', *Sov. J. Nucl. Phys.* 4, 182 (1967).
92. V. F. Rybachenko, *Sov. J. Nucl. Phys.* 5, 635 (1967).
93. M. Buttiker, *Phys. Rev. B* 27, 6178 (1982).
94. M. Buttiker and R. Landauer, *Phys. Rev. Lett.* 49, 1739 (1982).
95. P. Gueret, A. Baratoff, and E. Marclay, *Europhys. Lett.* 3, 367 (1987).
96. R. Landauer and Th. Martin, *Rev. Mod. Phys.* 66, 217 (1994).
97. A. F. M. Anwar, R. B. LaComb, and M. Cahay, *Superlatt. Microstruct.* 11, 131 (1992).
98. M. Cahay, K. T. Dalton, G. S. Fisher, and A. F. M. Anwar, *Superlatti. Microstruct.* 11, 113 (1992).
99. A. F. M. Anwar, On Modelling of Quantum Well Devices, Ph.D. Dissertation, Clarkson University, Potsdam, NY, 1988.
100. E. R. Brown, T. C. L. G. Sollner, C. D. Parker, W. D. Goodhue, and C. L. Chen, *Appl. Phys. Lett.* 55, 1777 (1991).
101. E. R. Brown, J. R. Söderström, C. D. Parker, L. J. Mahoney, K. M. Molvar, and T. C. Mcgill, *Appl. Phys. Lett.* 58, 2291 (1991).
102. Y. Hu and S. P. Stapleton, *Appl. Phys. Lett.* 58, 167 (1991).
103. Y. Hu and S. P. Stapleton, *J. Appl. Phys.* 31, 23 (1992).
104. C. L. Fernando and W. R. Frensley, *Phys. Rev. B* 52, 5092 (1995).
105. E. R. Brown, in "Device Research Conference," 1991, Co, USA.
106. Y. P. Li, A. Zaslavsky, D. C. Tsui, M. Santos, and M. Shayegan, *Phys. Rev. B* 41, 8388 (1990).
107. J. Han and F. S. Barnes, *IEEE Trans. Electron Devices* 38, 237 (1991).
108. M. M. Jahan and A. F. M. Anwar, *Solid State Electron.* 38, 429 (1995).

www.aspbs.com/enn

Room-Temperature Ballistic Nanodevices

Aimin M. Song

University of Manchester Institute of Science and Technology, Manchester, England

CONTENTS

1. Introduction
2. Ballistic Rectifiers and Semiclassical Model
3. Artificial Functional Nanomaterials
4. Three-Terminal Ballistic Junction Devices
5. Near Room-Temperature Operations of Quantum-Effect-Based Devices
6. Conclusions
Glossary
References

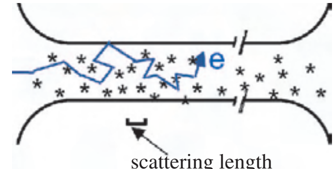
1. INTRODUCTION

One of the most important physical parameters to describe the quality of a piece of semiconductor material is the electron scattering length l_e. Also referred to as the mean-free path, it stands for the average distance between the randomly distributed scatterers in the material, such as lattice defects, impurities, and phonons. The electron mean-free path is typically a few nanometers (1 nm $= 10^{-9}$ m) in silicon and about 100–200 nm in high-quality compound semiconductors such as GaAs. The electrical resistance of a semiconductor device is closely associated with the scattering length l_e. Conventional semiconductor devices are much larger than the electron scattering length. As a result, electrons have to encounter a large number of random scattering events to travel from one device lead to another. In the last two decades, advanced semiconductor technologies have allowed the fabrication of devices that are smaller than the electron scattering length. In such devices, electrons may travel from one electric lead to another without encountering any scattering event from randomly distributed scatterers such as impurities. Instead, the electrons are scattered only at the device boundaries, that is, moving like billiard balls. Such electron transport is referred to as ballistic transport [1–3]. There had been a long debate in history on whether there is any electric resistance in such ballistic electron devices. Some had believed that the absence of scattering means that the resistivity inside the device is zero. Although this is, as will be discussed later, wrong because the concept of resistivity does not hold any longer in the ballistic transport regime, it was realized quite clearly that the electron transport in a ballistic device is very different from the transport in a traditional, macroscopic semiconductor device.

Consider a conventional semiconductor device that is much larger than the electron scattering length. Any electric current induced by an applied voltage consists of electrons diffusing in an electric field, as shown in Figure 1. The diffusive transport can be described by Ohm's law, which states a linear relation between the current and electric field (or applied voltage). However, if we examine the electron transport at the microscopic scale, we would find that every electron is accelerated between two subsequent scattering events by the electric field. As a result, the velocity component of the electron in the opposite direction of the electric field increases constantly until the next scattering event occurs. Since the current carried by the electron is equal to its velocity multiplied by the electron charge e, the current is time dependent, even in the presence of a constant electric field. This is seemingly in contradiction to Ohm's law. Actually, Ohm's law is valid at the macroscopic scale because of the existence of the large number of scattering events, which counteract the acceleration effect of the electric field, causing the average velocity of all electrons to be proportional to the strength of the electric field. The above discussion reveals that, at a scale that is smaller than the average distance between impurities or other scatterers, Ohm's law does not hold, and that new transport properties are expected to arise in these ballistic devices of a dimension smaller than l_e.

A number of new device concepts have been generated based on ballistic electron transport in the last two decades. Among these novel devices are only a few that have been demonstrated to work at room temperature. Quite understandably, being capable of operating at room temperature is an important criterion for practical applications. In this chapter, we review these recent breakthroughs in room-temperature ballistic devices. The theoretical framework to treat ballistic electron transport, namely, the scattering

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.



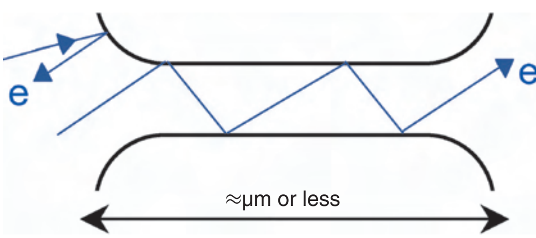
Diffusive electron transport in a macroscopic device

Figure 1. Schematic view of the diffusive electron transport in a macroscopic semiconductor device. The device size is much larger than the electron scattering length. As a result, the electrons experience a large number of scattering events from randomly distributed scatterers such as lattice defects and phonons (represented by the asterisks), and the electron transport is hence referred to as diffusive.

approach developed by Landauer and Büttiker, will be introduced. We will also discuss the promising possibilities for practical applications of these devices from the point of view of their new device properties. The chapter is organized as follows. After a short introduction to the ballistic transport in Section 1.1, an overview of the general aspects of nonlinear ballistic electron transport is presented in Section 1.2, which forms a basis for the few types of room-temperature nanodevices and nanomaterials discussed in Sections 2–4. In Section 5, promising results on near room-temperature operations of quantum-effect-based nanodevices are discussed. Finally, Section 6 provides conclusions and an outlook of future developments.

1.1. Overview of Ballistic Electron Transport

Ballistic electron transport is normally realized in mesoscopic semiconductor structures. The word "mesoscopic" refers to dimensions between the atomic scale and the macroscopic scale. The simplest mesoscopic ballistic device is a narrow, uniform semiconductor channel, as shown in Figure 2. At a finite applied voltage, a current flows through the channel. If the length of the channel is shorter than l_e, the electron transport inside the channel is ballistic. The electrons, when moving from one side of the channel to the other, do not experience any random scattering event from scatterers such as lattice defects, impurities, and phonons, but are scattered only by the designed device boundary. This



Ballistic electron transport in a mesoscopic device

Figure 2. Schematic view of the ballistic electron transport in a mesoscopic semiconductor device, which is smaller than the electron scattering length. In the device, the electrons do not experience any scattering event from random scatterers such as lattice defects and phonons, and are only scattered by the designed device boundary.

means that, even at a finite applied voltage, the electric field, and therefore the voltage drop inside the central part of the channel, which has a constant width, have to be zero in order to fulfill the requirement of the conservation of current: the current flow into one end of the channel must be the same as the current flow out of the channel at the other end. Therefore, inside the channel, current flows without the existence of a finite electric field. The striking phenomenon appears to be similar to the superconductivity effect, but occurring in a completely different physics system under a different physical condition. Such an effect of current flow without the presence of an electric field has been observed regardless of the length and width of the channel, as long as the channel is shorter than the electron mean-free path. Nevertheless, to drive a current through the entire device, a finite voltage has to be applied. The applied voltage will only drop at the entrance and exit of the channel, where finite resistance arises because most of the electrons are backscattered and only a fraction of the electrons can enter the channel [4]. As a consequence, the electric field exists only in the close vicinity of the entrance and exit of the channel. Clearly, local quantities such as resistivity and conductivity lose their meaning, and can no longer be defined in a ballistic semiconductor device. Only nonlocal quantities such as the total resistance or conductance can be used.

The conductance of a ballistic semiconductor device G is determined by an important nonlocal quantity: the transmission coefficient $T_{\beta\leftarrow\alpha}$ of the ballistic electrons going from one electric contact (also often referred to as the electric lead, carrier reservoir, or device terminal) α to anther contact β. If more than one transverse quantum confinement mode in the device is occupied, $T_{\beta\leftarrow\alpha}$ is the sum of the transmission probabilities of the electrons at each occupied transverse quantum confinement mode i:

$$T_{\beta\leftarrow\alpha} = \sum_i t_{\beta\leftarrow\alpha,\,i} \tag{1}$$

As described by the Landauer–Büttiker formula [5–7], the current flow through the narrow channel is given by

$$I = \frac{2e}{h}\sum_i t_{\beta\leftarrow\alpha,\,i}(\mu_\alpha - \mu_\beta) \tag{2}$$

where μ_α and μ_β are the chemical potentials of the electrons at the left and right contacts of the channel. If the electrons at each transverse mode transfer through the channel perfectly (without being reflected back), which was found to be the case if the boundary of a narrow channel is smooth and if the temperature is sufficiently low [8, 9], the overall conductance of the channel becomes quantized:

$$G = N\frac{2e^2}{h} \tag{3}$$

Here, N stands for the total number of the occupied transverse quantum confinement modes by the electrons in the channel, and spin degeneracy is assumed. This quantized conductance is fundamentally similar to the quantized Hall conductance at a high magnetic field: both are direct manifestations of one-dimensional electron transport [10]. Therefore, the quantum Hall effect has also been

elegantly described in the theoretical framework of the Büttiker–Landauer formalist [11].

It has been expected that devices based on ballistic electron transport have superb performance. Since electrons are free from random scattering events, ballistic electron devices may have a very high intrinsic working speed and a quick response. Secondly, the strongly temperature-dependent phonon scatterings, which in some cases lead to serious degradation in device performance in a traditional electron transport device, do not exist in a ballistic device by its nature. Therefore, ballistic devices in general are less temperature dependent.

It is easiest to realize ballistic electron transport at cryogenic temperatures using III–V compound semiconductor materials because of the long electron mean-free path obtained under both conditions. For example, l_e exceeding 160 μm has been observed in a high-quality GaAs–AlGaAs heterostructure at temperature $T = 0.1$ K [12]. Therefore, it is even possible to use conventional photolithography to fabricate devices smaller than l_e and realize ballistic transport at low temperatures. There have been a great number of studies on ballistic electron transport, and many striking phenomena were discovered in various types of ballistic devices. Examples are the above-mentioned quantized conductance in a point contact [8, 9], electron focusing experiments [13–16], negative bend resistance [17], quenched and negative Hall effects [18–20], and lateral hot ballistic electron devices [21, 22].

New transport properties realized in the ballistic transport regime have inspired inventions of novel device concepts. Many of the new concepts are based on the fact that the only scatterings that the ballistic electrons experience are those from the designed device boundaries. By simply tailoring the boundary of a ballistic device, the electron transport can be, to a large extent, modified and controlled. It is hence possible to generate new device functionalities by simply tailoring the device shape. In contrast, due to the nature of the diffusive electron transport, the properties of a traditional device are not sensitive to a change in the device boundary.

Despite great experimental efforts, only a small number of successful experiments have demonstrated ballistic transport at room temperature. From the application point of view, it is highly desirable that new types of ballistic devices are capable of operating at room temperature. At room temperature, the electron mean-free path in a III–V semiconductor heterostructure is typically from 100 to 200 nm. Device dimensions smaller than or at least comparable to l_e at $T = 300$ K are therefore required. This is, however, very difficult if using a GaAs–AlGaAs heterostructure due to the large depletion length (around 100 nm) close to the etched device boundary. The much shorter depletion length (below 30 nm) in, for example, InGaAs–InAlAs or InGaAs–InP material systems, makes it possible to fabricate much smaller devices that work at room temperature. In general, it is also easier to achieve a higher sheet carrier density in an InGaAs–InAlAs heterostructure (up to around 1×10^{12} cm^{-2}) than in a GaAs–AlGaAs system. The higher carrier density not only reduces the depletion length, but also increases the Fermi velocity, and hence enhances the electron mean-free path.

Figure 3 schematically shows a device fabricated by Hirayama and Tarucha using an AlGaAs/InGaAs/GaAs modulation-doped heterostructure [23]. This is, to the author's knowledge, the first device that has been demonstrated to show ballistic electron transport properties at room temperature. The four straight insulated lines (represented by the dark lines in the figure) were made by a focused ion beam (FIB) of Ga. The FIB beam diameter was about 80 nm. The FIB scanned lines separated the four electric contacts, numbered 1, 2, 3, and 4.

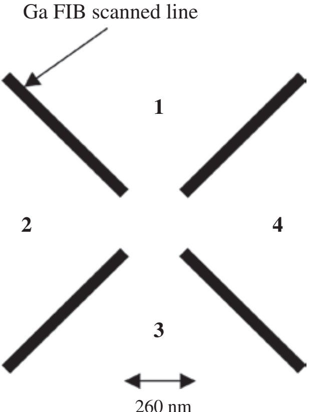


Figure 3. Schematic view of a device fabricated by Hirayama and Tarucha [23]. The four straight insulating lines (represented by the dark lines in the figure) were made by a focused Ga ion beam. Four electric contacts, numbered 1–4, were separated by the insulating lines.

During the experiments, Hirayama and Tarucha measured the so-called four-terminal bend resistance $R_{12,43} = V_{43}/I_{12}$ by sending a current I_{12} through contacts 1 and 2 and detecting the voltage between 4 and 3. If the electron transport is diffusive, the bend resistance $R_{12,43}$ would be positive. However, the experiment showed that $R_{12,43}$ was negative in the ballistic transport regime, which can be understood by considering the transmission coefficients of the ballistic electrons. Suppose that $I_{12} < 0$, and that electrons are ejected into the junction from contact 1. Instead of turning the corner between contacts 1 and 2 and transmitting into contact 2, the electrons enter the junction like billiard balls, and tend to go straight into contact 3. This results in an accumulation of electrons in contact 3, and therefore a positive V_{43} (instead of a negative voltage in a diffusive device) between contacts 4 and 3. The negative bend resistance is hence an indication of the existence of ballistic electron transport.

In the experiments, negative $R_{12,43}$ was observed from $T = 1.5$ K all the way up to room temperature, although the magnitude of the negative bend resistance was reduced by a factor of about 8 at $T = 290$ K. The result was interpreted in terms of the temperature-dependent mean-free path. At low temperature, l_e is much larger than the size of the junction in the center, and purely ballistic transport was realized. At $T = 290$ K, however, l_e is only about 135 nm, which is shorter than the length of the junction (260 nm), leading to the decrease of the magnitude of the negative bend resistance. Nevertheless, since the mean-free path at room temperature is still comparable to the length of the junction, some electrons still travel ballistically without being scattered when passing through the junction, although most electrons will experience at least one scattering event.

This is the reason why the bend resistance is still negative at room temperature. Furthermore, as Hirayama and Tarucha pointed out, at a higher temperature, more electrons will gain thermal energy above the Fermi level. These higher energy electrons might enhance the ballistic nature of the system. Recently, Hieke et al. similarly showed negative bend resistance in small cross junctions made from an InGaAs–InP modulation-doped quantum-well wafer [24]. The structures were fabricated by electron-beam lithography and deep dry etching. The experiments also indicated the partly ballistic electron transport in their devices at room temperature. Another important result of this study is the demonstration of low depletion length in such InP–InGaAs etched structures.

1.2. Nonlinear Ballistic Electron Transport

In this review, we focus on the recently demonstrated new types of ballistic devices that have been shown not only to operate at room temperature, but also to have great potentials in practical applications. All three types of devices are based on nonlinear ballistic electron transport, which is not at all simply by coincidence.

Nonlinear transport is particularly important in mesoscopic structures, and many striking phenomena discovered in ballistic devices are in the nonlinear regime, as shown in a number of experiments [25–40] and theoretical investigations [41–48]. First of all, a very small applied voltage suffices to drive the electron transport into the nonlinear regime. This is because of the small device dimension, and the fact that, in general, the applied voltage drops quite locally in the regions where the dimension of the electron path is changed, such as the opening and exit of a narrow channel, as mentioned earlier. Secondly, in a multiple-terminal ballistic device, the angular distribution of the electrons may largely determine the electron transmission probabilities between the terminals, and consequently, also the transport properties of the device. In contrast to the electron transport in a macroscopic semiconductor device, where the large number of scattering events completely randomize the direction of the electron velocities, a voltage applied to a ballistic device will change the angle of every electron in a predictable manner. As will be shown later, this has a direct influence on the transmission probabilities, and therefore also on the current distribution in the device. This is a new type of nonlinear effect which does not exist in a macroscopic device. In particular, it was recently realized that the geometrical symmetry of a ballistic device can also have a very pronounced influence on the nonlinear device properties [33–37]. One such example, which will be shown in Section 2, is called a ballistic rectifier. It is a completely new type of rectifying device because the nonlinear effect was based neither on a doping junction nor on a barrier structure. The ballistic rectifier has been demonstrated to operate at room temperature and at frequencies up to at least 50 GHz. A semiclassical model of the device is shown in Section 2.2.

Based on a similar working principle to the ballistic rectifier, a new class of nanostructured artificial material has also been constructed, as will be shown in Section 3. The nanomaterial turns out to be in analog with the so-called photogalvanic crystals. It was discovered more than 30 years ago that some natural crystals can generate a direct electron current under uniform illumination [49–51]. Referred to as the photogalvanic effect, the phenomenon has been identified to be a macroscopic manifestation of the absence of centrosymmetry in the elementary crystal structure, despite the macroscopic homogeneity of these crystals. In this case, light serves as a source of external fluctuation by imposing an unbiased, time-periodic electric field on the crystals, and the generation of direct electron current can be viewed as an extremely fast rectification at the frequency of the light. Recently, similar geometries to those found in photogalvanic crystals were realized to be useful for biomolecular separation devices consisting of arrays of symmetry-breaking microscopic obstacles [52–55].

Furthermore, it is naturally expected that, if the device dimension is comparable to the electron wavelength, quantum effects will take place. For a narrow channel or a device with a narrow constriction such as a quantum point contact, an applied voltage changes both the phase of the electrons and the electron distribution among the transverse quantum confinement modes in the channel, hence leading to a change in the conductance. In Section 4, three-terminal ballistic devices based on such a nonlinear effect will be discussed. The effect can also be well described with an extended Büttiker–Landauer formula.

In Section 5, the recently discovered quantum behavior of nanometer-sized ballistic rectifiers at temperatures very close to room temperature will be presented. Quantum devices generally have the advantages of being ultrasensitive and ultrafast. The results indicated the promising possibility of realizing room-temperature operations of quantum-effect-based nanoelectronic devices.

After the review on these nonlinear, room-temperature ballistic devices, ranging from basic device physics and modeling to the device performance and prospects of real applications, we conclude with an outlook in Section 6 on expected future developments in this yet new, but promising field.

2. BALLISTIC RECTIFIERS AND SEMICLASSICAL MODEL

In this section, a new type of semiconductor rectifier is introduced. The device, referred to as a ballistic rectifier, is based on a completely different working principle from a semiconductor diode since no doping junction or barrier structure has been used. The rectifying effect relies on the nature of the ballistic electron transport, and also on tailoring the device geometry. After the experimental results shown in Section 2.1, a semiclassical model of nonlinear ballistic electron transport will be introduced in Section 2.2, which yields an analytical description of the ballistic rectifying effect.

2.1. Experimental Results

There were quite a number of studies on mesoscopic semiconductor cross junctions in the early days of the realization of ballistic transport. A variety of novel phenomena have been observed in such devices, such as electron focusing

[13–15], negative bend resistances [17], and the quenched or negative Hall effect [18, 19]. While these experiments were performed in the linear transport regime, a ballistic rectifier operates in the nonlinear regime. Furthermore, a triangular scatterer was introduced to break the device symmetry which, as shown below, plays a key role in generating the rectifying effect. Figure 4(a) is an atomic force microscope image of the central part of one of the first ballistic rectifiers [34]. The dark areas in Figure 4(a) are the regions that were etched away and became insulating for the electrons. As a result, the triangular antidot was defined in the cross junction, which consists of the two narrow channels (labeled "source" S and "drain" D) and the two wide channels (labeled "upper" U and "lower" L). The triangular antidot acted as an artificial scatterer which, as shown in the following, induced a pronounced nonlinear effect.

The modulation-doped GaAs–AlGaAs heterostructure, from which the ballistic rectifier was fabricated, contains a two-dimensional electron gas (2DEG) located 37 nm below the wafer surface. The sheet density of the 2DEG is about 5×10^{11} cm^{-2} and the mobility is about 5×10^5 cm^2/V · s at a temperature $T = 4.2$ K. Standard electron-beam lithography and wet chemical etching are used in the fabrication. Since the electron mean-free path $l_e = 5.8$ μm is longer than the central part of the device, the electron transport is ballistic at 4.2 K.

Like a typical four-terminal measurement, the (input) current is applied through leads S and D, and the output voltage is detected via L and U. The typical electron trajectories, illustrated by the arrows in Figure 4(a), suggest an accumulation of electrons in the lower lead, and hence a negative voltage between L and U. Obviously, this cannot happen when the applied voltage or current is zero since the same number of electrons will go along the opposite directions of the arrows. For a nonzero applied current, $I_{SD} \neq 0$; however, as shown by a detailed model [47, 56], the electron transmission along the arrows in Figure 4(a) is changed by the applied electric field, while the transmission in the reversed direction is virtually unaffected (details in Section 2.2). This gives rise to a finite negative voltage between L and U, V_{LU}.

Even though the mirror symmetry along the S–D axis is broken by the triangular antidot, the symmetry along the central L–U axis remains. As a consequence, the output voltage V_{LU} is expected to remain the same if the sign of the applied current I_{SD} is changed, that is,

$$V_{LU}(I_{SD}) = +V_{LU}(-I_{SD}). \quad (4)$$

Correspondingly, $R_{SD,LU}(I_{SD}) = -R_{SD,LU}(-I_{SD})$, where $R_{SD,LU} = V_{LU}/I_{SD}$ is the four-terminal resistance. Such a picture is almost perfectly supported by the experimental V_{LU} versus I_{SD} curves at $T = 4.2$ K (solid line) and 77 K (dotted line) shown in Figure 5. The slight deviation from Eq. (4) is attributed to the unintentional breaking of the desired symmetry along the L–U axis by the imperfection of the fabrication. The imperfection is expected to contribute mainly a linear term to the V_{LU} versus I_{SD} curves, in a way similar to a change of one of the four resistors in an otherwise balanced resistor bridge.

The experimental results demonstrate that new device functionalities can be realized by simply tailoring the shape of a ballistic device because the only scattering events that the electrons undergo are from the device boundary. The device shown in Figure 4, called a ballistic rectifier, operated similarly to a bridge rectifier [see Fig. 4(b)], but the working principle is completely different. First, only a single device is used here rather than four diodes in the bridge rectifier. The mechanism is also entirely different since no doping junction or barrier structure along the current direction is used in the ballistic rectifier. The pronounced nonlinear effect comes from the breaking symmetry and the properties of the ballistic electron transport. Furthermore, as will be shown in Section 2.2, the ballistic rectifier has no intrinsic voltage or current threshold, and can be used for the detection of very weak signals without the need of an external bias.

The nonlinearity in the I–V characteristics is remarkably different from that of a two-terminal nonlinear device, such as a diode, where the nonlinearity always appears on top of a large linear term. For the ballistic rectifier with the four-terminal geometry, there can be, however, no linear term present in the relationship between the output voltage and the applied current, as shown by Eq. (4). The striking nonlinearity is a result of the broken symmetry along the S–D axis and the mirror symmetry along the U–L axis.

As the temperature was increased from 4.2 K, the electron scattering from phonons increased, resulting in a lower

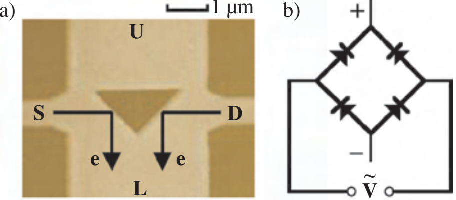


Figure 4. (a) Atomic force microscope image of the central part of a ballistic rectifier. The dark areas were etched away, forming the triangular antidot in the middle and the four channels S, D, L, and U. The current is applied via S and D, and the output voltage is measured between L and U. The device operates similarly to a bridge rectifier (b), but the ballistic rectifier is based on a completely new working principle. Adapted from [34].

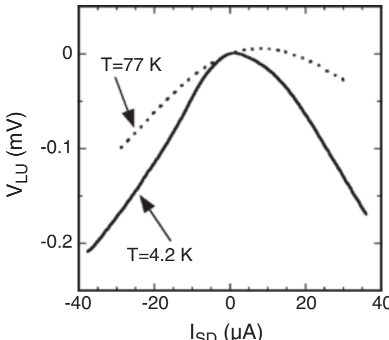


Figure 5. The output voltage between the lower and upper leads V_{LU} as a function of the input current through the source and drain leads I_{SD} measured at $T = 4.2$ K (solid line) and 77 K (dotted line). Ideally, the curves should be perfectly symmetric with respect to the vertical axis at $I_{SD} = 0$. However, the imperfection in the fabrication caused the slight asymmetry in the curves. Adapted from [34].

electron mobility and a shorter mean-free path. At $T = 77$ K, a weaker rectification than that at 4.2 K was observed, as shown by the dotted line in Figure 1(a). The less pronounced nonlinear behavior was attributed to the shorter l_e (about 1–2 μm) at 77 K. Since l_e was comparable to the distance between the openings of the *S* or *D* channels and the antidot, the electron transport was only partially ballistic.

To function at room temperature, a device size smaller than l_e (about 100 nm) at $T = 300$ K is required. It is very difficult to use wet chemical or dry etching to fabricate such small devices using a GaAs–AlGaAs heterostructure, due to the large depletion length close to etched device boundaries. The much shorter depletion length of the InGaAs–InP material system makes it possible to fabricate much smaller devices that work at room temperature [57]. The wafer that was used to fabricate the room-temperature ballistic rectifiers was a modulation-doped $In_{0.75}Ga_{0.25}As$–InP quantum-well structure [58] grown by a metalorganic vapor-phase epitaxy at a pressure of 50 mbar and a temperature of 600 °C. Si was used as the dopant because of its slow carryover effect. The structure consists of a semi-insulating (001) InP substrate, a 250 nm, not intentionally doped buffer layer, a 9 nm thick $In_{0.75}Ga_{0.25}As$ quantum well, a 10 nm thick, undoped InP spacer layer, and a 35 nm thick InP layer doped by Si to a level of about 6×10^{17} m^{-3}. On top of the doped layer, an additional 10 nm thick, undoped InP layer was grown. At room temperature, the carrier density of the two-dimensional electron gas in the $In_{0.75}Ga_{0.25}As$ quantum well is 4.7×10^{15} m^{-2} and the mobility 1.2 m^2/V · s, corresponding to $l_e = 140$ nm. Figure 6 shows the scanning electron microscope images of two room-temperature ballistic rectifiers with different sizes. Since l_e is only slightly shorter than the distance from the *S* or *D* channel to the triangular antidot, a large number of electrons can still travel through this distance without being scattered. As shown by Hirayama and Tarucha [23], if the electron transport is partially ballistic, the ballistic effect may still be observed, although weaker than that at low temperatures. Here, instead of a dc current, an ac voltage (1 kHz) was applied to the device, the dc (or average) voltage between the lower and upper terminals was measured. Indeed, a rectification efficiency of about 14% at room temperature, roughly half of that at 100 K, was observed in the smallest ballistic rectifier, as shown in Figure 7. This is one of the very few types of novel nanodevices that have so far been shown to work at room temperature.

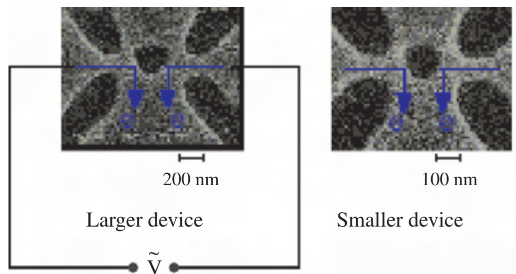


Figure 6. Scanning electron microscope images of two room-temperature ballistic rectifiers with different sizes, fabricated from an InGaAs–InP heterostructure. Reprinted with permission from [57], A. M. Song et al., *Jpn. J. Appl. Phys.* 40, L909 (2001). © 2001, Japan Society of Applied Physics.

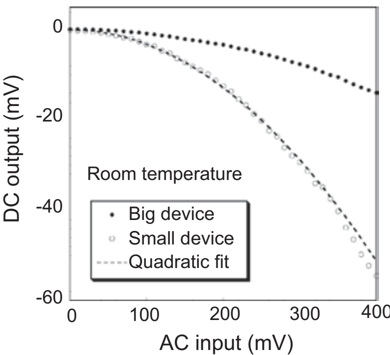


Figure 7. Room-temperature operation of the ballistic rectifiers shown in Figure 6. The frequency of the input ac signal is 1 kHz. The dashed line represents a quadratic fit to the experimental result of the smaller rectifier. Reprinted with permission from [57], A. M. Song et al., *Jpn. J. Appl. Phys.* 40, L909 (2001). © 2001, Japan Society of Applied Physics.

Apart from meeting the criterion of functioning at room temperature, for possible real applications, any new type of nanodevice is desirable to function at gigahertz frequencies. To investigate the speed of the ballistic rectifiers, high-frequency experiments up to 50 GHz have been performed. Because of the in-plane nature of the ballistic rectifier—the electrical contacts are laterally separated rather than placed on the surface and the back side (substrate)—the parasitic capacitance between contacts is substantially lower than in a conventional vertical device of the same size. Furthermore, the new working mechanism does not rely on any minority carrier diffusion or barrier structure, two factors that often limit the speed of conventional semiconductor diodes. The ballistic rectifier was therefore expected to function at very high frequencies. Figure 8 shows the dc output of the device versus the power of 50 GHz signal at room temperature. The 50 GHz signal was transmitted to the device via a coplanar probe. The power (horizontal axis) in Figure 8 was the output power from the signal source. Due to the impedance

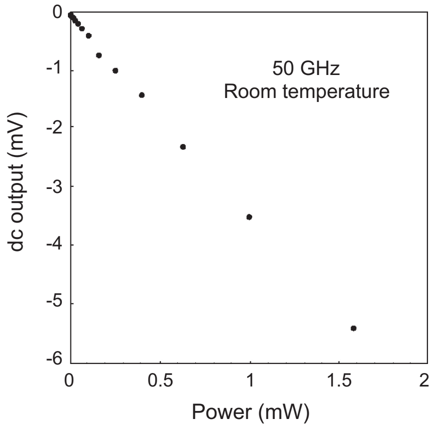


Figure 8. Output dc voltage of the smaller ballistic rectifier shown in Figure 6(b) versus the power of the applied 50 GHz signal at room temperature. Reprinted with permission from [57], A. M. Song et al., *Jpn. J. Appl. Phys.* 40, L909 (2001). © 2001, Japan Society of Applied Physics.

mismatch between the signal source (50 Ω) and the device (a few kiloohms), as well as some power loss due to the cables at 50 GHz, effectively only a small fraction of the power from the signal source was actually applied to the device. It was estimated that the real sensitivity of the device at 50 GHz was a few hundred millivolts of dc output per milliwatt input, which is virtually as high as that of a commercial microwave detection diode [57]. The ballistic rectifiers were, however, not specifically optimized. Much improvement can be expected in the future, by, for example, further reducing the device size. For a device about half of the one shown in Figure 7, one can expect not only a much higher efficiency, but also a stable performance that is insensitive to temperature changes up to room temperature.

As will be shown in the next section, theoretical studies on the ballistic rectifier yielded a quadratic response to input signals, which is very different from the exponential current–voltage characteristic of a conventional semiconductor diode. In Figure 7, a fit of $y = ax^2$ (where y represents the dc output voltage, x the ac input voltage, and a a constant) was performed. The remarkably good fit supports the theoretical prediction of the quadratic response to input voltage, Eq. (18), which will be derived in Section 2.2. This unique characteristic allows for generating second-harmonic signals without producing third or higher harmonics. The polarization of the generated second-harmonic signal (along the vertical direction) is perpendicular to that of the input signal (horizontal direction), which makes the separation of the second harmonic from the input signal much easier. Because of the quadratic characteristic, one also expects that the dc output is a linear function of the applied microwave power, which is demonstrated in Figure 8. Although we have not been able to test the devices at frequencies higher than 50 GHz so far, it is expected, from the working principle, that the cutoff frequency will be much higher, possibly up to hundreds of gigahertz or even in the terahertz regime. Recent numerical simulations confirmed this expectation, and showed excellent performance of the ballistic rectifier as a frequency doubler or power detector in the terahertz regime [59].

2.2. Semiclassical Model

As discussed in Section 1, Ohm's law does not hold in ballistic devices. Instead, the Büttiker–Landauer formalism [5–7], which treats transport in mesoscopic conductors as a transmission problem for carriers at the Fermi level, is widely used to describe the linear transport behavior of these conductors [13–15, 17–20]. For a multiterminal ballistic conductor at $T = 0$, the current in lead α in the linear transport regime can be written as

$$I_\alpha = \frac{2e}{h} \sum_{\beta \neq \alpha} T_{\beta \leftarrow \alpha}(\mu_\alpha - \mu_\beta) \tag{5}$$

Here, I_α is the current through lead α, and μ_α is the chemical potential of the electrons in contact (carrier reservoir) α. Such a standard formula is for linear electron transport only because the transmission coefficient $T_{\beta \leftarrow \alpha}$ for electrons from terminal (contact) α to terminal β is not considered a function of the applied voltage. To model the nonlinear (rectifying) effect observed in the ballistic rectifiers and the artificial nonlinear materials to be discussed in Section 3, the standard Büttiker–Landauer formula has been extended to the nonlinear regime. By doing so, a detailed model has been obtained, which provides analytical descriptions of the ballistic rectification effect in the framework of semiclassical ballistic electron transport [47].

Consider a general case where the electron transport is in the nonlinear regime, at a finite temperature, and in a finite magnetic field. The current through a lead α of a mesoscopic conductor, which is connected via perfect leads to a number of carrier reservoirs, can be written as

$$I_\alpha = \frac{2e}{h} \sum_{\beta \neq \alpha} \int [f(E-\mu_\alpha) - f(E-\mu_\beta)] \times T_{\beta \leftarrow \alpha}(E, B)\, dE \tag{6}$$

Here, $f(E-\mu_\alpha) = [\exp((E-\mu_\alpha)/k_B T) + 1]^{-1}$ is the Fermi–Dirac distribution function, and in general, the transmission coefficient for carriers from lead α to lead β, $T_{\beta \leftarrow \alpha}(E, B)$, is a function of the electron energy E and magnetic field B. At $k_B T = 0$ and $B = 0$, Eq. (6) becomes

$$I_\alpha = \frac{2e}{h} \sum_{\beta \neq \alpha} \overline{T}_{[\beta,\alpha]}(\mu_\alpha - \mu_\beta) \tag{7}$$

Here, $\overline{T}_{[\beta,\alpha]}$ equals $\int_{\mu_\beta}^{\mu_\alpha} T_{\beta \leftarrow \alpha}(E)dE/(\mu_\alpha - \mu_\beta)$ if $\mu_\alpha > \mu_\beta$, and $\int_{\mu_\alpha}^{\mu_\beta} T_{\alpha \leftarrow \beta}(E)dE/(\mu_\beta - \mu_\alpha)$ otherwise. This shows that only the transmissions from the reservoirs with higher chemical potentials to the reservoirs with lower chemical potentials need to be considered. Based on Eq. (7), the ballistic rectifying effect can be modeled in the framework of semiclassical ballistic electron transport.

Similar to the model for ballistic electron transport in a cross junction by Beenakker and van Houten [60], only the angular distribution of the ballistic electrons ejected from S and D was taken into account to determine the electron transmission probabilities into L or U leads. This is a particularly good approximation for this specific device geometry because the transmission probabilities for electrons from S (or D) to L and U are virtually completely determined by the ejection angles of the electrons.

First, consider the device at equilibrium, that is, when no voltage is applied. The angular distribution of the electrons ejected from a narrow constriction, for example, of the S and D channels, is given by

$$P(\theta) = \frac{1}{2}\cos\theta \tag{8}$$

Here, θ is the ejection angle with respect to the channel direction. Therefore, even at $V_{SD} = 0$, the electrons ejected out of S are, to some extent, collimated, as shown in Figure 9(c). This is actually a property of ballistic electrons, caused by the adiabatic electron transport at sufficiently low temperatures and in channels with smooth boundaries at the entrances and exits [14, 20, 61, 62].

Let θ_0 be the minimum ejection angle for an electron from S or D to be scattered by the triangular antidot to the lower lead. For the specific device in Figure 4, which is to

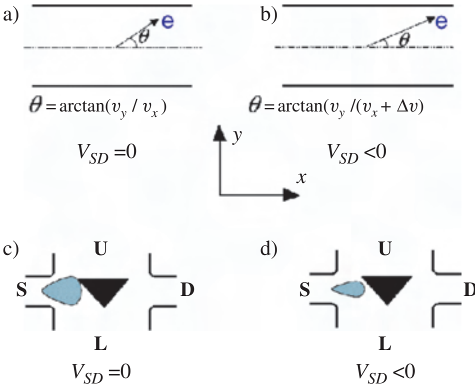


Figure 9. (a) The sketch shows that, at $I_{SD} = 0$, an electron is moving in the source channel with the angle $\arctan(v_y/v_x)$, where v_x and v_y are the velocity components along the direction of the channel and in the perpendicular direction. (b) When a negative voltage V_{SD} is applied, $I_{SD} < 0$, corresponding to a finite drift velocity Δv of the electrons in the source channel. This leads to a smaller ejection angle of the electron $\arctan[v_y/(v_x + \Delta v)]$. Overall, this results in a collimation effect induced by the applied voltage, and the angular distribution of the electron ejected out of S at $V_{SD} = 0$ shown in (c) becomes narrower at $V_{SD} < 0$, as schematically shown in (d). Consequently, the electrons more likely will be scattered by the triangular antidot to the lower lead, and have less chance to go through the gap between the antidot and the edge of the upper channel. This results in an accumulation of electrons in the lower lead, and consequently a negative voltage V_{LU}.

be modeled below as an example, $\theta_0 \approx \pi/4$. From Eq. (8), it is easy to obtain the transmission coefficients at zero bias:

$$T_{L\leftarrow S} = T_{L\leftarrow D} = T_{D\leftarrow L} = T_{S\leftarrow L} = N_{SD}\int_{-\pi/2}^{\theta_0} P(\theta)\, d\theta = \frac{N_{SD}(1+\sin\theta_0)}{2} \tag{9a}$$

$$T_{U\leftarrow S} = T_{U\leftarrow D} = T_{D\leftarrow U} = T_{S\leftarrow U} = N_{SD}\int_{\theta_0}^{\pi/2} P(\theta)\, d\theta = \frac{N_{SD}(1-\sin\theta_0)}{2} \tag{9b}$$

Here, N_{SD} is the number of occupied transverse quantum confinement modes at the Fermi level E_F in channels S and D.

To obtain an analytical description of the ballistic rectifying effect, certain approximations have to be taken. Ideally, the potential profile in the device at a finite I_{SD} should be calculated self-consistently in order to obtain the angular distribution of the ballistic electrons, and hence the transmission coefficients. Such a calculation is, however, very complex. Instead, it is noticed that the applied voltage mainly drops at the entrances and exists of the S and D channels [4]. Consider that, at $V_{SD} = 0$, the velocity components of a right-moving electron in the source channel are v_x and v_y, respectively, in the direction of the channel and in the perpendicular direction, as illustrated in Figure 9(a). The angle of ejection $\theta(I_{SD} = 0) = \arctan[v_y/v_x]$ will change if a finite V_{SD} is applied. For example, if V_{SD} is negative, the current in the source channel $I_S < 0$, meaning that the electrons in the source channel have a finite mean velocity or drift velocity, $\Delta v > 0$. As a good approximation, assume that the velocity component in the x direction of all of the electrons in channel S is increased by Δv, while the velocity component in the perpendicular direction v_y is essentially not affected. The ejection angle of the electron changes from $\arctan[v_y/v_x]$ at $I_{SD} = 0$ to a smaller angle $\arctan[v_y/(v_x + \Delta v)]$, as shown in Figure 9(b). Overall, this leads to a collimation effect induced by the applied voltage, which effectively narrows the angular distribution of the electron ejected out of S at $V_{SD} < 0$, as schematically shown in Figure 9(d).

The changes of the transmission coefficients, which fundamentally determine the output V_{LU}, can be directly evaluated once the angular distribution of the electrons is known. Since the angular distribution of the electrons ejected out of S becomes narrower, the electrons will have a greater chance to be scattered into the lower lead by the triangular antidot, and therefore the transmission probability for the ejected electrons to transmit from S to L will increase. In the mean time, the transmission probability from S to U will decrease since the probability for the electrons to have a large enough ejection angle to transmit through the gap between the antidot and the edge of the upper channel will reduce. The above analysis shows that the transmission probabilities of electrons could be directly linked to the electron drift velocity, and hence I_{SD}. It can be shown that

$$T_{L\leftarrow S}(I_{SD} < 0) - T_{L\leftarrow S}(I_{SD} = 0) = \frac{N_{SD}}{2}(\sin\theta_e - \sin\theta_0) \tag{10a}$$

$$T_{U\leftarrow S}(I_{SD} < 0) - T_{U\leftarrow S}(I_{SD} = 0) = -\frac{N_{SD}}{2}(\sin\theta_e - \sin\theta_0) \tag{10b}$$

In the equations, θ_e is determined by

$$\theta_e = \theta_0 + \arcsin\left[(\Delta v/v_F)\sin\theta_0\right] \tag{11}$$

where v_F is the Fermi velocity [47]. Assuming a hard wall potential for the source and drain channels because of the relatively large channel width, it can be shown that

$$\frac{\Delta v}{v_F} = -\frac{\hbar}{N_{SD} e E_F} I_{SD} \tag{12}$$

Here, the changes in the number of occupied modes in the source and drain channels with applied bias are neglected. This is, however, a good approximation for the four-terminal measurements of the ballistic rectifier. As will be shown by the following analysis, the output of a ballistic rectifier is determined by the relative changes of the transmission coefficients $\Delta T/T$ rather than the absolute values of the changes in the transmission coefficients ΔT.

Shown in Figure 10 is the dependence of the transmission coefficients on the electric current of the lead from which the electrons are ejected. Therefore, Eq. (7) can then be written as

$$I_\alpha \approx \frac{2e}{h}\sum_{\beta\neq\alpha} \overline{T}_{[\beta,\alpha]}(I_{[\beta,\alpha]})(\mu_\alpha - \mu_\beta) \tag{13}$$

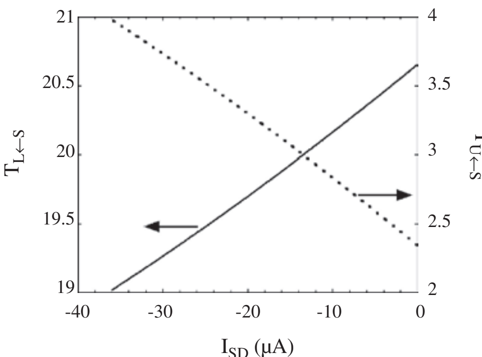


Figure 10. Calculated transmission coefficients $T_{L\leftarrow S}$ and $T_{U\leftarrow S}$ as a function of the negative source–drain current at $T = 0$. Adapted from [47].

where $\overline{T}_{[\beta,\alpha]}(I_{\{\beta,\alpha\}})$ equals $\overline{T}_{\beta\leftarrow\alpha}(I_\alpha)$ if $\mu_\alpha > \mu_\beta$ and $\overline{T}_{\alpha\leftarrow\beta}(I_\beta)$ otherwise [47]. With the values of the transmission coefficients determined, the V_{LU} versus I_{SD} characteristic is obtained by calculating the four-terminal resistance $R_{SD,LU} = V_{LU}/I_{SD}$. For the case of $V_{SD} < 0$,

$$R_{SD,LU} = \frac{h[T_{L\leftarrow S}(I_S)T_{D\leftarrow U}(I_U) - T_{D\leftarrow L}(I_L)T_{U\leftarrow S}(I_S)]}{2e^2 D} = \frac{h[T_{L\leftarrow S}(I_{SD})T_{D\leftarrow U}(0) - T_{D\leftarrow L}(0)T_{U\leftarrow S}(I_{SD})]}{2e^2 D} \quad (14)$$

where $I_S = I_{SD}$ and $I_L = I_U = 0$. Therefore, because the net current in L and U is zero, $T_{D\leftarrow L}$ and $T_{D\leftarrow U}$ are not sensitive to the applied current I_{SD}, and their values at zero bias are employed. In Eq. (14), D is a function of the individual transmission coefficients, and is found to be insensitive to the applied voltage or current [63]. For this specific device, $\theta_0 \approx \pi/4$, $T_{S\leftarrow D} = T_{D\leftarrow S} = 0$, and $T_{L\leftarrow U} \approx N_{LU}/3$, where N_{LU} is the number of occupied transverse quantum confinement modes in channels L and U. With these relations and taking into account $N_{LU} \gg N_{SD}$, it can be shown that

$$D = N_{SD}[2T_{L\leftarrow S}T_{U\leftarrow S} + T_{L\leftarrow U}N_{SD}] \approx \frac{N_{SD}^2 N_{LU}}{3} \quad (15)$$

Substituting Eqs. (9), (10), and (15) into Eq. (14) yields

$$V_{LU} = R_{SD,LU}I_{SD} = \frac{3h}{4e^2}\frac{\sin\theta_e - \sin\theta_0}{N_{LU}}I_{SD} < 0 \quad (16)$$

Note that θ_e is a function of I_{SD}, as shown by Eqs. (11) and (12). Combining the above with the case of $V_{SD} > 0$, where V_{LU} is also found to be negative, we have

$$V_{LU} = -\frac{3h}{4e^2}\frac{\sin\theta_e - \sin\theta_0}{N_{LU}}|I_{SD}| \quad (17)$$

The V_{LU} versus I_{SD} curve, calculated analytically for the device shown in Figure 4(a), is plotted with the solid line in Figure 11. It is in very good agreement with the experimental result (dashed line), especially in the low-current region. No adjustable parameters have been used at all in Eq. (17). At high currents, electron heating, not included in the above model, might reduce the rectifying effect. Moreover, the

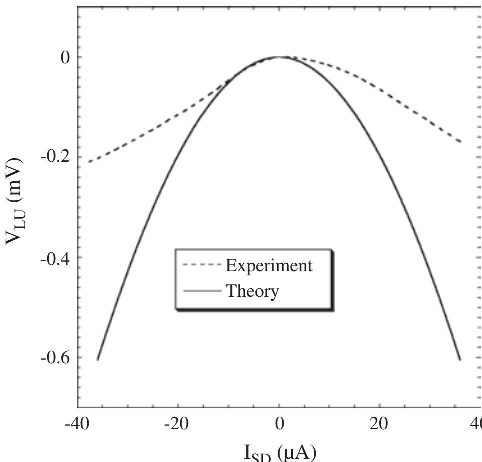


Figure 11. Theoretical (solid line) and experimental (dashed line, the same as that in Fig. 5) V_{LU} versus I_{SD} curves for the device shown in Figure 4(a). Although no fitting parameters have been used, the agreement between the experiment and the model is remarkably good. Reprinted with permission from [71], A. M. Song, *Appl. Phys. A* 75, 229 (2002). © 2002, Springer-Verlag.

length of the L and U channels (5 μm) is comparable to the electron mean-free path. This means that some electrons will be scattered by impurities or phonons, and change directions before they can travel through the channels and enter the electrical contact regions (carrier reservoirs). This certainly causes a reduced output.

The curves in Figure 10 display the dependence of $T_{L\leftarrow S}$ and $T_{U\leftarrow S}$ on I_{SD}. From $I_{SD} = 0$ to -37 μA, $T_{L\leftarrow S}$ increases by about 6.4%, while $T_{U\leftarrow S}$ decreases by about 34%. From Eq. (14), it is clear that, although the magnitude of the transmission from S to U is much less than that from S to L, it has a much stronger influence on $R_{SD,LU}$, and therefore contributes much more to the observed negative output V_{LU}.

It is also interesting to note that, in the limit as $I_{SD}| \to 0$, V_{LU} can be shown to have a quadratic response to I_{SD} [47]:

$$V_{LU} = -\frac{3\pi\hbar^2}{4e^3 E_F N_{SD} N_{LU}} I_{SD}^2 \quad (18)$$

Although the modeling was performed under the condition of $T = 0$, it has been shown that Eq. (18) holds as well at a finite temperature for sufficiently small currents [64], meaning that there is no threshold or turn-on voltage for the ballistic rectifier. With respect to applications, it is interesting to note that there are a number of practical advantages associated with the quadratic response, as have been introduced in Section 2.1 and will be discussed in Section 3. Based on the above picture, the nonlinear ballistic electron transport at finite magnetic fields can also be well described [65].

3. ARTIFICIAL FUNCTIONAL NANOMATERIALS

The concept of the ballistic rectifier has been extended from a single device to a new type of artificial functional material [66]. Figure 12 shows the atomic force microscope images

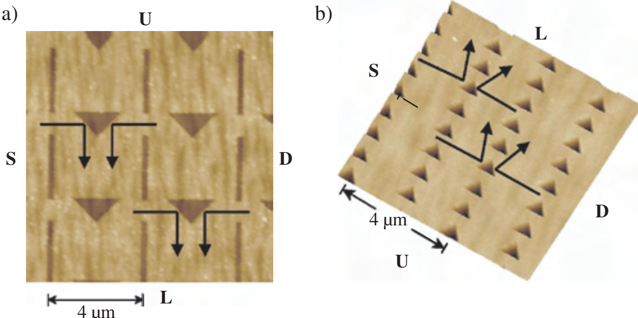


Figure 12. Atomic force microscope images of artificial functional materials of two different designs, fabricated using a GaAs–AlGaAs heterostructure. Note that the L terminal is on the upper-right side in (b). The arrows indicate the typical electron trajectories in the artificial materials.

of two examples of the artificial material that have been made. The working principle of the nonlinear materials is similar to that of the ballistic rectifier, and the materials were also fabricated using standard electron-beam lithography and wet chemical etching. In Figure 12(a), the artificial material was constructed by simply duplicating a ballistic rectifier over space. The duplicated individual rectifiers were connected in series and parallel, forming a network or an "artificial lattice." Apart from the above design, a material shown in Figure 12(b) was also fabricated, which only consisted of triangular antidots. In the material, not only is the intentionally broken symmetry by the triangular antidots important, but also the introduced offset between neighboring columns: the triangles were placed in such a coherent way that neighboring triangles support each other by simultaneously scattering electrons and forming channels through which electrons were ejected, as shown by the arrows in Figure 12(b).

In principle, these artificial materials can be made over a whole wafer. One apparent advantage is that the material is suitable for applications involving large signals, as a much higher voltage or current can be applied. Compared with the fabrication of a diode which involves multiple-step mask alignment, the production of the nanomaterial involves only one step of lithography. Therefore, the recently invented nanoimprint technology [67] can be used to mass produce these materials. It allows us to avoid the bottleneck of extremely long e-beam exposure time, a challenging problem in producing a large number of nanodevices efficiently and economically. With the nanoimprint technology, a master stamp (mold) with e-beam lithography can be used to replicate patterns over whole wafers of up to 6 in. in a matter of minutes [68]. Furthermore, for radio-frequency (gigahertz or even terahertz) applications, such as RF detection (the capability of the material to function at up to at least 50 GHz will be discussed later), it is often desirable to have a detector with a large active region in order to eliminate the need for building antennas. Moreover, as the nanomaterial is intrinsically nonlinear, individual devices can be made by simply cutting pieces from the material, according to the requirements for different applications.

The measurements on the artificial materials are very similar to those on the ballistic rectifier. An 80×80 μm^2 piece of the artificial material shown in Figure 12(b) has been measured at $T = 4.2$ K. To study how the nonlinear effect changes with the electron concentration or the electron Fermi wavelength, a layer of metal gate was evaporated on top of the artificial material after etching the triangular antidots. Negative gate voltages V_{GU} were applied between the gate and U to lower the carrier concentration of the 2DEG and increase the Fermi wavelength of the electrons. The lowered carrier concentration was directly indicated by the increase of the two-terminal source–drain resistance R_{SD}, as shown in Figure 13(c). From the figure, the pinch-off gate voltage is about -2.25 V. Figure 13(a) shows the V_{LU} versus I_{SD} curves measured at different gate voltages V_{GU} at $T = 4.2$ K. Similar to the experimental results of the ballistic rectifiers in Figure 5, the curves are not perfectly symmetric with respect to the vertical axis at $I_{SD} = 0$, which is a consequence of the imperfect device fabrication. The overall downward bending of the curves clearly indicates the expected rectifying effect. To better illustrate the influence of the gate voltage on the rectifying effect, the average output V_{LU} at $I_{SD} = +37$ μA and $I_{SD} = -37$ μA is plotted as a function of the applied gate voltage V_{GU} in Figure 13(b). The curve shows that the rectifying effect was enhanced dramatically by increasing the applied negative gate voltage. In particular, the average output increased rapidly when V_{GU} was approaching the pinch-off gate voltage of about -2.25 V.

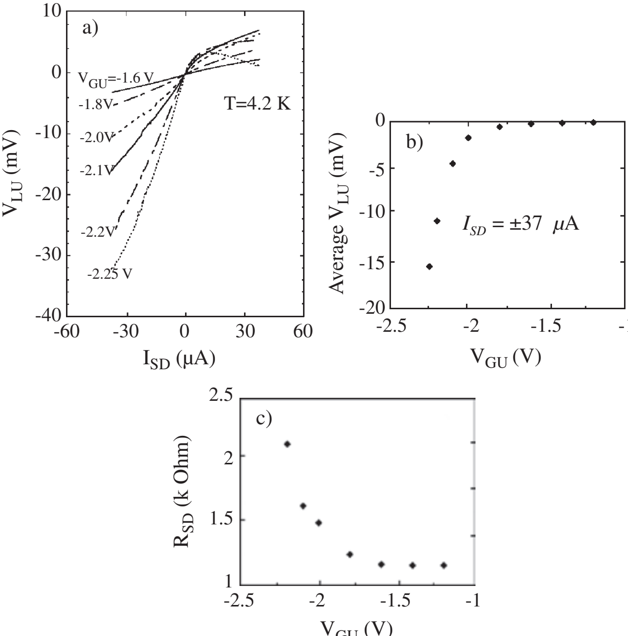


Figure 13. (a) V_{LU} versus I_{SD} curves at different voltages between the gate G and terminal U, V_{GU}, measured in an 80×80 μm^2 piece of the artificial material shown in Figure 12(b) at $T = 4.2$ K. (b) The average output voltage V_{LU} at $I_{SD} = +37$ μA and $I_{SD} = -37$ μA as a function of V_{GU}. (c) The two-terminal source–drain resistance R_{SD} as a function of V_{GU}.

The rectifying effect observed in the artificial material shown in Figure 13 can be qualitatively explained in the same physical picture. Similar to the voltage-induced collimation effect shown in Figure 9, the angular distribution of the electrons ejected out of the narrow gaps between neighboring triangular antidots at $V_{SD} = 0$ illustrated in Figure 14(a) becomes narrower at $V_{SD} < 0$, as schematically shown in Figure 14(b). The angular distribution is tilted upwards a little because of the asymmetry of the geometry [62, 69]. This leads to a higher probability for the electrons to be deflected downwards by the next triangular antidot on the side, as schematically shown in Figure 14(b), and fewer electrons will be able to transmit upwards. This induces an accumulation of electrons at the lower contact, and therefore a negative output voltage V_{LU}.

The pronounced gate–voltage dependence shown in Figure 13(a) and (b) is expected from the above model for the ballistic rectifier, although strictly speaking, Eq. (18) does not hold for the artificial material. However, since the working principle is the same, it could be used to qualitatively analyze the experimental results of the artificial material. For the material shown in Figure 12(b), by decreasing the voltage on the top gate, the electron density, and therefore E_F, reduces. From Eq. (18), the output voltage $|V_{LU}|$ will increase. Moreover, the lowered E_F also means less electron occupation of the transverse confinement modes in the gaps between neighboring triangular antidots. This can be regarded as reducing N_{SD} in Eq. (18), which also contributes to the dramatic increase of $|V_{LU}|$ in Figure 13.

Both of the artificial materials shown in Figure 12(a) and (b) could work only at temperatures below 77 K because of the large "lattice constants." It is possible to achieve room-temperature operations of the artificial material by fabricating nanomaterials with artificial lattice constants comparable to, or shorter than, the electron mean-free path [70]. Figure 15(a) is an atomic force microscope image of one of the nanomaterials that operates at room temperature, fabricated from the same InGaAs–InP heterostructure that was used to fabricate the room-temperature ballistic rectifiers. The lattice constants of the nanomaterial are $a_x =$ 800 nm and $a_y = 300$ nm in the lateral and vertical directions, respectively. Both the base and the height of the triangular antidots are 150 nm. The experimental data at room temperature are shown in Figure 15(b), in which the dc output between contacts L and U was measured as a function of the applied low-frequency (1 kHz) ac signal. The experiment

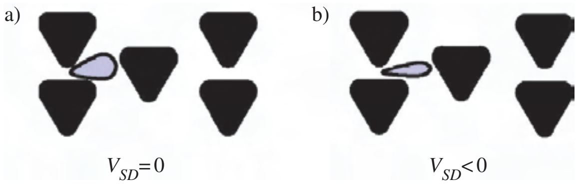


Figure 14. Similarly to the voltage-induced collimation effect shown in Figure 9, the angular distribution of the electrons ejected out of the narrow gaps between neighboring triangular antidots at $V_{SD} = 0$ (a) becomes narrower at $V_{SD} < 0$, as schematically shown in (b). The angular distribution is tilted upwards a little because of the asymmetry of the geometry.

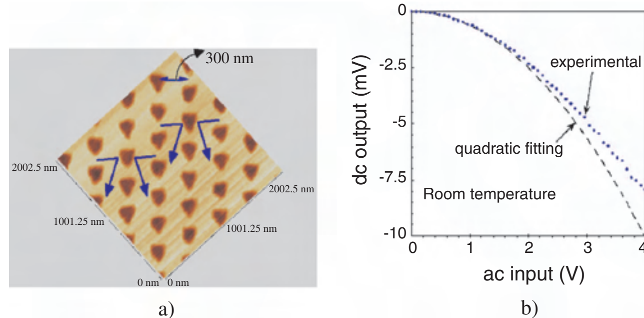


Figure 15. (a) Atomic force microscope image of one of the nanomaterials that operates at room temperature. The arrows represent typical electron trajectories. The lattice constants of the "artificial lattice" are $a_x = 800$ nm and $a_y = 300$ nm in the lateral and vertical directions, respectively. Both the base and the height of the triangular antidots are 150 nm. (b) Experimental data of a 30 μm × 30 μm piece of the nanomaterial shown in (a) at room temperature, in which the dc output between the lower and upper contacts was measured as a function of the applied low-frequency (1 kHz) ac signal. The dashed line represents a quadratic fit. Reprinted with permission from [71], A. M. Song, *Appl. Phys. A* 75, 229 (2002). © 2002, Springer-Verlag.

was performed using a 30 μm × 30 μm piece of the nanomaterial [71]. Note that the feature size of the nanomaterial is still far from ideal, that is, both a_x and a_y would ideally be shorter than l_e. The room-temperature operation shown in Figure 15(b) may be greatly improved by fabricating nanomaterials with a smaller feature size using the state-of-the-art nanolithography.

Similar to a ballistic rectifier, the in-plane nature of the nanomaterial ensures that the parasitic capacitance between contacts is much lower than in a conventional vertical device of the same size. The nanomaterial thus has an extremely high working speed, as is evident in high-frequency experiments. The curve in Figure 16 shows the output dc voltage between the contacts L and U as a function of the RF power of a 50 GHz signal, which was transmitted to the

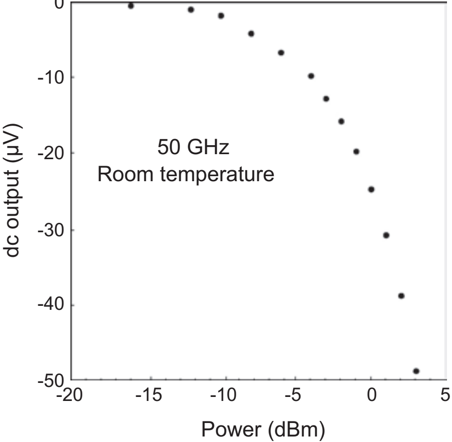


Figure 16. Output dc voltage of the nanomaterial versus the power of the applied 50 GHz signal at room temperature. The RF signal was applied to the material via coplanar probes. Reprinted with permission from [71], A. M. Song, *Appl. Phys. A* 75, 229 (2002). © 2002, Springer-Verlag.

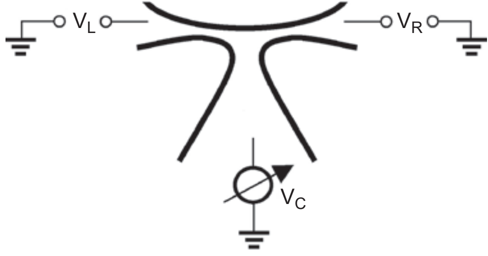
nanomaterial via coplanar probes. The ex periment was performed at room temperature. Similar results were obtained in other materials with different structural parameters.

Because of the experimental limitations, the material has not been tested at frequencies higher than 50 GHz. However, from the working principle, it is expected that the cutoff frequency will be much higher, possibly up to hundreds of gigahertz or even in the terahertz regime. The large area, in combination with the in-plane layout of the material, enables straightforward and efficient coupling to high-frequency microwaves. Moreover, microwaves traveling along the normal direction of the nanomaterial surface yield the most efficient coupling, while the coupling has to be achieved in specific ways (including using antennas) in a conventional vertical device. From the material point of view, this is the first artificial electronic nanomaterial that has been demonstrated to operate at room temperature.

The possibility of generating the dc output out of an external noise signal applied to the artificial material was also investigated. As expected, a guided net flow of electrons along the upward direction was induced, which generated a negative direct voltage between L and U [72]. In many aspects, such a ratchet effect in the artificial nanomaterial resembles the photovoltaic effect in some natural crystals [49–56, 66]. For the artificial crystals, however, one can design (adjust) the size and shape of the symmetry-breaking elemental units, as well as the lattice constants. Such flexibility may also provide a new route in further studies of the photovoltaic effect.

4. THREE-TERMINAL BALLISTIC JUNCTION DEVICES

The ballistic rectifiers and the artificial materials discussed in Sections 2 and 3 are four-terminal structures. Very recently, similar nonlinear functionalities were also discovered in another type of ballistic electron device which contained three terminals [38–40, 48]. Figure 17 schematically shows one such device. Referred to as a three-terminal ballistic junction (TBJ), Y-branch, or T-branch switch, the devices were fabricated by electron-beam lithography and chemical etching using either modulation-doped GaAs–AlGaAs heterostructures [40] or modulation-doped



A three-terminal ballistic junction

Figure 17. Sketch of a three-terminal ballistic junction device. The three branches are marked L (left), R (right), and C (center). In the experiment, a voltage V is applied to the left (L) and right (R) terminals in a push–pull fashion: $V_L = -V$ and $V_R = +V$. The output voltage V_C is measured at the lower terminal.

InGaAs–InP quantum-well wafers [38, 39]. In Figure 17, the three branches are marked L (left), R (right), and C (center). In the experiment, a voltage V is applied to the left (L) and right (R) terminals in a push–pull fashion: $V_L = -V$ and $V_R = +V$. The output voltage V_C is measured at the lower terminal. For simplicity, here we consider a symmetric TBJ device. If the electron transport is diffusive, that is, if the device size is much larger than the electron mean-free path, one would expect $V_C = 0$, regardless of the value of the input voltage. However, this turned out to be not true if the electron transport is ballistic, and V_C was found to be always negative as long as $V \neq 0$. For small applied voltages, experiments showed that V_C was a quadratic function of the applied voltage V, which is in good agreement with the theory developed by Xu [48].

In his model, Xu applied the scattering theory, and obtained an analytical expression of the nonlinear effect. The key mechanism was identified to be the nonlinear properties of each of the three narrow constrictions or point contacts: the conductance of each point contact is a function of the applied voltage (or chemical potential difference) across it.

To illustrate this, we first take a look at an individual, ideal (so-called) saddle-point contact, which is connected to two reservoirs (numbered 1 and 2) at both ends, and is spatially separated from other point contacts. The electrostatic potential of the saddle point contact is expressed as

$$v(x, y) = v_0 - \frac{1}{2}m^*\omega_x^2 x^2 + \frac{1}{2}m^*\omega_y^2 y^2 \tag{19}$$

where v_0 is the electrostatic potential at the saddle, x along the channel direction, and y in the transverse direction. The transverse quantum confinement energy levels are equally spaced by $\hbar\omega_y$. If the etched device boundary is smooth, so-called the adiabatic approximation can be applied. At zero temperature, the electrons that have energies higher than $v_0 + (1/2)\hbar\omega_y$ may pass through this single point contact without being reflected, while electrons with lower energies will have a zero possibility of passing the point contact (total reflection). At zero applied bias $V = 0$, this leads to a multiple-step dependence of the total transmission coefficient $T_{1\leftarrow 2, V=0} = T_{2\leftarrow 1, V=0} \equiv T_{12, V=0}$ on the chemical potential μ_F, with $T_{12, V=0} = 1, 2, 3, \ldots$, when μ_F is equal to or higher than $v_0 + (1/2)\hbar\omega_y$, $v_0 + (3/2)\hbar\omega_y$, $v_0 + (5/2)\hbar\omega_y, \ldots$, respectively. This actually results in the experimentally observed quantized conductance in a single quantum point contact [8, 9].

At a relatively high temperature where the thermal energy $k_B T$ is larger than the level spacing between the transverse quantum confinement modes, semiclassical approximations can be applied. The multiple-step dependence of the transmission coefficient on μ_F is smoothed out, and at zero bias voltage, T_0 becomes a linear function of μ_F:

$$T_{12, V=0} = \frac{\mu_F - v_0 + \frac{1}{2}\hbar\omega_y}{\hbar\omega_y} = \frac{\mu_F - v_0}{\hbar\omega_y} + \frac{1}{2} \tag{20}$$

as schematically shown in Figure 18.

Under the nonequilibrium condition (at a finite bias voltage V), the electrochemical potentials on both sides of the single point contact μ_1 and μ_2 become different, that is,

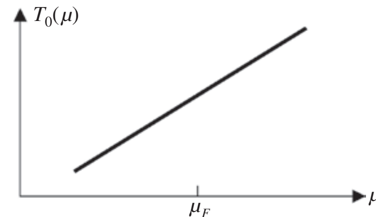


Figure 18. Linear relation between the transmission coefficients of a single, separated saddle-point contact and the chemical potential. This relation is obtained at relatively high temperatures, that is, the thermal energy k_BT is larger than the level spacing between the transverse quantum confinement modes, so that the semiclassical approximation can be applied.

$\mu_1 - \mu_2 = -eV$. Because of the linear dependence of the transmission coefficient on the chemical potential, the current can be expressed as

$$I = \frac{2e}{h}(\mu_1 - \mu_2)\overline{T_{12}} = -\frac{2e^2}{h}V\overline{T_{12}} \tag{21}$$

where $\overline{T_{12}}$ is the average transmission coefficient between μ_1 and μ_2:

$$\overline{T_{12}} = \frac{1}{2}\left[\frac{\mu_1 + \mu_2 - 2v_0}{\hbar\omega_y} + 1\right] \tag{22}$$

Clearly, if the applied voltage is dropped evenly on both sides of the point contact, which is a very good approximation of a single, symmetric point contact, then $\mu_1 = \mu_F - eV/2$ and $\mu_2 = \mu_F + eV/2$. Hence, such a simple model yields $\overline{T} = T_0$, and shows that the conductance is not sensitive to a change of the applied voltage. Note that this result is obtained under the conditions of a symmetric point contact, $|eV| < \mu_F$, and being at a relatively high temperature such that the thermal energy k_BT is larger than or comparable to the energy spacing between the transverse quantum confinement modes. Nevertheless, the above simplified model of a single point contact can be used to reveal the mechanism of the nonlinear effect of the TBJ devices, as will be discussed later.

For a TBJ device shown in Figure 17, the situation is not as simple because these three point contacts are closely connected. Even for a symmetric TBJ, as was observed experimentally, the voltage on branch C is not the average of the voltages applied to the left and right branches (which is zero in the push–pull fashion), indicating that the voltage drop in the whole device is neither even nor symmetric. Since a self-consistent calculation on the voltage drop in such an open system is very complex to perform, for simplicity, one may assume that the applied voltage is evenly dropped at the left and right terminals: $\mu_L = \mu_F + eV$, $\mu_R = \mu_F - eV$. Here, as was in the experiments, a push–pull bias fashion is assumed, and the total applied voltage is $\mu_L - \mu_R = 2\,eV$. Also, for simplicity, the potential at the saddle v_0 is assumed to be zero in all of these three point contacts, which is a good approximation if the channels are not very narrow.

If the terminal on the right is positively biased and the terminal on the left is negatively biased, that is, $V_L = -V$ and $V_R = +V$ (define such a bias polarity as $V > 0$), the net electron flows into the branch C are all from the terminal L, corresponding to a current

$$I_{C\leftarrow L} = \frac{2e}{h}(\mu_L - \mu_C)\left[\frac{\mu_L + \mu_C}{2\hbar\omega_{y(C,\,L)}} + \frac{1}{2}\right]t_{C\leftarrow L} \tag{23}$$

where $t_{C\leftarrow L}$ is the probability of the electrons transmitting from L to C. Here, $t_{C\leftarrow L}$ is assumed to be energy independent, which is a good approximation at a low applied voltage. $\hbar\omega_{y(C,\,L)}$ stands for the subband spacing of the transverse confinement modes in terminal L or in terminal C, of which $\hbar\omega_y$ is larger, and therefore of which the smaller number of occupied transverse confinement modes limits the current flow $I_{C\leftarrow L}$.

Because terminal C is a voltage probe, the net current flow in it is zero. Since $\mu_R < \mu_C, \mu_L$, the current by the electron flowing out of terminal C to R, $I_{R\leftarrow C}$, is equal to $I_{C\leftarrow L}$, and is given by

$$I_{R\leftarrow C} = \frac{2e^2}{h}(\mu_C - \mu_R)\left[\frac{\mu_C + \mu_R}{2\hbar\omega_{y(R,\,C)}} + \frac{1}{2}\right]t_{R\leftarrow C} \tag{24}$$

Similarly, here, $\hbar\omega_{y(R,\,C)}$ is the subband spacing of the transverse confinement modes in terminal C or terminal R, of which $\hbar\omega_y$ is larger, and therefore of which the smaller number of occupied transverse confinement modes limits the current flow I_{RC}.

The geometric symmetry of the device and the time-reversal symmetric lead to the relations $\hbar\omega_{y(C,\,L)} = \hbar\omega_{y(R,\,C)} \equiv \hbar\omega_{\max}$ and $t_{R\leftarrow C} = t_{L\leftarrow C} = t_{C\leftarrow L}$. Comparing Eqs. (23) and (24) yields very simple relations between the output voltage V_C, which is measured with respect to the electric ground, and the applied voltage:

$$V_C{}^2 - \frac{2\mu_F - \hbar\omega_{\max}}{e}V_C - V^2 = 0 \tag{25}$$

An analytical expression of the nonlinear effect in a symmetric TBJ device is obtained:

$$V_C = \frac{2\mu_F - \hbar\omega_{\max}}{2e} - \sqrt{\frac{(2\mu_F - \hbar\omega_{\max})^2}{4e^2} + V^2} < 0 \tag{26}$$

Similarly, it can be shown that the same result holds for $V < 0$ as shown in Figure 19. This explains the experimental discovery that the output voltage V_C is always negative. Remarkably, there are no geometric parameters of the device at all in Eq. (26) and the output voltage is only determined by the input voltage and the Fermi energy. The three-terminal ballistic junctions are therefore quite robust devices, and it is relatively easy to obtain a reproducible nonlinear behavior.

Note that, in the limit of a small applied voltage, $|V| \ll \mu_F/e$, the first term in Eq. (25) can be neglected because $|V_C| < |V| \ll \mu_F/e$. Therefore,

$$V_C = -\frac{e}{2\mu_F - \hbar\omega_{\max}}V^2 \approx -\frac{e}{2\mu_F}V^2 \tag{27}$$

by further assuming that $\mu_F \gg \hbar\omega_{\max}/2$, as derived by Xu in [48]. Equations (26) and (27) were obtained by assuming a parabolic confinement potential, Eq. (19), which is a good approximation for a narrow channel width. For

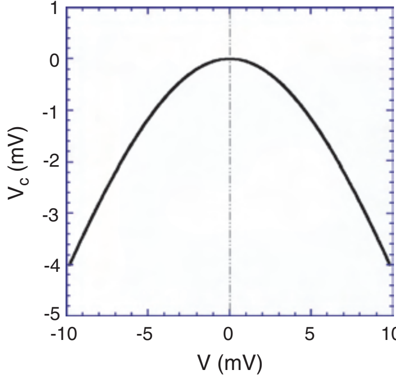


Figure 19. Theoretical results of the nonlinear output of a symmetric TBJ device. Typical parameters in the experiments were used in the calculation.

larger devices, however, a different V_C versus V relation is expected because of the different potential profiles in the device channels.

Similar to the ballistic rectifiers and nanomaterials discussed in Sections 2 and 3, TBJ devices are expected to be capable of high-frequency applications, allowed for by the ballistic transport nature and the low parasitic capacitance due to the planar device layout. This has been demonstrated in a number of experiments up to 20 GHz [73–75]. Recently, other interesting properties of the TBJ devices were predicted by Xu [76].

5. NEAR ROOM-TEMPERATURE OPERATIONS OF QUANTUM-EFFECT-BASED DEVICES

The above devices were demonstrated to function at room temperature, and the working principle is based on the semiclassical electron transport. Given the fact that the feature sizes of these room-temperture nanodevices are already comparable to the electron Fermi wavelength, it is expected to observe quantum effects not only at cryogenic, but also at elevated temperatures. In this section, we present the recent discovery of quantum effects associated with individual transverse confinement modes in some of the above devices at temperatures up to about 200 K. For the devices described in this chapter, the thermal energy k_BT at room temperature is comparable to or larger than the typical energy spacing of the transverse quantum confinement modes. It is therefore difficult to observe any quantum effect that is a direct manifestation of an individual transverse energy level at room temperature. However, it is worthwhile to point out that these devices reviewed in the previous sections are, even though not directly, fundamentally based on the wave nature of the electrons: the transverse quantum confinement levels in the narrow channels or point contacts. Otherwise, in the purely classical case, N_{SD} and N_{LU} in Eq. (18) would be infinity, leading to a zero output, even if the electron transport is ballistic. For the TBJ devices, it may appear that there was no evidence of the need for quantum transport in Eqs. (26) or (27). However, the derivations of these expressions are entirely based on Eq. (20), which actually represents the extent of the lateral quantum confinement in the point contacts. The nonlinear behavior is entirely based on the changes in the number of the occupied transverse quantum levels by the applied voltage.

Devices that display distinct quantum effects directly related to individual energy levels are expected to be ultrafast and sensitive because of the nature of the phase coherence of the electrons. A typical example of a superconductor counterpart is the superconducting quantum interference device (SQUID). Many types of semiconductor devices that directly manifest quantum effects have been realized at low (most often liquid helium) temperatures. However, it has not yet been possible to fabricate any such quantum semiconductor device that works at room temperature, largely due to the limitation in the device fabrication.

For the ballistic rectifier shown in Figure 6 and the nanomaterial shown in Figure 15(a), evidence of quantum behavior was recently observed at temperatures up to 200 K [72]. This suggests promising possibilities of realizing and operating quantum devices at room temperature in the near future by, for example, further reducing the device size. Figure 20(a) shows the experimental result of the ballistic rectifier shown in Figure 6(a), measured at a series of temperatures. Whereas, at room temperature, the negative output voltage V_{LU} was as expected in the physical picture of semiclassical ballistic transport, rather surprisingly, V_{LU} changed sign when the temperature was lowered to about 70 K. At even lower temperatures, the device generated a pronounced oscillatory output as a function of the

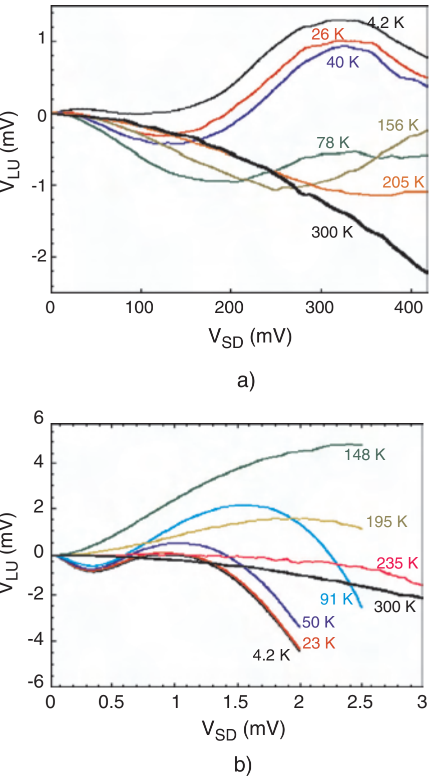


Figure 20. Normal (at room temperature) as well as reversed and oscillatory outputs (at low temperatures) of a small ballistic rectifier (a) and an artificial nanomaterial (b) as functions of the applied source–drain voltage.

applied voltage. Separate experiments showed that lower temperatures were needed to observe output reversal if the sample was illuminated by a light-emitting diode (LED) before the measurements. After the illumination, the electron concentration was increased by up to 30%, and the electron Fermi wavelength became shorter. A number of devices with different geometric parameters have also been measured, and the general trend is that, the larger the device, the lower the temperature at which the output reversal starts to be seen. Both facts suggest that quantum effects play an important role in causing the reversed and oscillatory output. This may be expected because the width of the source and drain channels, from which the ballistic electrons are ejected, is comparable to the electron Fermi wavelength (about 30–40 nm).

Recently, Fleischmann and Geisel developed a quantum model on a ballistic rectifier, and predicted reversed and oscillatory output [77–79]. The mechanism was attributed to the large difference between the channel widths of the upward and downward electron transmissions from S and D in the ballistic rectifier, that is, the gap between the triangular antidot and the edge of the upper channel is much narrower than the gap between the triangle and the edge of the lower channel. They argued that, when the applied voltage is small, the number of occupied modes of transverse quantum confinement in the (lower) wider gap may increase because of the applied voltage, while the number of occupied modes in the (upper) narrower gap remains constant due to the larger energy separation of the lateral confinement energy levels. This gives rise to the "normal" rectifying effect ($V_{LU} < 0$). If the applied voltage is high enough to open up a new mode in the upper gap, however, it was predicted that the output V_{LU} should undergo a change of sign, that is, output reversal.

While the result predicted by the model appeared to be in good agreement with the experimental data in Figure 20(a), such a mechanism for the output reversal does not seem to exist in the nanomaterials shown in Figure 15(a). This is because, in the nanomaterial, the width of the channel, which the electrons have to pass through to transmit upward after being ejected from a narrow gap between two neighboring triangles in a vertical column, is identical to the channel width of the downward electron transmission. As a result, the changes in the occupied lateral confinement modes in the channels of both the upward transmissions and the downward transmissions, induced by the applied voltage, are essentially identical. However, when measuring the nanomaterials with various geometric parameters, very similar reversed as well as oscillatory outputs were observed, which are unexpected from the model in [77]. This indicates that a different physical process is responsible for the observed oscillatory and reversed output.

Figure 20(b) shows the low-temperature experimental results of the artificial material, the geometric parameters of which are very similar to those shown in Figure 15(a). For this particular nanomaterial, the output reversal persists at up to about 200 K. Further observations, such as the dependence on device size and the influence of illuminations on the devices at low temperatures, also strongly suggest that the phenomena are a manifestation of quantum effects.

In Figure 20(a) and (b), the output is "normal" at room temperature, which has been understood in the framework of the classical ballistic transport in Section 2.2. To understand the output reversal and the oscillatory output, it was noticed that both the width of the source and drain channels in the ballistic rectifier and the width of the gaps between neighboring triangles in a vertical column in the nanomaterial are comparable to the electron Fermi wavelength, which is about 35 nm in the samples. By taking into account the finite depletion depth at the etched device boundaries, the real widths of the lateral quantum confinement are even narrower than those shown in the scanning electron micrographs. If temperature is lowered from room temperature, the electron transport may hence go from classical regime to the quantum transport regime.

Consider the ballistic rectifier first. At sufficiently low temperatures, the transverse modes in the source and drain channels become well resolved. From the width of the source and drain channels, it was estimated that there were between one and four occupied modes at low temperatures. In this estimate, the depletion depth, about 10–20 nm at each edge of the etched channels, has been taken into account. Similar to the case of a quantum point contact in the adiabatic transport regime, each mode contributes to a specific angular pattern of the electron flow from S. The number of lobes (or branches) in the pattern of electron flow corresponds to the number of maximums in $|\psi|^2$ (where ψ is the electron transverse wavefunction), as a direct result of the adiabatic transport. Therefore, for the Nth transverse mode, the number of lobes is N. Such branch-like patterns of electron flow were recently imaged using a low-temperature scanning probe microscope [80, 81]. For simplicity, assume that there is only one occupied mode in the source and drain channels at zero applied voltage, which corresponds to the angular pattern of the electron flow shown in Figure 21(a), but the following discussion applies to other (if not too large) numbers of initially occupied modes. If a small negative source–drain voltage is applied, the voltage-induced collimation effect leads to a narrower angular pattern of the electron flow out of S, which is similar to the case at higher temperatures shown in Figure 9(d). As a result, the downward transmissions of the electrons deflected by the triangular antidot increase, resulting in a normal negative output V_{LU}. If the applied negative voltage is decreased further, however, the right-moving electrons from S will eventually occupy the second lateral confinement mode [25]. The second mode gives rise to a two-branch angular distribution on top of the single-branch pattern of the first occupied mode. In total, there are three lobes of electron flow, as shown in Figure 21(c). The widening of the angular distribution clearly results in an increased probability (or percentage) for the electrons to transmit upward, and a reduced probability of the downward transmissions. As shown by a detailed model and analysis in [47] (see Eq. (6) and the discussions), the output of a ballistic rectifier is determined by the relative changes of the transmission coefficients $\Delta T/T$, rather than the absolute values of the changes ΔT. Therefore, the increased probability for the electrons to transmit upward and the reduced probability of the downward transmissions, which are discussed above and illustrated in Figure 21(c), will lead to an increase in V_{LU}, and may even

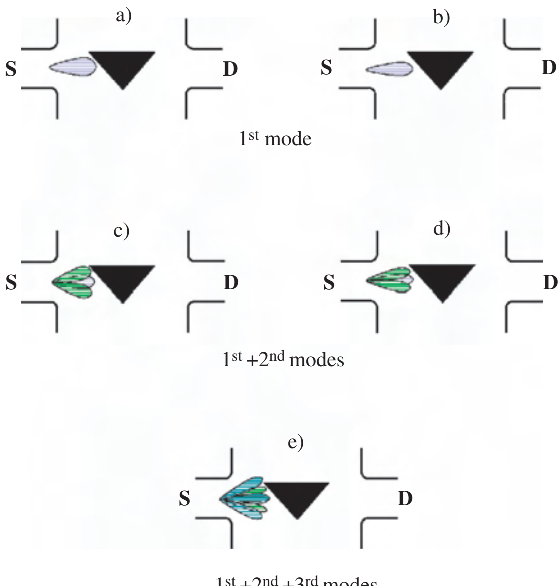


Figure 21. Schematic model for the output reversal and oscillatory output in the ballistic rectifier at low temperatures. As an example, suppose that there is only one lateral confinement mode occupied in the source channel when the applied voltage V_{SD} is zero. As a result of the adiabatic transport at low temperatures, the electrons ejected out of S have a narrow (rather than uniform) angular distribution (a). With decreasing V_{SD}, the angular distribution first becomes narrower due to the collimation effect induced by applied electric field, as shown in (b). By decreasing V_{SD} further (more negative), eventually the second lateral confinement mode becomes occupied, which gives rise to a two-branch angular distribution, on top of the single-branch pattern of electron flow of the first occupied mode. In total, there are three lobes in the electron flow (c). The total angular distribution will become narrower again with continuing decreasing V_{SD} because of the voltage-induced collimation effect (d), until at some point, the third transverse confinement mode starts to be occupied, giving rise to three additional lobes as shown in (e). Overall, by decreasing the applied voltage V_{SD} from zero, the angular distribution of the electrons ejected out of S changes alternately by subsequently narrowing (b), widening (c), narrowing (d), widening (e), and so on. Consequently, both the upward and downward transmission probabilities are oscillatory functions of V_{SD}, and this induces the oscillatory output V_{LU} which is shown in Figure 20.

cause output reversal (from negative to positive). If the negative applied voltage is decreased further (more negative), the applied voltage or electric field should again cause a narrowing of the overall angular pattern of the electron flow, as shown in Figure 21(d). The right-moving electrons hence have more and more chance to be scattered downward by the triangle on the right-hand side. This contributes again to a decrease in V_{LU}. A further decrease of the applied voltage will, at some point, cause the third transverse confinement mode to be occupied, giving rise to three additional lobes of electron flow, as shown in Figure 21(e), resulting in an increase in V_{LU}. Overall, by decreasing the applied voltage V_{SD} from zero, the angular distribution of the electrons ejected out of S changes alternately by subsequently narrowing (b), widening (c), narrowing (d), widening (e), and so on. Consequently, both the upward and downward transmission probabilities for the electrons ejected from S are oscillatory functions of V_{SD}, and this induces the oscillatory output.

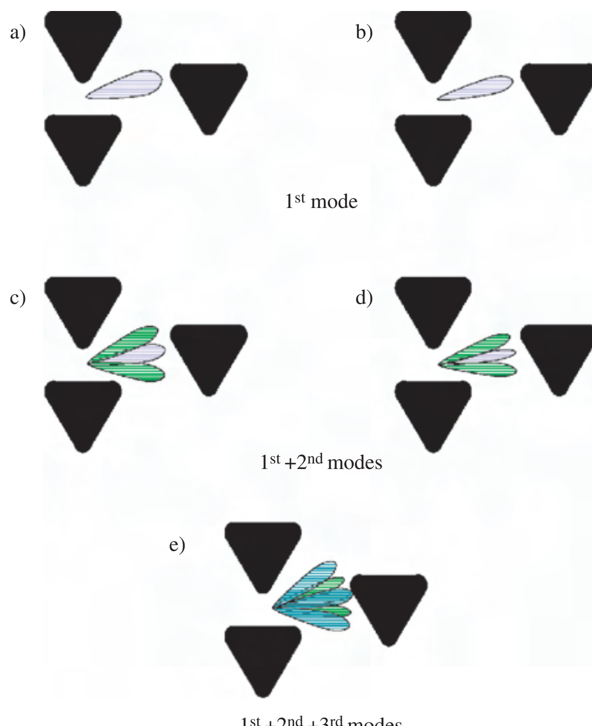


Figure 22. Schematic model for the output reversal and oscillatory output in the nanomaterial, which is similar to the case of the ballistic rectifier shown in Figure 21. For simplicity, only a small, representative part of the nanomaterial is sketched here. By decreasing the applied voltage to the nanomaterial V_{SD} from zero, the angular distribution of the electrons ejected right out of the gaps changes alternately by subsequently narrowing (b), widening (c), narrowing (d), widening (e), and so on because of the interplay between the voltage-induced collimation and the opening of additional lateral confinement modes in the gaps. As a result, the reversed and oscillatory output was observed in the experiments.

The reversed and oscillatory output in the nanomaterial can be explained very similarly, as sketched in Figure 22. At low temperatures, similar to the case of the ballistic rectifier, by reducing the applied voltage V_{SD} from zero (more negative), the angular distribution of the electrons ejected out of each gap changes alternately by narrowing (b), widening (c), narrowing (d), widening (e), and so on, as shown in Figure 22. This similarly results in an oscillatory modulation to the transmission probabilities for the electrons to be transmitted upward and downward, and hence leads to the oscillatory output V_{LU} in the nanomaterial, as shown in Figure 20(b).

The ballistic rectifier and the nanomaterials are still much larger than the smallest devices that can be fabricated with state-of-the-art nanolithography. By, for example, reducing the device size further, it is promising that these and other new types of room-temperature quantum devices would be realized in the near future, and lead to exciting applications.

6. CONCLUSIONS

Since the invention of the integrated circuit in 1959, the remarkable progress of the semiconductor industry has been described by Moore's law: every 12–18 months, the

integration density of integrated circuits doubles. While such a trend has been sustained for over three decades, it has been realized for many years that the miniaturization of conventional semiconductor devices will soon reach its limit, and the following generations of electronic devices will have to be built with nanoscience and technology, and be based on completely new concepts. From an application point of view, there is substantial industrial interest in new types of nanodevices that operate at room temperature, and at gigahertz or even terahertz frequencies. Despite great efforts in explorations of novel nanodevices that can be reproducibly fabricated for mass production, and that can function at room temperature, there has been very little success. One important reason is that, in order to work at room temperature, most electronic nanodevices proposed so far require a dimension of about a few nanometers, which is at the limit of even the best state-of-the-art facilities today. For example, although the first single-electron transistor (SET) was realized more than ten years ago [82], it is still very challenging to fabricate SETs that operate at room temperature [83–85]. Apart from the operation temperature, another key issue is the working speed, which is desired to be higher than 10 GHz for the next generation of integrated circuits. A great effort has been made to design SETs that operate at high speed [86].

The investigations on novel nanoelectronic devices reviewed above have provided working devices at room temperature, as well as gigahertz frequencies. Since the required device dimensions for ballistic transport are determined by the electron mean-free path of the host material, which is fairly long (about 150 nm) in III–V heterostructures, such and other room-temperature ballistic devices can be fabricated reproducibly with the currently available techniques in the industry. Although ballistic electron transport has been studied for more than two decades, new concepts of room-temperature ballistic devices are still being discovered, such as the devices discussed in this chapter. Currently, it is still very challenging to realize ballistic electron transport in silicon materials at room temperature because of the very short mean-free path (about 10 nm) [87]. Such a precision of device fabrication is nevertheless expected to be achieved by the industry in the near future. Apart from the bottom-up approaches, such as the promising possibilities of using nanowires to build novel room-temperature devices [88, 89], the top-down approach of design and fabrication of new types of ballistic devices such as the examples introduced in this Chapter, seems to be one of the most feasible routes to provide new building blocks for the next generation of electronics.

GLOSSARY

Ballistic rectifier A new type of nanometer-scale nonlinear device that functions like a bridge rectifier. Unlike a conventional diode, the working principle is based on the ballistic electron transport and symmetry breaking, and does not require the use of a doping junction or barrier structure to achieve nonlinear functionality.

Ballistic transport A type of electron movement contrary to the diffusive transport. If the dimension of an electronic device is smaller than the electron scattering length (mean-free path), electrons may travel from one electrode to another without encountering any scattering event from randomly distributed scatterers, such as impurities. Instead the electrons are scattered only at the device boundary, that is, moving like billiard balls. Such an electron movement is called ballistic transport.

Electron-beam lithography A standard process to create nanometer-scale structures. The resolution is generally much higher than optical lithography, well below 100 nanometers. The pattern definition is achieved by a high-energy electron beam focused on a thin layer of resist that is chemically sensitive to an electron beam. The resist layer is then developed in a suitable solution, in which either the exposed areas or the unexposed areas are selectively dissolved.

Electron mean-free path An important quantity to describe the electron transport in a material. Also often referred to as the electron scattering length, it stands for the average distance between the randomly distributed scatterers in a semiconductor material, such as defects, impurities, and phonons.

Heterostructure A multilayer semiconductor material created by growing two or more different types of semiconductors on top of each other. Each layer is normally quite thin, around tens of nanometers, and the degree of the interface homogeneity can reach the atomic level by using epitaxial growth technique, such as molecular beam epitaxy.

Mesoscopic scale The dimension between the atomic scale and macroscopic scale, typically shorter than one micron but longer than one nanometer. Ballistic electron transport devices, for instance, normally require mesoscopic dimensions.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (SFB 348), the Alexander von Humboldt Foundation, the Swedish Research Councils for Engineering and Natural Sciences, the Swedish Foundation of Strategic Research, and the European Commission through LTR research project NEAR. The bulk of the experimental work described here was carried out in laboratories of Ludwig-Maximilians University München and Lund University. Writing of this review chapter was supported by UMIST. The author thanks Max Bichler, Jörg P. Kotthaus, Armin Kriele, Axel Lorke, Anneli Löfgren, Stephan Manus, Ivan Maximov, Pär Omling, Lars Samuelson, Werner Seifert, Ivan Shorubalko, Martin Streibl, Werner Wegscheider, Hongqi Xu, and Herbert Zirath for their highly appreciated contributions to the work.

REFERENCES

1. C. W. J. Beenakker and H. van Houten, in "Quantum Transport in Semiconductor Nanostructures" (H. Ehrenreich and D. Turnbull, Eds.), Solid State Physics, Vol. 44, p. 1. Academic, Boston, 1991.
2. S. Datta, "Electronic Transport in Mesoscopic Systems." Cambridge University Press, Cambridge, 1995.
3. J. Imry, "Introduction to Mesoscopic Physics." Oxford University Press, New York, 1997.

4. S. Ulreich and W. Zwerger, *Superlattices Microstruct.* 23, 719 (1998).
5. R. Landauer, *Philos. Mag.* 21, 863 (1970).
6. M. Büttiker, *Phys. Rev. Lett.* 57, 1761 (1986).
7. M. Büttiker, *IBM J. Res. Dev.* 32, 317 (1988).
8. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, *Phys. Rev. Lett.* 60, 848 (1988).
9. D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, *J. Phys. C, Solid State* 21, L209 (1988).
10. For a review, see, for example, R. J. Haug, *Semicond. Sci. Technol.* 8, 131 (1993).
11. M. Büttiker, *Phys. Rev. B* 38, 9375 (1988).
12. G. R. Facer, B. E. Kane, A. S. Dzurak, R. J. Heron, N. E. Lumpkin, R. G. Clark, L. N. Pfeiffer, and K. W. West, *Phys. Rev. B* 59, 4622 (1999).
13. H. van Houten, C. W. J. Beenakker, P. H. M. van Loosdrecht, T. J. Thornton, H. Ahmed, M. Pepper, C. T. Foxon, and J. J. Harris, *Phys. Rev. B* 37, 8534 (1988).
14. L. W. Molenkamp, A. A. M. Staring, C. W. J. Beenakker, R. Eppenga, C. E. Timmering, J. G. Williamson, C. J. P. M. Harmans, and C. T. Foxon, *Phys. Rev. B* 41, 1274 (1990).
15. J. Spector, H. L. Stormer, K. W. Baldwin, L. N. Pfeiffer, and K. W. West, *Appl. Phys. Lett.* 56, 1290 (1990).
16. U. Sivan, M. Heiblum, C. P. Umbach, and H. Shtrikman, *Phys. Rev. B* 41, 7937 (1990).
17. G. Timp, H. U. Baranger, P. Devegvar, J. E. Cunningham, R. E. Howard, R. Behringer, and P. M. Mankiewich, *Phys. Rev. Lett.* 60, 2081 (1988).
18. M. L. Roukes, A. Scherer, S. J. Allen, H. G. Craighead, R. M. Ruthen, E. D. Beebe, and J. P. Harbison, *Phys. Rev. Lett.* 59, 3011 (1987).
19. C. J. B. Ford, S. Washburn, M. Büttiker, C. M. Knoedler, and J. M. Hong, *Phys. Rev. Lett.* 62, 2724 (1989).
20. C. W. J. Beenakker and H. van Houten, *Phys. Rev. B* 39, 10445 (1989).
21. A. Palevski, C. P. Umbach, and M. Heiblum, *Appl. Phys. Lett.* 55, 1421 (1989).
22. C. P. Umbach, A. Palevski, M. Heiblum, and U. Sivan, *J. Vac. Sci. Technol. B* 7, 2003 (1989).
23. Y. Hirayama and S. Tarucha, *Appl. Phys. Lett.* 63, 2366 (1993).
24. K. Hieke, J. O. Wesström, E. Forsberg, and C. F. Carlström, *Semicond. Sci. Tech.* 15, 272 (2000).
25. L. P. Kouwenhoven, B. J. Vanwees, C. J. P. M. Harmans, J. G. Williamson, H. van Houten, C. W. J. Beenakker, C. T. Foxon, and J. J. Harris, *Phys. Rev. B* 39, 8040 (1989).
26. A. Palevski, C. P. Umbach, and M. Heiblum, *Appl. Phys. Lett.* 55, 1421 (1989).
27. R. Taboryski, A. K. Geim, M. Persson, and P. E. Lindelof, *Phys. Rev. B* 49, 7813 (1994).
28. R. I. Hornsey, A. M. Marsh, J. R. A. Cleaver, and H. Ahmed, *Phys. Rev. B* 51, 7010 (1995).
29. D. R. S. Cumming and J. H. Davies, *Appl. Phys. Lett.* 69, 3363 (1996).
30. A. Messica, A. Soibel, U. Meirav, A. Stern, H. Shtrikman, V. Umansky, and D. Mahalu, *Phys. Rev. Lett.* 78, 705 (1997).
31. N. K. Patel, J. T. Nicholls, L. Martin-Moreno, M. Pepper, J. E. F. Frost, D. A. Ritchie, and G. A. C. Jones, *Phys. Rev. B* 44, 13549 (1991).
32. A. Kristensen, H. Bruus, A. E. Hansen, J. B. Jensen, P. E. Lindelof, C. Marckmann, J. Nygrd, C. B. Sorensen, F. Beuscher, A. Forchel, and M. Michel, *Phys. Rev. B* 62, 10950 (2000).
33. A. Lorke, S. Wimmer, B. Jäger, J. P. Kotthaus, W. Wegscheider, and M. Bichler, *Physica B* 249–251, 312 (1998).
34. A. M. Song, A. Lorke, A. Kriele, J. P. Kotthaus, W. Wegscheider, and M. Bichler, *Phys. Rev. Lett.* 80, 3831 (1998).
35. H. Linke, W. Sheng, A. Löfgren, H. Q. Xu, P. Omling, and P. E. Lindelof, *Europhys. Lett.* 44, 341 (1998).
36. H. Linke, T. E. Humphrey, A. Löfgren, A. O. Sushkov, R. Newbury, R. P. Taylor, and P. Omling, *Science* 286, 2314 (1999).
37. H. Ueno, K. Moriyasu, Y. Wada, S. Osako, H. Kubo, N. Mori, and C. Hamaguchi, *Jpn. J. Appl. Phys.* 38, 332 (1999).
38. K. Hieke and M. Ulfward, *Phys. Rev. B* 62, 16727 (2000).
39. I. Shorubalko, H. Q. Xu, I. Maximov, P. Omling, L. Samuelson, and W. Seifert, *Appl. Phys. Lett.* 79, 1384 (2001).
40. L. Worschech, H. Q. Xu, A. Forchel, and L. Samuelson, *Appl. Phys. Lett.* 79, 3287 (2001).
41. P. F. Bagwell and T. P. Orlando, *Phys. Rev. B* 40, 1456 (1989).
42. L. I. Glazman and A. V. Khaetskii, *Europhys. Lett.* 9, 263 (1989).
43. O. Heinonen and M. D. Johnson, *Phys. Rev. Lett.* 71, 1447 (1993).
44. M. D. Johnson and O. Heinonen, *Phys. Rev. B* 51, 14421 (1995).
45. M. Büttiker, *J. Phys. Condens. Matter* 5, 9361 (1993).
46. S. Komiyama and H. Hirai, *Phys. Rev. B* 54, 2067 (1996).
47. A. M. Song, *Phys. Rev. B* 59, 9806 (1999).
48. H. Q. Xu, *Appl. Phys. Lett.* 78, 2064 (2001).
49. A. M. Glas, D. von der Linde, and T. J. Negran, *Appl. Phys. Lett.* 25, 233 (1974).
50. V. I. Belinicher and B. I. Sturman, *Sov. Phys. Usp.* 23, 199 (1980).
51. B. I. Sturman and V. M. Fridkin, "The Photovoltaic and Photorefractive Effects in Noncentrosymmetric Materials." Gordon and Breach, Philadelphia, 1992.
52. D. Ertaş, *Phys. Rev. Lett.* 80, 1548 (1998).
53. T. A. J. Duke and R. H. Austin, *Phys. Rev. Lett.* 80, 1552 (1998).
54. C. F. Chou, O. Bakajin, S. W. P. Turner, T. A. J. Duke, S. S. Chan, E. C. Cox, H. G. Craighead, and R. H. Austin, *Proc. Natl. Acad. Sci. USA* 96, 13762 (1999).
55. A. van Oudenaarden and S. G. Boxer, *Science* 285, 1046 (1999).
56. A. M. Song, A. Lorke, A. Kriele, J. P. Kotthaus, W. Wegscheider, and M. Bichler, "Proceedings of the 24th International Conference on Physics of Semiconductors" (D. Gershoni, Ed.), World Scientific, Singapore, 1998.
57. A. M. Song, P. Omling, L. Samuelson, W. Seifert, I. Shorubalko, and H. Zirath, *Jpn. J. Appl. Phys.* 40, L909 (2001).
58. P. Ramvall, N. Carlsson, P. Omling, L. Samuelson, W. Seifert, M. Stolze, and Q. Wang, *Appl. Phys. Lett.* 68, 1111 (1996).
59. J. Mateos, B. G. Vasallo, D. Pardo, T. González, J. S. Gallo, S. Bollaert, Y. Roalens, and A. Cappy (to be published).
60. C. W. J. Beenakker and H. van Houten, *Phys. Rev. Lett.* 63, 1857 (1989).
61. K. L. Shepard, M. L. Roukes, and B. P. Vandergaag, *Phys. Rev. Lett.* 68, 2660 (1992).
62. R. I. Hornsey, *J. Appl. Phys.* 79, 9172 (1996).
63. H. U. Baranger and A. D. Stone, *Phys. Rev. Lett.* 63, 414 (1989).
64. A. M. Song, S. Manus, M. Streibl, A. Lorke, J. P. Kotthaus, W. Wegscheider, and M. Bichler, *Superlattices Microstruct.* 25, 269 (1999).
65. A. M. Song, A. Lorke, J. P. Kotthaus, W. Wegscheider, and M. Bichler, *Superlattices Microstruct.* 25, 149 (1999).
66. H. Linke and A. M. Song, Electron ratchets—Nonlinear electron transport in semiconductor dot and antidot structures, in "Electron Transport in Quantum Dots" (J. P. Bird, Ed.). Kluwer Academic, 2003.
67. S. Y. Chou, P. R. Krauss, and P. J. Renstrom, *Science* 272, 85 (1996).
68. B. Heidari, I. Maximov, E. L. Sarwe, and L. Montelius, *J. Vac. Sci. Technol. B* 17, 2961 (1999).
69. R. Akis, D. K. Ferry, and J. P. Bird, *Phys. Rev. B* 54, 17705 (1996).
70. A. M. Song, P. Omling, L. Samuelson, W. Seifert, I. Shorubalko, and H. Zirath, *Appl. Phys. Lett.* 79, 1357 (2001).
71. A. M. Song, *Appl. Phys. A* 75, 229 (2002).
72. A. Löfgren, I. Shorubalko, P. Omling, and A. M. Song, *Phys. Rev. B* 67, 195309 (2003).
73. L. Worschech, F. Fischer, A. Forchel, M. Kamp, and H. Schweizer, *Jpn. J. Appl. Phys.* 40, L867 (2001).

74. R. Lewen, I. Maximov, I. Shorubalko, L. Samuelson, L. Thylen, and H. Q. Xu, *J. Appl. Phys.* 91, 2398 (2002).
75. I. Shorubalko, H. Q. Xu, I. Maximov, D. Nilsson, P. Omling, L. Samuelson, and W. Seifert, *IEEE Electron Device Lett.* 23, 377 (2002).
76. H. Q. Xu, *Appl. Phys. Lett.* 80, 853 (2002).
77. R. Fleischmann and T. Geisel, *Phys. Rev. Lett.* 89, 16804 (2002).
78. M. Büttiker and D. Sánchez, *Phys. Rev. Lett.* 90, 119701 (2003).
79. R. Fleischmann and T. Geisel, *Phys. Rev. Lett.* 90, 119702 (2003).
80. M. A. Topinka, B. J. LeRoy, S. E. J. Shaw, E. J. Heller, R. M. Westervelt, K. D. Maranowski, and A. C. Gossard, *Science* 289, 2323 (2000).
81. R. Crook, C. G. Smith, M. Y. Simmons, and D. A. Ritchie, *J. Phys. Cond. Matter* 12, L735 (2000).
82. T. A. Fulton and G. J. Dolan, *Phys. Rev. Lett.* 59, 109 (1987).
83. H. Ishikuro, T. Fujii, T. Saraya, G. Hashiguchi, T. Hiramoto, and T. Ikoma, *Appl. Phys. Lett.* 68, 3585 (1996).
84. L. Zhuang, L. J. Guo, and S. Y. Chou, *Appl. Phys. Lett.* 72, 1205 (1998).
85. Yu. A. Pashkin, Y. Nakamura, and J. S. Tsai, *Appl. Phys. Lett.* 76, 2256 (2000).
86. R. J. Schoelkopf, P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, *Science* 280, 1238 (1998).
87. A. M. Song, M. Missous, P. Omling, A. R. Peaker, L. Samuelson, and W. Seifert, *Appl. Phys. Lett.* 83, 1881 (2003).
88. M. T. Björk, B. J. Ohlsson, T. Sass, A. I. Persson, C. Thelander, M. H. Magnusson, K. Deppert, L. R. Wallenberg, and L. Samuelson, *Nano Lett.* 2, 87 (2002).
89. M. S. Gudiksen, L. J. Lauhon, J. Wang, D. C. Smith, and C. M. Lieber, *Nature* 415, 617 (2002).

www.aspbs.com/enn

Scanning Probe Techniques for Semiconductor Nanostructures

Thomas Ihn

ETH Zurich, Zurich, Switzerland

CONTENTS

1. Introduction
2. Instrumentation
3. Kelvin Probe Force Measurements of Two-Dimensional Electron Gases
4. Scanning Gate Measurements on Mesoscopic Systems
5. Scanning Probe Experiments in the Quantum Hall Regime

Glossary

References

1. INTRODUCTION

In recent years a number of experiments have been performed utilizing scanning probe techniques for the investigation of the electronic properties of semiconductor nanostructures and similar systems like two-dimensional electrons on metallic surfaces or carbon nanotubes and other molecular nanostructures. Most of these experiments fall into one of the following three classes:

- Scanning tunnelling microscopy experiments, in which the tunnelling current between the conducting tip of a scanning tunnelling microscope and an electronic system is measured. Prominent examples are local density-of-states measurements on two-dimensional electron gases on metallic surfaces [like on the silver (111) surface] or semiconducting surfaces (like InAs surfaces), or even on individual carbon nanotubes.
- Scanning near field optical microscopy experiments, which allow one to measure optical excitations of individual nanostructures.
- Local probe experiments coupling a metallic sensor capacitively to a buried electronic system (e.g., the electrons in semiconductor heterostructures in which the two-dimensional electron gas is laterally patterned into quantum point contacts, quantum wires, or quantum dots). Examples for such experiments are scanning gate measurements, subsurface charge accumulation experiments, measurements with a scanning single-electron transistor, or experiments related to the Kelvin probe force microscopy.

This chapter concentrates on the last class of experiments and tries to give an overview of this field of experimental research in solid state physics. Most of these experiments had to be performed in a cryogenic environment at temperatures below 4.2 K. The reason for this lies in the relevant quantum mechanical energy scales in semiconductor nanostructures, like the Fermi energy (typically of the order of 10 meV), electrostatic charging energies (typically below 1 meV), and confinement energies (typically below 100 μeV). In addition, magnetic fields of up to about 10 T have been applied in order to be able to investigate systems in the quantum Hall regime.

In conventional low-temperature magnetotransport experiments on mesoscopic systems the measured conductance does not give direct insight into microscopic properties within the system. Within the Landauer–Büttiker description of the conductance in nanostructures the conductance is determined from the *asymptotic* behavior of the scattering states in the leads (i.e., from the corresponding transmission coefficients). This means that no microscopic information about the states *within* the nanostructure can be directly obtained from such measurements. On the other hand, detailed microscopic theories (e.g., based on the complete scattering wave functions) have to be invoked for the calculation of the conductance. These theories have to make assumptions about local potential landscapes in the structures in order to lead to quantitative results.

Scanning probe techniques can give more detailed information about the interior of a mesoscopic system, help to find the right model parameters for theoretical calculations, confirm theoretically predicted microscopic details of quantum systems, or even reveal theoretically unanticipated microscopic features.

Although this combination of magnetotransport experiments and scanning probe techniques seems to be a straightforward extension of the available experimental techniques,

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (391–413)

it turned out that the realization of these experiments is nontrivial and the number of successful experiments is smaller than one might have expected a couple of years ago. In the following, we try to review these experiments and related theoretical approaches.

In the second section the required scanning probe microscopy (SPM) instrumentation will be discussed. This includes the design criteria and a particular implementation of a suitable cryo-SFM (scanning force microscope) operating down to 300 mK and in magnetic fields up to 9 T. The third section describes a basic Kelvin probe force experiment on a two-dimensional electron gas in a GaAs/AlGaAs heterostructure. It serves as an example how a metallic tip couples capacitively to a buried electronic system and gives an exemplary demonstration of the effects that have to be taken into account. The fourth section covers scanning gate experiments on semiconductor nanostructures. As an example, scanning gate measurements on a quantum wire are discussed in detail. The last section reviews the investigation of edge channels in the quantum Hall regime with scanning probe techniques.

2. INSTRUMENTATION

2.1. Low-Temperature Scanning Force Microscopes

Since its invention by Binnig et al. [1], the SFM has become a standard tool for the investigation of conducting and insulating surfaces on the atomic scale. On the instrumentation side many low temperature SFMs have been built for more than a decade by researchers (see [2] for a historical overview). The first cryo-SFMs have become commercially available only recently, reaching base temperatures around 5 K under ultra-high-vacuum (UHV) conditions. Lower temperatures have been reached with home-built microscopes in non-UHV setups [3–7]. Such cryo-SFMs are desirable for many applications in physical research. The physics of mesoscopic semiconductor structures is a particular field of application.

2.2. Design Criteria

Here we discuss cryo-SFMs needed for the investigation of buried two-dimensional electron gases and laterally defined nanostructures on the microscopic scale. The required lateral resolution is given by the Fermi wavelength of the electrons which is 30–50 nm in typical structures. Only one or two nanostructures with a size of about 1 μm are located on a large 5 mm$\times$5 mm chip. Finding such a structure at low temperatures without optical access to the microscope is a major challenge. Large scan ranges of several μm are therefore important and a positioning system is needed for coarse positioning the structure at low temperatures.

The low temperatures are needed due to the small energy scales relevant for such structures. Elastic impurity or boundary scattering dominates the transport properties of electrons at temperatures below 4.2 K. At these temperatures quantum phenomena in electron transport such as the Shubnikov–de Haas effect or the quantum Hall effect can be observed. For the investigation of phase-coherence effects such as weak localization or conductance fluctuations temperatures below 1 K are typically required. Charging effects in quantum dots (i.e., the Coulomb blockade effect) become important at even lower temperatures. However, the cooling power of cryostats decreases with base temperature that can be achieved. While the cooling power of a ^{4}He cryostat is typically many mW at 2 K, a ^{3}He cryostat has a cooling power of the order of 0.1 mW and the cooling power of dilution refrigerators can be even smaller. Operating an SFM requires a large number of cables leading to a significant heat load which may diminish the time for scanning at base temperature in single shot systems (e.g., most ^{3}He-systems) or may not allow one to reach base temperature in dilution refrigerators, if thermal anchoring of the cables is not efficient.

Low-dimensional semiconductor structures show a rich variety of phenomena such as the quantum Hall effect only at elevated magnetic fields. Therefore a transport setup with a cryo-SFM has to include a superconducting magnet supplying magnetic fields of the order of 10 T and all the SFM components have to be made of nonmagnetic materials.

A feedback mechanism is needed for keeping the local sensor at a well-controlled distance from the sample surface. Typical semiconductor nanostructures are sensitive to light (persistent photoeffect) and therefore a nonoptical cantilever deflection detection method is needed. Such a method has the additional advantage that it leads to less involved setups.

The whole microscope has to be small enough in order to fit into the cryostats. While the sample space needed for a transport experiment at cryogenic temperatures is of the order of 1 cm^3, a SFM is much bigger. The size constraint is mainly dominated by the magnet bore which in turn is a main factor entering the prize of the magnet.

A modular design for such a microscope has to be preferred. Several components with different functionality and maintenance requirements are combined in an SFM. Easy access and exchange possibility has to be provided for all the components. Mechanical stability of the whole assembly is crucial for stable operation. Only the highest quality of all components will warrant satisfying performance of the SFM at cryogenic temperatures.

2.3. SFM Operated in a ^{3}He System

In the following we describe a SFM operating at temperatures between 350 K down to 300 mK and in magnetic fields up to 9 T. The microscope is operated in vacuum; however, samples and tips cannot be prepared and exchanged under UHV conditions, in contrast to other setups [2, 8]. Piezoelectric quartz tuning forks were employed for nonoptical tip–sample distance control in the dynamic operation mode. Faster response is achieved by employing a phase-locked loop for driving the mechanical oscillator.

Figure 1 shows a schematic drawing of the cryostat setup in which the microscope is implemented. The cryostat is sitting in a hole in the concrete lab foundation. The whole concrete block is mechanically decoupled from the rest of the building thereby providing a mechanically very quiet environment. Effective insulation of the cryostat from remaining vibrations is achieved by suspending the whole ^{3}He insert

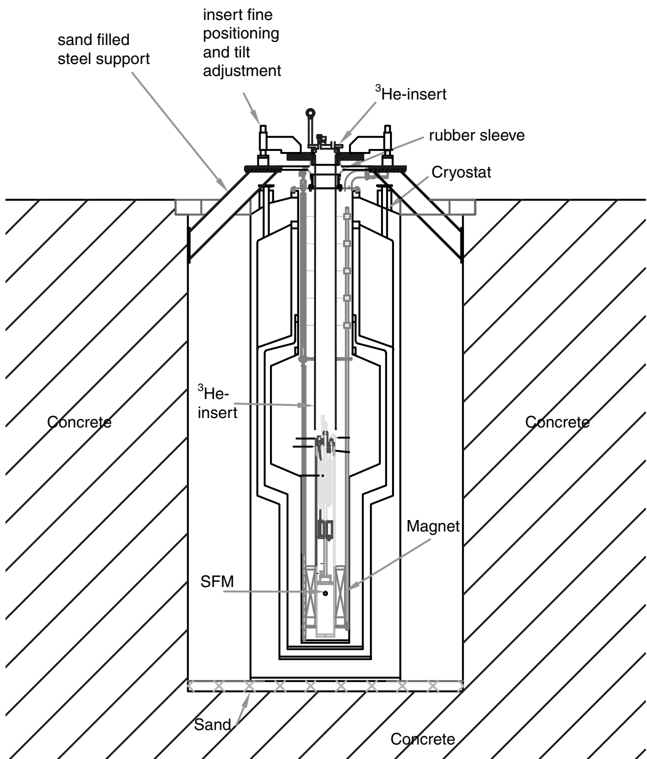


Figure 1. Schematic setup of the whole experimental setup including the cryostat, the ^{3}He insert, the superconducting magnet, and the microscope. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

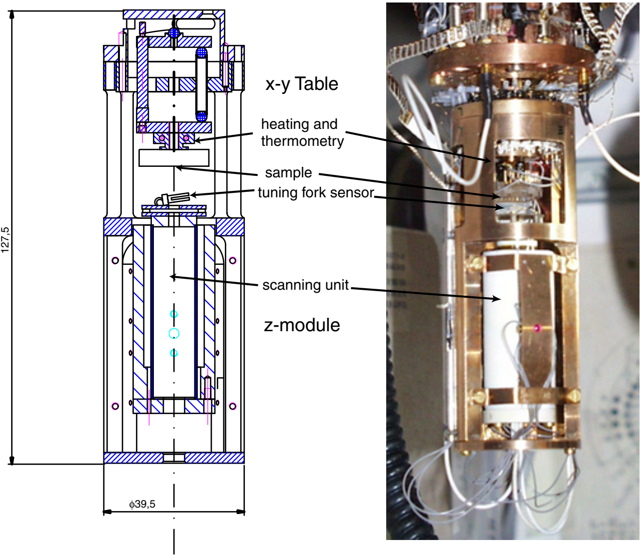


Figure 2. Schematic drawing of the microscope (left) and photography (right). Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

from a special low-vibration platform resting on four sand-filled steel supports into the ^{4}He dewar which also hosts the superconducting magnet. The only mechanical coupling between the dewar and the ^{3}He insert is made via a soft rubber sleeve needed in order to make the cryostat ^{4}He-tight. The ^{3}He insert can be fine positioned parallel to the lab floor and adjusted into exactly vertical position. Pumping lines and the He-recovery lines are plastic pipes laid through a box of sand in order to minimize mechanical coupling between pumps and the system. For the same reason, pumps are not located on the same concrete platform that hosts the cryostat setup. Microphony of the cryostat is reduced using a rubber mat which is tightly tied up to the cryostat body.

Figure 2 shows a schematic drawing of the completely home-built microscope comprising the two main units, the z-module, and the x–y table. This assembly is firmly attached to the ^{3}He pot of the commercial vacuum loading ^{3}He insert. A vacuum beaker that can be pumped down to about 10^{-5} mbar at room temperature surrounds the refrigerator and the microscope (cf. Fig. 1). The microscope can be operated in vacuum at temperatures down to 280 mK. About 50 cables, 6 of which are coaxial lines and 12 are suitable for high voltages up to 1000 V, connect this insert to the electronic setup outside the cryostat. Experiments at base temperature can be performed for about four days before it becomes necessary to recondense ^{3}He.

The z-module of the microscope (cf. Fig. 2) allows coarse tip–sample approach using a slip–stick drive moving the scan-piezo up or down. It hosts the scan-piezo tube at the end of which the tip–sample interaction sensor is mounted.

The outer frame of the z-module is machined out of nonmagnetic CuBe. The slip–stick z-motor has been constructed following the design of [2, 8]. It uses six shear-piezo stacks, which had to be fabricated by ourselves initially but are now available commercially [105]. Each stack consists of six shear-piezo plates and a sapphire endplate. A macor prism is clamped between the six stacks using a CuBe sheet as a spring. AlO plates are glued onto the macor prism in the places where the sapphire endplates of the stacks touch.

The macor prism has a central hole hosting the scan-piezo tube. The latter is made of a lead-zirconate-titanate ceramic [106] with nonmagnetic copper electrodes. The tube scanner has a length of 2 inches and an outer diameter of 0.5 inch. It gives a lateral scan range of 52.2 μm in the x–y direction and a z-range of 5 μm at a temperature of 290 K. At base temperature the lateral range is 8.8 μm and the z-range is 0.85 μm.

A sensor mount is attached to the end of the scan-piezo with tapped through holes for screwing the sensor plate onto it. The screws also provide electrical contact of the sensor plate to six coaxial cables.

The x–y table (cf. Fig. 2) is shown in detail in Figure 3. It consists of three main parts.

An x–y motor allows coarse positioning of the sample. It has a travel range of about 5 mm corresponding to the typical size of a semiconductor chip used in the experiment. The motor is built on a Shapal [107] platform. This ceramic material was used because of its comparably large thermal conductivity of 100 W/(mK) at room temperature, its thermal expansion coefficient of 5.2×10^{-6} K^{-1} at room temperature which is very close to that of piezoceramic materials, and the possibility of machining it. Three thin-walled piezo-tubes [108] with sapphire balls at their ends are fed through holes

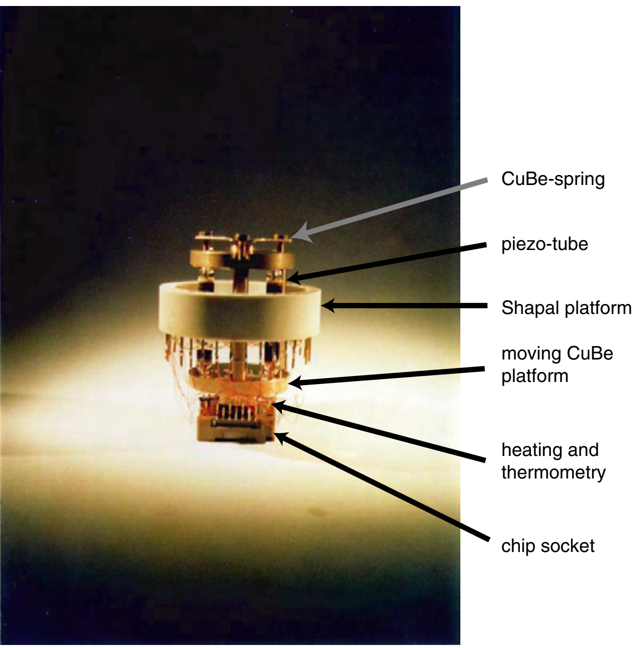


Figure 3. Photography of the x–y table. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

in the Shapal platform and firmly attached with insulating epoxy glue [109]. These tubes stick out above and below the platform. They can bend in x- and in y-direction similar to a conventional scan-piezo due to their four outer electrode segments. The moving x–y platform made of CuBe is clamped with a CuBe spring from top and bottom against the sapphire balls. Suitable sawtooth voltages applied to the electrodes of the piezo-tubes lead to slip–stick motion of the moving platform. At room temperature, about 30 V are sufficient to move the stage; at 100 V a single step amounts to about 1 μm. The motion in the x- and y-direction was found to be orthogonal with high accuracy.

Mounted on this moving CuBe platform there is a Cu block hosting two thermometers and a heating wire used for measuring and controlling the sample temperature. The heater allows us to evaporate the water film from the sample surface before cooldown. It further gives us the possibility to keep the sample warmer than its surroundings during the cooling process in order to avoid freezing contaminations on the sample surface.

On top of the thermometry block we have mounted a commercial 32-pin chip socket for ceramic chip carriers. The pins are connected to pins on the Shapal platform. The samples are prepared in chip carriers which can then simply be plugged into the socket. The sample mount unit can be used for standard magnetotransport measurements independent of the SFM operation.

2.4. Types of Sensors

There are a number of candidates for use as sensors in a scanning probe experiment on nanostructures. The most obvious one is the tip of a scanning tunnelling microscope [10]. Piezoresistive cantilevers have been proposed [11] and applied for low temperature scanning force microscopy [3, 6, 7, 12–18]. However, these cantilevers are expensive and not easily available. Recently, piezoelectric quartz tuning forks have been employed in a scanning near-field optical microscope designed for operation at low temperatures [19]. They have been successfully employed for atomic force microscopy (AFM) [20], scanning near field optical microscopy [19, 21–24], magnetic force microscopy [25], and acoustic near field microscopy [26]. A very involved and specialized sensor is a scanning single-electron transistor [27]. In the following we will briefly introduce the four sensors mentioned.

Probably the simplest sensor allowing capacitive coupling to a buried electronic nanostructure is a metallic tip as conventionally used for scanning tunnelling microscopy (STM). Such tips are commercially available or they can be electrochemically etched from tungsten or platinum/iridium wires. Used in conventional tunnelling mode topographic images can be obtained. However, since in typical semiconductor nanostructures the electronic system is buried underneath an insulating barrier of considerable thickness (typically more than 35 nm), relatively high voltages have to be applied in this imaging mode and irreversible local charging of the structure may result at low temperatures.

STM tips have been successfully employed for the so-called "subsurface charge accumulation imaging" [10] which can be—most simplified—described as a local capacitance measurement. In order to reach a sufficient sensitivity to very small local capacitance changes, stray capacitances have to be minimized by mounting a field-effect transistor (FET) at low temperature in close proximity to the tip. With this FET a low-temperature capacitance bridge can be realized [28] which can measure changes in capacitance as small as about 10 aF. A low-temperature current preamplifier has been employed for the measurement of shot noise in a scanning tunnelling microscope [29].

Piezoresistive cantilevers have been developed at Stanford University [11] in order to avoid more elaborate methods of cantilever deflection measurement, such as laser beam deflection or optical interferometry. This simplifies the design of a cryogenic scanning force microscope considerably. Such cantilevers are now available from Park Scientific Instruments. Piezoresistive cantilevers are fabricated from silicon wafers. The top layer of the cantilever is doped by ion implantation and shows a piezoresistive effect which is used to measure the cantilever deflection. These cantilevers have typical spring constants between 1 and 20 N/m. The resonance frequencies are between 25 and 240 kHz depending also on the cantilever preparation procedure. The total resistance of the piezoresistive layer is of the order of 2 kΩ and voltages of the order of several volts are used for measuring the resistance. This leads to a considerable joule heating of the sensor with a power of the order of a few milliwatts. The relative resistance change $\Delta R/R$ is proportional to the cantilever deflection Δz. That is,

$$\frac{\Delta R}{R} = \alpha \cdot \Delta z$$

where the proportionality constant α depends on the geometry of the lever. Typical values are in the range of

$(1–4) \times 10^{-6}$ nm^{-1}. A Wheatstone bridge is used to measure the resistance change with a resolution at least in the range of 10^{-6}. On-chip fabrication of the bridge is possible and can help to compensate the temperature dependence of the resistance. Vertical resolution of 0.1 Å has been demonstrated at a bandwidth of 1 kHz [11]. Force sensitivities of 8.6 fN/$\sqrt{\text{Hz}}$ have been reported [30].

Piezoresistive cantilevers have found many applications in the field of scanning probe techniques. They were, for example, used for magnetic force microscopy [31], infrared imaging [32], biosensing [33], parallel imaging, and lithography [34, 35] and are still subject of research [30]. An interesting self-sensing cantilever concept with an integrated field-effect transistor has been developed in [36].

For capacitive coupling to a buried electronic system coated [7] and uncoated [3, 6] cantilevers have been used. A metal coating (e.g., made of iridium) makes sure that the tip surface is an equipotential.

Piezoelectric tuning forks are an alternative to the piezoresistive cantilevers. They have initially been developed for the realization of very small and stable oscillators in watches. Due to their use in industry they are cheap and easily available. The forks are fabricated from wafers of α-quartz with the optical axis (c-axis) approximately normal to the wafer plane. Most of them have a (lowest) resonance frequency $f_0 = 2^{15}$Hz and quality factors in vacuum between $Q = 20{,}000$ and 100,000. In scanning force microscopy and related techniques tuning fork sensors offer the possibility of nonoptical detection of the tip oscillation in the dynamic mode via the piezoelectric effect [19–26, 37–39]. It was demonstrated that atomic resolution is possible with these unconventional and very stiff sensors [40]. The fundamental limits to force detection with quartz tuning forks were discussed by Grober and co-workers in [41]. Due to the high oscillator quality of tuning forks the power loss on resonance can be far less than 0.1 μW, which makes them ideal for use in low cooling power cryostats.

For experiments on semiconductor nanostructures a conductive tip is required for coupling capacitively to the buried electron gas. One possible realization of such a sensor is depicted in Figure 4. It is the same type of tuning fork sensor previously discussed and calibrated in [39, 42]. The commercially available tuning fork [110] is first prepared by carefully removing its protecting lid on the lathe. The fork is then soft-soldered onto a small circular piece of printed circuit board at an angle of about 10°. A 15 μm PtIr wire is glued with conducting epoxy [111] to the end face of one tuning fork prong parallel to the direction of its vibrational bending motion and to the copper post which serves as the electrical contact to the tip. The free end of the wire is then etched electrochemically resulting in a sharp tip which serves as a probe for the sample surface. Tip radii of about 30 nm have been confirmed using scanning electron microscopy and transmission electron microscopy techniques. The length of the tip measured from the edge of the tuning fork should not be longer than about 300 μm since the thin wire becomes unstable otherwise due to bending vibrations occurring during scans.

The tip prepared on the tuning fork can be used as a tunnelling tip (STM mode) or as the tip in dynamic SFM

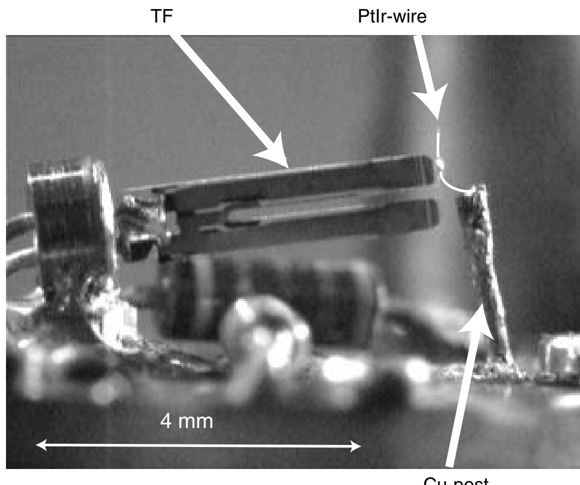


Figure 4. Tuning fork sensor used in our cryo-SFM. The casing of a commercial tuning fork (TF) is soldered on a small circular piece of printed circuit board (PCB). The PtIr wire is glued to the end of one prong of the fork. One end of the wire is connected electrically to the copper post (Cu post); the other end serves as the tip. The PCB can in addition host small electronic components, like the resistor that can be seen in the foreground. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

operation mode without any modifications on the scanning head. This gives simultaneous access to two complementary imaging modes at low temperatures without the need for tip exchange.

Given that the added mass (i.e., the glue and the wire material) is small compared to the mass of a single tuning fork arm and that the relative mechanical coupling of the two prongs is strong compared to the influence of the added mass and the added spring constant (introduced by the wire), tuning fork sensors can be modelled as harmonic oscillators (in complete analogy to conventional cantilevers) with an equation of motion

$$\ddot{x} + \frac{\omega_0}{Q}\dot{x} + \omega_0^2 x = \frac{\alpha_p}{m}U_0 + \frac{1}{m}F_{ts}(x) \tag{1}$$

where x is the deflection of one prong from the equilibrium position, ω_0 is the resonance frequency, and Q is the quality factor of the oscillator. The piezoelectric coupling constant α_p describes the coupling between the mechanical motion and the charges on the tuning fork electrodes, U_0 is the driving voltage, m is the effective mass of the oscillator, and $F_{ts}(x)$ is the tip–sample interaction force. To arrive at this model, the coupling between the center of mass and the relative motion of the prongs which can be estimated to be small has been neglected.

The mechanical oscillation of tuning fork sensors is measured via the piezoelectric effect of the quartz crystal. The induced piezoelectric charge on one tuning fork electrode is given by $q_p = 2\alpha_p x$ and the corresponding piezoelectric current is $I_p = 2\alpha_p \dot{x}$. In addition, a current of magnitude $I_c = C_0 \dot{U}_0$ flows through the capacitance C_0 between the tuning fork contacts. The total current as a function of

frequency $f = \omega/2\pi$ is therefore given by $I(\omega) = i\omega C_0 U_0 + 2i\alpha_p \omega x(\omega)$. If we neglect the tip–sample interaction force we can determine $x(\omega)$ from the harmonic oscillator model [Eq. (1)] and thus the admittance $G(\omega)$ is found to be

$$G(\omega) = \frac{I(\omega)}{U_0} = i\omega\left(C_0 + \frac{2\alpha_p^2/m}{\omega_0^2 - \omega^2 + i\omega\omega_0/Q}\right) \quad (2)$$

This admittance corresponds to the equivalent circuit shown in the inset of Figure 5b. Resonance curves like the one shown in Figure 5b can be excellently fitted with this equation.

In order to determine the mechanical tuning fork oscillation amplitude, the piezoelectric coupling constant α_p has to be determined in a calibration measurement [42]. This can be achieved by measuring the mechanical motion of one prong with a laser interferometer simultaneously with the electrical admittance as shown in Figure 5. From a combination of both measurements an effective mass $m =$ 0.332 mg, a quality factor $Q = 61734$, and a piezoelectric coupling constant $\alpha_p = 4.26$ μC/m were determined. The effective mass calculated from the density of quartz and the dimensions of a tuning fork arm turns out to be 0.36 mg, in good agreement with the measured value. It can be verified with low-temperature force–distance measurements with varying oscillation amplitude that at 300 mK, α_p has within a few percent the same value as that determined at room temperature.

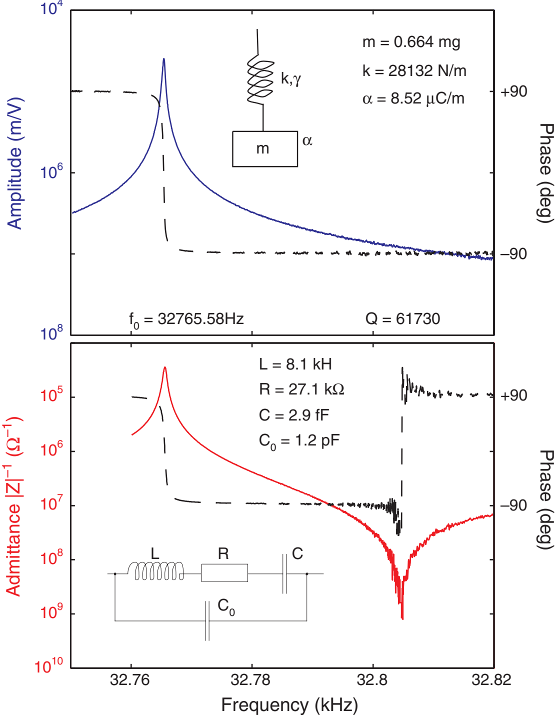


Figure 5. (a) Mechanical resonance measured at room temperature at a pressure of 6×10^{-7} mbar with an optical interferometer. Inset: Image of the tuning fork. (b) Electrical tuning fork resonance measured simultaneously. Inset: Equivalent circuit for the piezoelectric quartz tuning fork resonator. The solid and dashed lines are the respective amplitude and phase. Adapted with permission from [42], J. Rychen et al., *Rev. Sci. Instrum.* 71, 1695 (2000). © 2000, American Institute of Physics.

The sensors are completely equivalent to other tuning fork sensors [38] with one prong firmly attached to the support which acts essentially as an extremely stiff piezoelectric cantilever. In other respects there are important differences: for tuning forks as described quality factors Q of up to 250,000 are found under the UHV conditions occurring in the evacuated sample space at a temperature of 300 mK. These values are 1–2 orders of magnitude larger than those reported for the other type of tuning fork sensors [38]. For this reason, such high-Q sensors have to be operated as part of a phase-locked loop. Details can be found in [39].

The idea behind the dynamic mode SFM operation of the tuning fork is the same as for normal cantilevers. The elastic interaction of the tip with the sample surface will shift the resonance frequency via the presence of force gradients. Inelastic tip–sample interactions will alter the Q-value of the oscillator. The sensitivity of the sensor to force gradients can be calibrated experimentally with small amplitude calibration measurements [43]. It is found that the force gradient is related to the frequency shift via

$$\Delta f = \eta\left(\frac{\partial F_z}{\partial z}\right) \quad (3)$$

where $\partial F_z/\partial z$ is the gradient of the tip–sample interaction force and $\eta = 1$ Hz/(N/m) is the proportionality constant. The calibration constant η is independent of the details of the tip shape and depends only on the mechanical properties of the tuning fork.

Recently, a new sensor has been developed based on piezoelectric tuning forks which combines the benefits of soft cantilevers and piezoelectric detection [44]. The microfabricated cantilever is mounted between the two prongs of a tuning fork as shown in Figure 6. The tuning fork simultaneously drives the cantilever and detects its changing oscillation properties when a surface is present. Such sensors may be of future use for scanning probe experiments at very low temperatures.

Perhaps the most sophisticated sensors employed for scanning on shallow electronic semiconductor structures are the scanning single-electron transistors. Single-electron transistors (SETs) are tunnelling devices with a typical size of

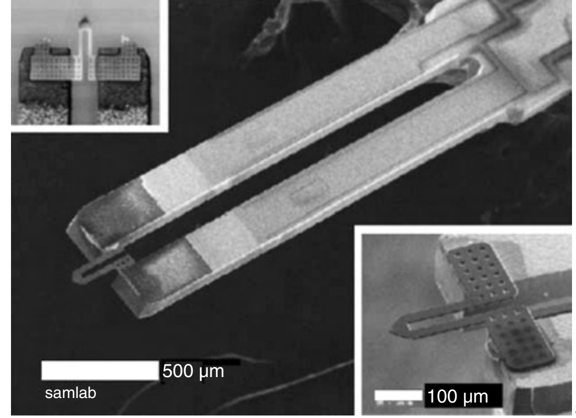


Figure 6. Microfabricated cantilevers mounted between the prongs of a commercial tuning fork. Reprinted with permission from [44], T. Akiyama et al., *Appl. Surf. Sci.* 210, 18 (2003). © 2003, Elsevier.

100 nm. The electrical current through SETs is governed by the Coulomb-blockade effect. The physics behind these devices has been reviewed in [45]. Electrons tunnel from a source electrode onto a small metallic island and subsequently from the island into the drain electrode. The magnitude of the current shows a strong modulation as a function of the electrostatic potential of the island relative to the source and drain electrodes. The period of the oscillation is given by the condition that exactly one elementary charge is additionally induced on the island. SETs can therefore be used as very sensitive electrometers for measuring local electric fields originating from localized charges, or local variations in the electrochemical potential of an electronic system acting as the gate for the island [46, 47].

Scanning SETs have been developed at the Bell Labs in the late 1990s [27, 48–51]. In this approach the SET is fabricated at the end of a glass fiber which has a shallow conical taper terminating at the tip in a flat nearly circular area of about 100 nm diameter [27]. Such a sensor is schematically shown in Figure 7. The preparation of such fibers is the same as that for scanning near field optical microscopy and has been described in [52]. The three metallic electrodes (i.e., source, drain, and island) are fabricated from 10–20 nm thick aluminum. Source and drain electrodes are evaporated onto the fiber from opposite directions onto the side of the fiber. Natural shadowing keeps these electrodes separate. An *in-situ* exposure to oxygen creates the aluminum oxide tunnel barriers of a few nanometer thickness. The island is evaporated in a last step from a direction along the fiber axis. Due to the conical taper of the fiber the aluminum film

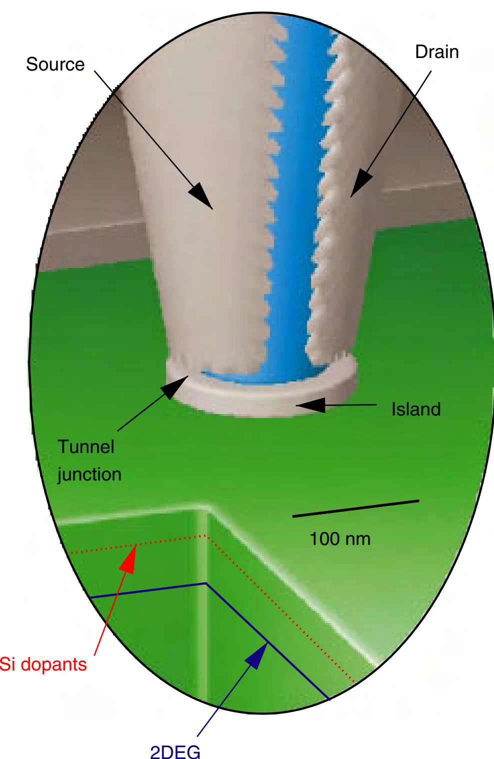


Figure 7. Schematic drawing of a scanning SET sensor. Reprinted with permission from [27], M. Yoo et al., *Science* 276, 579 (1997). © 1997, American Association for the Advancement of Science.

forms only at the end face of the tip. At low temperatures these aluminum SETs become superconducting. A magnetic field can be applied to suppress the superconductivity.

In a very simplified model the SET current may be modelled as [27]

$$I_{\text{SET}} = A \cdot \sin \frac{2\pi Q_i(x, y)}{e}$$

where $Q_i(x, y)$ is the total charge induced on the island and A is the amplitude of the current oscillation at a given source–drain bias. When the SET sensor is scanned at constant distance to the surface of a semiconductor nanostructure in the x- and y-direction, the charge induced on the island can be expressed as

$$Q_i(x, y) = Q_i^{(0)}(x, y) + \sum_n C_{ni}(x, y)(\phi_n - \phi_i)$$

where the $C_i(x, y)$ are the capacitance coefficients of the SET island to other metallic electrodes in the system (including, e.g., the electronic nanostructure to be investigated, if it is connected to external leads) and $Q_0(x, y)$ is the induced charge on the island due to the presence of fixed charges in the system (e.g., dopants or other charged impurities). The electrostatic potentials ϕ_i are given related to the applied voltages V_i and the chemical potentials of the electrodes μ_i via

$$-eV_i = \mu_i(x, y) - e\phi_i(x, y)$$

Inserting these relations into the equation for the SET current we obtain

$$I_{\text{SET}} = A \cdot \sin \frac{2\pi\left[Q_i^{(0)}(x, y) + \sum_n C_{ni}(x, y)(V_{ni} + \mu_{ni}(x, y)/e)\right]}{e} \tag{4}$$

with $V_{ni} = V_n - V_i$ being the voltage between electrode n and the island and $\mu_{ni}(x, y) = \mu_n(x, y) - \mu_i$ being the difference in chemical potentials between electrode n and the island (sometimes referred to as the work function difference between the two materials). From this formula it can be seen that the SET current senses local electric charges [via $Q_i^{(0)}(x, y)$], the local chemical potential $\mu_n(x, y)$, and the local capacitance coefficients $C_{ni}(x, y)$. It is clear that in a well defined experiment care must be taken that only one of these quantities depends significantly on position.

Since this sensor is not sensitive to the tip–sample force or force gradient, the height (z-) dependence of the capacitances C_{ni} or the induced charges $Q_i^{(0)}$ can be used for approaching the sensor to the surface [27].

3. KELVIN PROBE FORCE MEASUREMENTS OF TWO-DIMENSIONAL ELECTRON GASES

In this section a method for the local characterization of two-dimensional electron gases will be described. While conventional magnetotransport experiments (i.e., the Hall effect and the measurement of Shubnikov–de Haas oscillations in

the magnetoresistance) allow the determination of an average electron density in the system, the method described here leads to the measurement of a local electron density at the position of the conducting SFM tip.

The sample was a GaAs/AlGaAs heterostructure with a shallow two-dimensional electron gas (2DEG) confined at the heterointerface which is 34 nm below the sample surface. The electrons are supplied by a Si δ-doped plane 17 nm remote from the heterointerface. Four ohmic gold–germanium contacts have been diffused into the samples at the edges of a 5×5 mm^2 piece of wafer giving electrical access to the two-dimensional electron gas.

Experiments were performed at a temperature of 300 mK with a SFM setup employing a tuning fork sensor. A variable DC voltage was applied at the 2DEG contacts while the SFM tip was kept grounded. In this way the electrostatic force acting between the metallic tip and the 2DEG can be controlled. Like in Kelvin-force microscopy [53] this electrostatic force can be measured with the SFM tip. The frequency shift of the tuning fork sensor measures the additional force gradient created by the applied voltage. In contrast to a metallic sample the 2DEG in the heterostructure has a small density of states and a small Fermi energy. Therefore, small voltages applied between the 2DEG and the tip can change the local electron density significantly, leading locally to a total depletion of the electron gas in the extreme case.

Figure 8 shows measured frequency shifts as a function of the applied tip–sample voltage for various tip–sample separations. A clear asymmetry can be observed leading to a weaker (stronger) frequency shift at positive (negative) voltages. For a given voltage, the frequency shift decreases with increasing tip–sample separation.

Qualitatively, we expect a locally enhanced electron density under the tip for negative voltages (in this case the tip is positive relative to the 2DEG) and a reduced density for positive voltages eventually leading to complete local depletion. Therefore we expect an ideal capacitor model with a voltage independent capacitance to work only for negative voltages. For positive voltages depletion of the 2DEG leads to a hole in the capacitor plate which is a change in geometry of the capacitor and must therefore lead to a reduced capacitance. A deviation from the ideally expected parabolic shape of the frequency shift versus voltage is the natural consequence.

Following this qualitative picture we assume a voltage independent capacitance for negative voltages and fit only the left branches of the curves with a parabola according to the equation

$$\Delta f = \eta F_z' = \frac{\eta}{2}\left(\frac{\partial^2 C}{\partial z^2}\right)[U_{ts} + \Delta\mu_{ch}/e]^2 \tag{5}$$

where $F' = \partial F_z/\partial z$ is the gradient of the tip–sample interaction force and $\Delta\mu$ is the work function difference between the tip and the sample. The result is shown in Figure 9 for a tip–sample separation of 15 nm. The information deduced from these fits is twofold. On one hand, we obtain the tip–sample voltage at which the parabola has its maximum. This quantity will be discussed later. On the other hand, we get the curvature of the parabola which is proportional to $\partial^2 C(z)/\partial z^2$. A plot of this quantity as a function of tip–sample separation is shown in Figure 10. It can be seen that the curve can be well characterized by having a z^{-2} dependence. According to a comprehensive comparison of different models for tip–sample arrangements with different tip geometries performed by Belaidi and co-workers [54] such a behavior is consistent with the model of a metallic sphere above a metallic plane. Given that this slope persists up to tip–sample separations of about 700 nm we have to conclude that the tip diameter is at least of the same order of magnitude in this experiment.

We now turn to the discussion of the asymmetry of the Kelvin measurement. The explicit parabolic voltage dependence of the frequency shift can be eliminated from the curves by normalizing Δf_{meas} by Δf_{fit}, (i.e., by plotting the

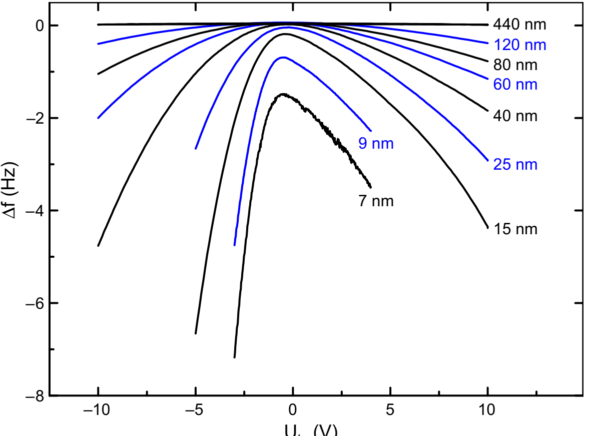


Figure 8. Frequency shift of the tuning fork resonance as a function of the voltage applied between the PtIr tip and the buried two-dimensional electron gas for various tip–sample separations. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

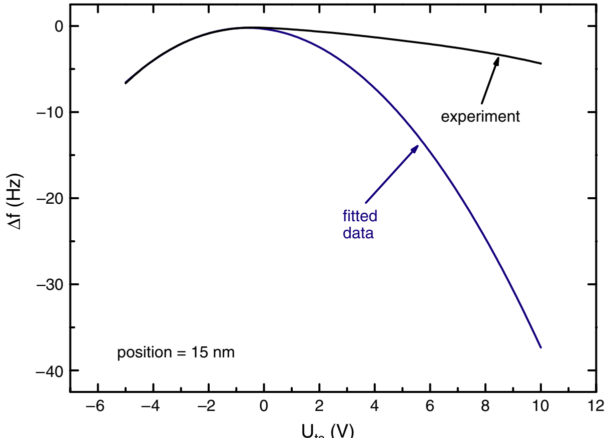


Figure 9. Frequency shift as a function of tip–sample voltage for a tip–sample separation of 15 nm. A parabola was fitted to the left branch of the measured curve. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

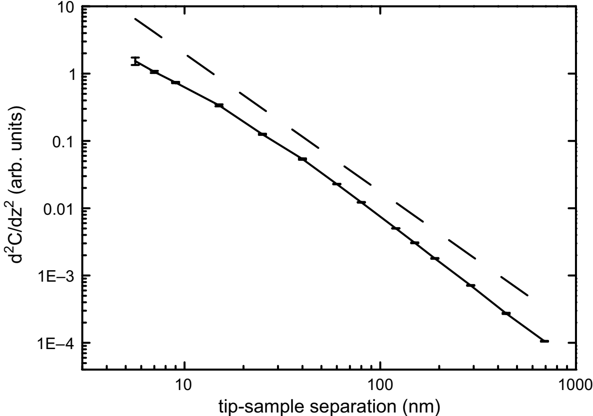


Figure 10. Second derivative of the tip–sample capacitance as a function of tip–sample separation as determined from the curvature of the fitted parabolae (cf. Fig. 9). The straight solid line has the slope of a z^{-2}-dependence. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

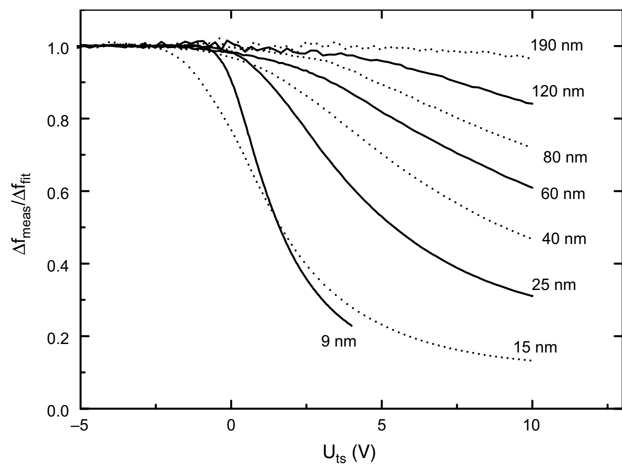


Figure 12. Normalized $\partial^2 C(z)/\partial z^2$ as a function of the tip–sample voltage for various tip–sample separations showing the depletion of the two-dimensional electron gas under the tip. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

ratio of the two as a function of voltage). Figure 11 shows the result of the normalization procedure for the curve taken at 15 nm tip–sample separation. By virtue of the normalization procedure and in accordance with our qualitative picture, $\partial^2 C/\partial z^2$ is constant for negative voltages. At a voltage of about −2.5 V, depletion starts under the tip. With increasing voltage the area of the depleted region grows continuously reducing $\partial^2 C/\partial z^2$. We empirically define the onset voltage for depletion U_{depl} in the way indicated in Figure 11. At this voltage a linear fit made in the turning point of the decaying $\partial^2 C/\partial z^2$ curve crosses unity. The same analysis has been applied to all curves in Figure 8 for different tip–sample separations and the result is shown in Figure 12. From these curves we can extract the dependence of U_{depl} on tip–sample separation (see Fig. 13). The series of points can be quite well fitted with a linear function. This finding can be interpreted within a suitable model.

We briefly come back to the second type of information that can be obtained from the parabolic fit shown in Figure 9, namely, the voltage at which the parabola has its maximum. Conventionally this voltage is interpreted as the difference in work functions of the two materials involved. Figure 14 shows this quantity as a function of tip–sample separation. Although the error bars are rather large, a slight trend for an increase at smaller tip–sample separation becomes evident, in contradiction to the notion of a constant material-dependent work function difference $\Delta\mu_{ch}$. The origin of this behavior will be traced back to the presence of

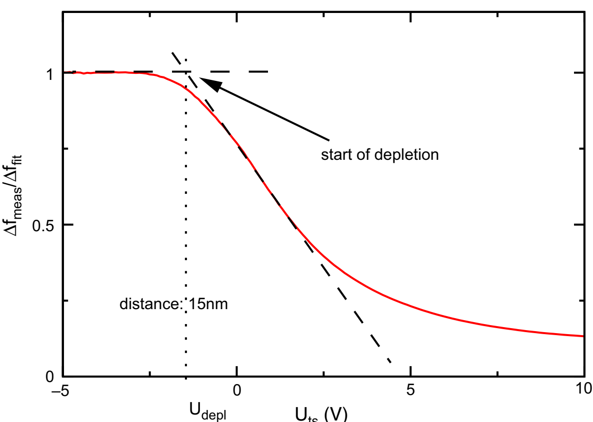


Figure 11. Normalized $\partial^2 C(V)/\partial z^2$. At negative tip–sample voltages this quantity is unity. The reduction at positive voltages is caused by total depletion of the 2DEG under the tip. Such curves allow the definition of a depletion voltage. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

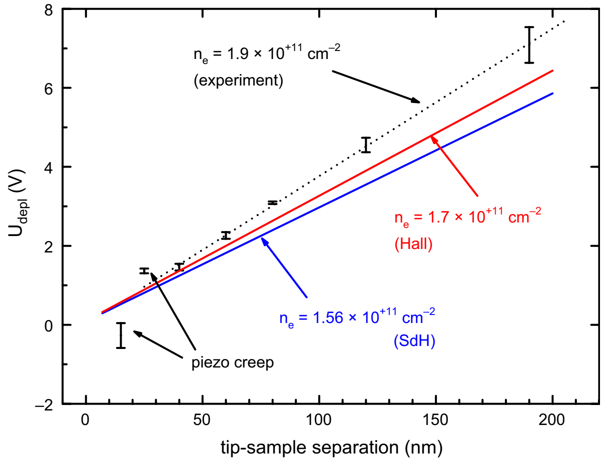


Figure 13. Local electron densities compared to the Hall and Shubnikov–de Haas densities. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

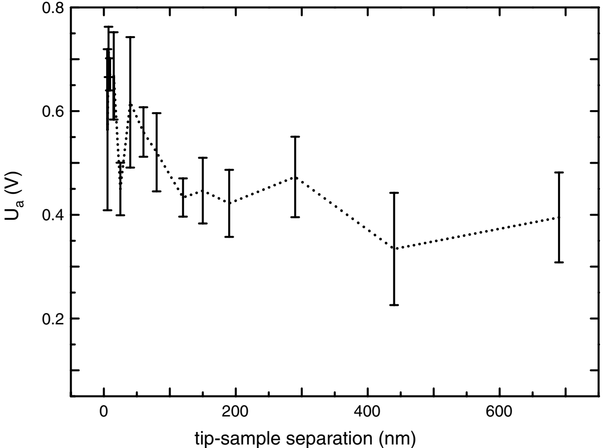


Figure 14. Contact voltage difference between the PtIr tip and a GaAs sample as a function of tip–sample separation. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

fixed localized charges in the system (surface charges, doping plane).

A general model for the experimental system starts from Poisson's equation

$$\nabla[\epsilon(\vec{r})\epsilon_0\nabla\phi(\vec{r})] = -\rho_{\text{ion}}(\vec{r}) \tag{6}$$

with the spatially varying dielectric function $\epsilon(\vec{r})$, fixed ionic charges $\rho_{\text{ion}}(\vec{r})$ accounting, for example, for ionized donors or surface charges and the electrostatic potential $\phi(\vec{r})$. The two-dimensional electron gas is, for simplicity, considered to be a perfect metal plate and also the tip is regarded as a perfect metal of arbitrary shape. The boundary conditions for the problem are given on the metallic surfaces S_i (i.e., the tip surface and the 2DEG) by

$$\phi(\vec{r})|_{S_i} = \phi_i$$

where the ϕ_i is the electrostatic potential of electrode i. The resulting charge density induced in the two-dimensional electron gas is given by

$$\sigma_{\text{2DEG}}(\vec{r}) = \sigma^{(0)}_{\text{2DEG}}(\vec{r}) + c_{\text{2DEG-tip}}(\vec{r})\phi_{\text{2DEG-tip}} \tag{7}$$

where the surface charge $\sigma^{(0)}_{\text{2DEG}}$ and the coefficient $c_{\text{2DEG-tip}}$ are given by integral expressions involving Green's function of the problem and therefore depend on position $\vec{r}$ on the electrode. Between the electrostatic, the chemical, and the electrochemical potential there is the general relation $\mu_{elch} = \mu_{ch} - e\phi$, such that $\phi = (\mu_{elch} - \mu_{ch})/(-e)$. The applied voltage between two electrodes is the difference of their electrochemical potentials: $V_{ij} = (\mu_{elch,i} - \mu_{elch,j})/(-e)$. Applying these relations to Eq. (7), we obtain

$$\sigma_{\text{2DEG}}(\vec{r}) = \sigma^{(0)}_{\text{2DEG}}(\vec{r}) + c_{\text{2DEG-tip}}(\vec{r})(U_{ts} + \Delta\mu_{ch}/e) \tag{8}$$

where $\Delta\mu_{ch}$ is the work function difference between the tip material and the two-dimensional electron gas and U_{ts} is the applied voltage. At this point we can nicely discuss the influence of the tip on the system. If $U_{ts} + \Delta\mu_{ch}/e = 0$, the electron density is simply given by $\sigma^{(0)}_{\text{2DEG}}(\vec{r})$. This quantity depends, however, on the presence and the exact position of the tip via the boundary conditions of the problem. This means that even if the difference in the work functions between the materials is exactly cancelled by the applied voltage, the presence of the tip can change the local electron density in the 2DEG via the dependence of $\sigma^{(0)}_{\text{2DEG}}(\vec{r})$ on tip position (i.e., via screening).

Application of a suitable voltage U_{ts} at constant tip position will change the local electron density proportional to the differential capacitance function $c_{\text{2DEG-tip}}(\vec{r})$. This function depends on the exact shape of the tip but will for a typical conical tip be similar to a Lorentzian distribution with an extremum below the tip apex [3]. A negative voltage on the tip relative to the (grounded) electron gas will deplete the electrons and the electrostatic model with a voltage independent $c_{\text{tip-2DEG}}(\vec{r})$ will be applicable until complete depletion is reached. Beyond this point $c_{\text{tip-2DEG}}$ will become voltage dependent, since the boundary conditions of the problem change: the 2DEG will behave like a metallic plate with a hole under the tip whose size depends on U_{ts}.

The measured frequency shift in the experiment is proportional to the force gradient in the z-direction felt by the scanning tip. Within our model we find for the force gradient

$$\frac{\partial F_z}{\partial z} = \frac{1}{2}C''(z)\left(U_{ts} + \frac{\Delta\mu_{ch}}{e} + \frac{Q_0^{(0)''}(z)}{C''(z)}\right)^2 - \frac{\left(Q_0^{(0)''}\right)^2}{2C''(z)}$$

where $C''(z) = \partial^2 C_{\text{2DEG-tip}}/\partial z^2$ is the second derivative of the capacitance coefficient between the tip and the 2DEG and $Q_0^{(0)}$ is the charge induced on the tip when the 2DEG is grounded. This induced charge is due to the presence of the fixed ionic charges in the system. The quantity $\Delta\mu_{ch}$ is the difference in the chemical potentials between the tip material and the electron gas (also referred to as the work function difference between the two materials) and U_{ts} is the tip–sample voltage. According to this equation, the force gradient (and thus the measured frequency shift) is parabolic in the voltage and the apex of the parabola is at

$$U_a = -\frac{\Delta\mu_{ch}}{e} - \frac{Q_0^{(0)''}(z)}{C''(z)} \tag{9}$$

This result makes clear that the apparent work function difference given by the apex of the parabola may in fact depend on the z-position of the tip if fixed charges are present in the system, in agreement with the results shown in Figure 14.

In analogy to the remarks made about the depletion of the electron gas, $C(z)$ and also $C''(z)$ will become functions of U_{ts} as soon as the two-dimensional electron gas begins to deplete under the tip.

In order to illustrate the somewhat abstract considerations we demonstrate them with a simple parallel plate capacitor model sketched in Figure 15. It is expected to lead to reasonable results, if the tip radius is large compared to the tip–sample separation. Two planes of fixed charges are present in this model: the doping plane with charge density σ_d is located at a distance d_1 from the two-dimensional electron gas and a surface charge σ_s is located directly at

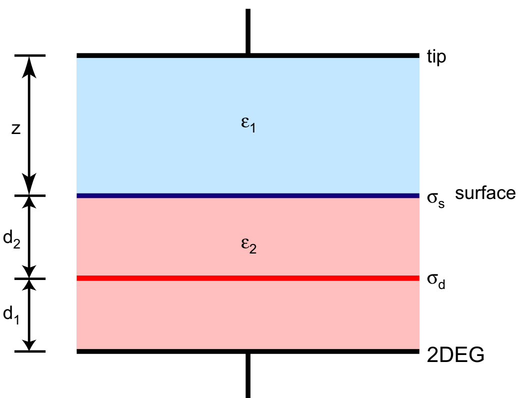


Figure 15. Plate capacitor model for the tip–sample system. Part of the capacitor is filled with the AlGaAs/GaAs dielectric which has a dielectric constant of $\varepsilon_2 = 12.53$ [55]. The other part is the vacuum gap characterized by $\varepsilon_1 = 1$.

the dielectric interface of the sample (ε_2) to vacuum (ε_1). The separation of the tip from the surface is z. In such a system the charge density in the two-dimensional electron gas is given by

$$\sigma_{\text{2DEG}} = -\underbrace{\frac{\varepsilon_1\varepsilon_2\varepsilon_0}{\varepsilon_1(d_1+d_2)+\varepsilon_2 z}}_{C(z)/A}\cdot(U_{ts}+\Delta\mu_{ch}/e) + \underbrace{\frac{\varepsilon_1\sigma_d d_2+\varepsilon_2(\sigma_d+\sigma_s)z}{\varepsilon_1(d_1+d_2)+\varepsilon_2 z}}_{Q^{(0)}_{\text{2DEG}}(z)/A:=-en_s(z)}$$

which has exactly the structure of Eq. (8). We start the discussion assuming $U_{ts} + \Delta\mu_{ch}/e = 0$. For very large tip–sample separation z, the charge density in the two-dimensional electron gas is given by $-en_s^\infty = \sigma_d + \sigma_s$. For very small z it changes to $n_s^0 = \sigma_d d_2/(d_1 + d_2)$. The total voltage that one has to apply across this capacitor in order to deplete an electron density $n_s(z)$ is given by

$$U_{\text{depl}} = -\frac{\Delta\mu_{ch}}{e} + \frac{en_s(z)A}{C(z)} = -\frac{\Delta\mu_{ch}}{e} - \frac{\varepsilon_1\sigma_d d_2 - \varepsilon_2 en_s^\infty z}{\varepsilon_0\varepsilon_1\varepsilon_2} \tag{10}$$

The depletion voltage depends linearly on the tip–sample separation in accordance with the measurement depicted in Figure 13. From the slope $dU_{\text{depl}}/dz = en_s^\infty/(\varepsilon_0\varepsilon_1)$ the local electron density can be determined to be 1.9×10^{11} cm^{-2}. As a comparison we plot curves into Figure 13 which are predicted by Eq. (10) with densities n_s determined from Hall and Shubnikov–de Haas measurements. Given the simplicity of the model, the agreement is reasonable, especially if we take into account that the 2DEG density can be inhomogeneous across a macroscopic sample.

If we evaluate the z-dependence of the apparent work function difference given by Eq. (9) for the plate capacitor model, we find

$$U_a = -\frac{\Delta\mu_{ch}}{e} - \frac{\sigma_s(d_1+d_2)+\sigma_d d_1}{\varepsilon_1\varepsilon_0}$$

independent of z. Inserting our sample parameters, $\varepsilon_1 = 1$, $d_1 + d_2 = 34$ nm, $d_1 = d_2 = 17$ nm, $\sigma_d = 1.6 \times 10^{-3}$ Cm^{-2} (estimated) and $\sigma_s = -1 \times 10^{-3}$ m^{-2} (estimated) we obtain a value of -0.76 V for the second term, indicating that the presence of fixed charges in the sample can have an appreciable influence on the apparent work function difference measured in the experiment. However, given the uncertainties in the estimated values, one should not take the exact value too quantitatively, but rather consider it as an estimate of the order of magnitude of the effect. This z-independent result has to be compared with the experimental results shown in Figure 14 which indicate a small z dependence. Obviously, a more realistic model including the actual tip shape would cure this problem. In fact, the tip shape does influence the geometric capacitance and the dependence of $\partial^2 C/\partial z^2$ on the tip–sample separation. Looking at the order of magnitude of the apparent work function difference in the measurement, we can see that the estimated influence of the fixed charges in the system can be quite significant and is of the same order as typical work function difference s between materials.

More involved models for the tip–sample system would take a more realistic shape of the tip into account, as we mentioned. Further improvement can be gained by taking the screening properties of the two-dimensional electron gas into account self-consistently. This would include density of states contributions to the capacitance which were not taken into account in the presented analysis.

A related experiment has been performed with piezoelectric cantilevers on a two-dimensional electron gas in a magnetic field [7]. In this experiment, the variation of the chemical potential of a two-dimensional electron gas was measured locally as a function of magnetic field B. Oscillations periodic in $1/B$ were detected corresponding exactly to the filling factors determined from transport experiments.

4. SCANNING GATE MEASUREMENTS ON MESOSCOPIC SYSTEMS

4.1. Existing Experiments

Electronic transport in mesoscopic semiconductor structures at low temperatures is nonlocal and phase coherent [56]. In a conventional transport experiment, in which the resistance or conductance of a sample is typically measured as a function of gate voltage or magnetic field, the measured conductance is given by the transmission probabilities between the leads connecting the system to the measurement circuit [57]. In general, it is not possible to reconstruct the scattering potential within the mesoscopic system from the knowledge of the transmission matrix. However, additional information about the quantum states within the device can be obtained from measurements which apply a local perturbation to the system [3, 4, 6, 13–17, 58, 59]. While in [58] the local potential perturbation was fixed and the two-dimensional electron gas in a parabolic quantum well could be displaced, in [3, 4, 6, 13–17, 59] the local electrostatic potential induced by the tip of a scanning force microscope was utilized. In the following, we briefly review the experimental work published in recent years.

In the first paper reporting this "scanning gate technique" Eriksson and co-workers [3] investigated the ballistic electron flow through a quantum point contact at a temperature of 4.2 K. They found that the resistance change caused by the tip-induced potential decays exponentially with distance from the point contact in agreement with elastic scattering by the disorder potential in the collector electron gas. In addition, they measured the angular distribution of the transmitted electrons showing the collimation of the ballistic electron beam close to the point contact. Calculations were presented in this paper indicating that the spatial resolution of the technique is comparable to the electron gas depth below the surface. Similar experiments on a quantum point contact have been reported in [13].

Another measurement on a quantum point contact was performed by Crook and co-workers [16] at 4.2 K where the authors were able to map the transmitted lateral modes with remarkable spatial resolution of about 40 nm although their electron gas was buried 100 nm below the surface. This experiment is the analog of the "wave function mapping " experiment reported in [58]. It exemplifies the ultimate potential of scanning gate techniques to map the states in the interior of a mesoscopic system.

Recently, the electron flow past a quantum point contact was imaged with high spatial resolution achieved at 4.2 K on a GaAs/AlGaAs heterostructure with an only 57 nm thick cap layer above the electron gas [17]. It was shown in the experiment and also modelled theoretically that the classical contribution of the electron flow branches in the two-dimensional collector as shown in Figure 16, a result which had not been anticipated by other experiments. In addition, quantum interference effects were observed in these highly sensitive measurements.

Experiments on two constrictions in series for a "magnetic steering experiment" were reported in [6, 14]. Classical cyclotron orbits of ballistic electrons in a magnetic field with the radius $R_c = \hbar k_F/(eB)$ could be mapped out with a spatial resolution of about 200 nm.

Scanning gate microscopy (SGM) has been applied to the investigation of electron transport in carbon nanotubes [60–62] at room temperature. The authors were able to "directly image individual scattering sites in semiconducting single-wall nanotubes" [61] and metallic single-wall nanotubes [62] or an inhomogeneous potential landscape [60], again in semiconducting single-walled tubes. They showed that the resistance image of a nanotube shows hot spots in locations where the electric field drop was found to be high with electrostatic force microscopy (EFM) [61]. In some cases these hot spots occurred in positions where the nanotube had a kink allowing one to correlate information from three different sources (topography, EFM, and SGM) to extract a unified picture of the system.

A very recent publication by Woodside and McEuen reports low temperature (0.6 K) scanning gate studies of single-walled carbon nanotubes in which one or more quantum dots form leading to the Coulomb blockade effect in the conductance [59]. These authors find Coulomb blockade oscillations as a function of the position of the scanning tip as shown in Figure 17 allowing them to locate the quantum dots along a certain nanotube and to estimate their sizes.

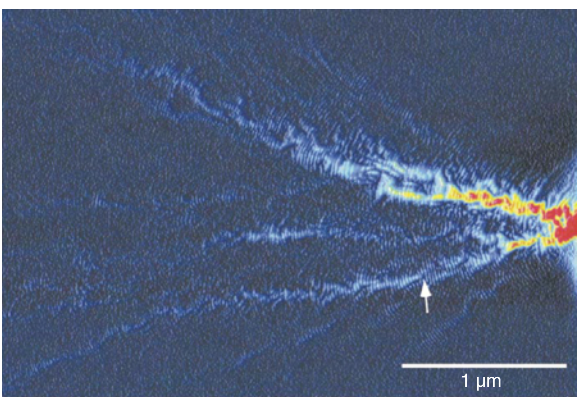


Figure 16. Coherent branched flow of electrons past a quantum point contact. Reprinted with permission from [17], M. Topinka et al., *Nature* 410, 183 (2001). © 2001, Macmillan Magazines Ltd.

4.2. Measurements on a Quantum Wire

The scanning gate approach has also been employed for the investigation of the transport properties of a semiconductor quantum wire at cryogenic temperatures [63]. We will discuss this experiment in more detail in the following.

The sample is based on a GaAs/AlGaAs heterostructure in which the heterointerface is buried 34 nm below the surface. A Hall-bar geometry, a part of which can be seen in Figure 18, was patterned using photolithography. At the temperature $T = 1.7$ K the bulk electron density measured in this system is $n_S = 4.3 \times 10^{11}$ cm^{-2}, corresponding to a Fermi wavelength of 38 nm. The mobility at $T = 1.7$ K is $\mu = 10^6$ cm^2/V s corresponding to a mean free path of $l_e^{2D} = 10.8$ μm and a diffusion constant of $D = 1.52$ m^2/s.

The quantum wire has been patterned using the so-called AFM lithography [64, 112] directly onto the Hall-bar structure. In Figure 18 the white lines on the Hall bar are the SFM-written oxide lines which deplete the 2DEG below

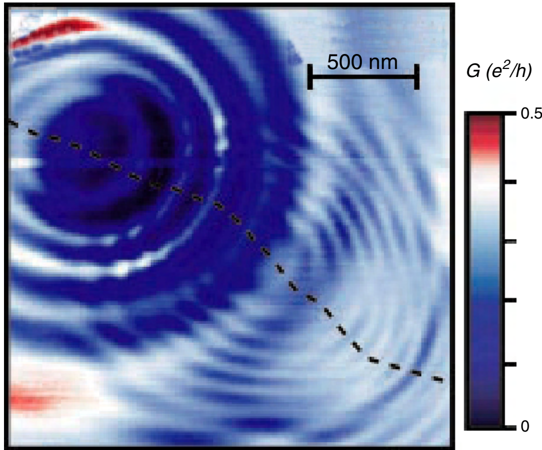


Figure 17. Coulomb oscillations as a function of tip position measured on a carbon nanotube at $T = 6$ K. The dashed line indicates the location of the nanotube. Reprinted with permission from [59], M. Woodside and P. McEuen, *Science* 296, 1098 (2002). © 2002, American Association for the Advancement of Science.

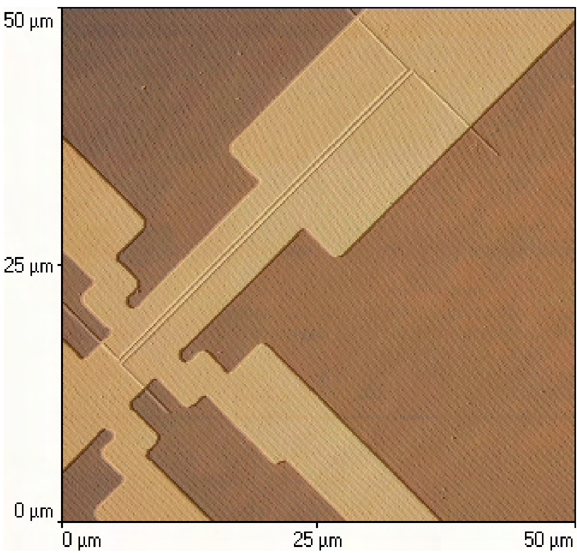


Figure 18. SFM image of the Hall bar with the oxide lines defining the quantum wire. Reprinted with permission from [63], T. Ihn et al., *Physica E* 12, 691 (2002). © 2002, Elsevier.

them and thereby form a long quantum wire flanked by two separately contacted regions of the 2DEG which we used as in-plane gates. The wire length is $L = 40$ μm and its lithographic width is $W = 400$ nm. The lateral depletion length is typically 15–20 nms such that the number of modes in the channel can be estimated to be $N = 20$. This number can be confirmed by Shubnikov–de Haas measurements which count the number of modes which become successively depopulated as the magnetic field is increased.

The sample was mounted in the sample holder of a low-temperature scanning force microscope [39]. This microscope utilizes piezoelectric quartz tuning forks with a metallic tip attached to one prong as the force sensor. Operation characteristics at low temperatures and sensor calibration have been described and are reported in [42, 43]. The microscope was cooled to $T = 1.7$ K in the He gas flow of a variable temperature insert. The resistance of the wire was measured at an ac current of 20 nA in a four-probe configuration. The measurement frequency was 421 Hz and a lock-in time constant of 10 ms was used in order to obtain a reasonably large output bandwidth. The tungsten tip of the SFM was kept grounded while it was oscillating with an amplitude of about 1 nm normal to the surface at the tuning fork resonance frequency of about 32 kHz. The basic idea of the measurement is shown in Figure 19. The tip induces a local scattering potential in the electron gas. The wire resistance is changed depending on the position of the tip. Resistance images of 256 × 256 points were taken at scan speeds of 500 nm/s.

Figure 20 shows the resistance image of a 5×5 μm^2 area obtained with a voltage of +200 mV on one of the in-plane gates. The fact that the wire can be electronically imaged indicates that when the tip is above the wire the resistance is enhanced. This resistance contrast is due to the fact that tungsten and the heterostructure have a work function difference of the order of 100 mV which causes the tip to induce a local repulsive electrostatic potential in the 2DEG

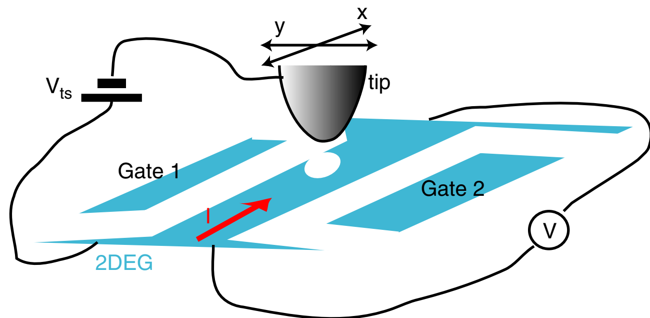


Figure 19. Principle of scanning gate measurements. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

which is scanned along when the tip moves. The stripe of enhanced resistance is not visible in the topmost quarter of the image. From the simultaneously measured topography we know that the wire ends there.

Cross sections through this image as indicated by three horizontal lines in Figure 20 are shown in Figure 21. As the tip moves across the wire the resistance goes from its base value of 3 kΩ through a pronounced peak of 9.4 kΩ. We estimate the number of modes which become locally depleted, ΔN, with a simple model which regards the repulsive tip potential as the cause of an additional resistance ΔR which adds to the background wire resistance R_0. In the simplest case ΔR can be modelled by a short piece of wire with transmission $T = 1$ which is not mode-matched to the rest of the wire. In this case $\Delta R = h/(2e^2)(1/(N_0 - \Delta N) - 1/N_0)$ with $N_0 = 20$ being the number of modes in the wire in the absence of the tip. This gives a substantial depletion of $\Delta N = 18$ modes if the tip is centered above the wire.

From the width of the two lower peaks of about 400 nm which is close to the geometrical width of the wire we estimate that the tip induced potential perturbation cannot be much wider than that.

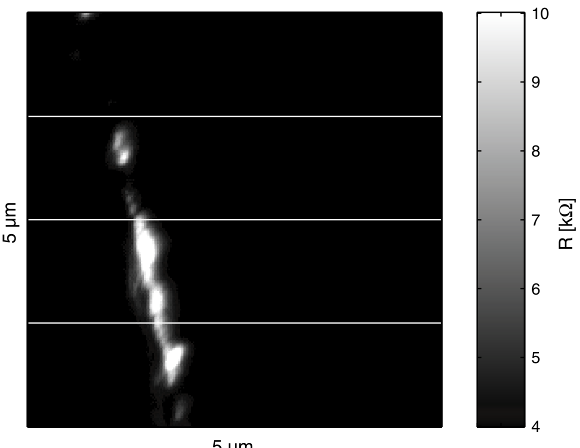


Figure 20. Resistance image of the quantum wire measured at $T = 1.7$ K. Reprinted with permission from [63], T. Ihn et al., *Physica E* 12, 691 (2002). © 2002, Elsevier Science.

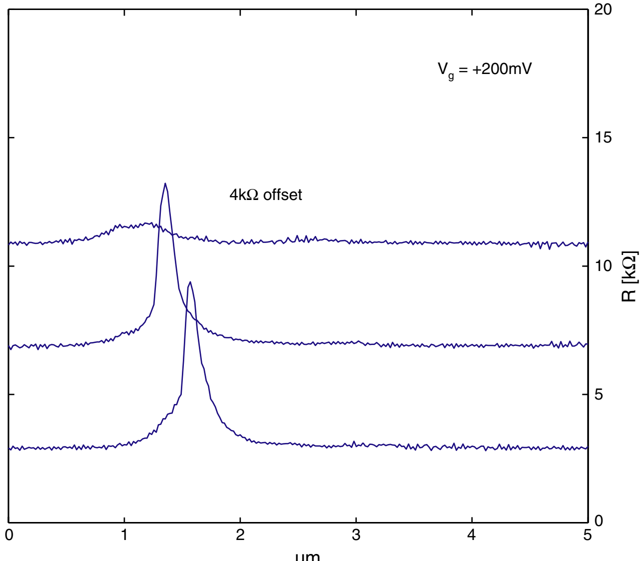


Figure 21. Cross sections through the resistance image shown in Figure 20. Reprinted with permission from [63], T. Ihn et al., *Physica E* 12, 691 (2002). © 2002, Elsevier Science.

A very striking feature of the image in Figure 20 is the variation of the resistance along the wire direction. In a perfect wire with no potential fluctuations and perfectly smooth boundaries no such variation would be expected irrespective of the exact geometrical shape of the tip-induced potential. From a comparison of the resistance variations with the surface topography we find no correlation between the two that could explain the former as the result of a varying tip–2DEG separation. We therefore conclude that the variations of the resistance along the wire direction reflect the roughness of the potential landscape in the electron gas.

There are two distinct effects in mesoscopic systems that could cause the observed resistance fluctuations as the perturbation is moved along the wire. The first is the ballistic chaotic motion of *classical* electrons in a spatially varying potential landscape. In order to develop a better understanding of how this mechanism would appear in our measurements we have calculated the classical transmission through wire structures with a given potential landscape as a function of the position of an additional external perturbation. The results which will be discussed in the next section show that the complex electron dynamics does indeed lead to fluctuations in the resistance image.

The second effect which could produce resistance fluctuations along the wire is the *quantum interference* of phase coherent paths which leads to fluctuations in the transmission as a function of tip position. Theories exist that predict conductance fluctuations in a disordered sample when the position of a single impurity is moved [65–68]. Given our sample parameters and the measurement temperature the phase coherence length can be estimated to be $\ell_\varphi \approx 8$ μm but the observability of phase coherence effects is limited by energy averaging through the Fermi-distribution function, [i.e., the thermal length $\ell_T = \sqrt{\hbar D/(kT)} \approx 2$ μm]. Fluctuations will have the order of magnitude $\Delta R = R^2 \Delta G \approx R^2 e^2/h(\ell_T/\ell_\varphi)(\ell_\varphi/L)^{3/2} \approx 100\ \Omega$. The characteristic length scale for these fluctuations is the Fermi wavelength of the electrons.

Figure 22 shows an image of the wire measured in a different measurement configuration, where a dc current was applied and the wire resistance was measured at the resonance frequency of the oscillating tip effectively resulting in a local transconductance measurement. This image taken with increased spatial resolution exhibits amazingly regular stripelike patterns on a length scale of less than 100 nm. Given the considerations about phase coherence the tentative explanation in terms of interference effects seems reasonable. Such quantum interference effects in scanning gate measurements will be further discussed in the next section.

4.3. Theoretical Considerations

Models for conductance fluctuations have been based on different approaches. Among them are the Anderson tight-binding Hamiltonian approach [68, 69], one-dimensional models incorporating randomly spaced delta scatterers [70, 71], and a scattering matrix approach [72].

In order to obtain some insight into the physical processes leading to the observed resistance images different types of theoretical models can be applied. A model aiming at classical effects in the resistance, for example, can be based on a calculation of the classical transmission of a wire within a billiard model following the general approach introduced by Beenakker and van Houten in [73]. If a fully quantum mechanical treatment is desired, a model can be based on the picture of noninteracting electrons in a multimoded wire. It can, for example, be similar to the scattering matrix approach used by Cahay and co-workers [72] for calculating the fluctuations in the resistance of a multimoded wire with randomly placed delta scatterers. In the following, we describe the scanning gate measurement based on such a quantum description.

We consider the Hamiltonian of a two-dimensional quantum wire with confinement-potential $V_c(y)$ and some scattering potential $U(x, y)$. The Schrödinger equation reads

$$\left[-\frac{\hbar^2}{2m^\star}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)+V_c(y)+U(x,y)\right]\psi(x,y)=E\psi(x,y)$$

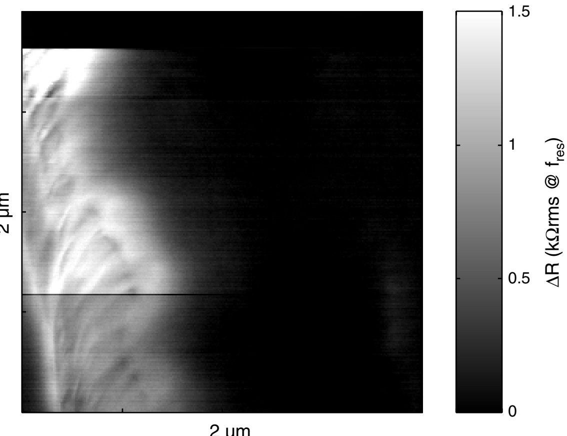


Figure 22. Interference effects in the quantum wire. Reprinted with permission from [63], T. Ihn et al., *Physica E* 12, 691 (2002). © 2002, Elsevier Science.

The tip-induced potential in the scanning gate experiment has a finite extent in the plane large compared to the Fermi wavelength. As a simplified example we consider the scattering potential

$$U(x, y) = \delta(x - x_i) \cdot \frac{\gamma}{b\sqrt{\pi}} e^{-[(y-y_i)/b]^2} \tag{11}$$

which takes into account the finite extent in the y-direction but neglects the extent in the x-direction. This allows us to illustrate some key features of the quantum system without going through the much more elaborate theory for the potential extended in all two directions. The resistance of the wire induced by the scatterer can then be calculated from the conductance formula of Landauer and Büttiker [74–76]:

$$G = \frac{e^2}{h} \mathrm{Tr}(\hat{t}\hat{t}^{\dagger})$$

For specific calculations we have used an infinite square well potential modelling the wire potential. Within this framework we first discuss the influence of the width of the tip-induced potential in the y-direction on cross sections through a resistance image along constant x (c.f. Fig. 21). The two extreme limits are intuitively clear: when the tip-induced potential is much wider than the wire, the wire will act as the probe for the tip induced potential and the cross section will have the width and shape of the tip-induced potential. On the other hand, if the wire width is large compared to the tip-induced potential, details of the internal electronic structure of the wire will become visible. Figure 23 shows the result of a calculation for a wire with width $W = 100$ nm in which four modes are occupied, corresponding to a Fermi energy $E_F = 14$ meV. For $b = 100$ nm the resistance shows a broad unstructured Gaussian peak when the tip crosses the wire. This peak becomes narrower at $b = 50$ nm and develops a flat top at 30 nm (i.e., when the width of the tip-induced potential becomes smaller than the wire width). Reducing b further results in additional fine structure reflecting the mode structure of the wire. With four occupied modes the highest mode has four maxima of the probability amplitude which is reflected in the resistance cross section. The condition for the observability of the mode structure in the wire is that the width of the tip-induced potential has to be smaller than the Fermi wavelength λ_F. Comparing these results to the cross sections through the resistance image shown in Figure 21, we conclude that the width of the tip-induced potential in the experiment was of the order of the wire width, but certainly not much larger.

This model can also be used to illustrate the possibility of imaging individual modes in a quantum wire or quantum point contact. Corresponding experiments on a quantum point contact have been published recently by Crook and co-workers. However, in their experiments, the two-dimensional electron gas with the quantum point contact is buried by 100 nm below the surface and a model for screening by mobile carriers in the remote doping plane had to be invoked in order to understand the experimental resolution of about 40 nm. Figure 24 shows calculated resistance changes due to a Gaussian tip potential ($b = 10$ nm) for a 100 nm wide wire with Fermi energies $E_F = 2.24, 5, 9$ meV corresponding to the occupation of one, two, and three modes. It can be seen that the number of local maxima in the resistance cross sections corresponds to the number of maxima in the probability density of the highest occupied mode, confirming the statement that the mode structure of narrow constrictions can be mapped with scanning gate techniques.

We continue the discussion of quantum effects in scanning gate measurements by a discussion of the case of many scatterers in the quantum wire. For simplicity, we use again δ-shaped potentials in the x-direction with Gaussian shape

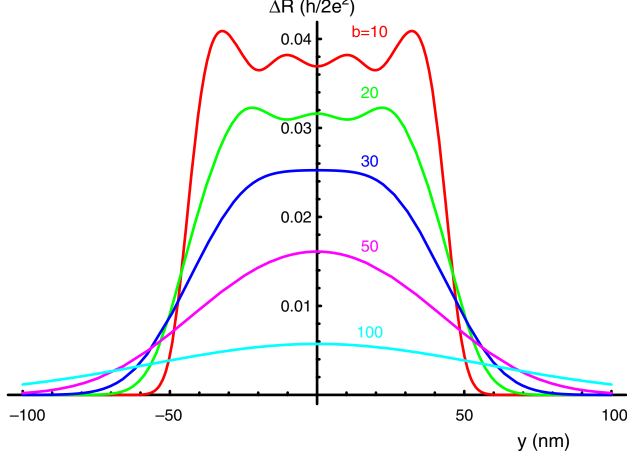


Figure 23. Calculated resistance change in a homogeneous wire due to scatterers with different widths $b = 100, 50, 30, 20, 10$ nm in y-direction and constant strength $\gamma = 4000$ meV nm^2 as a function of the y-position of the scatterer. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

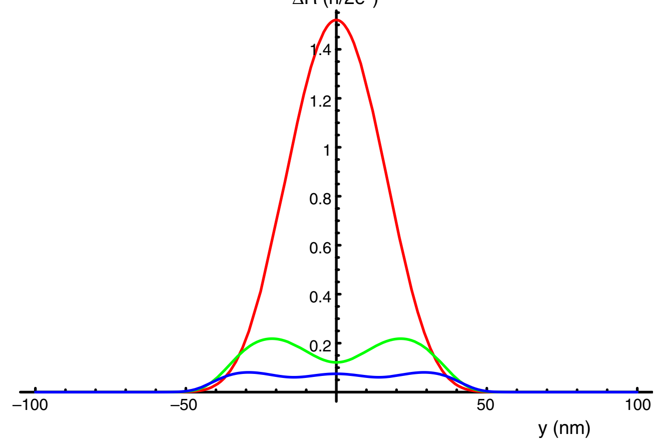


Figure 24. Calculated resistance change in a homogeneous wire with different Fermi energies $E_F = 2.24, 5, 9$ meV and hence different numbers of modes ($N = 1, 2, 3$, respectively). Here the width of the scatterer has been taken to be $b = 10$ nm; its strength is $\gamma = 4000$ meV nm^2. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

in y. In this case the formalism can easily be extended to the case of an arbitrary number of impurities. Qualitatively we now expect interference effects between electron waves to arise due to multiple scattering. Figure 25 shows a grayscale resistance image calculated for two fixed impurities in addition to the moving tip. In some sense, all three "perturbing potentials" form two neighboring cavities with transmission resonances changing when the sizes or shapes of the cavities are changed by moving the tip. Indeed, interference patterns can be seen along the wire axis. In contrast to the expectation for a very wide two-dimensional system, where interference effects have been predicted to occur on the scale of the Fermi wavelength λ_F [65–68], in a wire with a relatively small number of modes more than one frequency contributes to these fluctuations because the individual modes contribute with different Fermi wavelengths. In addition, it can be seen that not all interference stripes are exactly normal to the wire axis. Both findings are also in agreement with the measurement shown in Figure 22 where different frequencies of the fluctuations are observed as well and the stripes' orientation deviates from being normal to the wire axis. All this supports the interpretation of the measurements in terms of interference effects. The quantum interference effects look quite regular (i.e., they come in clearly discernible stripes), and they occur at a smaller length scale.

The aspect of a finite phase coherence length was not taken into account in the model. The obvious question is whether phase coherence is needed over the whole length of the wire to make the observation of interference effects possible. In a conventional transport experiment—where the conductance is measured as a function of magnetic field—the answer would certainly be "yes" because otherwise, the connection of many phase coherent sections in series fluctuating independently would average sufficiently. In *local* scanning gate experiments this argument is no longer true. There is always only *one* fluctuating segment (i.e., the one in which the tip is scanning), and no averaging takes place. Phase coherence effects can therefore be observed even in systems with characteristic lengths $L \gg \min(l_\varphi, l_T)$, provided that temperature is low enough to maintain phase coherence over length scales larger than the mean free path.

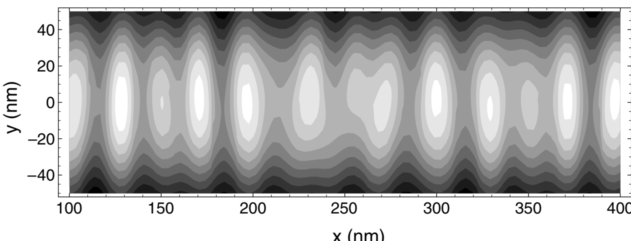


Figure 25. Calculated resistance change ΔR in a homogeneous 100 nm wide wire with four occupied modes ($E_F = 14$ meV). One fixed Gaussian impurity with $b = 50$ nm and strength $\gamma = 500$ meV nm^2 was positioned at $(x = 0, y = 40$ nm). A second Gaussian impurity had the parameters $b = 30$ nm, $\gamma = 500$ meV nm^2, and $(x = 500, y = -20)$. The scanned tip-induced potential was modelled with a second Gaussian potential with $b = 40$ nm and $\gamma = 4000$ meV nm^2. Reprinted with permission from [9], T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, Berlin, 2003. © 2003, Springer-Verlag.

This argument is also of relevance for the observations in [17], where scanning gate measurement were performed in the vicinity of a quantum point contact. It was observed that interference fringes do not disappear with increasing separation from the point contact. The explanation given in [17] is essentially identical to the argument presented. It implies that the phase coherence length cannot be read from a spatial resistance image obtained from scanning gate measurements. However, interference effects should disappear with increasing temperature. Scanning gate measurements at various temperatures could therefore allow one to determine the phase coherence length and to discriminate between fluctuations of the resistance due to classically chaotic motion and phase coherent effects.

Another aspect neglected in the model is interaction effects. In principle it is conceivable that due to potential inhomogeneities, electrons localize in the wire and form—in the extreme case—quantum dots exhibiting the Coulomb-blockade effect, similar to the situation in carbon nanotubes in the experiments of [59]. However, from the overall conductance of our wires which is smaller than e^2/h and the excellent sample quality we can conclude that interaction effects will play a negligible role in our experiments.

5. SCANNING PROBE EXPERIMENTS IN THE QUANTUM HALL REGIME

5.1. Edge Channels

The importance of the sample edge for the quantum Hall effect was pointed out soon after its discovery [77] on the basis of a model of noninteracting electrons [78]. Later on interactions were considered by using self-consistent descriptions of the sample edge [79–81]. Figure 26 shows the general notion of the self-consistent potential and the electron density at the sample edge in equilibrium. In the bulk of the electron gas (left in the figure) the Landau levels have an

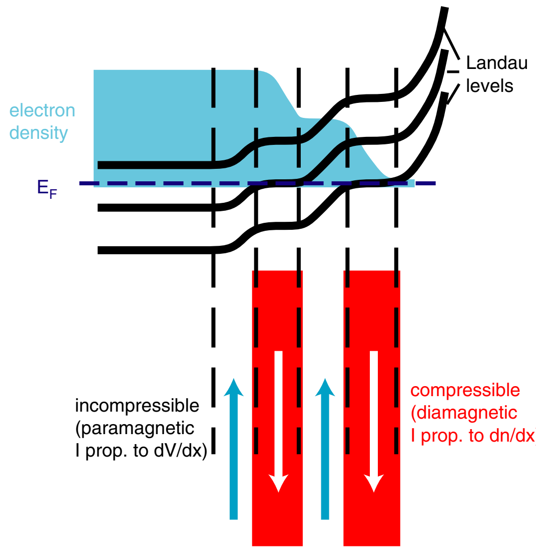


Figure 26. Self-consistent edge channel structure. Reprinted with permission from [82], T. Ihn et al., *Microelectron. Eng.* 63, 81 (2002). © 2002, Elsevier Science.

energetic spacing of $\hbar\omega_c$, where $\omega_c = eB/m$ is the cyclotron frequency, B is the magnetic field normal to the plane of the electron gas, and m is the effective mass of the electrons. The electron density in the bulk is constant. Near the edge the Landau levels have to follow the increasing electrostatic confinement potential and the electron density has to go to zero. Due to the density of states in a magnetic field with peaks in energy at the positions of the ideal Landau levels, the density of states at the Fermi energy oscillates strongly as a function of position leading to oscillating screening properties of the electron gas at the edge. Whenever a Landau level crosses the Fermi energy, screening is good and the potential flattens out. At the same time the electron density shows a strong gradient. These regions are called compressible strips because they have a finite density of states at the Fermi energy. The regions between compressible strips are called incompressible, since the density of states at the Fermi energy vanishes there. In these strips the potential has a strong gradient while the density is flat. Screening in the incompressible strips is poor. Geller and Vignale have calculated the equilibrium currents in the compressible and incompressible strips self-consistently [81]. They found that in the compressible regions the current is proportional to the density gradient and the current flow leads to a diamagnetic magnetic moment. In the incompressible regions the equilibrium current is proportional to the potential gradient and produces a paramagnetic effect. The subtle interplay between these two spatially separated types of current leads naturally to the oscillatory magnetization of a two-dimensional electron gas as a function of magnetic field [83].

Since the theoretical prediction of the existence of self-consistent edge channels and compressible and incompressible strips in the quantum Hall regime there has been a series of experimental attempts to measure local properties at sample edges in high magnetic fields. Electron–phonon interaction was used in first experiments [84]. Optical techniques with a spatial resolution of about 1 μm in the best case also supported the notion of edge channels [85–87]. Later experiments tried to detect the edge currents inductively, but evidence for bulk currents was found [88, 89]. Recently, edge strips were imaged using a metallic single electron transistor fabricated near the edge of a two-dimensional electron gas [90].

5.2. Scanning Probe Experiments

It is obvious to employ scanning probe techniques with their unprecedented potential of spatial resolution for such investigations and several experiments have been reported during the past few years using a scanning single-electron transistor [49], scanned potential microscopy [91], Kelvin probe techniques [12], and subsurface charge accumulation measurements [92–94]. In the following we give a brief overview of these experiments.

5.2.1. Scanning SET Measurements

The scanning single-electron transistor sensors described before have been employed for an investigation of GaAs/AlGaAs heterostructures with a two-dimensional electron gas 100 nm below the surface in [49]. The SET sensor was kept at a distance of 100–200 nm above the sample surface. Spatial resolution is estimated to be a few hundred nanometers.

The scanning SET was operated as part of a feedback loop keeping the SET current [Eq. (4)] and thereby the induced charge on the SET island constant at the operating point by controlling the voltage between the sensor and the 2DEG. This voltage is the measured output signal. Low-frequency (below 100 Hz) lock-in techniques were used in order to extract the small signals from transport through the 2DEG or the capacitance between the SET and the back-gate from the huge unwanted signals of biased surface gates and fluctuating surface dopant charges.

In the sample a back gate 5.4 μm below the 2DEG was used for tuning the electron density which was in the range of 10^{11} cm^{-2} at liquid He temperatures. Mobilities of about 4×10^6 were determined.

In so-called transparency measurements, the local compressibility of the two-dimensional electron gas (i.e., the local density of states at the Fermi level) was measured. This was accomplished by modulating the voltage between the back gate and the SET. The electron gas between these two "metallic" electrodes is kept at constant potential with respect to the sensor and therefore screens the ac field more or less, depending on the (local) filling factor. Compressible strips are found which screen the signal from the back gate well, while incompressible strips do not screen. The inhomogeneous local electron density can be deduced from such measurements by plotting contours of constant filling factors. In some places closed incompressible contours lead to isolated regions in the 2DEG which cannot be charged. Steplike features of the local chemical potential difference between 2DEG and SET could be measured with the sensor at a fixed position while an incompressible strip was made to pass the region below the tip by changing the back gate. Such steps reflect the steplike local self-consistent potential expected in the regions of the strips (see Fig. 26). It was also attempted to measure the Hall-voltage distribution between opposite edges of the sample. Around filling factor $\nu = 2$ a change from a linear to a flat voltage distribution was found. However, the full Hall voltage drop could not be measured due to screening effects by the top gate electrodes defining the edges. It remains to be discussed how far the presence of the metallic SET influences the distribution of potentials and currents, compressible and incompressible strips (e.g., by screening or by an insufficiently compensated work-function difference between 2DEG and SET).

5.2.2. Scanned Potential Microscopy

In another experiment, the quantum Hall effect has been locally investigated using a gold coated piezoresistive cantilever coupling capacitively to the two-dimensional electron system [91]. The 2DEG was again realized in a GaAs/AlGaAs heterostructure, 77 nm below the surface. Measurements were performed at temperatures between 0.7 and 1 K. An ac voltage was applied to the source contact of the Hall bar sample while the drain contact was grounded. This leads to a local electrostatic potential $V_{ac}(x, y)$ in the two-dimensional electron gas. The measured quantity in this

experiment is the force acting on the cantilever. It is dominated by the electrostatic force contribution given by

$$F = \frac{1}{2}C'(x,y)(V_{\text{tip}} - V_{\text{2DEG}} + \Delta\mu_{ch}(x,y)/e)^2 + Q_{\text{tip}}^{(0)'}(x,y)(V_{\text{tip}} - V_{\text{2DEG}} + \Delta\mu_{ch}(x,y)/e) \qquad (12)$$

in a classical capacitive model. Here, $C'(x,y) = \partial C(x,y,z)/\partial z$ is the z-derivative of the position dependent capacitance between tip and 2DEG, $Q_{\text{tip}}^{(0)'}$ is the z-derivative of the charge induced on the tip at zero bias due to fixed charges in the system (e.g., surface charges and ionized dopants), $\Delta\mu_{ch}(x,y)$ is the local work function difference between tip and sample, V_{tip} is the voltage on the tip, and $V_{\text{2DEG}} = -\mu_{elch}(x,y)/e$ is the local voltage in the 2DEG under the tip. In equilibrium the latter quantity is position independent. The idea of the measurement is to apply an ac transport current through the Hall bar sample at frequency ω which induces a position dependent $V_{\text{2DEG}}^{\text{ac, tr}}(x,y)$ reflecting the Hall voltage profile. The resulting ac force is

$$F_{\text{ac}}^{\text{tr}} = -\left[C'(x,y)(V_{\text{tip}} + \Delta\mu_{ch}(x,y)/e) + Q_{\text{tip}}^{(0)'}(x,y)\right]V_{\text{2DEG}}^{\text{ac, tr}}(x,y)$$

used for driving the cantilever at its resonance frequency at 120 kHz. The measured oscillation amplitude is proportional to the driving force (enhanced by the Q-factor of the oscillating cantilever on resonance) and therefore to the local electrochemical potential in the two-dimensional electron gas. At a fixed magnetic field the other quantities (capacitance and induced charge) may also contribute to the position dependence of the measured force, especially in the quantum Hall regime, where localization of electronic charge in the bulk of the Hall bar plays a crucial role and the time constants for establishing an equilibrium electrochemical potential in the bulk can be very long. A suitable reference measurement, in which the ac voltage is applied to all contacts of the Hall bar, implying that no transport current flows and no Hall voltage builds up, can be used for eliminating the undesired prefactor. In this case, the measured ac force, $F_{\text{ac}}^{\text{eq}}$, will be given by the expression for $F_{\text{ac}}^{\text{tr}}$ with the same prefactor, but a position dependent equilibrium voltage $V_{\text{2DEG}}^{\text{ac, eq}}$. Normalizing $F_{\text{ac}}^{\text{tr}}$ by $F_{\text{AC}}^{\text{eq}}$ gives

$$\frac{F_{\text{ac}}^{\text{tr}}}{F_{\text{ac}}^{\text{eq}}} = \frac{V_{\text{2DEG}}^{\text{tr}}(x,y)}{V_{\text{2DEG}}^{\text{eq}}(x,y)}$$

The normalized Hall voltage profile measured in this experiment at a given position along the Hall bar is found to be periodic in the inverse magnetic field $1/B$, in agreement with a general consideration of the effect of B on the voltage $V_{\text{2DEG}}^{\text{ac}}(x,y;B)$. The normalized Hall voltage profile at different filling factors ν around $\nu = 6$ shows a transition between a linear behaviour above and below the plateau in the Hall resistance and a profile with a flat region in the bulk with all the voltage dropping near the edges of the sample. This implies that sample edges and bulk do not equilibrate at these fields implying that the sample edges are effectively decoupled from the bulk. High dc transport currents with creating Hall voltages above the Landau level spacing $\hbar\omega_c$ lead to equilibration of edge channels. No compressible or incompressible strips were observed in this measurement, but the width of the edge state regions is found to increase with the number of edge channels with a width of about 300 nm per channel. In the flat region of the Hall resistance plateau the normalized Hall voltage profile was found to develop significant gradients within the bulk of the sample.

Related experiments for the measurement of Hall voltage profiles were performed by another group using a different technique [7, 12, 18, 95]. In contrast to the previous method not the force [Eq. (12)] is measured in this experiment, but the force gradient

$$F' = \frac{1}{2}C''(x,y)(V_{\text{tip}} - V_{\text{2DEG}} + \Delta\mu_{ch}(x,y)/e)^2 + Q_{\text{tip}}^{(0)''}(x,y)(V_{\text{tip}} - V_{\text{2DEG}} + \Delta\mu_{ch}(x,y)/e)$$

This is accomplished by driving the iridium coated piezoresistive cantilever externally on resonance and measuring the resonance frequency shift caused by the electrostatic force gradient with a feedback loop. Apart from this difference, the technique corresponds to the one presented already: a few Hz ac transport current is applied to the source contact of the 2DEG with the drain contact grounded. This induces a local low frequency variation of the electrochemical potential which in turn leads to a position dependent low frequency modulation of the frequency shift which can be detected using lock-in techniques. The position dependent prefactor of $V_{\text{2DEG}}^{\text{ac, tr}}$ is again eliminated by normalizing to the result of a reference measurements in which the ac voltage is applied to all the contacts of the electron gas. An advantage of not driving the cantilever with the tip–sample voltage lies in the fact that the very low modulation frequency of 3.4 Hz (compared to 100 kHz) can be used which minimizes RC effects in the charging of regions in the 2DEG with localized states.

Three different types of Hall potential profiles are observed in these experiments depending on the value of the filling factor ν (see Fig. 27): linear Hall potential drops are observed up to $\nu = 1.76$ when $\nu = 2$ is approached from lower filling factors. At filling factors around $\nu = 2$ the Hall potential drops rather arbitrarily in the bulk of the Hall bar. At filling factor $\nu = 2.14$ the potential becomes flat in the bulk and all the potential drop occurs near the edges of the sample. Proceeding to lower filling factors the same scenario repeats around filling factor $\nu = 4$. The behavior is consistent with the experimental results discussed before and it is related to the notion of compressible and incompressible regions in [18]. It is shown that the equilibrium center position of the innermost incompressible strip does indeed move with ν according to a theoretical prediction [96].

With this technique the Hall potential distribution was also investigated near potential probe contacts [95]. An incompressible strip is found at filling factors slightly above integer values which is not only present along the mesa edges, but also along the metal contact. It is concluded that the bulk is strongly decoupled from the edge and the ohmic contacts at these filling factors.

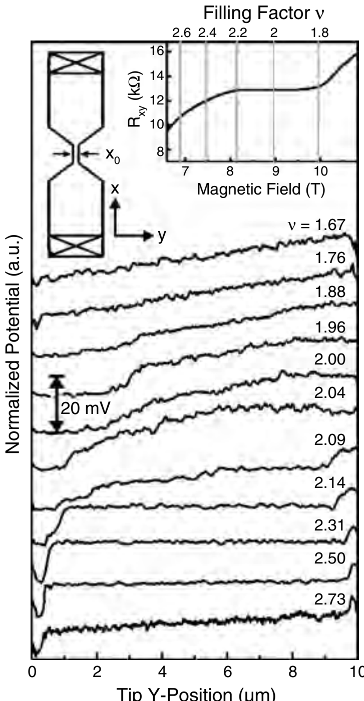


Figure 27. Normalized Hall potential profiles measured at different magnetic fields around filling factor $\nu = 2$. Reprinted with permission from [18], E. Ahlswede et al., *Physica B* 298, 562 (2001). © 2001, Elsevier Science.

5.2.3. Subsurface Charge Accumulation Imaging

The third set of experiments to be discussed here uses the so-called subsurface charge accumulation imaging in the quantum Hall regime at a temperature of 300 mK [10, 92–94, 97, 98]. This technique measures the local accumulation of charge in a two-dimensional electron gas in response to an ac voltage (100 kHz) applied between the conducting tip scanned at constant height above the surface and the 2DEG. The dc perturbation of the tip on the 2DEG is minimized by compensating the work function difference between the tip and the sample with a dc tip voltage of 0.4 V. The tip is connected to a sensitive charge detector which relies on a field-effect transistor mounted at cryogenic temperatures [28]. Signals in-phase and out-of-phase with the ac excitation are detected with lock-in techniques. Since most of the electric field lines between 2DEG and tip terminate far away from the tip apex there is a large background signal adding to the local signal of interest. Detailed fine structure on a length scale of about 100 nm can be resolved on the resulting images acquired in the quantum Hall regime, if the background signal is subtracted employing a bridge circuit [28]. Some of the observed structure in the experimental images can be explained using the idea of compressible and incompressible strips. The suppression of the local charging signal can arise due to two distinct reasons: either the local compressibility of the electron gas below the tip apex is low (low density of states at the electrochemical potential) or the conductivity of the 2DEG surrounding the tip position is too low to allow charging under the tip within half a period of the excitation (RC effect).

Possible models needed for an interpretation of the data are discussed in [98]. The simplest model entirely neglects RC effects and describes the measured capacitive signal as arising from a series connection of a plate capacitor with a geometric capacitance C_g and a quantum capacitance contribution $C_q = e^2\mathcal{D}(B)$ which is proportional to the local magnetic field dependent density of states $\mathcal{D}(x, y; B)$. The most involved model treats the electron gas on a discrete lattice with given local conductances and capacitances between neighboring sites. This conductance and capacitance pattern has to be optimized until the simulation fits the measured image.

The charge sensor is sensitive enough to detect individual charges entering the region under the tip. At suitable filling factors and tip–sample voltage a quantum dot can be formed under the tip which contains only a small number of electrons separated from the surrounding electron gas by a circular incompressible strip. Single electron charging of this dot has been observed [93] and the varying number of electrons in this dot as it is dragged along with the tip has been used for mapping the local potential in the Hall liquid (see Fig. 28).

5.2.4. Local Inter-edge-channel Tunnelling

In this section we discuss another type of local edge state imaging in the quantum Hall regime which is also based on low-temperature scanning probe techniques [82, 99]. In this approach the tunnelling current is measured between edge channels that are separately contacted. The tunnelling current is locally modified using a potential perturbation induced by the conducting tip of a scanning force microscope.

The samples are based on a GaAs/AlGaAs heterostructure with the two-dimensional electron gas residing 34 nm below the sample surface. Figure 29 shows a photograph of the structure which was prepared using photolithographic

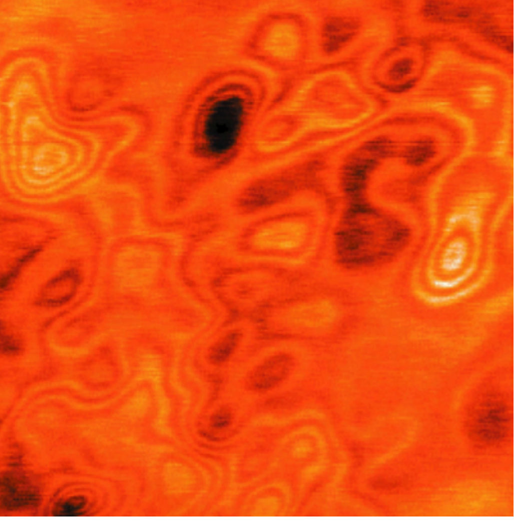


Figure 28. A 2.5×2.5 μm subsurface charge accumulation image taken at a filling factor $\nu \leq 1$. Reprinted with permission from [93], G. Finkelstein et al., *Science* 289, 90 (2000). © 2000, American Association for the Advancement of Science.

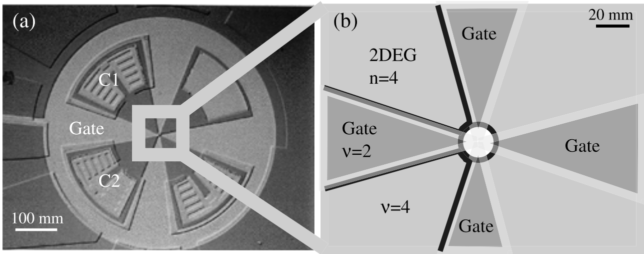


Figure 29. (a) Photograph of the circular mesa structure with the internal contacts C1 and C2. A small circular hole is etched into the center of the structure. (b) Schematic blow-up of the central part of the sample. For appropriate choice of V_G and B the indicated filling factors can be realized leading to the distribution of edge channels as indicated by the black lines. Reprinted with permission from [99], T. Ihn et al., *Physica E* 13, 671 (2002). © 2002, Elsevier Science.

techniques. The structure is essentially a circular mesa with a diameter of about 500 μm with a central hole of 20 μm in diameter. The sample is connected to the measurement setup via the internal ohmic contacts C1 and C2. A star-shaped gate electrode splits the two-dimensional electron gas into four ungated sectors.

The sample is mounted in a home-built low-temperature atomic force microscope utilizing a piezoelectric quartz tuning fork force sensor. The basic principle of the microscope and the characteristics of the tuning fork sensors have been described and can be found in [39, 42, 43]. Experiments were performed in the variable temperature insert of a standard ^{4}He cryostat at a temperature of 1.9 K. A voltage of 10 μV was applied between C1 and C2 and the two-terminal conductance was measured with a current–voltage converter.

Figure 29b shows a schematic magnification of the central part of the structure for illustrating the basic concept of our experiment. Using the appropriate combination of gate voltage, U_G, and magnetic field, B, we are able to set up a situation in which the bulk of the ungated electron gas has a Landau-level filling factor $\nu = 4$ (i.e., two spin-degenerate edge channels exist), while the gated regions have a filling factor of $\nu = 2$ supporting only one spin-degenerate edge channel. The latter will be able to circulate around the central hole, of the structure as shown in the figure. The other edge channel is not allowed under the gate. It will therefore come from C1 (or C2) along the edge of the gate, run in parallel with the $\nu = 2$ edge channel along the edge of the central hole, and then return to contact C1 (or C2) along the other edge of the gate. Since under these conditions the bulk regions of the gated as well as of the ungated 2DEG are insulating, the current from C1 to C2 involves tunnelling processes between the two edge channels. Spatially, tunnelling has to occur in regions where both edge channels run in parallel (i.e., in the vicinity of the edge of the central hole). Conceptually, this tunnelling transport is somewhat related to recent experiments on tunnelling through the edge states around an antidot [100]. However, the quantization of edge states around the central hole (the "antidot") plays no role in our structure. The central idea of the present experiment is that the tunnelling current between the two edge channels can be locally enhanced (or suppressed) by applying a local potential perturbation capacitively induced by the conducting tip.

In order to find the appropriate settings for V_G and B we characterized the structure by measuring the two-terminal conductance between contacts C1 and C2 as a function of these two parameters with the tip withdrawn from the surface of the structure. The result is shown as a grayscale image in Figure 30.

It can be seen that at $V_G < -0.08$ V the regions under the gate are completely pinched off and the conductance vanishes. At $V_G > -0.05$ V the grayscale plot is dominated by the Landau fan given by the 2DEG under the gate. The conductance oscillates as a function of B due to the Shubnikov–de Haas effect. The contribution of the ungated 2DEG can barely be seen in this range of V_G. However, in a very small range of gate voltage around $V_G \approx -0.07$ V vertical cuts through the plot reveal $1/B$-periodic oscillations which are independent of V_G but can hardly be seen on the grayscale plot (i.e., we observe the Shubnikov–de Haas oscillations of the ungated regions). They lead to the bright horizontal lines in the figure where the filling factors for the ungated regions are indicated. It can be seen from the figure that at $B \approx 5$ T and $V_G \approx 0.02$ V we achieve the desired $\nu = 2$ under the gate, while in the ungated regions we have $\nu = 4$.

While keeping the sample under these conditions the tip is now scanned across the sample surface near the edge of the inner hole. The tip was kept at a constant voltage $V_t = 0$ V, but due to the work function difference between the tip material PtIr and the sample there exists an effective electrostatic potential drop between tip and 2DEG which locally depletes the electron gas. Figure 31a shows a 5 μm × 5 μm topographical image of the edge of the central hole of the sample. In Figure 31b the conductance image can be seen measured simultaneously with the topography. The conductance of the sample is enhanced along a stripe of about 700 nm width which follows the curvature of the edge of the 2DEG. The image proves in a direct way that we can indeed enhance the tunnelling coupling between the two parallel edge channels by applying a local potential perturbation.

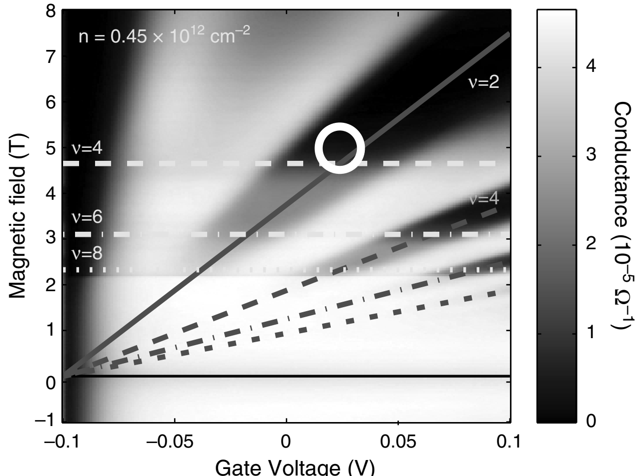


Figure 30. Conductance measured as a function of magnetic field and gate voltage. The bright horizontal lines indicate the filling factors in the ungated regions of the 2DEG. The gray Landau fan indicates filling factors in the gated regions of the 2DEG. Reprinted with permission from [99], T. Ihn et al., *Physica E* 13, 671 (2002). © 2002, Elsevier Science.

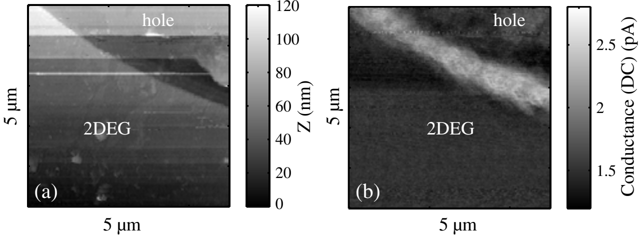


Figure 31. (a) Topographic image of the mesa edge defining the central hole etched into the 2DEG. (b) Conductance as a function of the tip position. The tip induced local potential leads to an enhanced tunnelling coupling of the two parallel edge channels and thereby to an increase in the conductance. Reprinted with permission from [99], T. Ihn et al., *Physica E* 13, 671 (2002). © 2002, Elsevier Science.

It is well established that the conductance or resistance of a sample in the quantum Hall regime is highly nonlocal [101, 102] and the phase-coherence length of electrons in edge channels can be macroscopically large. Therefore it is not *a priori* clear that the local perturbation induced by the presence of the tip will change the tunnelling coupling of the edge channels *locally*. Yet we argue that the self-consistent nature of the edge channel formation may well lead to a local enhancement of the tunnelling coupling. According to the adiabatic edge channel description [74] edge channels will follow equipotential lines along the sample edge. The local perturbing potential will affect the spatial run of the equipotential lines locally (i.e., on the scale of the screening length) and therefore we expect the edge channels to follow this perturbation on a local scale. This scenario is schematically depicted in Figure 32. Edge states form along the mesa edge as well as around the tip-induced anti-dot-like potential. When the tip is sufficiently remote from the edge (Fig. 32a), there is no mutual interaction between the two types of edges. When the tip is close enough to the mesa edge, the edge channel structure will change locally as illustrated in Figure 32b. The tunnelling coupling between the edge channels depends on the exponential overlap of the wave functions and therefore on the width of the incompressible stripe separating them. A spatial rearrangement of the self-consistent stripes will change the inter edge channel coupling. We suggest that the *local* change of the tunnelling coupling creates the observed conductance contrast in Figure 31.

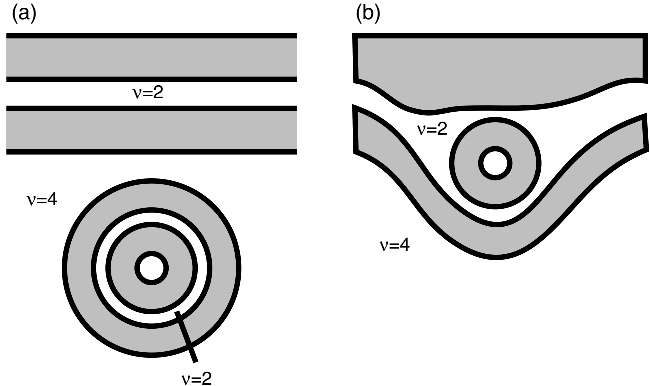


Figure 32. (a) Tip far away from the edge. Two edge states encircle the tip-induced potential hill. States at the sample edge are not affected. (b) Tip close to the edge of the sample. The tip induced potential locally changes the run of the edge currents and thereby modifies the tunnelling coupling between them. Reprinted with permission from [82], T. Ihn et al., *Microelectron. Eng.* 63, 81 (2002). © 2002, Elsevier Science.

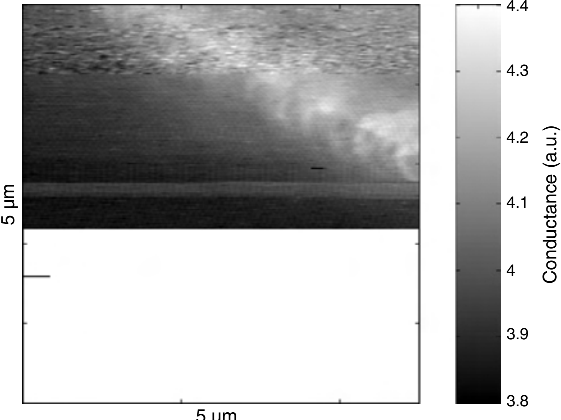


Figure 33. Image of the edge-channel coupling at enhanced spatial resolution. Reprinted with permission from [99], T. Ihn et al., *Physica E* 13, 671 (2002). © 2002, Elsevier Science.

The local nature of the conductance contrast makes the presented imaging method a promising tool for the local investigation of edge channel coupling and inter-edge-channel scattering (see [103] for a review and also [104]). As a matter of fact it can already be observed in Figure 31b that the bright stripe is not homogeneous but exhibits some internal structure. In order to illustrate this further we show in Figure 33 the conductance image obtained in the course of the same cooldown of the sample but with the tip in better condition. Unfortunately the tip shape can change during a cooldown (e.g., due to contaminations picked up from the sample surface). In some fortunate cases, however, the spatial resolution of the images is improved. In Figure 33 it is clearly visible that the extent to which the edge channel coupling can be enhanced depends strongly on the position of the tip even within the stripe of enhanced conductance. The image shows fringes leading away from the center of the stripe reminding one of the fascinating fine structure observed with other methods in the bulk of 2DEGs in the quantum Hall regime [10]. More investigation will help to further establish the presented imaging technique and to deepen our understanding of its interpretation.

GLOSSARY

AFM lithography Lithography technique using the tip of a scanning force microscope for patterning the surface of a sample. For example, with a conducting tip, local anodic oxidation of GaAs or Ti can be achieved when a suitable voltage is applied between the tip and the sample.

Cyclotron frequency Characteristic frequency for the motion of an electron in a magnetic field. If the magnetic field is applied normal to the plane of the electron motion, classical electrons move on circles. If the electron needs a time T for a single roundtrip, the cyclotron frequency is $2\pi/T$.

Landau levels Quantum mechanical energy levels for an electron in a magnetic field. In a two-dimensional system, Landau levels are equidistant in energy, like the states of a harmonic oscillator. However, each Landau level has a large degeneracy, i.e. it can accommodate a large number of electrons.

Mesoscopic system Electronic system in which several relevant length scales, like sample size, the phase coherence length and the mean free path are of the same order of magnitude. If the mean free path is small compared to the other two lengths, the system is called diffusive, otherwise ballistic.

Phase coherence length Characteristic distance over which quantum mechanical particles, e.g., electrons, show interference of their probability amplitudes. The phase coherence length is limited by dephasing processes, such as inelastic scattering.

Quantum Hall effect Arises in a two dimensional electron gas at low temperatures (typically below 4.2 K), when a strong magnetic field (of the order of 10 T) is applied normal to the plane of the electron motion. The Hall resistance is no longer rising linearly with the magnetic field strength, but exhibits plateaus at integer fractions of the fundamental resistance quantum h/e^2. The origin of the effect is the Landau quantization of electronic states (see Landau levels) and the localization of electrons due to potential inhomogeneties and interactions.

Quantum point contact Narrow constriction for electron flow with a conductance close to the conductance quantum $e^{2/h}$. The electron motion normal to the constriction is quantized into modes. Each mode traversing the point contact contributes one conductance quantum to the total point contact conductance. Therefore, the total conductance is quantized, i.e. it is no longer rising linearly with the width of the constriction, but exhibits plateaus at integer multiples of the conductance quantum.

Shubnikov-de Haas effect Oscillatory resistance of an electron gas in a magnetic field. The effect can be observed in metals (it was discovered in Bismuth), but is very pronounced in semiconductors at low temperatures. The effect is due to the Landau quantization of energy levels in a magnetic field (see Landau levels). It allows the determination of the electron density, for example, in two-dimensional electron gases.

Two-dimensional electron gas System in which electrons are free to move in two spatial directions, but quantized in the third. Two dimensional electron gases form, for example, at the interface between the semiconductors GaAs and AlGaAs, or the interface between Si and SiO_2 in a field effect transistor (MOS-FET).

REFERENCES

1. G. Binnig, C. Quate, and C. Gerber, *Phys. Rev. Lett.* 56, 930 (1986).
2. H. Hug, B. Stiefel, P. van Schendel, A. Moser, S. Martin, and H.-J. Güntherodt, *Rev. Sci. Instrum.* 70, 3625 (1999).
3. M. Eriksson, R. Beck, M. Topinka, J. Katine, R. Westervelt, K. Campman, and A. Gossard, *Appl. Phys. Lett.* 69, 671 (1996).
4. K. McCormick, M. Woodside, M. Huang, P. McEuen, C. Duruoz, and J. J. S. Harris, *Physica B* 249–251, 79 (1998).
5. D. Pelekhov, J. Becker, and J. G. Nunes, *Appl. Phys. Lett.* 72, 993 (1998).
6. R. Crook, C. Smith, C. Barnes, M. Simmons, and D. Ritchie, *Physica E* 6, 234 (2000).
7. P. Weitz, E. Ahlswede, J. Weis, K. Klitzing, and K. Eberl, *Appl. Surf. Sci.* 157, 349 (2000).
8. W. Allers, A. Schwarz, U. Schwarz, and R. Wiesendanger, *Rev. Sci. Instrum.* 69, 221 (1998).
9. T. Ihn, "Electronic Quantum Transport in Mesoscopic Semiconductor Structures," Springer Tracts in Modern Physics. Springer-Verlag, New York, 2004.
10. S. Tessmer, P. Glicofridis, R. Ashoori, L. Levitov, and M. Melloch, *Nature* 392, 51 (1998).
11. M. Tortonese, R. Barrett, and C. Quate, *Appl. Phys. Lett.* 62, 834 (1993).
12. P. Weitz, E. Ahlswede, J. Weis, K. Klitzing, and K. Eberl, *Physica E* 6, 247 (2000).
13. R. Crook, C. Smith, C. Barnes, M. Simmons, and D. Ritchie, *J. Phys.: Cond. Matter* 12, L167 (2000).
14. R. Crook, C. Smith, M. Simmons, and D. Ritchie, *Phys. Rev. B* 62, 5174 (2000).
15. R. Crook, C. Smith, M. Simmons, and D. Ritchie, *J. Phys.: Cond. Matter* 12, L735 (2000).
16. R. Crook, C. Smith, M. Simmons, and D. Ritchie, *J. Phys.: Cond. Matter* 13, L249 (2001).
17. M. Topinka, B. LeRoy, R. Westervelt, S. Shaw, R. Fleischmann, E. Heller, K. Maranowski, and A. Gossard, *Nature* 410, 183 (2001).
18. E. Ahlswede, P. Weitz, J. Weis, K. Klitzing, and K. Eberl, *Physica B* 298, 562 (2001).
19. K. Karrai and R. Grober, *Appl. Phys. Lett.* 66, 1842 (1995).
20. H. Edwards, L. Taylor, W. Duncan, and A. Melmed, *J. Appl. Phys.* 82, 980 (1997).
21. A. Ruiter, K. van der Werf, J. Veerman, M. Garcia-Parajo, W. Rensen, and N. van Hulst, *Ultramicroscopy* 71, 149 (1998).
22. W. Atia and C. Davis, *Appl. Phys. Lett.* 70, 405 (1997).
23. J. Salvi, P. Chevassus, A. Mouflard, S. Davy, M. Spajer, and D. Courjon, *Rev. Sci. Instrum.* 69, 1744 (1998).
24. D. Tsai and Y. Lu, *Appl. Phys. Lett.* 73, 2724 (1998).
25. M. Todorovic and S. Schultz, *J. Appl. Phys.* 83, 6229 (1998).
26. R. Steinke, M. Hoffmann, M. Böhmisch, J. Eisenmenger, K. Dransfeld, and P. Leiderer, *Appl. Phys. A* 64, 19 (1997).
27. M. Yoo, T. Fulton, H. Hess, R. Willett, L. Dunkleberger, R. Chichester, L. Pfeiffer, and K. West, *Science* 276, 579 (1997).
28. R. Ashoori, H. Stormer, J. Weiner, L. Pfeiffer, S. Pearton, K. Baldwin, and K. West, *Phys. Rev. Lett.* 68, 3088 (1992).
29. H. Birk, K. Oostveen, and C. Schönenberger, *Rev. Sci. Instrum.* 67, 2977 (1996).
30. J. Harley and T. Kenny, *Appl. Phys. Lett.* 75, 289 (1999).
31. C. Yuan, E. Batalla, M. Zacher, A. de Lozanne, M. Kirk, and M. Tortonese, *Appl. Phys. Lett.* 65, 1308 (1994).
32. P. Oden, P. Datskos, T. Thundat, and R. Warmack, *Appl. Phys. Lett.* 69, 3277 (1996).
33. D. Baselt, G. Lee, and R. Colton, *J. Vac. Sci. Technol. B* 14, 789 (1996).
34. S. Minne, P. Flueckiger, H. Soh, and C. Quate, *J. Vac. Sci. Technol. B* 13, 1380 (1995).
35. S. Minne, G. Yaralioglu, S. Manalis, J. Adams, J. Zesch, A. Atalar, and C. Quate, *Appl. Phys. Lett.* 72, 2340 (1998).
36. R. Beck, M. Eriksson, M. Topinka, R. Westervelt, K. Maranowski, and A. Gossard, *Appl. Phys. Lett.* 73, 1149 (1998).
37. P. Günther, U. C. Fischer, and K. Dransfeld, *Appl. Phys. B* 48, 89 (1989).
38. F. Giessibl, *Appl. Phys. Lett.* 73, 3956 (1998).
39. J. Rychen, T. Ihn, P. Studerus, A. Herrmann, and K. Ensslin, *Rev. Sci. Instrum.* 70, 2765 (1999).
40. F. Giessibl, *Appl. Phys. Lett.* 76, 1470 (2000).

41. R. Grober, J. Acimovic, J. Schuck, D. Hessmann, P. Kindlemann, J. Hespanha, A. Morse, K. Karrai, I. Tiemann, and S. Manus, *Rev. Sci. Instrum.* 71, 2776 (2000).
42. J. Rychen, T. Ihn, P. Studerus, A. Herrmann, and K. Ensslin, *Rev. Sci. Instrum.* 71, 1695 (2000).
43. J. Rychen, T. Ihn, P. Studerus, A. Herrmann, K. Ensslin, H. Hug, P. van Schendel, and H. Güntherodt, *Appl. Surf. Sci.* 157, 290 (2000).
44. T. Akiyama, U. Stauffer, and N. de Rooij, *Appl. Surf. Sci.* 210, 18 (2003).
45. "Single Charge tunnelling" (Grabert and Devoret, Eds.). Plenum Press, New York, 1991.
46. Y. Wei, J. Weis, K. Klitzing, and K. Eberl, *Appl. Phys. Lett.* 71, 2514 (1997).
47. S. Ilani, A. Yacoby, D. Mahalu, and H. Shtrikman, *Phys. Rev. Lett.* 84, 3133 (2000).
48. H. Hess, T. Fulton, M. Yoo, and A. Yacoby, *Solid State Commun.* 107, 657 (1998).
49. A. Yacoby, H. Hess, T. Fulton, L. Pfeiffer, and K. West, *Solid State Commun.* 111, 1 (1999).
50. N. Zhitenev, T. Fulton, A. Yacoby, H. Hess, L. Pfeiffer, and K. West, *Nature* 404, 473 (2000).
51. N. Zhitenev, T. Fulton, A. Yacoby, H. Hess, L. Pfeiffer, and K. West, *Physica E* 6, 242 (2000).
52. E. Betzig, J. Trautmann, T. Harris, J. Weiner, and R. Kostelak, *Science* 251, 1468 (1991).
53. M. Nonnenmacher, M. O'Boyle, and H. Wickramasinghe, *Appl. Phys. Lett.* 58, 2921 (1991).
54. S. Belaidi, P. Girard, and G. Leveque, *J. Appl. Phys.* 81, 1023 (1997).
55. Landoldt-Börnstein, "Semiconductors," Vol. 17a of *Numerical Data and Functionals Relationships in Science and Technology.* Springer-Verlag, Berlin, 1982.
56. H. U. Baranger and A. Stone, *Phys. Rev. B* 40, 8169 (1989).
57. M. Büttiker, *Phys. Rev. Lett.* 57, 1761 (1986).
58. G. Salis, B. Graf, K. Ensslin, K. Campman, K. Maranowski, and A. C. Gossard, *Phys. Rev. Lett.* 79, 5106 (1997).
59. M. Woodside and P. McEuen, *Science* 296, 1098 (2002).
60. S. Tans and C. Dekker, *Nature* 404, 834 (2000).
61. A. Bachtold, M. Fuhrer, S. Plyasunov, M. Forero, E. Anderson, A. Zettl, and P. McEuen, *Phys. Rev. Lett.* 84, 6082 (2000).
62. M. Bockrath, W. Liang, D. Bozovic, J. Hafner, C. Lieber, M. Tinkham, and H. Park, *Science* 291, 283 (2001).
63. T. Ihn, J. Rychen, T. Cilento, K. Ensslin, W. Wegscheider, and M. Bichler, *Physica E* 12, 691 (2002).
64. R. Held, T. Vancura, T. Heinzel, K. Ensslin, M. Holland, and W. Wegscheider, *Appl. Phys. Lett.* 73, 262 (1998).
65. B. Altshuler and B. Spivak, *JETP Lett.* 42, 447 (1985).
66. S. Feng, P. Lee, and A. Stone, *Phys. Rev. Lett.* 56, 1960 (1986).
67. S. Feng, P. Lee, and A. Stone, *Phys. Rev. Lett.* 56, 2772 (1986) [Erratum].
68. N. Giordano, *Phys. Rev. B* 36, 4190 (1987).
69. A. Stone, *Phys. Rev. Lett.* 54, 2692 (1985).
70. J. Sak and B. Kramer, *Phys. Rev. B* 24, 1761 (1981).
71. C. Lambert and M. Thorpe, *Phys. Rev. B* 27, 715 (1983).
72. M. Cahay, M. McLennan, and S. Datta, *Phys. Rev. B* 37, 10125 (1988).
73. C. Beenakker and H. van Houten, *Phys. Rev. Lett.* 63, 1857 (1989).
74. C. Beenakker and H. van Houten, in "Semiconductor Heterostructures and Nanostructures" (H. Ehrenreich and D. Turnbull, Eds.), Vol. 44 of *Solid State Physics*, pp. 1–228. Academic Press, San Diego, 1991.
75. S. Datta, "Electronic Transport in Mesoscopic Systems," vol. 3 of *Cambridge Studies in Semiconductor Physics and Microelectronic Engineering.* Cambridge Univ. Press, Cambridge, UK, 1995.
76. D. Ferry and S. Goodnick, "Transport in Nanostructures," no. 6 in Cambridge Studies in Semiconductor Physics and Microelectronic Engineering. Cambridge Univ. Press, Cambridge, UK, 1997.
77. K. von Klitzing, G. Dorda, and M. Pepper, *Phys. Rev. Lett.* 45, 494 (1980).
78. B. Halperin, *Phys. Rev. B* 25, 2185 (1982).
79. D. Chklovskii, B. Shklovskii, and L. Glazman, *Phys. Rev. B* 46, 4026 (1992).
80. D. Thouless, *Phys. Rev. Lett.* 71, 1879 (1993).
81. M. Geller and G. Vignale, *Physica B* 212, 283 (1995).
82. T. Ihn, J. Rychen, K. Ensslin, W. Wegscheider, and M. Bichler, *Microelectron. Eng.* 63, 81 (2002).
83. L. Bremme, T. Ihn, and K. Ensslin, *Phys. Rev. B* 59, 7305 (1999).
84. A. Kent, D. McKitterick, P. Hawker, and M. Henini, *Helv. Phys. Acta* 65, 331 (1992).
85. R. Knott, W. Dietsche, K. Klitzing, K. Eberl, and K. Ploog, *Semicond. Sci. Technol.* 10, 117 (1995).
86. A. Shashkin, A. Kent, P. Harrison, L. Eaves, and M. Henini, *Phys. Rev. B* 49, 5379 (1994).
87. R. van Haren, F. Blom, and J. Wolter, *Phys. Rev. Lett.* 74, 1198 (1995).
88. E. Yahel, D. Orgad, A. Palevski, and H. Shtrikman, *Phys. Rev. Lett.* 76, 2149 (1996).
89. E. Yahel, A. Tsukernik, A. Palevski, and H. Shtrikman, *Phys. Rev. Lett.* 81, 5201 (1998).
90. Y. Wei, J. Weis, K. Klitzing, and K. Eberl, *Phys. Rev. Lett.* 81, 1674 (1998).
91. K. McCormick, M. Woodside, M. Huang, M. Wu, P. McEuen, C. Duruoz, and J. J. S. Harris, *Phys. Rev. B* 59, 4654 (1999).
92. G. Finkelstein, P. Glicofridis, S. Tessmer, R. Ashoori, and M. Melloch, *Phys. Rev. B* 61, R16323 (2000).
93. G. Finkelstein, P. Glicofridis, R. Ashoori, and M. Shayegan, *Science* 289, 90 (2000).
94. G. Finkelstein, P. Glicofridis, S. Tessmer, R. Ashoori, and M. Melloch, *Physica E* 6, 251 (2000).
95. E. Ahlswede, J. Weis, K. v. Klitzing, and K. Eberl, *Physica E* 12, 165 (2002).
96. K. Lier and R. Gerhardts, *Phys. Rev. B* 50 (1994).
97. P. Glicofridis, G. Finkelstein, R. Ashoori, and M. Shayegan, *Phys. Rev. B* 65, 121312 (2002).
98. S. Tessmer, G. Finkelstein, P. Glicofridis, and R. Ashoori, *Phys. Rev. B* 66, 125308 (2002).
99. T. Ihn, J. Rychen, K. Ensslin, W. Wegscheider, and M. Bichler, *Physica E* 13, 671 (2002).
100. I. Maasilta and V. Goldman, *Phys. Rev. B* 57, R4273 (1998).
101. P. McEuen, A. Szafer, C. Richter, B. Alphenaar, J. Jain, A. Stone, R. Wheeler, and R. Sacks, *Phys. Rev. Lett.* 64, 2062 (1990).
102. M. Buttiker, *Phys. Rev. B* 38, 9375 (1988).
103. R. Haug, *Semicond. Sci. Technol.* 8, 131 (1993).
104. B. Alphenaar, P. McEuen, R. Wheeler, and R. Sacks, *Phys. Rev. Lett.* 64, 677 (1990).
105. PI Ceramic GmbH, Lindenstrasse, 07589 Lederhose, Germany.
106. EBL-4 by Staveley Sensors Inc., 91 Prestige Park Circle, East Hartford, CT 06108, USA.
107. supplied by Goodfellow, Ermine Business Park, Huntingdon, PE29 6WR, England.
108. EBL-4 by Staveley Sensors Inc, 91 Prestige Park Circle, East Hartford, CT 06108, USA.
109. EPO-TEK H77 by Polyscience AG, Riedstrasse 13, 6330 Cham, Switzerland.
110. NTF3238 from SaRonix, 141 Jefferson Drive, Menlo Park, CA 94025, USA.
111. EPO-TEK H20E by Polyscience AG, Riedstrasse 13, CH-6330 Cham, Switzerland.
112. AFM stands for atomic force microscope, synonym for scanning force microscope (SFM).

www.aspbs.com/enn

Scanning Tunneling Microscopy of Carbon Nanotubes

László P. Biró
Research Institute for Technical Physics and Materials Science, Budapest, Hungary

Philippe Lambin
Universitaires Notre-Dame de la Paix, Namur, Belgium

CONTENTS

1. Introduction
2. Scanning Tunneling Microscopy Theory
3. Scanning Tunneling Microscopy and Scanning Tunneling Spectroscopy of Multiwall Carbon Nanotubes
4. Scanning Tunneling Microscopy and Scanning Tunneling Spectroscopy of Single-Wall Carbon Nanotubes
5. Computer Modeling of Scanning Tunneling Microscopy Images
6. Structural Defects
7. Complex Tubular Structures
8. Functionalized Nanotubes
9. Conclusions
Glossary
References

1. INTRODUCTION

Carbon nanotubes are perhaps one of the most exciting members of the family of carbon nanostructures initiated by the discovery of fullerene almost two decades ago [1]. Their very promising properties [2, 3] brought single wall carbon nanotubes (SWCNTs), and multiwall carbon nanotubes (MWCNTs) to the focus of attention of researchers in various fields ranging from nanoelectronics to composite materials. In addition to the straight carbon nanotubes discovered by Iijima in 1991 [4], other more complex but tubular-type structures built of layers of sp^2 hybridized carbon, like Y-junctions, helical coils, and multiple coils, have been predicted theoretically [5–8] and observed experimentally by transmission electron microscopy (TEM) and scanning tunneling microscopy (STM) [9–13].

A SWCNT consists of one atomic layer thick sheet with a graphitelike structure (graphene sheets) seamlessly rolled into a cylinder (Fig. 1), while MWCNTs are built of a set of graphene cylinders with increasing diameter, concentrically placed in each other. The diameter and helicity of a defect-free SWCNT are uniquely characterized by the vector $\boldsymbol{C} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2 \equiv (n, m)$ that connects crystallographically equivalent sites on a two-dimensional graphene sheet, where $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are the graphene lattice vectors, and n and m are integers (Fig. 2). The interlayer spacing of MWCNTs is consistent with that of turbostratic graphite, 0.34 nm [14]. Typical diameter values for SWCNTs are around 1 nm, while MWCNTs have diameters in the range of a few tens of nanometers. Their structure confers the nanotubes with special properties, very closely determined by the way in which the C–C bonds in the graphene sheet are oriented with respect to the nanotube axis, and by the diameter of the tube [2, 3] (Fig. 2). Depending on the way in which the graphene sheet is wrapped into a cylinder, SWCNTs may be metallic or semiconducting [2]. One-third of all possible SWCNTs exhibit metallic behavior [i.e., their electronic density of state (DOS) at the Fermi energy is different from zero], while two-thirds show semiconducting behavior, with zero DOS around the Fermi energy. MWCNTs are constituted from a random alternation of semiconducting and metallic layers [15]. Due to the strong interdependence between atomic structure and electronic properties, atomic resolution topographic STM images and scanning tunneling spectroscopy (STS) are extremely useful in investigating and understanding the physical properties of carbon nanotubes. STM is the only tool able to give information about both the atomic arrangement and the electronic structure of the very same nanotube.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

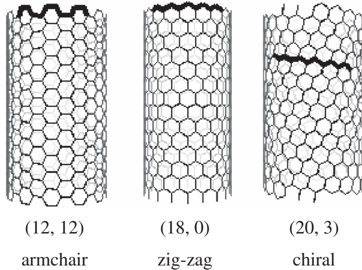


Figure 1. Schematic structure of the three main types of carbon nanotubes.

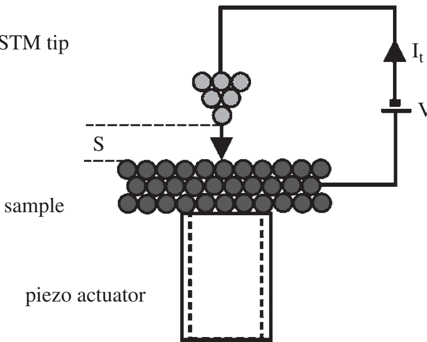


Figure 3. Schematic principle of operation of the STM (not to scale).

2. SCANNING TUNNELING MICROSCOPY THEORY

In principle, the concept of the STM is simple (Fig. 3): an atomically sharp, metallic tip is brought within a distance s, of the order of a few tenths of a nanometer, to a conducting surface. Due to their quantum mechanical behavior, electrons may tunnel with equal probability from the tip to the surface and vice versa. If an external bias V is applied between the tip and the surface a preferential tunneling direction is imposed which will yield an average current, I_t. The tip–sample distance s is regulated using a piezoactuator and a feedback loop. The tunneling current measured at a given moment is taken as input signal for the feedback loop.

A comprehensive overview of the theoretical models used to describe the tunneling through a one-dimensional potential barrier—the simplest case of tunneling—is given by Wiesendanger [16]. Two major conclusions arise from the various models:

(1) The transmission T of electrons through a one-dimensional potential barrier depends exponentially on the width s of the barrier:

$$T \propto e^{-2\chi s} \tag{1}$$

(2) The decay rate χ, characterizing the decrease of the probability of the electron to cross over the barrier, is

$$\chi = \frac{2\pi[2m(V_0 - E)]^{1/2}}{h} \tag{2}$$

where V_0 is the average height of the barrier, and E is the energy of the tunneling electron. Independently of the exact shape of the barrier, the strong exponential dependence of T with the barrier width s and the square root of the effective barrier height, $(V_0 - E)^{1/2}$, is typical of tunneling. The exponential dependence makes the tunneling channel be very narrow, thus making possible the atomic resolution.

The tunnel current flowing between the tip and the sample at a bias voltage V will be

$$I(V) \propto \int_0^{eV} \rho_s(E)\rho_t(E - eV)T(E)\,dE \tag{3}$$

where ρ_s is the sample electronic DOS, ρ_t is the DOS of the tip, and T is the transmission coefficient of the barrier. From formula (3) one may observe that the STM offers the opportunity to acquire not only topographic information about the sample but also information regarding the DOS of the sample. If the tip DOS is flat and the transmission coefficient T may be taken as constant, then Eq. (3) reduces to

$$I(V) \propto \int_0^{eV} \rho_s(E)\,dE \tag{4}$$

Then the quantity dI/dV, named differential conductance, will be proportional to $\rho_s(V)$, the sample electronic DOS. However, T may be regarded as constant only in the limit of small voltage variations; it gives deviation from this simple dependence at large bias values.

The real case of an STM measurement is different from the simple case of tunneling through a one-dimensional potential barrier. First of all, in the case of a sharp tip, one has a three-dimensional potential barrier instead of a one-dimensional one. The real potential barriers may strongly differ from the rectangular shape assumed in deducing the formulas (1) and (2). The most frequently used model for the interpretation of STM experiments is the model given by

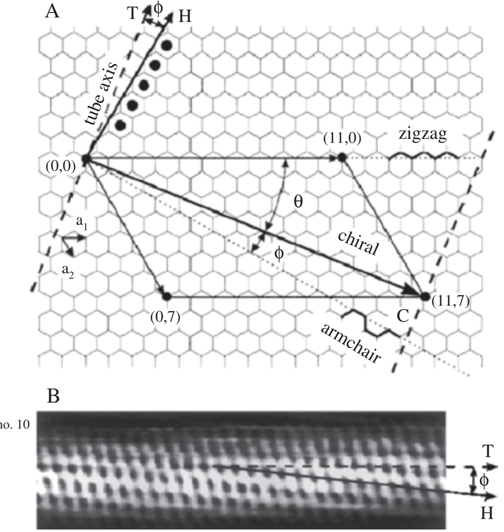


Figure 2. Relation between the rolling of the graphene sheet into a (11, 7) SWCNT (A) and the atomic resolution STM image of a (11, 7) nanotube (B). Reprinted with permission from [45], J. W. G. Wildöer et al., *Nature* 391, 59 (1998). © 1998, Macmillan Magazines Ltd.

Tersoff and Hamann [17]. In this theory the tip is treated as a single *s orbital* with a constant DOS and the tunnel current is proportional to the sample DOS computed at the location of the tip and at the Fermi energy at vanishingly small bias. A more comprehensive and detailed treatment of tunneling is given in the book on STM edited by Wiesendanger and Güntherodt [18].

In practical STM instruments, the positioning and scanning of the STM tip is achieved by piezoelectric actuators. The width of the STM gap is controlled by a feedback loop which keeps the value of the tunneling current at a setpoint value selected by the operator. An STM can operate in several regimes; the two most important ones are as follows:

- Topographic (constant current) imaging: the feedback loop is on, the image is generated from the values of the voltage applied to the piezoactuator to maintain a constant value of the tunneling current. Provided the electronic structure at the sample surface is homogeneous, the topographic profile of the surface will be generated.
- Current–voltage spectroscopy, frequently called scanning tunneling spectroscopy. The scanning and the feedback loop are switched off, the value of the tunneling gap is fixed, the bias voltage is ramped from $-U$ to $+U$, and the corresponding current variations are recorded. The function dI/dV gives information about the local DOS of the sample.

To understand the results of the STM imaging of carbon nanotubes it is useful to keep in mind a few particularities of the tunneling system specific to this case:

- For being imaged in an STM, a nanoscopic object like a carbon nanotube has to be placed on an atomically flat, conducting surface, the support. Highly oriented pyrolitic graphite (HOPG) or Au (111) terraces are the most frequently used substrates [19].
- Opposite from the case of tunneling into a bulk sample, when a carbon nanotube is imaged, the electrons have to cross two tunnel gaps: one between the STM tip and the nanotube, the other between the nanotube and its support [20, 21].
- The electronic structure of the carbon nanotube and its support may be different. As it follows from formula (3), the different values of $\rho_{s(\text{support})}(V)$ and $\rho_{s(\text{nanotube})}(V)$ will yield different values of tunneling current, thus having influence on the height values measured from topographic images [22].
- When the STM tip comes very close to the nanotube, or when switching from pure tunneling to a point contact imaging regime occurs [23], the nanotube may suffer elastic deformation under the STM tip.
- When imaging a nanometer sized object, important convolution effects arise between the object shape and the tip shape. This phenomenon yields the apparent lateral broadening of the imaged tube [24]. The diameter D is related to the apparent height h of the tube by the approximate relation $D = (8Rh)^{1/2}$, where R is the curvature radius of the tip [23]. The height of the topographic profile h is close to the diameter of the tube, and h is often used as a measure of d. But the apparent height depends on the adsorption distance of the nanotube above the substrate and on the difference between the tunneling distances above the tube and over the substrate arising from their different electronic properties. All these distances are not known with precision. Using the apparent height h of the nanotube seems to *underestimate* the actual diameter by as much as 0.2–0.5 nm [25], perhaps due to the radial compression of the nanotube exerted by the tip, or/and due to the second tunneling gap between tube and support the current must go through as already discussed. A fitting procedure of the tunneling current has been proposed [26] for the better determination of the diameter from the measured height.
- Distortion effects may arise from a “floppy” tip. Such a tip may be produced when the rigid STM tip picks up a carbon nanotube from the sample [27]. In this case the attached nanotube will act as the tip generating the image. Such a tip may cause in the topographic STM image reduced apparent height and dips on the sides of the imaged nanotube.

3. SCANNING TUNNELING MICROSCOPY AND SCANNING TUNNELING SPECTROSCOPY OF MULTIWALL CARBON NANOTUBES

Multiwall carbon nanotubes are a variety of nanotubes discovered by Iijima in 1991 [4]. STM was the ideal tool to characterize the topography and the electronic properties of these nano-objects. Topographic and voltage dependent STM imaging of arc grown MWCNTs in a bundle was first reported by Zhang and Lieber [28]. The three-dimensional structure of the bundles was investigated, and it was found that the large bundles consist of closely packed nanotubes with diameters on the order of 10 nm. A switchover from stable imaging when using bias voltages larger than 200 mV to unstable tunneling below this value was found. The observation was interpreted as experimental evidence for the existence of a conduction gap of the nanotube bundle around the Fermi energy. The first STS measurement on individual carbon nanotubes was reported by Olk and Heremans [29]. Their measurements were carried out in air on MWCNTs grown by the electric arc method transferred onto an Au substrate by ultrasonication in ethanol. Both semiconductor and metallic carbon nanotubes were found. The comparison of measured energy gap and experimental diameter values with gap values predicted by theory [30, 31] showed an increasing deviation with decreasing tube diameter. The deviations may originate from the unavoidable error introduced in the measured diameter by tip/sample convolution effects, the more complex structure of the tunneling region than in the well known case of a flat, homogeneous sample [20, 22], and effects arising from the multiple walls of the MWCNTs. On the other hand, it is not known yet to what extent the inner layers of a MWCNT may play a role during tunneling through the nanotube. Theoretical calculations show that the inner layers of a MWCNT may rotate in

the larger diameter outer layer [32]. However, the first STM experiment proving that atomic resolution is possible on a carbon nanotube reported by Ge and Sattler [33] was carried out on MWCNTs. The MWCNTs were produced *in-situ* by the condensation of evaporated carbon on a HOPG substrate. A similar atomic resolution image is reported as in the case of HOPG. The triangular lattice seen on the surface of HOPG is caused by the nonequivalence of the carbon atoms (in the Bernal graphite stacking) produced by interlayer interaction [34, 35]. Thus a similar atomic resolution pattern in the case of a MWCNT like that of HOPG is an indication of interlayer interaction. Additionally Ge and Sattler [33] found a periodicity of 16 nm along the tube axis superimposed on the atomic lattice. This was interpreted as arising from the misorientation of the outer two layers of the MWCNT, analogously with the generation of Moiré patterns known in geometric optics. The STM observation of Moiré patterns on HOPG [36] is well known experimentally; however, no clear theoretical description has been given yet.

Although discovered first, both the experimental and theoretical understanding of MWCNTs is behind that of SWCNTs. Atomic resolution images acquired on MWCNTs grown by chemical vapor deposition [22] were interpreted in a way similar to the work by Ge and Sattler [33]. Hassanien and co-workers [37] interpreted their atomic resolution images taken on MWCNTs like in the case of SWCNTs (i.e., like for a single shell system). In a more recent work Hassanien and co-workers compared atomic resolution images of SWCNTs, MWCNTs, and HOPG, and they concluded that the atomic resolution images on MWCNTs and HOPG are similar. Moreover, the difference between α and β sites (β sites do not have a neighbor in the layer below) was evidenced by STS, too [38].

Topology related changes in the local density of states near the ends of closed MWCNTs were investigated using spatially resolved scanning tunneling spectroscopy and tight binding calculations. Sharp resonant valence band states were found at the tube ends, dominating the valence band edge and filling the bandgap. Calculations show that the strength and position of these states with respect to the Fermi level depend sensitively on the relative positions of pentagons and their degree of confinement at the tube ends [39, 40]. In accordance with theoretical predictions [41], STS measurements on MWCNTs grown directly on the HOPG support for STM measurements showed that for MWCNTs of diameter exceeding 30 nm no differences can be evidenced between the nanotube and the HOPG [42].

STS was successfully used to investigate the effect of mechanical strain on the electronic structure of MWCNTs on a support. The spectroscopic data show that while small deformations do not have a marked effect on the local electronic structure, high strain, like that due to abrupt substrate morphology change, showed a drastic perturbation in the local electronic structure of the tube [43]. This experimental finding is an agreement with theoretical work on the effects of defects produced by strain in the STM image of SWCNTs and alterations in their transport properties [44].

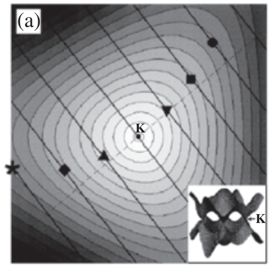
4. SCANNING TUNNELING MICROSCOPY AND SCANNING TUNNELING SPECTROSCOPY OF SINGLE-WALL CARBON NANOTUBES

The first atomic resolution topographic STM images and STS results on SWCNTs were reported simultaneously by two groups [45, 46] on SWCNTs grown by laser ablation, imaged in a low temperature STM on Au substrate. The atomic resolution images allow the determination of the chiral angle Φ between the tube axis and the centers of the closest row of hexagons appearing in the STM image as dark spots (Fig. 2).

The tube diameter and chirality allow the structural identification of the nanotube [2]. The STS results are in good agreement with the calculated densities of states of SWCNTs [2, 30, 47, 48]. Both metallic and semiconductor carbon nanotubes were found. The typical gap values found for semiconductor nanotubes with diameters in the 1.2 to 1.9 nm range were of the order of 0.5 eV, while the metallic carbon nanotubes with diameters in the range of 1.1 to 2.0 nm had subband separations of the order of 1.7 eV [45]. Both groups found that for semiconducting SWCNTs the measured gap values were proportional to $1/d_t$, where d_t is the diameter of the nanotube. Some differences found in the proportionality factors may arise from the differences in taking into account the systematic errors in the determination of the tube diameter due to convolution effects. In first approximation, the distances between van Hove singularities across the gap ΔE_{11} of a semiconducting SWCNT are found to respect the rule $\Delta E_{33} = 4 \times \Delta E_{11}$ and $\Delta E_{22} = 2 \times \Delta E_{11}$, while for metallic nanotubes, the rule is $\Delta E_{33} = 3 \times \Delta E_{11}$ and $\Delta E_{22} = 2 \times \Delta E_{11}$ [49, 50], where ΔE_{11} is the subband separation. This relation may be very useful in experimentally identifying metallic and semiconducting SWCNTs from often noisy STS data.

In the interpretation of the spectroscopic data it has to be taken into account that due to work function differences, the Au substrate may produce the doping of the nanotube by charge transfer [45]. The center of the gap that separates the first two van Hove singularities may be offset toward the valence band by as much as 0.2–0.3 eV. Apparently, there is no systematic difference between metallic and semiconducting tubes [45].

Sharp van Hove singularities found by STS in the DOS of SWCNTs imaged in atomic resolution allowed comparison of these data to tight-binding calculations for specific tube indices (n, m) [51]. The increasing deviation (rounded triangular contours in Fig. 4a) from isotropic dispersion of the π-bands of graphene around the **K** point of the first Brillouin zone (circular contours) with increasing distance from **K** is taken into account. The trigonal anisotropy causes the increase in the splitting of the second and third van Hove singularities as compared with the first one [52]. A remarkably good agreement was obtained between experiment and π-only calculations. Pronounced peaks in the DOS were also found at the end of an atomically resolved metallic SWCNT. Comparison of these data with calculations suggests that the topological arrangement of pentagons is responsible for the localized features in the experiment [51].



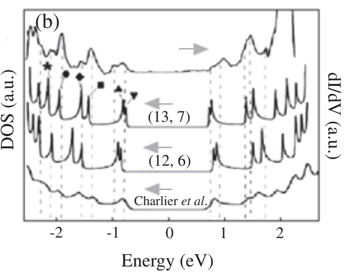


Figure 4. Comparison of measured dI/dV with calculated DOS for a metallic nanotube. (a) Energy dispersion of the π band of a graphene sheet near **K**. The solid lines correspond to the (13, 7) 1D bands obtained by zone folding. Symbols are located at the positions where van Hove singularities occur in these 1D bands. The inset depicts a three-dimensional view of the graphene π/π^* bands. (b) Comparison of the DOS obtained from experiment (upper curve) and a π-only tight-binding calculation for the (13, 7) SWCNT (second curve from top). The broken vertical lines indicate the positions of VHS in the tunneling spectra after consideration of thermal broadening convolution. The symbols correspond to the VHS shown in (a). The calculated DOS for a (12, 6) tube and an independent calculation for a (13, 7) tube [48] are included for comparison. Reprinted with permission from [51], Ph. Kim et al., *Phys. Rev. Lett.* 82, 1225 (1999). © 1999, American Physical Society.

Atomic resolution images of SWCNTs have been also obtained on the top of a rope [53–55]. The achieved resolution made it possible to compare the chiralities of the tubes that belong to the same rope. In some cases it was reported that the nanotubes have different chiralities and some of them may be twisted. In other cases, nearly uniform diameter and chirality were observed. Spectroscopic data taken on bundles show that the I–V curve may vary a lot along the same nanotube, and the spectra may exhibit a pronounced asymmetry [56]. The variation of the noise level of constant current topographic images taken at the same values of positive and negative bias confirm the asymmetric character of the spectra.

The STM tip is not only able to image the nanotube; it can be used as a "tool" to modify the nanotube. SWCNTs were cut to desired size by applying a voltage pulse of the order of 4 V. No polarity dependence was found [57]. After shortening to a length of 30 nm, the STS spectrum of the nanotube piece taken at 4 K showed nonequidistant, steplike features absent before the cutting. The observed effect was interpreted as arising from discrete electron stationary waves observed as periodic oscillations in the differential conductance as a function of the position along the tube axis, with a period that differed from that of the atomic lattice. This technique made it possible to map the square modulus of the electron wave functions that could be observed for several electron states at adjacent discrete energies. The measured wavelengths were found to be in good agreement with the calculated Fermi wavelength for armchair nanotubes [58]. Similar studies were also devoted to chiral metallic and semiconducting nanotubes shortened to a few nanometers [59, 60]. In contrast to the metallic nanotubes, no significant length dependence is observed in finite-sized semiconducting nanotubes down to 5 nm. The STS data obtained from the center of the shortened tubes showed a striking resemblance to the spectra observed before cutting [60].

The STM tip may be used not only to cut the nanotube but as a sliding contact along a nanotube extracted from a nanotube mat [61], too. Abrupt switching from a graphiticlike response to one that is highly nonlinear and asymmetric, including near-perfect rectification, was observed. The observation was interpreted in terms of existence of localized on-tube nanodevices formed by defects.

STS was used to investigate the effects of the ambient atmosphere on SWCNT samples. It was found that some semiconducting samples exhibit metallic behavior on exposure to oxygen, while in the case of other samples the effect of O_2 is restricted to the modification of the apparent density of states at the valence band edge [62]. The investigation of the electrical response of semiconductor SWCNTs to gas molecules showed that while NH_3 produces a sharp decrease in conductance, NO_2 has an opposite effect [63].

Statistical evaluation of nanotube diameters in a SWCNT mat (buckypaper) from the measured gap values and the comparison of these values with statistics obtained form TEM images demonstrated that the evaluation of diameters from gap values has to be done with caution because the gap is not always sharply defined in the STS spectrum [64]. When examined in a large enough bias interval, the positions of the peaks in DOS corresponding to the van Hove singularities form a fingerprint of the wrapping indices n and m of the nanotube, which can be used to better estimate the diameter [52]. Effects arising from interaction with the substrate [65] and possible charge transfer [45, 66] have to be taken carefully into account. These effects may shift the Fermi level, as mentioned, but do not affect too much the local DOS on the topmost part of the nanotube, which is the part that is scanned by the STM tip [65].

Curvature effects in carbon nanotubes studied analytically as a function of chirality show that the π orbitals are found to be significantly rehybridized in all tubes, so that they are never normal to the tubes' surface. This results in a curvature induced gap in the electronic band structure of the metallic tubes. The tilting of the π orbitals should be observable by atomic resolution scanning tunneling microscopy measurements [67]. Indeed, recent STS measurements confirm the existence of the predicted gaps in the case of metallic zigzag nanotubes and the presence of pseudo-gaps produced by tube–tube interaction in bundles [60, 68].

5. COMPUTER MODELING OF SCANNING TUNNELING MICROSCOPY IMAGES

As already emphasized in the previous paragraphs the interpretation of experimental data obtained in a complex tunneling system consisting of a supported nanoobject sandwiched between two tunneling gaps is far from being straightforward. Therefore computer modeling of the expected STM images is of great help in the correct evaluation of the experimental data. Modeling the atomic resolution images has been carried out by Meunier and Lambin. Their method is based on a tight-binding π-electron Hamiltonian for the case of a pointlike tip in combination with a self-supported nanotube [69]. In the calculated images

(see Fig. 5), the depressions (dark features) correspond to the centers of the hexagons of the honeycomb structure. Over these locations the STM tip gets closer to the nanotube to keep the current constant [69]. The light features are produced by the C–C bonds. The calculations confirm the different image symmetries observed experimentally (Fig. 5). In the case of armchair tubes, the honeycomb structure is easily observable; all the bonds look similar [58]. For zigzag tubes a strong anisotropy of the bonds is found. The ones parallel with the tube axis appear much more stronger [25]. For chiral tubes, stripes spiraling around the nanotube are observed [19]. These reverse orientation on bias reversal [70, 71] in the case of semiconducting tubes, as indeed observed experimentally in some occasions [72]. Calculated atomic resolution images of semiconductor SWCNTs show that the angle between the stripes spiraling around the nanotube and the tube axis changes gradually with the changing of the chiral angle (Fig. 6). The two distinct classes of semiconductor tubes $n - m = M(3) + 1$ and $n - m = M(3) - 1$ exhibit complementary behavior. The image acquired with a negative tip in the first case is similar to the image acquired with positive tip bias for the second group of tubes. The complementarity of the two families of tubes is produced by the change in the occupied/unoccupied character of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) when changing from $n - m = M(3) + 1$ to $n - m = M(3) - 1$, due to the phase change imposed by the periodic boundary condition along the circumferential direction [71]. These calculated STM images of SWCNTs may prove to be extremely useful for the quick identification of nanotubes while acquiring experimental data.

The method based on the tight-binding π-electron Hamiltonian is suitable for calculating the expected STM images of 5–7 defects and other defects on carbon nanotubes, too [69]. An overview of the method and of the main results was given recently [73]. In a recent work, Meunier and co-workers used this method to calculate the atomic resolution STM images of a tapered carbon nanotube tip [74]. They found that the STM images of tapers are dominated by a protruding pentagon inherent in the taper structure, which unfortunately does not allow for an easy identification of the atomic structure of the underlying nanotubes, because the electronic structure of the taper is markedly affected by the 5–7 defects.

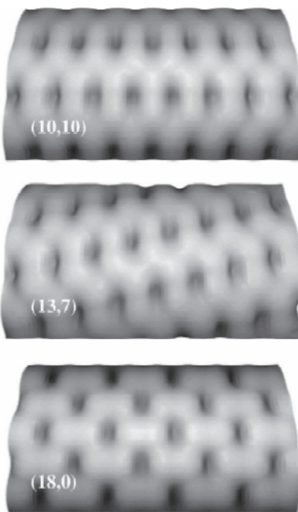


Figure 5. Constant current (topographic) STM images calculated using a tight-binding Hamiltonian and a pointlike tip for: an armchair (10, 10), a chiral (13, 7), and a zigzag (18, 0) nanotube. Reprinted with permission from [70], Ph. Lambin et al., *Carbon* 40, 1635 (2002). © 2002, Elsevier.

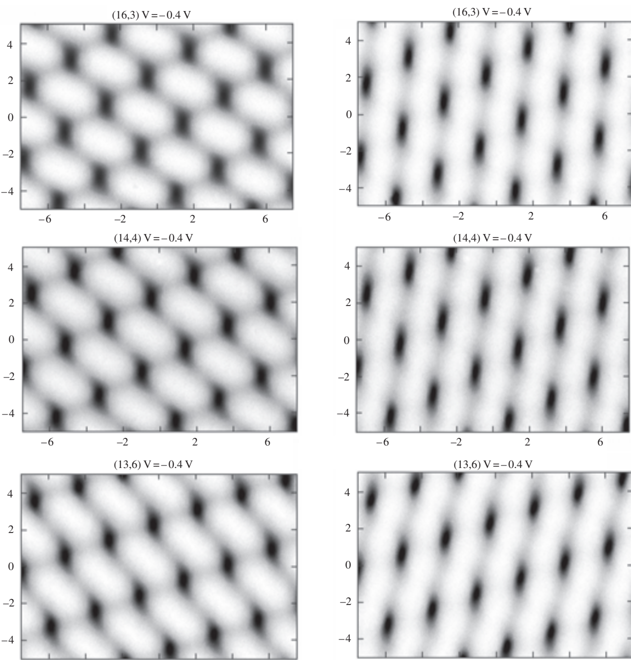


Figure 6. Tight-binding computed atomic resolution STM images of semiconductor nanotubes from the family $n - m = M(3) + 1$, with diameters close to each other: $d_{(16,3)} = 1.38$ nm; $d_{(14,4)} = 1.28$ nm; $d_{(13,6)} = 1.37$ nm, and increasing chirality: $\theta_{(16,3)} = 8.44°$; $\theta_{(14,4)} = 12.21°$; $\theta_{(13,6)} = 18°$, at positive and negative bias values.

A wave-packet dynamical method for the simulation of the geometric and charge spreading effects during tunneling in supported carbon nanostructures was developed and applied successfully for getting insight in the particularities of tunneling in two-dimensional (2D) representations [20] and more recently in full 3D tunneling through carbon nanotubes [75]. In this method the current density flowing through the STM tip–nanotube–support system is calculated based on the scattering of the wave packets incident on the barrier potential. The geometric effects are well described by electrodes and nanotube treated in the jellium-potential approximation (see Fig. 7). The STM topographic profile through a carbon nanotube was calculated which shows that as long as the electronic structure of the supported object and that of the support may be regarded as similar, the major image distortions arise from pure geometric tip-shape convolution [20] already discussed. With increasing differences in the electronic structure of the nanotube and that of the support larger distortions are expected. The incidence angle and energy dependent transmission of wave packets through single and bundled carbon nanotubes were calculated, effects arising from point-contact imaging and tip shape were analyzed, and STS curves were modeled [23]. The calculations show that for tip negative bias

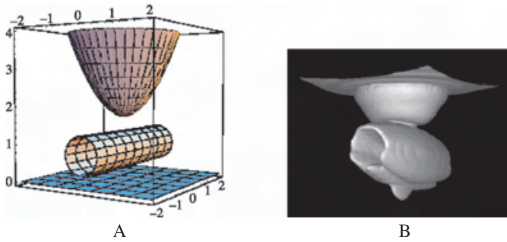


Figure 7. (A) Geometric model of STM tip, nanotube, and support surface. All dimensions are in nm. (B) Snapshot (at $t = 3.5$ fs) of the constant probability density surface of a wave packet tunneling through the tunnel junction shown in (A). Reprinted with permission from [35], G. I. Mark et al., in "Frontiers of Multifunctional Nanosystems" (E. Buzaneva and P. Scharff, Eds.), p. 43. Kluwer Academic, Dordrecht, 2002. © 2002, Kluwer Academic.

the angular dependence of the transmission is determined by the tip shape. The particular tip shape introduces an asymmetry on the negative side of the STS spectrum. For tip positive bias the angular dependence of the transmission depends strongly on the nature of the nanosystem in the STM gap. While the transmission of the STM tunnel junction without nanotubes can be well represented by a one-dimensional model, all other geometries cause a large normal-transverse momentum mixing of the wave packet. Recently an overview and comparison of the tight-binding and wave packet dynamical methods has been given [76].

The usefulness of the wave-packet dynamical method is completely revealed when using it in full 3D simulation [35, 75]. The computer simulation clearly shows that during tunneling the charge spreads along the nanotube. Thus, while the tunnel junction between the STM tip and the nanotube is a zero-dimensional junction, the tunnel junction between the nanotube and the support will be a linear, one-dimensional junction (Fig. 7). The consequences of the spreading will affect the tunneling situations when the nanotube is only partly supported or is placed over a nonhomogeneous substrate [75].

6. STRUCTURAL DEFECTS

As already reveled by early theoretic papers, structural defects, like 5–7 pairs, vacancies, isolated pentagons, and heptagons may have dramatic effects on the structure and electronic properties of carbon nanotubes [77–82]. Irrespective of its nature, a defect in a SWCNT gives rise to backscattering of the electron wave function. Interference of the incoming and backscattered waves leads to a standing wave pattern, which can extend far away from the defect due to the one-dimensionality of the nanotube. The corresponding oscillations of the DOS have been observed experimentally in armchair nanotubes where the Fermi wavelength of 0.73 nm is within the resolution easily achieved in spatially resolved STS experiments [83, 84].

The identification of defects and their detailed investigation by STM could lead to a better understanding of the way in which the defects are formed and how they can be used for the engineering of carbon nanotube devices. As already mentioned in the previous paragraphs, the theoretical models indicate that the defects should have a clear signature in the atomic resolution STM images and in spectroscopy. In a very thorough theoretic study of defects including Stone–Wales defects, which may arise due to tensile stress in SWCNTs, Orlikowski et al. focus on the expected STM images of pentagons, heptagons, and various combinations of 5–7 pairs, too [44]. An approach based on a tight-binding π-electron Hamiltonian for the evaluation of the tunneling current, as proposed by Meunier and Lambin [69], is used. The most prominent features observed in the STM images are a set of bright rings whose positions correlate with the location of the pentagons within the defect. These features are useful, as they enable the ready identification of many of the defects. By contrast, most of the defects have only a relatively modest effect on the transport properties of the nanotubes [44].

Based on the tight-binding approach, Krasheninnikov and co-workers calculated STM images for vacancies occurring on carbon nanotubes [85] and for defects expected after ion bombardment [86]. Vacancies of SWCNTs produce hillock-like features of about 0.1 nm in height and 1 nm in width in the STM images of the nanotubes, which are observable in low bias imaging ($V = 0.1$–0.4 V). Electronic superstructures similar to those in graphite near point defects produced by ion bombardment [87, 88] are evident in the simulated STM images.

Experimental investigation of small diameter kinked MWCNTs [83, 89] confirmed that both in topographic images and in STS measurements, the defects which must be present in the kink region give a detectable signature. In atomic resolution images of SWCNTs the signature of localized defects has been observed on straight SWCNTs [19]. Later, a detailed study by Ouyang and co-workers [90] focusing on atomic structure and electronic properties of intramolecular junctions in SWCNTs showed that about 10% of the SWCNTs grown by laser ablation incorporate defects which could be imaged reproducibly in an ultrahigh vacuum (UHV), low temperature STM. Metal–semiconductor junctions are found to exhibit an electronically sharp interface without localized junction states, whereas a more diffuse interface and low-energy states are found in metal–metal junctions. Tight-binding calculations for junction models based on observed atomic structures show good agreement with spectroscopy and provide insight into the topological defects forming the intramolecular junctions [90].

Two particular cases of "defects" have to be mentioned at the end of this section: (a) the defect concentrations and their STS signature occurring at tube ends, which has been discussed in previous sections [39, 51], and (b) the regular arrangement of defects, which yield novel architectures. This will be discussed subsequently.

7. COMPLEX TUBULAR STRUCTURES

As already pointed out in the Introduction, shortly after the discovery of straight carbon nanotubes, theoretical model calculations showed that with the incorporation of nonhexagonal rings in the hexagonal network, more complex structures can be built from tubular elements. The simplest one is a bent nanotube, a "knee" [7], produced by

incorporating a pentagon and a heptagon in the hexagonal network on opposite sides of a nanotube. Later the atomic structure and the expected electronic structure of more complex intramolecular junctions were explored by theoretical calculations [82, 91–94]. The possible application of these junctions in all-carbon nanoelectronic devices is foreseen. Calculations of the charge on the defect rings reveals an excess of electrons on the pentagon (+0.40 electron) and a deficiency on the heptagon (−0.25 electron) for several joints. This is attributed to the tendency of cyclic π-electron systems to achieve a stable configuration [93]. Quasioscillations of the local DOS with a triple period around the Fermi level were found in the metallic part of the (20,0)/(12,12) metal–semiconductor hybrid [93], very much as in the (10,0)/(6,6) junction [95] reported earlier. The LDOS and transmission coefficient calculations using a tight-binding algorithm suggest that the SWCNT bends can behave as M–S, M–M, or S–S heterojunctions (M = metallic, S = semiconductor) [94]. Electrical transport measurements carried out across the kinked and straight portions of a nanotube showed that while the straight part shows ohmic behavior, a strong rectifying effect is found across the knee [96]. Recent STM measurements on multiwall carbon nanotube knees produced by the arc discharge method indicate that increased electron density corresponding to the pentagon may be observed in topographic images even in the absence of atomic resolution [97], while at the location of the heptagon no such features are found.

Tori and coiled carbon nanotubes constructed by joining several knees head to tail were proposed as possible structures [7, 8, 98]. Scanning tunneling microscopy observation of coiled carbon nanotubes, following the relatively early TEM observation of the coiled carbon nanotubes [9], was first reported by Biró and co-workers [11, 22]. Taking into account that, as already discussed, nanotube junctions may behave like rectifiers, the electrons tunneling from the STM tip on the topmost part of a coil built from a succession of such rectifiers will have a small probability of reaching the substrate if the distance between the junctions is large enough. This may cause the STM tip to come into mechanical and electrical contact (point contact imaging) with the coil built of M–S junctions, while no such contact is expected for the M–M type coils [11]. Recent STS measurements on a coil (grown by the decomposition of fullerene directly on the STM substrate) which did not show any “camel-back” type feature, arising from point contact imaging, confirm the metallic DOS of the coil [99]. A regularly coiled SWCNT is showed in three-dimensional presentation in Figure 8.

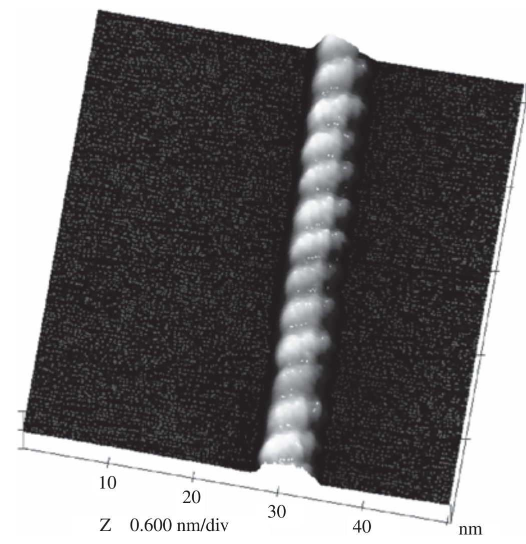


Figure 8. Topographic STM image of a regularly coiled metallic SWCNT in three-dimensional presentation.

With the increase of the number of the defects incorporated in a relatively short region of a tubular structure, branched tubular architectures can be built; the branching can be symmetrical [5, 6] or asymmetrical [100]. Such branched structures have been found by STM in carbon nanotube samples grown by fullerene decomposition (Fig. 9a) and by arc discharge, too (Fig. 9b). The Y-junctions grown by the first method are usually symmetrical [12], while the ones grown using the arc method were found to be less symmetrical, or asymmetrical [101]. In the arc grown tubes, the signatures of point defects, or knees, are observed in STM images preceding the branching point. It is speculated that these defects arise from growth instabilities, which usually are eliminated after the branching [101]. The branching occurs when a specific defect combination is achieved. Recent STS measurements and tunneling conductance maps confirm the existence of both M–M and M–S type joints within such Y-junctions [102]. These Y-junctions could constitute the basic building blocks of large, all-carbon nanocircuitry which needs only a few connection points with the macroscopic world, making it possible in this way to fully exploit the possibilities of carbon nanostructures to achieve new levels of integration. In some occasions one branch of the Y-junction can be a coil [103]. Nanotube networks with complex branching patterns where found after Cr filling. It was observed in the topographic STM images that strong apparent height variations along the tube axis are sensitively dependent on the imaging bias. These results suggest that Cr filling of carbon nanotubes results in strong electronic interactions between tube and filling materials [27].

Recently a new structural model, similar to the Haeckelite structure [104], was proposed for the coiled carbon nanotubes [105]. In contrast to earlier models of coiled tubes this approach allows nonhexagonal to hexagonal ring ratios to be over, or close to, unity. Using this model it

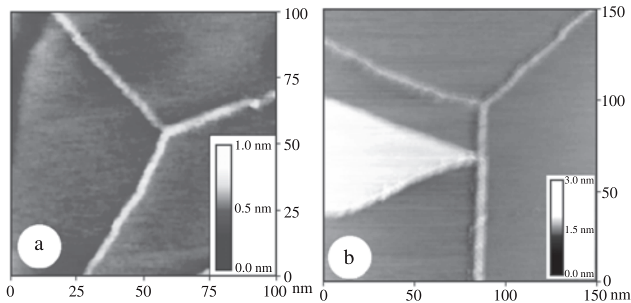


Figure 9. Constant current STM images of two Y branched carbon SWCNTs. (a) Grown by fullerene decomposition [12] and (b) grown by the electric arc method [101].

was possible to construct double helix type structures which were recently observed experimentally by STM [105] and TEM [106]. Figure 10a shows an experimental STM image of a double helix, while in Figure 10b and c two views of a model of double coil type structure generated as described by Biró and co-workers [105] are shown. The STM images in Figure 10d and e were computed for negative and positive tip polarity, using the tight-binding approach of Meunier and Lambin [69]. One can notice that in contrast to straight SWCNTs shown in Figure 5, due to the large number of nonhexagonal rings like in the case of tapers [74], no atomic resolution is possible. The tunneling current maxima show a very weak polarity dependence. The maxima are associated with the two hexagonal rings (highlighted) on the sides of the bonds where the two adjacent pentagons are joined. These rows of hexagons spiral along the tube axis in a double helix. Such structures of larger diameter, for which the calculation is not possible using the available computing facilities, are expected to be seen in the STM like the object in Figure 10a.

A new family of tubular carbon nanostructures was created by filling the inner hollow of SWCNTs with C_{60} balls [107]. These types of carbon nanostructures were named "peapods." STM images of peapods have been published recently [108]. Nothing special could be observed from the topography and the STS spectrum when electrons tunnel from the occupied states of the peapods to the tip. For a reverse bias of 1.5 V, by contrast, there is a large current when tunneling into the unoccupied states of the peapod; the topography shows maxima regularly spaced along the nanotube, which correspond to the fullerene interdistance. This effect was attributed to a large coupling of the LUMO t_{1u} state of the C_{60} with the electronic conduction bands of the nanotube.

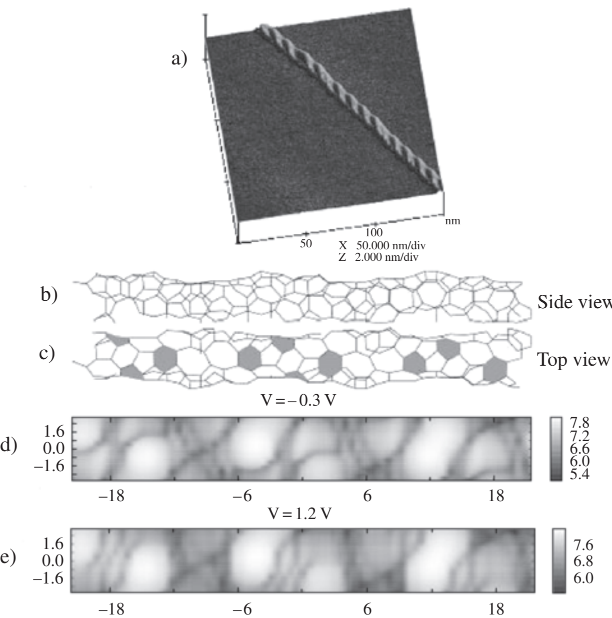


Figure 10. Experimental and calculated STM image of double coil type structures (the experimental image shows an object of much larger diameter than the calculated image). (a) Constant current topographic STM image in 3D presentation of a double coil; (b) structural model of a double coil structure in side view; (c) top view with the rings over which the bright maxima of tunneling current are found highlighted; (d) calculated top view STM image at a tip bias of −0.3 V; (e) calculated top view STM image at a tip bias of 1.2 V.

8. FUNCTIONALIZED NANOTUBES

The demand to functionalize carbon nanotubes (i.e., to attach to their outer wall covalently bonded groups of atoms, which can act as chemical anchoring points) was originally raised in the field of carbon nanotube reinforced composites, because the chemically inert, perfect nanotubes move in the matrix like the sword in the scabbard [109]. Due to its atomic resolution on individual objects, STM is a particularly well suited method for the investigation of the way in which the functional groups are attached to the nanotube. On the other hand, the presence of the functional groups often results in the nanotubes not coming close enough to the support on which they are placed; therefore the van der Waals attraction cannot fix them. This makes the STM imaging of functionalized nanotubes a very challenging task.

The first STM results on functionalized carbon nanotubes were reported by Kelly and co-workers [110]. Fluorination was carried out in a fluorine–helium atmosphere at various temperatures which yielded different degrees of functionalization up to 2C:1F. The fluorinated SWCNTs were imaged under ambient atmosphere using Pt/Rh tips. Fluorination caused the appearance of band shaped functionalized regions transversal to the tube axis, in which the SWCNT appears to have a larger diameter as compared to the nonfluorinated regions. Sharp delimitation of the border from fluorinated to nonfluorinated regions was found. No atomic resolution was possible over the functionalized regions, while atomic details were revealed between these regions. The fluorinated SWCNTs were used as precursors for attaching other chemical moieties, like butyl [110]. In contrast to the fluorinated nanotubes the butylated ones exhibit islandlike arrangement of the functional groups; strong electronic modulation effects are found in the vicinity of the butylated spots. Density-functional theory based calculations showed that depending on the fluorination pattern and tube chirality the fluorinated regions of the SWCNTs may have gaps varying from 2.7 to 0 eV [111]. The differences in the gap may well cause the appearance of the transversal bandlike features in the STM images of fluorinated tubes.

Recently MWCNTs functionalized by ball milling in NH_3 and H_2S atmosphere were investigated by STM and various other techniques [112]. Constant current STM images revealed that the functionalization by $-NH_2$ groups takes place in islands, around which the normal atomic resolution images characteristic for MWCNTs may be acquired. The STS spectra taken over the functional islands as compared to spectra typical for nonfunctionalized parts show and increase of the DOS above the Fermi level [112].

9. CONCLUSIONS

As the overview of the achievements of the first 10 years of STM investigation of carbon nanotubes shows, in this particular case the studied object and the method applied to study

it match in a unique way. Scanning tunneling microscopy and spectroscopy proved to be the most suitable methods for truly revealing the very characteristic electronic properties of carbon nanotubes arising from their one-dimensional character and nanometric size. A remarkable agreement between the theoretically computed results and experiments was achieved. It may be foreseen that in the attempts to achieve carbon nanotube based nanoelectronics for practical applications, STM and STS—probably in combination with atomic force microscopy—will play the role of "quality control tool" and will be a basic element in process monitoring. No doubt that a lot of effort will be needed to find ways in which these tools can be made widely applicable in industrial environments, where the low temperature, UHV scanning tunneling microscopes could be operated only with huge difficulties and high cost.

Many basic level questions concerning the STM investigation of multiwall carbon nanotubes are still open. The advance of this field is expected as suitable techniques emerge for the selective production of few wall carbon nanotubes [113].

The variety of novel carbon nanostructures like coils and multiple coils will pose a new challenge for both theoreticians and experimentalists. In several laboratories the selective synthesis of coiled nanocarbons was achieved already [114, 115]. It is expected that in the near future the coiled forms of carbon nanotube will be available on a regular basis, not only by accident.

A new class of carbon nanomaterials emerges with the discovery of the peapods. The first experimental results already indicate that the STM will be an extremely useful tool in getting deeper insight in the properties of these new materials.

Last, but not least, a very wide field of research opens up with the functionalization of carbon nanotubes. The first results already demonstrate STM and STS are the ideal tools to understand the way in which the chemical moieties are coupled to the nanotubes. A close collaboration between physicists, chemists, and materials scientists will be needed in this field.

Summarizing, one may say that the work done thus far is just the beginning of a quickly broadening path.

GLOSSARY

Branched tubular architectures Carbon nanotube junctions of Y, T, etc. shapes.

Carbon nanotube Allotrope of carbon discovered in 1991 by Iijima, constituted from a single atomic layer thick graphitic plate—called a graphene sheet—wrapped into a seamless tube.

Chiral angle The angle which characterizes the orientation of carbon–carbon bonds with respect to the orientation of the tube axis.

Coiled carbon nanotubes Regularly coiled carbon nanotubes looking like "telephone cables" of nanometer size.

Convolution A particular case of STM imaging when the radius of curvature of the tip is in the range of the radius of curvature of the object to be imaged. The acquired image will be equally influenced by the tip and the sample geometry. Deconvolution is not possible in a straightforward way, only by using dedicated software.

DOS Electronic density of states.

Energy gap The width of the region in which the density of states is zero (semiconductor tube) around the Fermi energy. For a metallic tube the region of finite density of states around the Fermi level is called "plateau region."

Functionalized carbon nanotubes Carbon nanotubes which have attached to their outer walls various chemical groups (functionalities).

Haeckelite structure Graphiticlike sheets or tubes containing a large fraction of regularly arranged nonhexagonal rings.

Helical coils Regularly coiled carbon nanotubes looking like "telephone cables" of nanometer size.

Interlayer spacing of MWCNTs The distance between two consecutive graphene layers.

Metallic carbon nanotube A nanotube which has a finite density of states at the Fermi level.

Multiple coils Carbon nanotubes built from several grown-together coils of identical pitch.

Multiwall carbon nanotube (MWCNT) Carbon nanotube consisting of several concentric SWCNTs with increasing diameter. The interlayer separation is similar to that in turbostratic graphite.

Peapods Carbon nanotubes filled with fullerenes.

Point contact imaging A particular case of STM imaging when due to the nonideality of the potentials and due to the small value of the tunneling gap, the potential barrier between the STM tip and sample is "punctured," so that the conduction mechanism switches from tunneling to ballistic conduction.

Semiconductor carbon nanotube A nanotube which has around the Fermi energy a so-called gap in which the density of states is zero.

Single-wall carbon nanotube (SWCNT) Carbon nanotube constituted from one graphitic cylinder.

STM tunnel junction A narrow channel between the STM tip and sample in which 90% of the tunneling current is flowing.

Stone–Wales defects Defect consisting of fused pentagonal and heptagonal rings.

STS Scanning tunneling spectroscopy, current–voltage spectroscopy acquired scanning the bias and measuring the corresponding tunneling current values, while the feedback loop and the scanning of the STM tip in the plane of the sample are switched off.

Topographic (constant current) imaging STM image acquired in a way that the tunneling current is maintained constant using a feedback loop and a piezoactuator.

Tunnel gap The geometric separation between the STM tip and the sample.

Van Hove singularities Singularities occurring in the density of states of carbon nanotubes.

Y-junctions Carbon nanotubes which have grown together in a Y-like fashion; structurally one the simplest models

implies the incorporation of six heptagons in the junction region.

ACKNOWLEDGMENTS

This work has been partly funded by the Inter-University Attraction Pole (grant IUAP P5/1) on "Quantum-size effects in nanostructured materials" of the Belgian Office for Scientific, Technical, and Cultural affairs and partly by the EU5 Center of Excellence ICAI-CT-2000-70029, and by OTKA grants T 30435 and T 043685 in Hungary. L.P.B. gratefully acknowledges the Belgian Fonds National de la Recherche Scientifique and the Hungarian Academy of Sciences for financial support.

REFERENCES

1. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
2. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, San Diego, 1996.
3. "Carbon Filaments and Nanotubes: Common Origins, Different Applications?" (L. P. Biró, C. A. Bernardo, G. G. Tibbetts, and Ph. Lambin, Eds.). Kluwer Academic, Dordrecht, 2001.
4. S. Iijima, *Nature* 354, 56 (1991).
5. G. E. Scuseria, *Chem. Phys. Lett.* 195, 534 (1992).
6. L. A. Chernozatonskii, *Phys. Lett. A* 172, 173 (1992).
7. B. I. Dunlap, *Phys. Rev. B* 46, 1933 (1992).
8. S. Ihara, S. Itoh, and J. Kitakami, *Phys. Rev. B* 48, 5643 (1993).
9. S. Amelinckx, X. B. Zhang, D. Bernaerts, X. F. Zhang, V. Ivanov, and J. B. Nagy, *Science* 265, 635 (1994).
10. J. Li, C. Papadopoulos, and J. M. Xu, *Nature* 402, 253 (1999).
11. L. P. Biró, S. D. Lazarescu, P. A. Thiry, A. Fonseca, J. B. Nagy, A. A. Lucas, and Ph. Lambin, *Europhys. Lett.* 50, 494 (2000).
12. L. P. Biró, R. Ehlich, Z. Osváth, A. Koós, Z. E. Horváth, J. Gyulai, and J. B. Nagy, *Mater Sci. Eng. C* 19, 3 (2002).
13. L. P. Biró, G. I. Márk, A. A. Koós, J. B. Nagy, and Ph. Lambin, *Phys. Rev. B* 66, 165405 (2002).
14. D. Reznik, C. H. Olk, D. A. Neumann, and J. R. D. Copley, *Phys. Rev. B* 52, 116 (1995).
15. Ph. G. Collins, M. S. Arnold, and Ph. Avouris, *Science* 292, 706 (2001).
16. R. Wiesendanger, "Scanning Probe Microscopy and Spectroscopy." Cambridge Univ. Press, Cambridge, UK, 1994.
17. J. Tersoff and D. R. Hamann, *Phys. Rev. B* 31, 805 (1985).
18. "Scanning Tunneling Microscopy III" (R. Wiesendanger and H.-J. Güntherodt, Eds.). Springer-Verlag, Berlin, 1993.
19. W. Clauss, *Appl. Phys. A* 69, 275 (1999).
20. G. I. Márk, L. P. Biró, and J. Gyulai, *Phys. Rev. B* 58, 12645 (1998).
21. Animated computer simulation showing the tunneling through a supported nano-object are available at *http://www.mfa.kfki.hu/int/nano/*.
22. L. P. Biró, J. Gyulai, Ph. Lambin, J. B. Nagy, S. Lazarescu, G. I. Márk, A. Fonseca, P. R. Surján, Zs. Szekeres, P. A. Thiry, and A. A. Lucas, *Carbon* 36, 689 (1998).
23. G. I. Márk, L. P. Biró, J. Gyulai, P. A. Thiry, A. A. Lucas, and Ph. Lambin, *Phys. Rev. B* 62, 2797 (2000).
24. L. P. Biró, S. Lazarescu, Ph. Lambin, P. A. Thiry, A. Fonseca, J. B. Nagy, and A. A. Lucas, *Phys. Rev. B* 57, 12490 (1977).
25. L. C. Venema, V. Meunier, Ph. Lambin, and C. Dekker, *Phys. Rev. B* 61, 2991 (2000).
26. Ph. Kim , T. W. Odom, J. Huang, and Ch. M. Lieber, *Carbon* 38, 1741 (2000).
27. F.-X. Zha, D. L. Carroll, R. Czerw, A. Loiseau, H. Pascard, W. Clauss, and S. Roth, *Phys. Rev. B* 63, 165432 (2001).
28. Z. Zhang and Ch. M. Lieber, *Appl. Phys. Lett.* 62, 2792 (1993).
29. Ch. H. Olk and J. P. Heremans, *J. Mater. Res.* 9, 259 (1994).
30. J. W. Mintmire, B. I. Dunlap, and C. T. White, *Phys. Rev. Lett.* 68, 631 (1992).
31. R. Saito, M. Fujita, G. Dreselhaus, and M. S. Dresselhaus, *Appl. Phys. Lett.* 60, 2204 (1992).
32. Y. K. Kwon and D. Tománek, *Phys. Rev. B* 58, R16001 (1998).
33. M. Ge and K. Sattler, *Science* 260, 515 (1993).
34. D. Tománek and S. G. Louie, *Phys. Rev. B* 37, 8327 (1988).
35. G. I. Márk, L. P. Biró, and Ph. Lambin, in "Frontiers of Multifunctional Nanosystems" (E. Buzaneva and P. Scharff, Eds.), p. 43. Kluwer Academic, Dordrecht, 2002.
36. J. Jxhie, K. Sattler, M. Ge, and N. Verkateswaran, *Phys. Rev. B* 47, 15835 (1993).
37. A. Hassanien, M. Tokumoto, S. Ohshima, Y. Kuriki, F. Ikazaki, K. Uchida, and M. Yumura, *Appl. Phys. Lett.* 75, 2755 (1999).
38. A. Hassanien, A. Mrzel, M. Tokumoto, and D. Tománek, *Appl. Phys. Lett.* 79, 4210 (2001).
39. D. L. Carroll, P. Redlich, P. M. Ajayan, J. C. Charlier, X. Blase, A. De Vita, and R. Car, *Phys. Rev. Lett.* 78, 2811 (1997).
40. T. Tamura and M. Tsukada, *Phys. Rev. B* 52, 6015 (1995).
41. M. S. Dresselhaus, in "Carbon Filaments and Nanotubes: Common Origins, Different Applications?" (L. P. Biró, C. A. Bernardo, G. G. Tibbetts, and Ph. Lambin, Eds.), p. 11. Kluwer Academic, Dordrecht, 2001.
42. L. P. Biró and G. I. Márk, in "Carbon Filaments and Nanotubes: Common Origins, Different Applications?" (L. P. Biró, C. A. Bernardo, G. G. Tibbetts, and Ph. Lambin, Eds.), p. 219. Kluwer Academic, Dordrecht, 2001.
43. D. Tekleab, D. L. Carroll, G. G. Samsonidze, and B. I. Yakobson, *Phys. Rev. B* 64, 035419 (2001).
44. D. Orlikowski, M. B. Nardelli, J. Bernholc, and Ch. Roland, *Phys. Rev. B* 61, 14194 (2000).
45. J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, *Nature* 391, 59 (1998).
46. T. W. Odom, J.-L. Huang, Ph. Kim, and Ch. M. Lieber, *Nature* 391, 62 (1998).
47. N. Hamada, S. Sawada, and A. Oshiyama, *Phys. Rev. Lett.* 68, 1579 (1992).
48. J.-C. Charlier and Ph. Lambin, *Phys. Rev. B* 57, R15037 (1998).
49. C. T. White and J. W. Mintmire, *Nature* 394, 29 (1998).
50. J.-C. Charlier and Ph. Lambin, *Phys. Rev B* 57, R15037 (1998).
51. Ph. Kim, T. W. Odom, J.-L. Huang, and Ch. M. Lieber, *Phys. Rev. Lett.* 82, 1225 (1999).
52. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 61, 2981 (2000).
53. E. D. Obraztsova, V. Yu. Yurov, V. M. Shevluga, R. E. Baranovsky, V. A. Nalimova, V. L. Kuznetsov, and V. I. Zaikovski, *Nanostruct. Mater.* 11, 295 (1999).
54. W. Clauss, D. J. Bergeron, and A. T. Johnson, *Phys. Rev. B* 85, R4266 (1998).
55. A. Hassanien, M. Tokumoto, Y. Kumazawa, H. Kataura, Y. Maniwa, S. Suzuki, and Y. Achiba, *Appl. Phys. Lett.* 73, 3839 (1998).
56. L. P. Biró, P. A. Thiry, Ph. Lambin, C. Journet, P. Bernier, and A. A. Lucas, *Appl. Phys. Lett.* 73, 3680 (1998).
57. L. C. Venema, J. W. G. Wildöer, H. L. J. Temminck Tuinstra, C. Dekker, A. G. Rinzler, and R. E. Smalley, *Appl. Phys. Lett.* 71, 2629 (1977).
58. L. C. Venema, J. W. G. Wildöer, J. W. Janssen, S. J. Tans, H. L. J. Temminck Tuinstra, L. P. Kouwenhoven, and C. Dekker, *Science* 283, 52 (1999).
59. T. W. Odom, J. L. Huang, Ph. Kim, and C. M. Lieber, *J. Phys. Chem. B* 104, 2794 (2000).

60. T. W. Odom, J. L. Huang, and C. M. Lieber, *J. Phys. Condens. Mater.* 14, R145 (2002).
61. Ph. G. Collins, A. Zettl, H. Bando, A. Thess, and R. E. Smalley, *Science* 278, 100 (1997).
62. Ph. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, *Science* 287, 1801 (2000).
63. J. Kong, N. R. Franklin, Ch. Zhou, M. G. Chapline, S. Peng, K. Cho, and H. Dai, *Science* 287, 622 (2000).
64. I. Wirth, S. Eisebitt, G. Kann, and W. Eberhardt, *Phys. Rev B* 61, 5719 (2000).
65. A. Rubio, *Appl. Phys. A* 68, 275 (1999).
66. Y. Xue and S. Datta, *Phys. Rev. Lett.* 83, 4844 (1999).
67. A. Kleiner and S. Eggert, *Phys. Rev. B* 64, 113402 (2001).
68. M. Ouyang, J.-L. Huang, C. L. Cheung, and Ch. M. Lieber, *Science* 297, 702 (2001).
69. V. Meunier and Ph. Lambin, *Phys. Rev. Lett.* 81, 5588 (1999).
70. Ph. Lambin, A. Loiseau, C. Culot, and L. P. Biró, *Carbon* 40, 1635 (2002).
71. C. L. Kane and E. J. Mele, *Phys. Rev. B* 59, R12759 (1999).
72. W. Clauss, D. J. Bergeron, M. Freitag, C. L. Kane, E. J. Mele, and A. T. Johnson, *Europhys. Lett.* 47, 601 (1999).
73. Ph. Lambin, and V. Meunier, in "Carbon Filaments and Nanotubes: Common Origins, Different Applications?" (L. P. Biró, C. A. Bernardo, G. G. Tibbetts, and Ph. Lambin, Eds.), p. 233. Kluwer Academic, Dordrecht, 2001.
74. V. Meunier, M. B. Nardelli, C. Roland, and J. Bernholc, *Phys. Rev B* 64, 195419 (2001).
75. G. I. Márk, A. Koós, Z. Osváth, L. P. Biró, J. Gyulai, A. M. Benito, W. K. Maser, P. A. Thiry, and Ph. Lambin, *Diam. Rel. Mat.* 11, 961 (2001).
76. G. I. Márk, L. P. Biró, and Ph. Lambin, in "Frotniers of Multifunctional Nanosystems" (E. Buzaneva and P. Scharff, Eds.), p. 243. Kluwer Academic, Dordrecht, 2002.
77. A. L. Macky and H. Terrones, *Nature* 352, 762 (1991).
78. B. I. Dunlap, *Phys. Rev. B* 46, 1933 (1992).
79. S. Ihara, S. Itoh, and J. Kitakami, *Phys. Rev. B* 48, 5643 (1993).
80. G. E. Scuseria, *Chem. Phys. Lett.* 195, 534 (1992).
81. L. A. Chernazatonskii, *Phys. Lett. A* 172, 173 (1992).
82. J.-C. Charlier, T. W. Ebbesen, and Ph. Lambin, *Phys. Rev. B* 53, 11108 (1996).
83. L. C. Venema, J. W. Jansen, M. R. Buitelar, J. W. G. Wildöer, S. G. Lemay, L. P. Kouwenhoven, and C. Dekker, *Phys. Rev. B* 62, 5238 (2000).
84. M. Ouyang, J.-L. Huang, and C. M. Lieber, *Phys. Rev. Lett.* 88, 066804 (2002).
85. A. V. Krasheninnikov, *Solid State Commun.* 118, 361 (2001).
86. A. V. Krasheninnikov, K. Nordlund, M. Sirviö, E. Salonen, and J. Keinonen, *Phys. Rev. B* 63, 245405 (2001).
87. L. P. Biró, J. Gyulai, and K. Havancsák, *Phys. Rev. B* 52, 2047 (1995).
88. J. R. Hahn and H. Kang, *Phys. Rev. B* 60, 6007 (1999).
89. D. Tekleab, R. Czerw, D. L. Carroll, and P. M. Ajayan, *Appl. Phys. Lett.* 76, 3594 (2000).
90. M. Ouyang, J.-L. Huang, C. L. Cheung, and Ch. M. Lieber, *Science* 291, 97 (2001).
91. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 53, 2044 (1996).
92. L. Chico, V. H. Crespi, L. X. Benedict, St. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* 76, 971 (1996).
93. V. Meunier, L. Henrard, and Ph. Lambin, *Phys. Rev. B* 57, 2586 (1998).
94. J. Han, M. P. Anantram, R. L. Jaffe, J. Kong, and H. Dai, *Phys. Rev. B* 57, 14983 (1998).
95. Ph. Lambin, A. Fonseca, J. P. Vigneron, J. B. Nagy, and A. A. Lucas, *Chem. Phys. Lett.* 245, 85 (1995).
96. Zh. Yao, H. W. Ch. Postma, L. Balents, and C. Dekker, *Nature* 402, 273 (1999).
97. Z. Osváth, A. A. Koós, Z. E. Horváth, J. Gyulai, A. M. Benito, M. T. Martínez, W. Maser, and L. P. Biró, *Mater. Sci. Eng. C* 23, 561 (2003).
98. S. Ihara and S. Itoh, *Phys. Rev. B* 48, 8323 (1993).
99. A. A. Koós. R. Ehlich, Z. E. Horváth, Z. Osváth, J. Gyulai, J. B. Nagy, and L. P. Biró, *Mater. Sci. Eng. C* 23, 275 (2003).
100. A. N. Andriotis, M. Menon, D. Srivastava, and L. Chernozatonskii, *Appl. Phys. Lett.* 79, 266 (2001).
101. Z. Osváth, A. A. Koós, Z. E. Horváth, J. Gyulai, A. M. Benito, M. T. Martínez, W. K. Maser, and L. P. Biró, *Chem. Phys. Lett.* 365, 338 (2002).
102. Z. Klusek, S. Datta, P. Byszewski, P. Kowalczyk, and W. Kozlowski, *Surf. Sci.* 507–510, 577 (2002).
103. L. P. Biró, R. Ehlich, Z. Osváth, A. Koós, Z. E. Horváth, J. Gyulai, and J. B. Nagy, *Diam. Rel. Mat.* 11, 1081 (2002).
104. H. Terrones, M. Terrones, E. Hernandez, N. Grobert, J.-C. Charlier, and P. M. Ajayan, *Phys. Rev. Lett.* 84, 1716 (2000).
105. L. P. Biró, G. I. Márk A. A. Kooós, J. B. Nagy, and Ph. Lambin, *Phys. Rev. B* 66, 165405 (2002).
106. M. Zhang, Y. Nakayama, and L. Pan, *J. Appl. Phys.* 39, L1442 (2000).
107. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Chem. Phys. Lett.* 315, 31 (1999).
108. D. J. Hornbaker, S.-J. Kahng, S. Misra, B. W. Smith, A. T. Johnson, E. J. Mele, D. E. Luzzi, and A. Yazdani, *Science* 295, 828 (2002).
109. L. S. Schadler, S. C. Giannaris, and P. M. Ajayan, *Appl. Phys. Lett.* 73, 3842 (1998).
110. K. F. Kelly, I. W. Chiang, E. T. Mickelson, R. H. Hauge, J. L. Margrave, X. Wang, G. E. Scuseria, C. Radloff, and N. J. Halas, *Chem. Phys. Lett.* 313, 445 (1999).
111. K. N. Kudin, H. F. Bettinger, and G. E. Scuseria, *Phys. Rev. B* 63, 045413 (2001).
112. Z. Kónya, I. Vesselenyi, K. Niesz, A. Demortier, A. Fonseca, J. Delhalle, Z. Mekhalif, J. B. Nagy, A. A. Koós, Z. Osváth, A. Kocsonya, L. P. Biró, and I. Kiricsi, *Chem. Phys. Lett.* 360, 429 (2002).
113. A. Peigney, P. Copuay, E. Flahaut, E. De Grave, R. E. Wandenberghe, and Ch. Laurent, *J. Phys. Chem.* 105, 9699 (2001).
114. M. Zhang, Y. Nakayama, and L. Pan, *Jpn. J. Appl. Phys.* 39, L1242 (2000).
115. C.-J. Su, D. W. Hwang, S.-H. Lin, B.-Y. Jin, and L.-P. Hwang, *Phys. Chem.Commun.* 5, 34 (2002).

www.aspbs.com/enn

Self-Assembled Monolayers on Semiconductor Surfaces

D. Zerulla
Heinrich-Heine-Universitaet Duesseldorf, Duesseldorf, Germany

CONTENTS

1. Introduction
2. Organic Self-Assemblies
3. Thiol Monolayers on III–V Semiconductor Surfaces
4. Nanostructuring of SAMs
5. Miscellaneous Self-Assembling Systems
6. Conclusion
Glossary
References

1. INTRODUCTION

Self-assembly is a very general principle in nature, as seen in the formation of for example, membranes from lipid molecules, or probably the most important paradigm, the living cell. Self-assembled monolayers (SAMs) are ordered molecular assemblies that are formed spontaneously by the adsorption of a surfactant. SAMs have recently attracted much attention because of the interesting opportunities for tailoring of solid surfaces. These monolayers exhibit unique chemical and structural properties. The formation of a SAM is initiated by bonding of a specific group of the organic molecule to the surface, followed by a rearrangement of the adsorbate layer with increasing coverage of the surface. This process is driven by intramolecular interactions between the adsorbed organics. Although self-assembly has attracted considerable attention over the last years, the subject has fairly old roots.

Pockels prepared monolayers at the air–water interface in the 19th century [1–4], followed by Rayleigh [5], Hardy [6], Deveaux [7], and others. Later, monolayers of amphiphilic molecules on the water surface were named after Langmuir [8, 9]. The first one to study the deposition of long-chain carboxylic acids on solid substrates was Blodgett [10, 11], who can be seen as the ancestor of our modern self-assembling systems, even if, around that time, amphiphilic monolayers were already used to control the wetting behavior of metal condenser plates in steam engines. Systematic research on self-assembling monolayers was performed later by Zisman [12] and Dewar [13]. For further information regarding the history of organic monolayers, we refer to [14, 15].

In the 1980s, the field of wetting and surface modification underwent a revival, and the great potential of thin organic films was recognized, which led to a very large number of publications in the following years [16–20].

Highly ordered organic films are very promising initial or intermediate states in order to achieve well-characterized tailoring of surfaces. Well-defined chemical modification and control of surfaces and interfaces are also essential for a variety of applications of semiconducting materials, as in photocatalysis, optoelectronic, and photovoltaic devices [21–23]. Recently, alkanethiolate films were also investigated on semiconductors because of their application potential for such purposes. But to date, still much less is known about alkanethiolate films on surfaces of compound semiconductors compared to the extensively studied surfaces of gold, silver, and copper. Passivation of GaAs surfaces using adsorbed thiols has been reported by Lunt et al. [24]. A study on alkanethiols on a chemically prepared InP(100) surface has been published by Gu et al. [25]. We will present very detailed results on the character, structure, passivating properties, and patterning of alkanethiols on III–V semiconductor surfaces in this chapter.

In contrast to the numerous publications in the field "self-assembling systems on coinage metals," the number of papers dealing with self-assembling on semiconductors is by far lower. In this chapter, the very detailed results on metals are sometimes used, to be compared with the results on semiconductors. This comparison allows us to judge if a self-assembling feature is a more general feature or a specific one for semiconductors. However, the main focus of this chapter is self-assembling and nanostructuring on semiconductors. Furthermore, this review is written from a surface-science perspective, with the focus on the fundamental principles governing the growth and structures

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.
Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (427–457)

of self-assembling monolayers. While this naturally puts the emphasis on chemically simple compounds (e.g., thiols), we also outline the rich opportunities in the areas of lateral patterning, chemical functionalization, chelating compounds, and some technological applications. For the used surface-science techniques, we refer to well-known textbooks.

The chapter is organized as follows. In Section 2, the basic facts about the systems, the sample preparation techniques, and self-assembly are discussed. Section 3 deals with thiols on different III–V semiconductors, the most discussed and, in our opinion, the most fruitful combination. This includes speed of assembly, passivating properties, and a very detailed discussion of the structure of these systems. Section 4 is dedicated to the structuring of self-assembling systems. We present different approaches to structuring, but also discuss in detail the mechanisms behind the different technologies. Section 5 briefly describes systems which have not been discussed in the previous sections in order to give an as complete as possible review on the topic. Section 6 closes this chapter with a general discussion, and we conclude with an outlook and some open issues.

2. ORGANIC SELF-ASSEMBLIES

The most often used organic species in self-assembly is the *n*–alkanethiol molecule. Alkanethiols CH_3–$(CH_2)_{n-1}$–SH (alkyl chain length n) are well known to form SAMs on metals like gold, silver, or copper [16, 26–31]. The self-assembly of the alkanethiols results in a unique arrangement regarding a rather well-defined polar angle of the chains with respect to the surface normal of the metallic substrates. The head group contains a sulphur atom which binds to the substrates by splitting off the hydrogen atom [32–34]. The backbone of the molecule is an alkyl chain which is, in its most simple form, terminated by a methyl end group. Since the bonds in the alkyl chain are only σ (single) bonds, they permit rotation around the bond axis, and therefore the chain has a high degree of freedom apart from the symmetric all-trans conformation. Instead of the terminating methyl group, other groups can be used as a substitute (e.g., –OH, –COOH). Depending on the chemical character of the groups, the assembled monolayer has a hydrophobic or more hydrophilic behavior.

Instead of the saturated alkyl chain, double bonds or even aromatics can be introduced. The head group may also consist of any molecule which binds to the substrate of interest. There have been numerous reports on the properties of such monolayers on metal surfaces, including their structure, chemical properties, and applications for the structuring of surfaces. In contrast to metal surfaces, very few investigations so far have considered the adsorption of alkanethiols on semiconductor surfaces. This lack of knowledge is also in contrast to the large body of investigations which deal with the modification of semiconductor surfaces using inorganic sulphur compounds [33–37]. Reports have been given on the influence of alkanethiols on Schottky barrier heights on metal–semiconductor contacts and on the passivating properties of adsorbed thiols on semiconductor surfaces [35, 37].

Besides thiols in *n*–alkanethiols, often-used head groups are sulphonic and phosphorus acids. Another popular example is organosilicon monolayers (e.g., alkyltrichlorosilane) on hydroxylated surfaces, which are well ordered, but typically do not exhibit the degree of long-range order observed for thiols on coinage metals and III–V semiconductor surfaces [38–41]. A compilation of the most often used SAMs and their structure is shown in Table 1. For more details, we refer to Section 5.

2.1. Substrates

With the advent of the transistor in 1947 (Bardeen and Shockley, Bell Laboratories), semiconductors became more and more important in technological development. Nowadays, practically no modern technological device runs without semiconductors. While silicon (Si) is the basic material for most electronic devices (e.g., processors) [42–49], III–V semiconductors are the materials of choice for optoelectronic devices. Recently, the development of, for example, organic light emitters (OLEDs) and other organic devices demonstrates the necessity of combining and structuring organic–semiconductor systems.

Semiconductor compounds consisting of elements of the third (B, Al, Ga, In, Tl) and fifth row (N, P, As, Sb, Bi) of the periodic system, that is, the so-called III–V semiconductors GaAs, InP, or $Al_xGa_{1-x}As$, are frequently used in

Table 1. Structure of some frequently used SAMs.

Structure	
HS CH_2 CH_2 CH_3 / CH_2 CH_2 CH_2	(a)
HS CH_2 CH_2 CH_2 / CH_2 CH_2 CH_2 OH	(b)
O // HS CH_2 CH_2 C / CH_2 CH_2 CH_2 OH	(c)
HS CH_2 CH_2 CH_2 / CH_2 CH_2 CH_2 SH	(d)
HS — (benzothiazole ring: N, S)	(e)
H_3C CH_2 CH_2 S—S CH_2 CH_2 CH_3 / CH_2 CH_2 CH_2 CH_2 CH_2 CH_2	(f)
Cl_3Si CH_2 CH_2 CH_3 / CH_2 CH_2 CH_2	(g)

Note: (a) *n*–alkanethiol, (b) ω–mercaptoalkanol, (c) ω–mercaptoalkane carboxylic acid, (d) α, ω–alkanedithiol, (e) mercaptobenzothiazole, (f) 1,1′–dialkyl–disulfide, (g) alkyltrichlorosilane. For the purpose of illustration, the chain length is set to $n = 6$.

modern optoelectronic devices, such as light-emitting diodes (LEDs), semiconductor lasers, optoelectronic modulators, or light detectors. The reason for using these materials in optics rather than the most widely employed Si is the difference in band structure: many III–V semiconductors have a direct bandgap, resulting in very efficient optical absorption or emission. Silicon is an indirect bandgap semiconductor which, in its crystalline form, has only a few optical applications. The specific optical properties of III–V semiconductors are given by their band-structure characteristics. The wavelength of luminescence (e.g., the color of the LED) is determined by the width of the forbidden gap between the valence and conduction bands. The efficiency of such devices is strongly related to the effective mass or density of states for the carriers in each band.

Because of the above-mentioned optoelectronic properties, the most prominent compound semiconductors are gallium arsenide (GaAs), gallium phosphide (GaP), gallium nitride (GaN), indium arsenide (InAs), and indium phosphide (InP) [50a, 50b]. Heterostructures of alternating thin layers of the above compounds are made in order to tune their optoelectronic properties [50b].

Most semiconductor surfaces used for research are made from high-quality single crystals which are produced by Czochralski or Bridgman techniques. In the case of III–V semiconductors, two surface orientations are of utmost importance. First is the (110) surface, which is the natural cleavage plane. Cleaving a III–V semiconductor along this plane results in surfaces with the highest grade of perfection. For more details, for example, surface relaxation, we refer to [50, 51]. Second is the (001) surface, which is the technologically used plane for etching processes in the electrooptical industry [35–37].

The crystals are typically doped (naturally or artificially) with carrier concentrations of 10^{-15}–10^{-18}/cm^3 in order to make them useful for technical applications. The influence of the carrier concentration on the binding of organic self-assembling monolayers is weak, and thus will be ignored in this review.

2.2. Sample Preparation and Self-Assembly

2.2.1. Sample Preparation and Self-Assembly from Solutions

The alkanethiols (most typically hexadecanethiol or dodecanethiol 95–99.5%) were dissolved in isooctane, ethanol (absolute), or hexadecane. The typical concentrations were in the range of 10^{-3}–10^{-5} M. More polar species (e.g., mercaptobenzothiazole, MBT; see Section 4) were dissolved in an aqueous solution with 5% ethanol. After immersion times of 24–36 h, the samples were rinsed with fresh solvent, dried under a nitrogen stream, and inserted into, for example, the spectrometer immediately. As substrates, either prenotched bars of Czochralski-grown and crystallographically oriented single crystals were used or, for the purpose of comparison, gold films were prepared using a Knudsen cell in UHV (base pressure: 2×10^{-10} mbar), evaporating gold (Goodfellow, 99.999 + %) onto a freshly cleaved slice of muscovite mica. The mechanism behind the self-assembly process on gold was first published by Poirer and Pylant [52]. The corresponding mechanism from dilute solutions on semiconductors will be investigated in Section 3.1, and for vapor deposition in the following paragraph.

2.2.2. Sample Preparation and Self-Assembly in UHV

The alkanethiol (e.g., dodecanethiol Aldrich, 98%) was let into the preparation chamber by means of a sapphire-sealed leak valve (Vacuum Generators VG, U.K.), which is inert to sulphur-containing species. In order to minimize the content of dissolved oxygen in the alkanethiol, the liquid was degassed by several repeated pumping, purging (with N_2, 99.999%, Messer Griesheim, Germany), and freezing cycles, before passing the leak valve. The valve and its surrounding should be slightly heated (35–45 °C) to prevent condensation and physisorption of the alkanethiol molecules on the inner sealing surfaces of the leak valve [53]. Besides spectroscopic techniques, a scanning tunneling microscope (STM) was used to investigate the adsorbed systems in UHV. Because of the active properties of the freshly cleaved InP(110) surface, all experiments (at least with submonolayer coverage) had to take place in UHV. Therefore, the STM head, which was constructed on a CF-100 flange, had to be mounted inside a UHV chamber. The base pressure in the recipient was $2 * 10^{-10}$ mbar using a combination of a turbo molecular and a rotary pump. For the sake of low vibration, this combination could be shut down. In this case, we used a liquid-nitrogen-cooled Titan sublimation pump and an ion getter pump instead. These were able to maintain good vacuum conditions without generating vibrations. When the TSP was cooled with liquid nitrogen, the pressure went down to $5 * 10^{-11}$mbar. The tungsten tips were self-made with the help of a two-step procedure (see [22, 54–58]). Because of additional control of the current direction of the last applied sine wave, the tips are also free from covering oxide layers. Following this procedure, the tips were rinsed with tridest H_2O, and were directly inserted into the vacuum chamber. A self-constructed system allowed not only changing the samples, but also changing up to seven tips at one time, without breaking the vacuum. The supporting X-ray photoelectron diffraction (XPD) and X-ray photoelectron spectroscopy (XPS) measurements were made with an Escalab 220 iXL (Vacuum Generators VG, U.K.) spectrometer, equipped with a monochromatized Al Kα source and a computer-controlled microstage.

The comparative SXPS spectra were recorded at the BESSY synchrotron storage ring in Berlin at the undulator beamline TGM-5, using a VG Baby-Aries spectrometer.

As an example for consecutive adsorption of thiols in UHV, the coverage and order of dodecanethiolate molecules on InP(110) surfaces was investigated by Zerulla et al. [53] and on gold by Widrig et al. [59] and Poirer et al. [60]. In the following paragraph, the coverage and structure of the adsorbed molecules are controlled by spectroscopic means and STM investigations depending on the dodecanethiol dose. Starting from the clean, cleaved semiconductor surface, the STM images show the different states of order in the self-assembling films with molecular/atomic resolution.

Additionally, STS was applied, in order to distinguish covered and uncovered regions with high lateral resolution. Further proof is given by (S)XPS and XPD spectra, obtained from measurements at the synchrotron storage ring BESSY and high-resolution laboratory sources, at all consecutive stages of adsorption. The first straightforward step in the investigation of vacuum-deposited alkanethiols on III–V semiconductors was to monitor the clean, freshly cleaved (UHV conditions) semiconductor surface itself via STM. Depending on the polarity of the tip with respect to the substrate surface, the STM image shows either the positions of the indium or the phosphorus atoms [61–63]. The surface was recorded with a positive tip, therefore showing the filled "dangling bonds" of the phosphorus atoms. The distances of the atoms were analyzed in the [001] and [1–10] directions of the surface. These measured distances of 590 and 420 pm are close to the ideal distances of 587 and 415 pm, respectively, which can be taken from [35–37, 64]. The difference between ideal and measured distances are well in the limit of the temperature-dependent reproducibility of the STM piezocrystals, which were calibrated before with the help of a highly ordered pyrolytic graphite (HOPG) surface. It should be mentioned that surface-relaxation effects have an influence on the height offset of the topmost phosphorus atoms with respect to the indium atoms (I–P), but no effect on the P–P distances within the first layer [35–37, 65].

Further STM investigations on a larger scale show only minor distortions of the surface in the form of small steps (typical height 1–6 atoms). Typically, the surfaces are free from defects for lateral dimensions as large as 1 mm^2. For the typical lateral scale of atomic-resolved STM graphs, this, in fact, allows for the investigation of adsorbates on regions without any substrate defects.

STS measurements of various positions on such freshly cleaved InP(110) surfaces lead to a composite I/U spectrum. The shown spectrum is a result of 256 independent recorded spectra, which have been added to suppress random noise. The main characteristic feature of these spectra is the very good visible bandgap, which is about 1.3 eV at room temperature in the case of InP [53]. In our experiments, the line shape of the spectra was independent of the exact lateral position of the STM tip; therefore, it still represents a more macroscopic feature of the semiconductor, and not an atomic behavior—even if the lateral resolution in the normal image mode achieved atomic resolution. Surface-sensitive (S)XPS studies of these cleaved surfaces, which are proof of their cleanness, are presented in [53, 64–67].

After the characterization of the clean InP(110) surface, a systematic, consecutive, dose-dependent treatment with dodecanethiol molecules was applied to the semiconductor surface. First, a dose of 1 Langmuir dodecanethiol was applied. The result of this treatment was the complete loss of resolution in the STM graphs. The reason for this behavior is probably that dodecanethiol molecules are bounded via the sulphur atoms to the indium atoms [66, 67]. The alkyl chains, on the other hand, only adhere loosely to the surface. Because of the sparse density of the molecules at 1 Langmuir (sticking coefficient <1), sufficient space remains between the molecules. When the STM tip scans across such a molecule, the alkyl chain is moved by the tip. Therefore, a clear STM image could not be obtained.

Nevertheless, the adsorbed molecules can be traced with the help of XP spectroscopy. The spectra show the expected C1*s*/S2*p* intensity ratio of 12:1 (corrected with the respective sensitivity factors and the spectrometer transmission function) for dodecanethiol, while the total signal strength (e.g., C1*s*) of the adsorbed molecules is far below the well-known intensity expected for a typical monolayer (see [32]).

The measured relative and absolute intensities are proof for nonoverlapping molecules with a sparse density at the surface.

This behavior changes if a larger dose of dodecanethiol is applied. It is possible at 100 Langmuir to get a good resolved graph again. The situation is displayed in Figure 1. It shows a dense monolayer of "lying" dodecanethiol molecules. Note that the chains are not randomly distributed, but mostly pointing into a crystallographic main axis (to prevent disturbances by the scanning tip, and therefore misinterpretations, the main axis of the InP crystal was not chosen parallel to the image boundaries, but turned 15° counterclockwise).

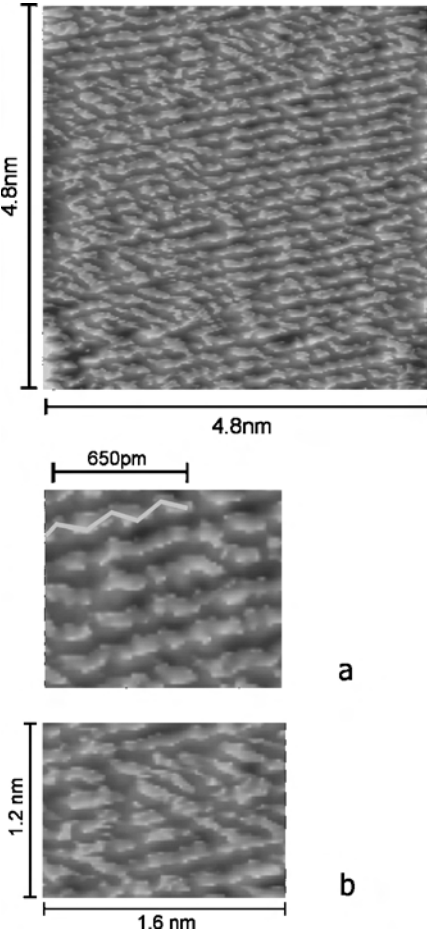


Figure 1. STM picture at 100 Langmuir dodecanethiol ($U_T = 2.4$ V; $I_T = 500$ pA). (a) STM-picture at 100 Langmuir dodecanethiol (enlarged). The all-trans conformation of a single molecule has been graphically emphasized. (b) STM picture at 100 Langmuir dodecanethiol (enlarged). The picture shows a region where some chains are misaligned. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

Figure 1a (enlarged part of Fig. 1) shows the structure of the alkyl chains (all-trans conformation), which can also be resolved in some regions. While most of the chains are pointing into the [001] direction of the substrate, upon closer inspection, it can be seen that some chains point partially into other directions (see Fig. 1b). This can easily be explained by the following fact: because of the σ bondings between the carbon atoms in the alkyl chains, turning of molecular parts around these bonding axes is permitted. While the all-trans conformation is the energetically lowest form, other conformations will occur at room temperature, whose total energies are only slightly higher. Once lying on the surface at high densities, these conformations are also stabilized by the neighboring molecules within the dense package. To confirm the above thesis, it is necessary to investigate all possible geometrical variations which might occur on the surface. Assuming that the orientation of all of the In–S bondings between the substrate and the thiol molecules point into the [001] direction of the substrate due to the InP(110) surface (we will discuss this later), the number of alternate conformations projected onto the surface is limited, at least if the permitted values of the C–C angles are low in number, because of the state of adsorption at the surface. This means that, besides the pure all-trans conformation which is shown in Figure 1a, it should be possible to find a limited number of molecules in slightly differing conformations (e.g., only one turn around one bonding axis, leading to a change in one angle in the carbon backbone in contrast to an all-trans conformation), and therefore a differing orientation of their main axis after the deviating bond. This case is shown in Figure 1b. Further geometric analysis of the STM graphs reveals that the measured bonding angles between the carbon atoms in all-trans conformation molecules are larger then 110° (about 130–140°). First, this is a result of the slightly changed conformation of the molecules, while being physisorbed at the surface. The necessary structure to reach a minimum of total energy for a physisorbed alkyl chain (carbon in sp^3 hybridization) does not have a perfect planar carbon backbone parallel to the semiconductor surface. Second, the STM imaging mode projects this three-dimensional molecule into the two-dimensional space, resulting in images of molecules which seem to deviate from the octahedrally bonded situation.

A further proof that the STM graph displays "lying" molecules is given by angular-resolved XPS measurements. Here, the ratio between the C1*s* and the S2*p* peak intensity does not change while varying the take-off angle. Besides, the pure C1*s* and S2*p* peak intensities at normal take-off angle do not allow for models other than "lying" molecules, especially if models of a higher thickness than a "lying" monolayer can be excluded due to the substrate (e.g., P2*p* or In4*d*) to adsorbate intensity ratio, for example, C1*s* (see adsorption step 3 in Figs. 4 and 5).

If the sulphur atoms were covered significantly by the alkyl chains, the (corrected) intensity ratio (C1*s*/S2*p*) would have to be higher than the measured ratio (12.2), which still displays approximately the stoichiometric ratio.

Increasing the dodecanethiol dose to 1000 Langmuir leads to the conditions shown in Figure 2. Inspection of this STM image reveals two different kinds of regions: the higher ones (brighter gray scales), which we interpret to consist

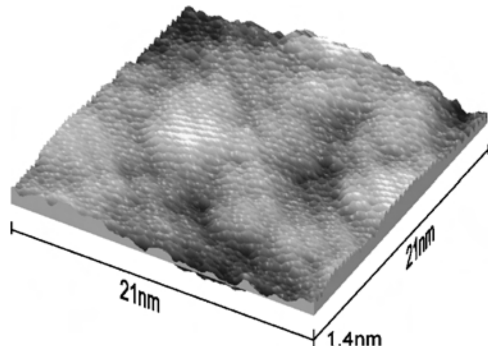


Figure 2. STM picture at 1000 Langmuir dodecanethiol ($U_T = 3.8$ V; $I_T = 250$ pA). Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

of "standing" molecules, and the lower ones (darker gray scales), which are still formed by "lying" molecules, as can be seen at lower doses. Note that the STM graph still shows molecular resolution simultaneously in both kinds of regions at room temperature.

To ensure that both the "standing" and the "lying" regions do indeed consist of alkyl chains, both regions have been investigated via STS. The spectra are shown in comparison with an SXPS spectrum (excited at 55 eV with synchrotron radiation), recorded from surface-bonded alkyl chains, prepared from dilute solutions of dodecanethiol (Fig. 3) [32, 66, 68]. Despite the differences in excitation, the energetic positions of the peaks are identical. Both regions can therefore be assumed to be covered with dodecanethiols. The following arguments can be found to support the interpretation.

- From the STS–SXPS comparison, one can conclude that the whole surface is covered by dodecanethiol.
- The average height offset of the bright and dark regions of approximately 1.0–1.4 nm (taken from STM measurements) fits into the model of lying minus standing regions—taking the length of one dodecanethiol molecule (1.84 nm) and the expected angle from the surface normal of 34° into account [66, 67].

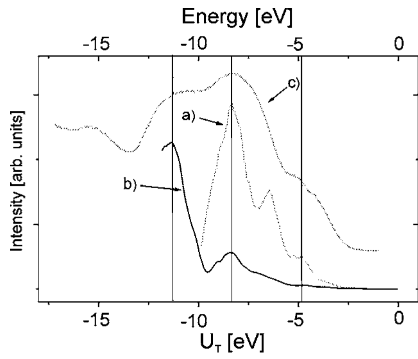


Figure 3. Locally resolved ST spectra of a DDT-covered region in comparison to an SXPS spectrum of a similar sample. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

- In the following, we will present additional information, derived from detailed XPS studies of the substrate (In3d, P2p) and the adsorbate (C1s, S2p), respectively.

In order to obtain additional reliable information on the surface coverage, we performed highly resolved XPS measurements for all peaks of interest at all stages of adsorption. As an example, we provide the detailed XP spectra for step 3 (1 kLangmuir dose) in Figure 4. The intensity information was extracted from these spectra by calculating the true peak area from an elaborate fitting procedure which uses an iterative background subtraction combined with convoluted Voigt profiles in a multipeak algorithm. The corresponding residuals are also given in each graph. The extracted intensities for all stages of adsorption are given in Figure 5. Note that the data are normalized individually. That is, for example, the ratio between carbon and sulphur is always greater than 12. In Figure 5, the highest point (at step 4) of this ratio is normalized to 1 in order to display all graphs in the same diagram. This allows a detailed comparison between the individual intensities. A measure of the reliability of this method is given by the substrate intensities (In3d, P2p). In principle, both intensities should be identical in pairs at all points. In comparison to the magnitude of changes of the adsorbate intensities, this is true. Now, we focus on the carbon/sulphur ratio. In first-order approximation, the intensity of both elements should rise while more molecules are being adsorbed on the surface. The difference between both intensities in the course of adsorption is due the fact that

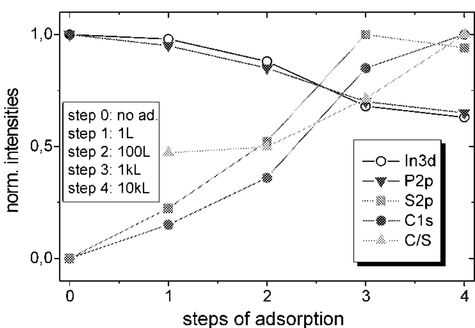


Figure 5. Comparison of the XPS intensities of adsorbate and substrate peaks and the carbon-to-sulphur ratio, derived from background-corrected peak areas at the different adsorption steps. Note that all measurements have been normalized individually. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

the carbon–alkyl chains are covering the binding sulphur atoms in a standing film, and therefore weaken the XPS intensity of the sulphur signal. This means that the higher the C/S ratio, the more "standing" is the ordered monolayer. Returning to the adsorption step at 1 kLangmuir this means that this film already shows a high degree of standing molecules, but compared to the last step (doses >10 kLangmuir), one finds that the degree of order is not at its maximum. This is exactly how we interpreted the STM graph before. The film is partially well ordered (standing), but has also some regions which have a lower density of adsorbed molecules (lying).

Well-ordered self-assembling monolayers, which can be compared to systems prepared from dilute solutions, are also found. The required minimum dose for a quasifaultless monolayer is between 5000 and 10,000 Langmuir dodecanethiol. An example for such a system is shown in Figure 6. The SAM is prepared at 10,000 Langmuir dodecanethiol on a freshly UHV-cleaved InP(110) surface. The array of "standing" molecules appears faultless, although, on closer inspection, a small deviation in brightness (=height) is detectable. This effect might arise from some residual stress in the SAM, which itself comes from the exceptional three-dimensional tight packing of the alkanethiol molecules

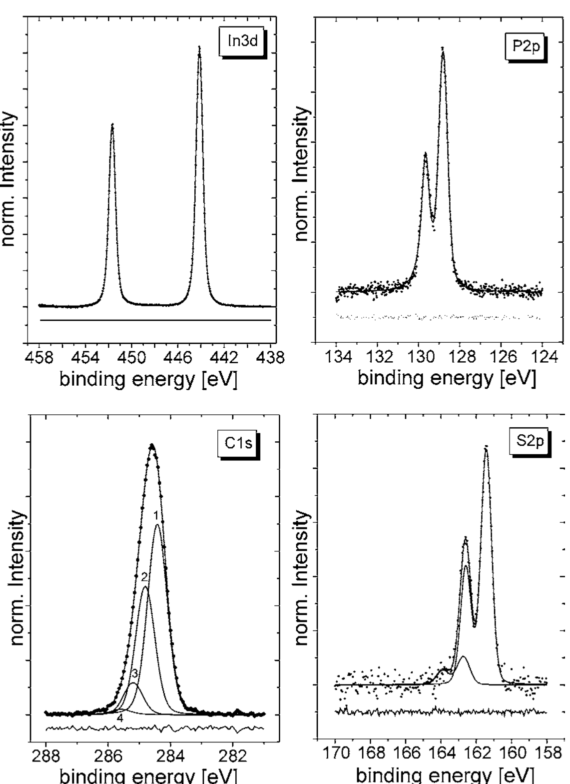


Figure 4. XPS spectra of DT on InP(110) at 1 kLangmuir. (a) In3d. (b) P2p. (c) C1s. (d) S2p. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

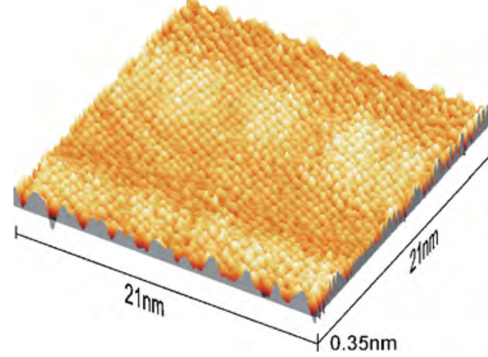


Figure 6. STM picture at 10 kLangmuir dodecanethiol ($U_T = 3.8$ V; $I_T = 250$ pA). Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

in the well-ordered "van der Waals stabilized" layer (see Section 3.3 and [66, 67]).

Note that neither the gray scale nor the *z*-axis scale of Figure 6 is equivalent to that of Figure 2, but is adjusted for more sensitivity in height display. Figure 7 shows an enlarged section of the former STM graph. Even on this scale, the monolayer appears well ordered. A small distortion of the graph can be found by comparing the resolution in the *X* and *Y* directions. This distortion could be caused by two points. First, the shape of the STM tip could be slightly asymmetric. Second, the tilted structure of the self-assembled monolayer in combination with the impact of the tip into the layer (Fig. 8) could cause such an effect. Because of some further tests with this tip and the knowledge of the inner structure of the monolayer [66, 67], we tend to interpret the distortion to be due to the "tilted" structure of the adsorbate, and not caused by imperfections in the shape of the tip.

The STM analysis of the adsorbate system is completed by a determination of the molecular spacing. For this purpose, not only was an exemplary determination on a "good fitting location" in real space made, but so was a complete two-dimensional fast Fourier transformation (FFT) [53]. The transformation allows us to judge the distances of all investigated molecules simultaneously, and to detect superlattices and repeating distortions in one image. Repeating structures in time (and therefore, due to the scanning, in space) are displayed as bright "spots" (accumulation points). The most prominent spots belong to the rectangular reciprocal surface unit cell of the underlying InP lattice, and are highlighted via white lines for better visibility. As far we can judge from numerous STM graphs, the adsorbed monolayer appears faultless and well ordered. But how can we be sure that the system under investigation is a unique, distortion-free monolayer system on larger scales? From STM measurements alone, it is practically impossible to obtain positive proof of the single-domain structure. If the area under investigation (STM) is successively made larger, there will be a certain point where the necessary molecular resolution will be lost. This is especially true if superlattices, and therefore domain boundaries, are missing, which has to be expected for single-domain structures. Proving the single-domain structure investigations on a macroscopic scale is necessary. For this, we refer to Section 3.3, where two different spectroscopic approaches have been chosen which are able to analyze structural aspects on macroscopic scales [66, 67].

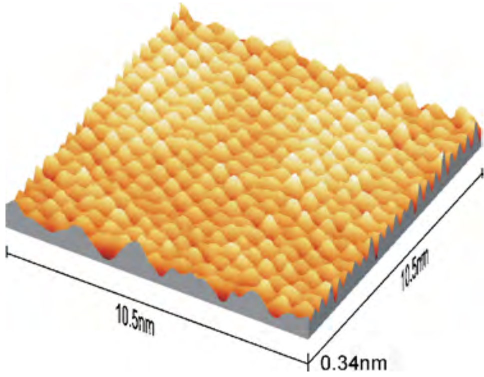


Figure 7. STM picture at 10 kLangmuir dodecanethiol (enlarged). Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

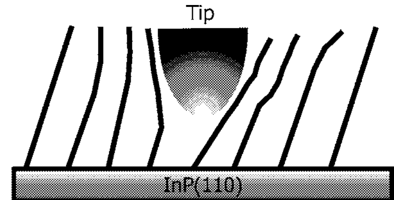


Figure 8. Model of the dodecanethiol–InP(110) system while measured by STM. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

Summarizing the recent results for UHV deposition, four main phases of adsorption on InP(110) were found. First, at doses < 10 Langmuir, the surface is only sparsely covered, still permitting the alkyl chains to move "sideways" on the surface; therefore, no stable STM image was presented. Second, approximately at 10 Langmuir, a dense layer consisting of lying molecules is formed. The contact between the alkyl chains is stabilizing enough to give good resolution, even at room temperature. Third, at doses of 1 kLangmuir, parts of the monolayer start to stand up. This is done neither statistically nor individually, but in well-defined domains, while other regions are still flat on the surface. A comparison between the height of the standing and the lying molecules in the same graph permits us to deduce this difference precisely. From that, we can learn how deep the tip penetrates the organic layer.

From other papers concerned with the behavior of tip–surface interactions on organic films, it is well known that the tip might intrude into the organic layer while scanning [22, 56, 69]. This behavior is also to be expected for the STM parameters used here in connection with the insulating properties of the adsorbate system under investigation (See Section 2). Further proof is given by the height information of the different regions in Figure 2. The average height difference of the "standing" and the "lying" regimes is about 0.8 nm. Note that the given *z* axis in the graph denotes the maximum distance between the highest and lowest point; therefore, the average distance between neighboring regions is significantly lower. If the tip did not intrude into the layer, this distance would be about 1.5 nm (the height of a 34° tilted dodecanethiol monolayer) minus the height of a lying molecule, which should be near 0.2 nm. The result of 1.3 nm is significantly higher than the measured value of 0.8 nm; therefore, the tip intrudes about 0.5 nm into the layer, as is depicted in Figure 8.

In the last stage (dose $>$10 kLangmuir), a macroscopic "single-domain" monolayer covers the semiconductor, which is virtually free from defects. All molecules have the well-defined tilt angle in the [001] direction of the InP (110) surface (see Section 3.3). The chain–chain distances, and therefore the tilt angle, reflect the lattice properties of the underlying semiconductor, which were additionally proved by an FFT analysis of the STM image and locally resolved tunneling spectra.

3. THIOL MONOLAYERS ON III–V SEMICONDUCTOR SURFACES

Thiols on III–V semiconductors are by far the most promising combination, as we will show in the following paragraphs of this section. Their ease of preparation is similar to thiols on gold (with some exceptions; see Section 2), and they form dense and therefore passivating (see Section 3.2) and highly ordered (see Section 3.3) monolayers which can be structured with the methods introduced in Section 3. But first, we will discuss the self-assembly and its speed, depending on the method of sample preparation, in the next paragraph.

3.1. Speed of Self-Assembly

While the adsorption under vacuum conditions is primarily a function of the differential pressure of the molecule species [53, 70], the adsorption from dilute solutions is easily traceable in the time domain [71]. Confining our systems to alkanethiols with lengths between 6 and 20 carbon atoms, the adsorption and self-assembling speeds are very similar. The methods to probe the degree of order or state of assembly can include in-line techniques which can be used without interrupting the self-assembling. This has been done by Grunze and Wöll, using second-harmonic generation (SHG) and sum-frequency generation (SFG) on the adsorption of thiols on gold [72].

The other class of experiments, that is, all surface-science techniques working under UHV conditions, needs an interruption of the assembling process. For gold and silver as substrates, this is not critical because they do not react with the atmospheric oxygen, but for semiconductors, this is not easily done. One method is to prepare as many samples as planned measurements, and to use each sample only one time. Despite the great efforts which are connected with this method, we chose to use it because it opened up the full choice of surface-science techniques [53, 64]. In the following, we present a time-dependent XPS analyzed study on the speed of self-assembly. Our concrete system is 1 mmol concentrated HDT in isooctane on InP (110).

The XPS intensities are presented in Figure 9. In the beginning, all signals, except the substrate-related signals, increase quickly until the alkyl chains start to cover the binding sulphur atoms of the thiol groups. This is the point

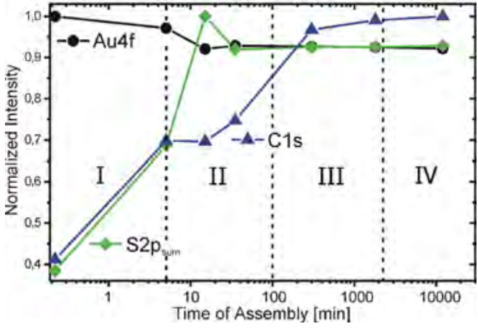


Figure 9. The self-assembling process is monitored via time-dependent XPS spectroscopy of the important substrate and adsorbate elements. Adsorbate: HDT in isooctane [64].

where the ordering starts. After 24–36 h, the system does not change anymore. In order to be on the safe side, we chose assembly times >36 h for our experiments on well-ordered SAMs.

3.2. Passivating Properties

III–V semiconductors are not immune to oxygen. A freshly cleaved (110) surface, and of course other surfaces as well, would oxidize within fractions of a microsecond if not protected against atmosphere. Furthermore, if chemical etching is needed in the production of, for example, nanostructured devices (see Section 4), the monolayer should be able to resist the etching on the time scale of minutes. In the following, we will discuss the passivating properties of self-assembling monolayers on the example of alkanethiols on InP(110). First, we briefly note the preparation procedure which is necessary to make a monolayer free of defects (for more details regarding the preparation, we refer to Section 2). If the self-assembling takes place in a dilute solution, it is necessary to purge the solution first with nitrogen (or Ar) for about 20 min. Otherwise, the residual oxygen in the organic solvent immediately reacts with the substrate surface [32, 64]. But, if the prenotched crystal bars are cleaved in a degassed solvent, which contains μmol · mmol concentrated thiols, oxygen is absent, even in an extremely surface-sensitive SXPS measurement [32, 64]. In order to test the passivating properties of samples with fully assembled monolayers, we exposed the samples to atmosphere and humidity on time scales of hours, days, and even months.

While no changes of the samples were spectroscopically detectable within days, as judged from the SXPS sensitivity scale, on the time scale of a month, slight changes were detectable [64]. But even this slight change proved that at least 90% of the samples were not affected. These results show that readily assembled monolayers can be handled without further precautions in technological processes.

Then, we exposed similar samples to electrochemistry [64, 73]. At the start, a potential which was suitable for etching of the semiconductor was applied and changed to cyclic, while in parallel the current voltage (*CV*) ratio and the capacity were recorded. The results showed that the monolayer protects the semiconductor surface effectively [64, 73]. After some minutes, the monolayer showed the first degradation, but from the current density, it was clear that this is only a minor effect, starting at the edges and at imperfections (e.g., cleavage steps) of the substrate (for more details, refer to [64, 73]).

Summarizing the results, alkanethiol monolayers (for length >6) protect III–V semiconductor (110) surfaces well enough to be used as masks, especially if nanostructured (see Section 4).

3.3. Structure via Spectroscopy

In this section, we will analyze the complete three-dimensional structure of the adsorbate–substrate system from a multitude of spectroscopic approaches. First, we describe investigations of the X-ray absorption near-edge structure (XANES) of adsorbed hexadecanethiol monolayers on InP (110).

We will demonstrate that the self-assembly on III–V semiconductor surfaces results in a unique arrangement of the alkyl chains with a rather well-defined polar angle of the chains with respect to the surface normal. However, on the metal single-crystal surfaces investigated so far, the molecular orientation in the azimuthal direction is not uniform. This is because the surface symmetry permits the formation of several domains of differing azimuthal molecular orientation. Here, we report on self-assembled monolayers, which possess a very high degree of molecular order in both the polar and azimuthal orientation of the organic molecules with respect to the substrate. This unique feature has been observed during investigations of III–V (110) surfaces, which were performed in order to explore the self-assembling properties of compound semiconductor surfaces. These conclusions have been derived from the X-ray absorption near-edge structure XANES measurements taken at the carbon K edge (CK XANES), which have been carried out to determine the main structural features of the alkyl chains on InP (110). First, core-level photoemission XPS studies were performed in order to characterize the adsorbed organic layers of interest, which were prepared by adsorption from millimolar solutions of thiols in isooctane for durations of 24–36 h, according to Section 4.4. InP (110) faces were prepared by cleavage of prenotched bars in solution (see Section 4.4). As we will discuss in more detail in Section 4, essentially only one sulfur species is present on the freshly prepared SAMs in XPS and SXPS experiments. Also, taking into account the complete absence of oxygen-related features, even after contact with the atmosphere, we conclude that alkanethiols form stable and contamination-free monolayers on InP (110). Examples of other investigations on similar semiconductor surfaces can be found in [24, 25]. The XANES measurements in a partial electron yield mode on SAMs of adsorbed alkanethiols on InP 110 were carried out at the beamlines HE-TGM2 and PM5 of the BESSY I storage ring in Berlin. Figure 10 displays angular-resolved CK XANES of a monolayer, prepared by adsorption of hexadecanethiol (HDT). The absorption spectra were recorded at several incidence directions of the soft X-ray photons with respect to the surface normal polar angle (−60 to 80°). Here, the polarization vector of the linearly polarized synchrotron radiation was fixed in the azimuthal orientation to allow only variation parallel to the (110) plane. The spectra exhibit two main absorption peaks. The peak at the higher photon energy is attributed to σ^* C–C related resonance. The strong feature at the lower photon energies near the edge is mainly due to the σ^* C–H orbital [66, 67]. However, the well-resolved splitting that occurs at certain angles suggests a more complicated origin of this feature, which will not be discussed here. Notice the very good matching of the background for all of the spectra at both the lowest and highest photon energies. This is evidence for the reliability of the spectra following the flux-normalization procedure. The spectra show significant angular-dependent changes. In particular, the intensities of these features do not proceed in parallel, but they show pronounced maxima at quite different incidence angles, exhibiting nearly an inverse behavior. However, we will start focusing on the angular dependence of σ^* C–C related absorption, which contains information on the spatial orientation of the carbon backbone of the adsorbed molecules. Therefore, a peak shape analysis has been performed after subtracting an arctan-like step function which emulates the CK edge. In our special case, the exact position and width of this edge are not critical to the fit. The data in Figure 11 display the evolution of the absorption of the σ^* C–C resonance of the CK XANES of HDT (Fig. 10) in dependence on the photon incidence direction as characterized by the polar and azimuth angles. The data have been derived from σ^* C–C intensities obtained from the fits, following flux normalization of the recorded absorption spectra.

The solid line represents a fit curve to the experimental absorption values which will be explained later. The graph shown in Figure 11 reveals a clear minimum, as well as a maximum separated by about 90°. From the angular

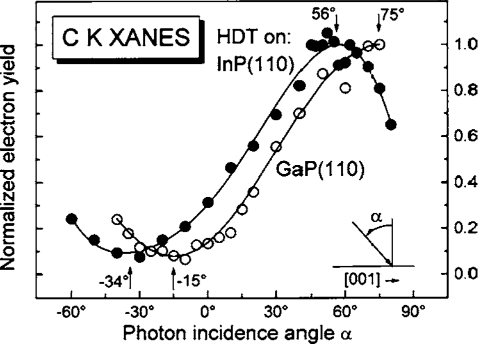


Figure 10. Series of normalized XANES absorption spectra of an adsorbed HDT monolayer on InP (110) with different polar angles. The retard potential of the partial yield detector was set to 150 V. Reprinted with permission from [66], D. Zerulla and T. Chassé, *Chem Phys. Lett.* 311, 8 (1999). © 1999, Elsevier Science.

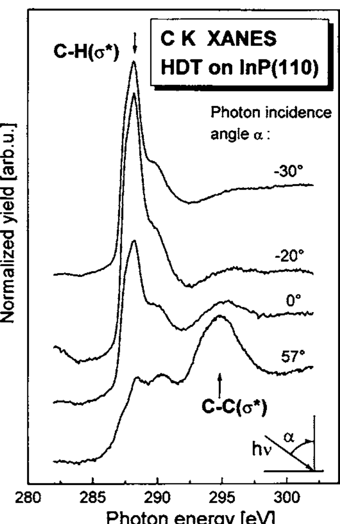


Figure 11. Plot of the absorption strength of the C–C(σ^*) resonances of HDT monolayers on InP and GaP in dependence on the polar angle. Normalization was performed as described in the text, and the maxima were set to 1. Error bars given for 0 and 80°. Reprinted with permission from [67], T. Chassé et al., *Surf. Rev. Lett.* 6, 1179 (1999). © 1999, World Scientific Publishing.

position of these extrema, the tilt angle can be determined to be about 34° from normal. Furthermore, we suggest from the nearly vanishing absorption in the minimum and high absorption in the maximum that a high degree of order has to be present in the layer. Note that the degree of linear polarization of the synchrotron beam itself is near 85%, and therefore not permitting, for example, the complete vanishing of the adsorption in the minimum, even if the order in the layer is perfect. In order to shed light on the azimuthal orientation of the chains, additional intensity data were obtained from a set of measurements taken at normal incidence (polar angle 0°), but varying azimuth angles of the sample 0–360° by rotating the sample with respect to its surface normal. The normalized intensities shown in Figure 12 in this manner also represent different orientations of the polarization vector with respect to the crystallographic directions in the surface plane. The graph clearly shows two striking absorption maxima at azimuth angles close to about 0 and 180°, and low absorption for polarization vector orientations far off these preferred angular directions. This very pronounced preference for strong X-ray absorption at polarization vector orientations about parallel to the (110) plane of InP (110) is rather exceptional. Although the graph of the experimental data is not exactly symmetric, this result provides strong evidence for a preferred tilt of the alkyl chains from normal toward the [001] direction. Because of the high anisotropy of this plot and the number of tested samples, we may essentially rule out averaging effects due to the possible existence of several domains possessing different tilt directions. This is important for the quantitative evaluation of the tilt angle because the data will only reveal the correct tilt angle if the azimuth polarization vector is chosen to correspond to a maximum in Figure 12, as has been done by recording the data displayed in Figure 11. In fact, the solid line in Figure 11 results from a fit to the experimental data following a $\cos^2$ function in dependence on the angle, as suggested by the matrix elements. This fit function has been slightly modified by a low-order polynomial in order to

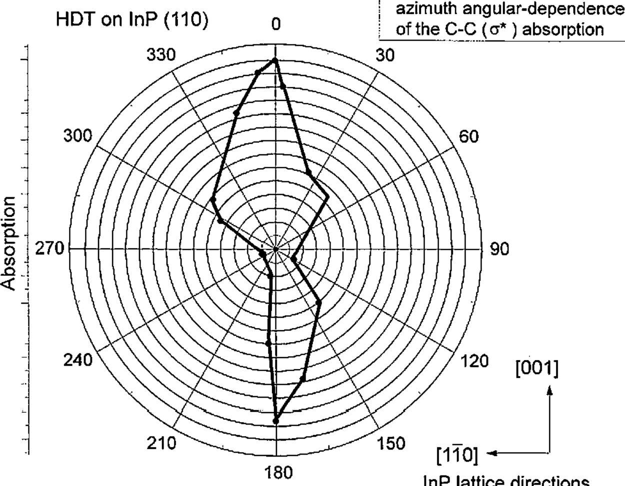


Figure 12. Plot of the changing absorption strength of the σ^* C–C resonance by varying the azimuthal angle while maintaining the polar angle at 0°. Reprinted with permission from [66], D. Zerulla and T. Chassé, *Chem Phys. Lett.* 311, 8 (1999). © 1999, Elsevier Science.

take into account an intensity decrease at highest incidence angles due to limited sample size. From the positions of the maximum and minimum in Figure 11, we can thus directly derive the tilt angle of the alkyl chains to be 34°+/−3°. The error has been estimated by taking into account both the sensitivity to intensity changes near normal incidence (Fig. 11), as determined from the slope of the fit curve, and the difference of the strength of the two opposite absorption maxima in Figure 12. Note, however, that the uncertainty in the azimuth angle of about 10° is much larger, unfortunately.

Thus, we cannot rule out slight deviations from an exact alignment within the (110) plane on this scale. Nevertheless, the experimental data presented above have clearly demonstrated the well-defined orientation of the adsorbed monolayer, resulting in an exceptionally high order of the adsorbed monolayer. To the best of our knowledge, a similarly pronounced preference of molecular orientation and high molecular order of alkanethiols has not been observed before on any other surface so far, including metal surfaces.

This strikingly high perfection of orientation of the adsorbed alkanethiol layers has to be discussed in terms of both adsorbate–substrate bonding and surface structure of the substrate. First, a preferential bonding of sulphur atoms to the surface indium atoms has been observed in several photoemission studies of inorganic sulphur species on InP (110) and InP (001) surfaces [35–37]. In particular, the formation of In–SH and P–H surface bonds has been reported in the case of the dissociative adsorption of H_2S on InP(110) [33, 34]. Second, we have performed synchrotron light-excited SXPS studies with alkanethiols on InP (110). These very surface-sensitive measurements also support the thesis of S–In bonding [73].

The rectangular surface unit cell of this zinc-blende type of substrate is characterized by In–In/P–P spacings along the [001] and [1–10] surface directions of 587 and 415 pm, respectively. The surface indium and phosphorus atom In–P bond lengths, 254 pm, are linked to form In–P zigzag chains within the InP (110) surface plane, which extend along the [1–10] direction. Thus, following structural arguments regarding the space requirements of the alkyl chains and their well-known tendency to form dense structures, a tilt of the alkyl chains toward [001] is indeed much more likely than toward [1–10]. These simple arguments fit very well with our observations regarding the tilt direction, which are schematically summarized in Figure 13. Further, a dependence of the tilt angle on the lateral distance of the sulphur–substrate bonding sites on the surface, and thus on the substrate lattice constant, has to be expected. In fact, we have observed such a trend by investigating thiol adsorption on different semiconductor surfaces, and the results are presented in [67]. Besides this purely geometrical consideration, we have to be aware of the structural inequivalence of the two opposite [001] surface directions due to the asymmetry of the In–P all-trans chains, which exhibit an indium and a phosphorus side. This structural asymmetry is also known to be responsible for the opposite spatial orientations of the empty and filled dangling bond orbitals of the surface indium and phosphorus atoms, respectively. Hence, it is reasonable to suggest that the experimentally observed preference of just one orientation of the alkyl chain has to be attributed to the

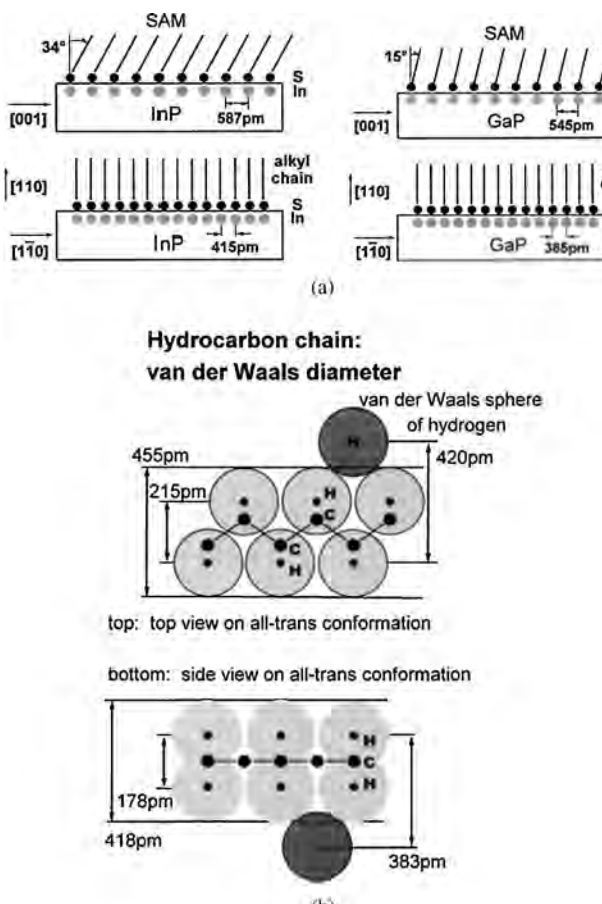


Figure 13. (a) Schematic structural model for the tilt orientation of the alkyl chains of HDT on InP(110) and GaP(110) (side views). Note that details of the local adsorption geometry are not reflected by this drawing. The surface indium and gallium atoms are included in order to emphasize the bonding to the sulphur atoms. The surface phosphorus atoms are not shown. (b) Schematic model of a hydrocarbon chain in an ideal all-trans conformation in combination with the hydrogen van der Waals radii. Space requirements due to the van der Waals interactions are provided by numbers. The reduction of these van der Waals dimensions by approaching an optimum registry is also indicated. Reprinted with permission from [67], T. Chassé et al., *Surf. Rev. Lett.* 6, 1179 (1999). © 1999, World Scientific Publishing.

orientation of these dangling bonds, most probably to the empty indium dangling bonds. Thus, structure as well as bonding constraints evidently force the adsorbed monolayer to adopt a uniquely well-defined structure, which allows us to propose the existence of single domains of the monolayer on InP (110), only limited by high-cleavage steps or other large defects of the cleaved substrate surface. Summarizing our results, we have observed the formation of a passivating, self-assembling organic monolayer on InP (110) by the adsorption of alkanethiols from the liquid phase. A tilt angle with respect to the sample normal of 34° and an azimuthal orientation parallel to the (110) plane of the InP lattice have been derived from the quantitative analysis of the XANES. The exceptionally well-defined orientation and high order within the adsorbed monolayer, as may be deduced from the strength and anisotropy of the X-ray absorption features, make this interface a candidate for a single-domain monolayer-adsorption system.

We will now extend the XANES measurements and discuss them in more detail. The structure of alkanethiol monolayers on III–V(110) surfaces was studied by analyzing the X-ray absorption near-edge structure (XANES) of the carbon K edge. Pronounced absorption maxima were observed for special orientations of the polarization vector of the radiation as revealed from angular-dependent measurements, suggesting a rather well-defined molecular axis of the alkyl chains. From quantitative evaluations of these angular dependences, the chains were found to be tilted from the normal toward the [001] direction of the (110) surfaces by 34 and 15° in the case of HDT adsorption on InP and GaP, respectively. The similarities as well as the differences in tilt angles between the substrates are discussed in terms of constraints imposed by the surface structure and lattice constants, as well as the space requirements of the van der Waals spheres of the adsorbed thiols. A unique feature observed on these monolayers is the nearly complete alignment of the alkyl chains with respect to the azimuthal orientation. We suggest that this adsorbate system represents the case of a single-domain orientation within the organic monolayer.

The spectra were recorded in a partial electron yield mode at a pressure of $5–8 * 10^{-10}$ mbar. The retard potential of the partial yield detector was typically set to 50 V. All spectra were normalized with respect to the incident photon flux and the independently recorded background curve, which represents influences due to carbon contaminations on the optical elements of the beamline. The intensity calibration was performed by normalizing to the intensity difference between the spectrum background below the edge and at the highest energies, well above the resonances of interest. The orientation of the polarization vector of the soft X-rays was varied systematically by changing the incidence angle of the radiation with respect to the crystallographic orientations of the substrates in order to take advantage of the polarization dependence of the dipole matrix elements, which govern the absorption process.

Recent photoemission studies of the adsorption of alkanethiols (chain lengths $n = 10–18$) on InP(110) have demonstrated that alkanethiols form stable and contamination-free monolayers on this semiconductor surface [32]. This conclusion has been drawn from the complete absence of oxygen-related features in the spectra, even after contact with atmosphere, and the existence of essentially only one distinct sulphur species in freshly prepared thiol layers, which has been attributed to In–S bonds at the interface [32–37, 53, 66, 67]. Note that, from the chemical point of view, based on the experience outlined above as well as from the steric limitations due to the large space requirements of the alkyl chains, which will be discussed below, P–S bonds have been excluded via SXPS measurements, and will not be considered here [64]. In order to derive information on the intramolecular bonding, and most importantly on the structure of the adsorbed monolayer, we have investigated the X-ray absorption near-edge structure of the carbon K edge from alkanethiol SAMs on the cleavage faces of the III–V semiconductors InP and GaP.

Absorption spectra were recorded from SAMs of hexanethiol (HT) up to hexadecanethiol. Most complete datasets were gathered for the longer alkanethiols, which are well known to establish highly ordered SAM domains on metallic substrates like Au(111). In this paragraph, we will focus on the interpretation of these spectra, which were measured on the cleavage faces of InP and GaP. Figure 10 displays a set of CK XANES of a monolayer prepared by adsorption of HDT on InP(110). The absorption spectra in the figure were recorded at several incidence directions of the soft X-ray photons with respect to the surface normal, while keeping the azimuthal orientation of the polarization vector strictly parallel to the [001] surface direction. The spectra exhibit three absorption peaks. The broad peak at the highest photon energy, close to 295 eV, is attributed to the σ^* resonance, which is due to absorption by electron excitation from the C1*s* core level into the antibonding orbitals of the C–C bonds (hereafter called C–C(σ^*) resonance). The sharp peak at the photon energy of about 289 eV, close to the CK edge, mainly corresponds to the C–H(σ^*) resonance [68]. The third peak feature, which may be observed at about 2 eV higher energy compared to the latter, is much smaller in intensity. This peak feature appears much stronger in the absorption spectra for the shorter thiols, suggesting an origin due to the C–S bonds at the end of the alkyl chains. We emphasize the very good matching of the background for all of the spectra at both the lowest and highest photon energies, which is evidence for the reliability of the flux normalization procedure. Both C–C(σ^*) and C–H(σ^*) resonances show significant angular-dependent changes, which are reflected by the evident peak shape changes between the spectra. In particular, the intensity of the C–C(σ^*) resonance decreases when going from positive toward negative polar angles with respect to the surface normal. Note that there is no symmetry of the intensities with respect to the surface normal. Similar spectra of HDT monolayers on GaP(110) are shown in Figure 14. The overall shape of the spectra and the evolution of the intensities exhibit the same general trends as described above for HDT on InP(110). A noteworthy exception to this evident similarity is the smaller energy difference between C–C and C–H related resonances in the case of HDT on GaP. This energy difference is reduced by 0.7 eV (+/−0.3 eV) compared to HDT on an InP substrate. However, this difference calls for further statistical analysis of the whole body of experimental spectra and accompanying theoretical considerations, which will not be provided here. The C–C(σ^*) resonances contain the information on the spatial orientation of the carbon backbones of the adsorbed alkanethiols, which are of particular interest for the structure of the SAM. According to the dipole matrix elements for X-ray absorption, the intensity of the CK absorption resonance will be highest if the *p* contribution to the spatial orientation of the particular molecular orbital is parallel to the polarization of the X-ray light [74, 75]. The significant dependence of the resonance intensities on the polar angle of the photon incidence, and thus also the orientation of the polarization vector, as exhibited by the CK resonances in Figures 10, 11, and 14, indicates a preferential orientation of the carbon bonds of the organic adsorbates on both InP and GaP(110). Quantitative evaluation of the polar and azimuthal angles

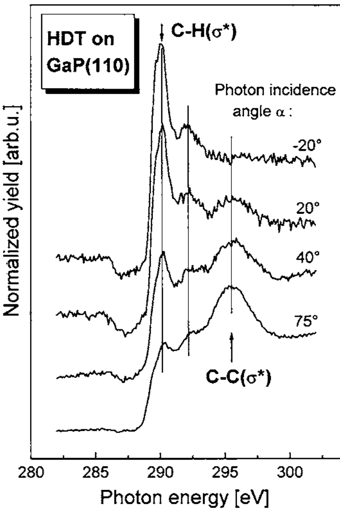


Figure 14. Normalized XANES absorption spectra of an adsorbed HDT monolayer adsorbed on GaP(110) at different polar angles. The retard potential of the partial yield detector was set to 50 V. Polarization vector and azimuth orientation parallel to [001]. Reprinted with permission from [67], T. Chassé et al., *Surf. Rev. Lett.* 6, 1179 (1999). © 1999, World Scientific Publishing.

to be associated with this preferential orientation cannot be derived from the data in Figure 11 alone without detailed knowledge of the corresponding azimuth dependences.

We have recently shown, for HDT monolayers on InP(110), that a very pronounced anisotropy of the C–C(σ^*) resonance can be observed by rotating the azimuthal orientation of the polarization vector around the surface normal at normal photon incidence [66, 67]. In brief, the observed azimuthal dependence of the C–C(σ^*) intensity was strongly peaked along the two opposite [001] surface directions. The intensity nearly vanished far off these surface directions, in particular along [1–10]. Hence, a predominant orientation of the alkyl chains of HDT toward [001] was predicted [53, 66, 67]. In this manner, the anisotropies exhibited in Figures 10, 11, and 14, which were obtained for the particular azimuth with the highest C–C(σ^*) intensity, directly reflect a highly preferred orientation of the C–C bonds with respect to the surface normal, that is, a tilt angle of the alkanethiol chain. A peak shape analysis of the CK spectra has been performed in order to obtain quantitative data. An arctan-like step function was used to emulate the CK edge. In the special case of the C–C (σ^*) resonance and determination of its intensity, the exact position and width of this edge are not critical to the fit because this high-energy peak is already situated well above the threshold. The data in Figure 11 display the evolution of the absorption of the C–C (σ^*) resonance of the CK XANES of HDT (Figs. 10, 14) as a function of the photon incidence direction in terms of the polar and azimuth angles. The experimental data represent the intensities derived from the analysis of the spectral shapes, following appropriate flux normalization of the recorded absorption spectra. Closed and open circles denote data obtained from InP(110) and GaP(110), respectively. The solid lines represent fit curves to the experimental absorption values for SAMs on these two surfaces. The applied model functions for the fits represent the angular

dependence by the term [$\cos^2(90° - (\alpha + \beta))$], essentially, including a single tilt angle $\beta > 0$ of the alkyl chains and the variable photon incidence angle α. The sign of the incidence angle α was arbitrarily chosen to obtain the maximum and minimum for $\alpha > 0$ and $\alpha < 0$, respectively.

The choice of this type of fit function is directly related to the polarization dependence of the absorption matrix elements and Fermi's golden rule [74]. We were able to neglect any effects due to azimuth orientation in the $\cos^2$ function because the experimental data to be analyzed were taken for the favored high-symmetry azimuth orientation, characterized by the polarization vector within the tilt plane, which is determined by the surface normal and the preferred [001] direction. The fit function has been slightly modified by a low-order polynomial in the case of the InP data in order to take into account an intensity decrease at highest incidence angles due to limited sample size. We further emphasize that no angular averaging was included in the fit functions (for example, regarding a range of tilt angles β or several distinct values for α), as might originate from differently oriented domains or from the absence of a preferential orientation. Nevertheless, the fit model matches the experimental data very well over the whole range of investigated incidence angles. We may conclude that a strongly preferred orientation of the alkyl chains exists with respect to the polar angle as well. The pronounced intensity maxima occurring only at certain azimuth and polar angles fix the observed preferential orientation in spherical coordinates by two independent angles, demonstrating the existence of a rather well-defined molecular axis of the hydrocarbon chain of the adsorbed thiol, which extends toward this direction. The data from InP show a clear minimum, as well as a maximum separated by about 90° (Fig. 11). From the angular positions of these extrema, the tilt angle β of this molecular axis of the HDT chain on InP(110) can be determined to be about 34°(+/−3°) from the surface normal toward [001]. In a completely analogous manner, we may determine a tilt angle of 15°(+/−3°) from the normal for HDT on GaP(110). We cannot draw any conclusion on preferred tilts toward the two opposite [001] directions, [001] or [00–1], from the XANES data alone. Although indistinguishable by LEED patterns, these two directions are not equivalent on III–V(110) surfaces because of the existence of the asymmetric In(Ga)P zigzag chains in the surface plane, which extend along [1–10]. This structural asymmetry is also known to be responsible for the opposite spatial orientations of the empty and filled dangling bond orbitals of the surface indium and phosphorus atoms, respectively. Hence, it is reasonable to suggest that the experimentally observed preference of just one orientation of the alkyl chain has to be attributed to the orientation of these dangling bonds, most probably to the empty indium dangling bonds. In fact, by combining analysis of the asymmetric features in the XPD patterns of InP with the XANES data, we have been able to identify the absolute tilt orientation as pointing from the indium side (sulphur-bonding site near indium due to In–S bonds) toward the phosphorus side of the same zigzag chain, as will be discussed in detail elsewhere [53, 64].

An independent check of the orientation of the hydrocarbon chain may be performed by analysis of the intensity evolution of the C–H (σ^*) resonances in the CK XANES. Figure 15 shows a comparison of the angular dependences of C–C (σ^*) and C–H (σ^*) resonances of HDT on InP(110), derived from the same set of spectra. The C–H (σ^*) resonance intensities were obtained by fitting Gaussians to the sharp low-energy peak only, after subtracting the arctan-like threshold function, characterized by a position at 289.5 eV and a full width at half maximum of 0.5 eV, and by excluding the contributions from the second, small resonance. Because of this fitting procedure, the C–H (σ^*) intensity is much more prone to error than the C–C (σ^*) resonance. Evidently, these intensities evolve roughly inversely. This result may also be confirmed for GaP by a glance at the spectra displayed in Figures 10 and 14 for InP and GaP, respectively.

An inverse behavior is in accordance with the above-mentioned proposal of a rather well-defined molecular axis. The individual bond angles C–C–H will be close to the tetrahedral angle due to the ideally adopted sp^3 hybridization of the carbon valence electrons. But in the case of an alkyl chain that extends along an axis, that is, in an ideal all-trans conformation, the C–H bond directions are expected to be 90° off the average C–C orientation, corresponding to the molecular axis. The slight deviation of the minimum/maximum positions of the two curves from a 90° difference may be related to some distortions of the chains. However, we have mentioned the much larger sources of errors for the C–H (σ^*) related resonance, and thus we cannot make any quantitative statements on the significance of this suggestion. Another argument for the presence of a high degree of molecular order in the HDT monolayer may be derived from the high anisotropy of the minimum/maximum intensity ratios of C–C (σ^*) angular dependencies. The absorption curves in Figure 11 were normalized by setting the maxima of the fit curves to 1.0 arbitrarily. Note that the degree of linear polarization of the synchrotron beam itself is near 85%. Therefore, complete vanishing of the absorption may not be expected, even if the order in the layer was perfect. Thus, we can also suggest, from the rather low values of absorption intensity in the minima of Figure 11, that a high degree of order regarding the orientation of the alkyl chain axis should be expected in the organic HDT layers on both investigated

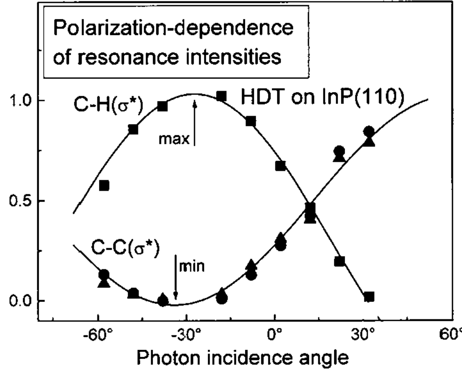


Figure 15. Comparison of the polar angle dependences of the intensities of C–C(σ^*) and C–H(σ^*) resonances for HDT on InP(110). Reprinted with permission from [67], T. Chassé et al., *Surf. Rev. Lett.* 6, 1179 (1999). © 1999, World Scientific Publishing.

substrates. Furthermore, we have performed simulations of the angular dependence in relation to simple models of the intramolecular structure, which are shown in Figure 16. The degree of polarization of the X-rays was assumed to be 100%, and the angular dependence was modeled by a bare $\cos^2$ function of the angle between the possible C–C bond orientations and the polarization vector of the radiation. An ideal all-trans conformation was adopted as a simplified model of a hydrocarbon. The highest absorption and the maximum anisotropy are obtained, as expressed by the minimum/maximum ratio (min = max = 0), for a hypothetical linear carbon atom chain. A tetrahedral C–C–C angle of 109.5° of a carbon zigzag chain within the plane of the variation of the polarization vector reduces and increases the absorption in the maximum and minimum, respectively, resulting in a dramatically reduced anisotropy (min = max = 0.5). The anisotropy would completely vanish (min = max = 1) in the other hypothetical case of a C–C–C angle of 90°, as shown in Figure 16. However, a minimum/maximum intensity ratio of 0.5 and larger would be in strong contrast to the much lower experimental values (min = max $<$ 0.2) which may be taken from Figure 11. Another idealized hypothetical model considers a rotation of the C–C zigzags by 90° with respect to the plane determined by the polarization vector and molecular axis. The total intensity of the maximum is reduced within this model, but interestingly, the anisotropy remains high (min / max = 0). While this behavior would fit the observed high anisotropy in the polar angle dependence, it would be in contrast to the azimuthal anisotropy. Therefore, we may rule out the extreme cases of high-symmetry orientations of all-trans conformations with respect to [001] or [1–10]. In contrast, twisted alkyl chains or, to a certain extent, even some random orientations of individual C–C bonds of the chains fit the experimental observations much better. This result does not at all contradict the generally observed extremely well-defined orientation of the molecular axis. So far, the experimental data and the analysis presented above have clearly demonstrated the well-defined orientation of the adsorbed alkanethiol monolayers (HDT), resulting in an exceptionally high order of the adsorbed monolayers. To our knowledge, a similarly pronounced preference of molecular orientation and high molecular order of alkanethiols has not been observed before on any other surface, including metal surfaces. This strikingly high perfection of orientation of the adsorbed alkanethiol layers has to be discussed in terms of both adsorbate–substrate bonding and surface structure of the substrate. Preferential bonding of sulphur atoms to the surface indium atoms has been observed in several photoemission studies of sulphur species on InP(110) and InP(001) surfaces [35–37], and surface-sensitive core and valence level photoemission measurement of several alkanethiols on III–V(110) also support the thesis of S–In(Ga) bonding [33–37]. Thus, structural constraints by the substrates will be imposed by the different In–In (Ga–Ga) distances in the surface, which provide the links to the adsorbed thiols. The III–V(110) surfaces of zincblende structure possess a rectangular surface unit mesh. The unit mesh dimensions correspond to 587 and 544 pm along the [001], and to 415 and 385 pm along the [1–10] surface directions of InP(110) and GaP(110), respectively. This reflects the 7.9% larger lattice constant of InP. The unit mesh dimensions also correspond to the In–In (Ga–Ga) distances in the surface plane. Following straightforward structural arguments regarding the space requirements of the alkyl chains and their well-known tendency to form dense structures, a tilt of the alkyl chains toward [001] is indeed much more likely than toward [1–10]. These simple arguments fit our observations regarding the tilt direction very well, and are summarized in Figure 13(a) schematically. Further, a dependence of the tilt angle on the lateral distance of the sulphur–substrate bonding sites on the surface, and thus on the substrate lattice constant, has to be expected. This effect is clearly reproduced in our experiments by the smaller tilt angle of 15° on GaP compared to 34° on InP. From these data, we may estimate a vertical distance between the average molecular axes of 487 pm (InP) and 525 pm (GaP), in reasonable agreement with qualitative expectations.

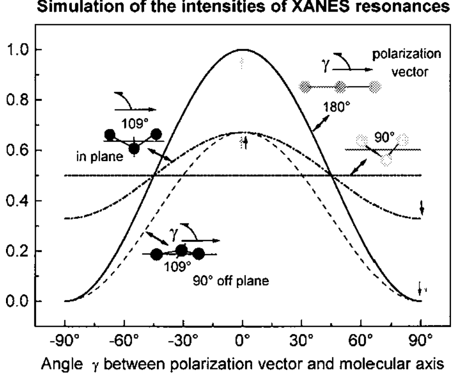


Figure 16. Simulation of the influence of intramolecular arrangement of the carbon atoms and the relative orientation of the C–C plane with respect to the polarization vector during an angular scan for hypothetical orientations of ideal all-trans conformation. Reprinted with permission from [67], T. Chassé, et al., *Surf. Rev. Lett.* 6, 1179 (1999). © 1999, World Scientific Publishing.

So far, this discussion has ignored the actual numbers for the space requirements of the alkyl chains, which essentially originate from the lateral dimensions of the covalently bound alkyl chains and the van der Waals radii of the hydrogen atoms (120 pm) of the organic molecule. The hydrogen atoms are mainly responsible for the intermolecular interactions between the hydrocarbon chains. The relevant data may be obtained from the schematic structures shown in Figure 13(b). It is evident that a problem exists regarding the space requirements of the chains parallel to [1–10], which is most serious for GaP because of the smaller lattice constant. However, by tilting or/and twisting the alkyl chains with respect to the molecular axis, an optimum registry of the van der Waals spheres of hydrogen atoms from neighboring chains may be obtained, which may significantly reduce the minimum van der Waals chain–chain distance from 418 down to 370 pm in the optimum case. Hence, by twisting the chains and adopting the optimum registry of hydrogen atoms between neighboring chains, the structural constraints imposed by the substrates may be essentially fulfilled. We suggest that the actually adopted tilt angles (34 versus 15°), as well as the resulting minor differences in average chain distances (487 versus 525 pm), may be affected by the need

to reach an optimum registry in this direction too. We conclude that structural constraints will not prevent any Ga (In) surface atom from forming covalent bonds to the sulphur heads. This is an important result in terms of passivation of the surface indium (gallium) atoms. Otherwise, missing alkanethiols would permit easy oxidation of the unpassivated In (Ga) in air. Note that passivation of the P-surface sites has to proceed along another route. We further speculate that the need for optimum registry may additionally contribute to the excellent order of the chains in the layers. Summarizing our results, we have demonstrated the formation of a passivating, self-assembling organic monolayers on InP(110) and GaP(110) by adsorption of alkanethiols from the liquid phase. An exceptionally well-defined orientation and high order within the adsorbed monolayers on these semiconductor surfaces has been found by employing the angular dependences of CK XANES resonances. Tilt angles toward the [001] directions of the (110) surfaces with respect to the sample normal of 34 and 15° for InP and GaP, respectively, were obtained from quantitative analysis of the XANES. The considerable difference in lattice constants permitted us to draw some conclusions concerning the substrate influence in the case of these semiconductors. The bonding to the surface and surface structure, as well as intermolecular interactions, impose constraints, which evidently force the adsorbed monolayers to adopt a uniquely well-defined structure. The constraints result in single domains of the monolayers on the investigated III–V (110) surfaces, only limited by high-cleavage steps or other large defects of the cleaved substrate surface.

The sensitivity of XANES to determine the spatial orientation of the bonding axis (especially the C–C and C–H bonds at the CK edge) has been used to construct a model of the complete alkanethiol monolayer with respect to the underlying InP(110) lattice. The angular-dependent (in azimuth and vectorial angle) datasets have been measured, while averaging an area (spot size) of several mm^2, hence representing a truly macroscopic feature. Although the integration of spectral features over large regions has been made, the data reveal sharp angular-dependent changes. In particular, these changes are not symmetric with respect to different surface directions, which has to be expected for an oriented, single-domain system. More precisely, the following conclusions have been drawn from the XANES measurements.

- The angle between the surface normal and the tilted alkyl chains is 34° on InP(110) and 15° on GaP(110) [66, 67].
- The projection of this angle onto the surface lies along the [001] axis of the substrate, which has been confirmed by varying the azimuth angle parallel to the surface from 0 to 360° in 24 steps [66].
- Within the (001) plane, the chains are only tilted to one side of the surface normal, which is confirmed by varying the vectorial angle from −70 to +70° from the normal in more than 30 steps [53, 67].

From SXPS measurements which have been tuned for maximum surface sensitivity, it additionally has been proved that the chemical binding of SAM and the substrate takes place between the thiol sulphur and the indium, while the hydrogen atoms which are deprotonated in the binding process are connected to the phosphorus atoms [65]. The set of XPS intensities also supports the thesis of a well-ordered system. The C/S ratio is at its maximum, while the corresponding substrate intensities are at their minimum. To demonstrate the cleanness of the self-assembled system, we have presented an XPS-survey spectrum [32, 66, 67]. Only peaks which are related to the substrate (In, P) or the adsorbate (C, S) are present. No oxygen or other contaminants are detectable, which is proof of a good surface preparation.

We are now ready to sketch a complete model of the adsorbate–substrate system. However, to eliminate all possible objections to our conclusions, we will further prove the exceptional order of the SAMs by using X-ray photoelectron diffraction on exactly the same UHV-adsorbed alkanethiol film (either prepared in a solution or by vacuum deposition). For that, we made two sets of angular-dependent measurements at room temperature on spot sizes of several mm^2. The first one, with angular variations within the (001) plane, shows no significant deviations in intensity. But the second set (consisting of two independent series), which is made in the (1–10) plane, shows a clear peak at 34°, but not at −34° from normal (see Fig. 17). This peak comes from the so called "channeling" of the photoelectrons along the alkyl chains, and therefore describes the orientation of the chains. Note two important additional pieces of information, which can be derived from these measurements. First, the graph is not symmetric with respect to the surface normal (only one

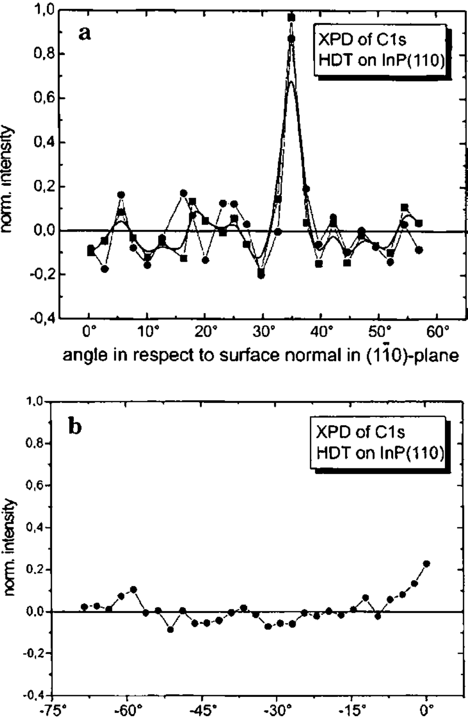


Figure 17. XPD experiment at the energy of the C1*s* peak: (a) along the (1–10) plane, (b) along the (001) plane of the InP lattice. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

peak), and second, the peak is absent in the perpendicular plane. This again allows only one conclusion: the adsorbed monolayer is a single-domain system. Furthermore, from our analysis, one may conclude that fully assembled monolayers, prepared from solutions, and UHV-adsorbed monolayers at high doses (>10 kLangmuir) do form the same kind of structure on InP(110) surfaces. In order to complete the model of the spatial geometry of the adsorbate–substrate system, we are in need of the exact orientation of the substrate lattice with respect to the adsorbate. Having the exact orientation in space of the alkyl chains, the only missing parameter is the orientation of the substrate. From the previously shown XPD measurements on preoriented InP single crystals, it is clear that the alkyl chains lie in the (1–10) plane. The symmetry properties of the semiconductor lattice further allow the determination of "tilt direction" with respect to the sequence and spacing of the nonaxial-symmetric substrate lattice in the denoted plane. Hence, a second set of XPD measurements has been made along the (1–10) plane. In order to investigate the orientation of the substrate and not the adsorbate, the P2*p* peak instead of the C1*s* peak was chosen. The result of this experiment is depicted in Figure 18. Because of the unequivocal result, an interpretation using simple first-order scattering is sufficient. The matching of the first-order scattering model and the XPD peaks is illustrated in Figure 19. Taking all ascertained data from the measurements, we can now complete the structure of the single-domain adsorbate–substrate system. The resulting model, which is additionally confirmed by computer simulations, is shown in Figure 20.

The STM/STS results lead to further insights into the behavior of self-assembling monolayers on III–V(110) surfaces. From the preceding spectroscopic investigations [32, 66, 67], which were carried out with alkanethiols adsorbed from solutions, it was clear that the organic films build up a very highly ordered and dense layer on large scales. The data derived from these spectroscopic measurements include the molecular tilt angle of 34° [InP (110)] and 15° [GaP (110)] from the surface normal into the [001] direction of the substrate. An explanation for this behavior can be given solely by geometrical considerations. The rectangular surface unit cell of this zinc-blende type of substrate is characterized by In–In (P–P) spacing along the [001] and [1–10] surface directions of 587 and 415 pm, respectively. The surface

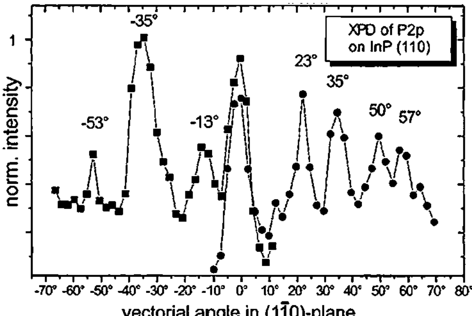


Figure 18. XPD experiment at the energy of the P2*p* peak along the (1–10) plane of the InP lattice. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

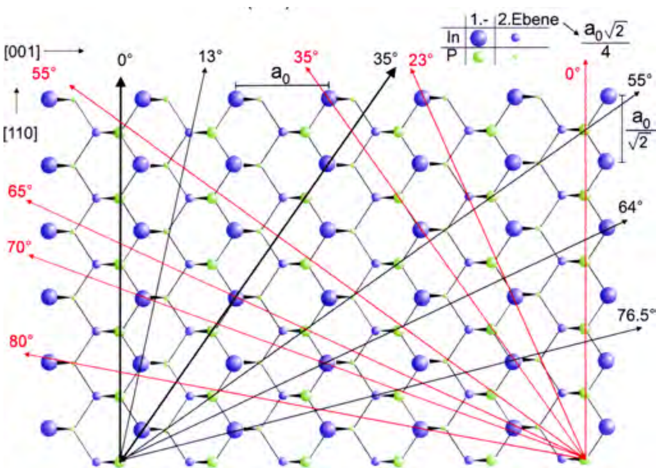


Figure 19. Illustration of the first-order scattering within the (1–10) plane. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

indium and phosphorus atoms (In–P bond length 254 pm) are linked to form In–P zigzag chains within the InP(110) surface plane, which extend along the [1–10] direction. Thus, following structural arguments regarding the space requirements of the alkyl chains and their well-known tendency to form dense structures, a tilt of the alkyl chains toward [001] is indeed much more likely than toward [1–10]. These simple arguments agree very well with our above-mentioned observations regarding the tilt direction.

Furthermore, a dependence of the tilt angle on the lateral distance of the sulphur–substrate bonding sites on the surface, and thus on the substrate lattice constant, had to be expected. In fact, we observed such a trend by investigating thiol adsorption on different semiconductor surfaces [53, 67].

Besides this purely geometrical consideration, we have to be aware of the structural inequivalence of the two opposite [001] surface directions due to the asymmetry of the In–P

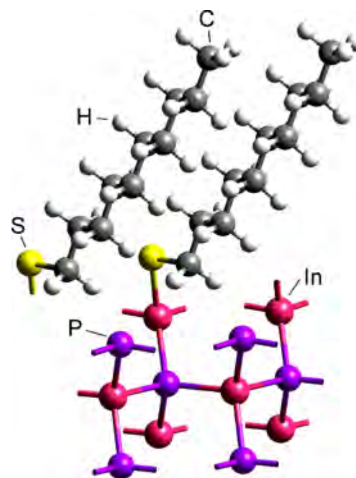


Figure 20. Model of the adsorbate–substrate system. Reprinted with permission from [53], D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002). © 2002, American Chemical Society.

zigzag chains, which exhibit an indium and a phosphorus side. This structural asymmetry is also known to be responsible for the opposite spatial orientations of the empty and filled dangling bond orbitals of the surface indium and phosphorus atoms, respectively. Hence, it is reasonable to suggest that the experimentally observed preference of just one orientation of the alkyl chain has to be attributed to the orientation of these dangling bonds, that is, to the empty indium dangling bonds, which was proved by SXPS measurements [53, 67]. Thus, structure as well as bonding constraints evidently force the adsorbed monolayers to adopt a uniquely well-defined structure, which allows us to propose the existence of single domains of the monolayers on InP(110), only limited by high-cleavage steps or other large defects of the cleaved substrate surface.

Summarizing all data, we have proved macroscopically, from pure spectroscopic investigations [32, 66, 67], as well as microscopically from STM/STS measurements, the spatial structure of the complete adsorbate–substrate system. We have shown on self-assembling systems prepared from dilute solutions, as well as on systems prepared in UHV, that the whole layer under investigation is highly ordered and oriented into a specific direction, and therefore is a single-domain system on InP (110) and GaP (110).

Additionally, from the comparison of different preparation techniques (UHV deposition versus dilute solutions), it was possible to shed more light on the role of solvent molecules in the mechanism of self-assembly and the necessary doses. From the measurements with low doses, it is obvious that the preorientation of the alkyl chains, pointing toward the crystallographic [001] axis of the substrate, is induced by the empty (In) dangling bonds on the surface, while the tilt angle itself is a result of the dense package and the three-dimensional space requirements of the alkyl chains at higher doses.

The different stages of adsorption also give an explanation of the comparable high doses needed to obtain a densely packed layer despite the high reactivity, and therefore high sticking coefficient, of the surface. The reason for the requirement of higher doses is that the lying molecules on the semiconductor surface hinder the contact between new incoming molecules and the substrate surface itself. This is also the main reason for the differences in adsorption kinetics in comparison to adsorptions from solutions. There, the alkanethiols are expected to be kept afloat by the surrounding solvent molecules, and that is why they do not block the surface at low coverages. The corresponding results on gold surfaces are quite similar, but do not lead to single-domain systems; see, for example, [76].

4. NANOSTRUCTURING OF SAMs

In principle, self-assembling monolayers are stable against etching, act as lithographic resists, and can be combined with functional groups, to mention only a few examples. Because of the potential technical applications of such organized assemblies in a variety of fields, for example, chemical/optical sensors, these systems have been studied with increased attention for a number of years, and different techniques to structure the monolayers have been developed.

In this section, we will discuss different techniques that enable us to write structures with or into a self-assembled monolayer. Furthermore, we want to shed light on the mechanisms behind the different techniques in terms of changes in the molecular structure. Coarsely, the methods could be divided into techniques which remove parts of the monolayer (Sections 4.1, 4.2) or techniques which add molecules to the surface (Sections 4.3, 4.4).

4.1. X-Ray and *E*-Beam Lithography

Here we will discuss two methods whose origins are quite different. We do this because the final mechanism of the removal of molecules at the surface is the same in both techniques.

First, we discuss the influence of long-term X-ray irradiation on the structure and chemical state of a ready-assembled alkanethiol monolayer on InP (110) and Au (111). For this, we have performed an XPS study in order to investigate these effects. For the general preparation of the monolayer systems, we refer to Section 2. However, because we feel that it is an important point for the interpretation of the experiments, we have checked the surface quality on the gold films and on InP(110) using an ultrahigh-vacuum scanning tunneling microscope. According to these investigations, the gold surfaces are composed of large flat terraces, exhibiting typical dimensions of 200 nm $\times$ 200 nm. The terraces are separated by steps with a height of a few atomic distances. The largest part of the surface belongs to the terraces, which are almost free of defects. On the contrary, only a few single-cleavage steps, typically separated by distances of several micrometers, were found on InP(110), which thus exhibits very large, atomically flat terraces [64].

Thiol monolayers (1–decanethiol (DT), Aldrich, 96%; dodecanethiol (DDT), Fluka, 97%; hexadecanethiol, Fluka, 95%; octadecanethiol (ODT), Fluka, 97%) were prepared in a μM solution in isooctane (Merck, SupraSolv, for organic trace analysis). Before the solvent was used, it was purged with nitrogen for about 30 min to remove the dissolved oxygen. The indium phosphide crystals were cleaved in the above-described solutions. After immersion times of approximately 36 h, the samples were removed, rinsed with fresh solvent, dried under a gentle stream of nitrogen, and inserted into the spectrometer immediately.

The spectra were fitted by a least squares procedure with a model peak shape (Gaussian–Lorentzian convoluted functions) after background subtraction from experimental data using a Shirley-type function (using the software Unifit 2.1 by Hesse et al. [77]).

The samples were characterized by the core-level spectra of the substrates (Au4f, In3d, P2p) and of the organic films (C1s, $S2p$). However, for the purpose of this chapter, we mostly focus on the sulphur core levels of the organic adsorbates. First, we will present the data obtained for the adsorbed alkanethiol monolayers on gold. Figure 21 shows a set of $S2p$ spectra recorded from a DT layer on gold. From bottom to top, the spectra represent increasing accumulated irradiation time by X-rays up to a total of 8.5 h. The bottom spectrum in this figure, which corresponds to the shortest irradiation time, is dominated by a single, well-resolved

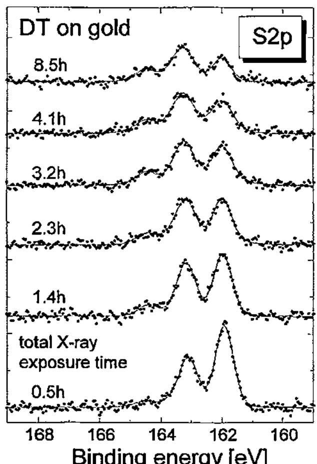


Figure 21. XPS S2p spectra of DT on gold for different X-ray exposure times. The formation of a new chemical species and the loss of sulphur can be seen directly from the spectra as a function of time. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

$S2p$ doublet structure. There is no signal from oxidized sulphur species observable, which would occur on the highest binding energies (BEs) on the left-hand side in Figure 21. But with an increasing interaction time of the X-rays with the adsorbed organic layer, the typical $2p_{3/2}/2p_{1/2}$ doublet branching ratio for XPS core-level spectra becomes increasingly distorted from this value, and an additional peak structure grows on the high-binding-energy side of the main peak structure. These changes of the spectral shape are evidence for the appearance of another sulphur species. This conclusion is supported by the peak-shape analysis of the spectra. Figure 22 shows two representative experimental spectra, together with the results of a peak-shape analysis using two doublet peaks for the fits of each sulphur spectrum. The corresponding fit components, the background lines, and the fit residuals have also been included in the figure. The fit-sum curves match the experimental data well, and the fit residuals appear to be randomly distributed, in this manner demonstrating the successful description of both experimental datasets by these two doublets. The initially dominating S2p component (Fig. 22, bottom spectrum) at a binding energy ($S2p_{3/2}$ with respect to the Fermi level) of 161.92 eV evidently represents the sulphur head group of the thiolate, which bonds to the gold substrate (hereafter called sulphide component). The second sulphur component at the higher binding energy of 163.36 eV has reached the same relative intensity as the initial one in the top spectrum of Figure 22. It may be related to radiation-induced changes of the chemical bonding of the respective sulphur atoms. We have tested this hypothesis by switching off the X-ray tube for periods up to 12 h. After having resumed the irradiation, we could get about the same intensities from the sample surfaces as before the stop of the irradiation. This test has demonstrated that the chemical transformations that are indicated by the peak-shape changes are, in fact, X-ray induced (including X-ray excited secondary processes), and not solely due to storage of the layers in the UHV environment.

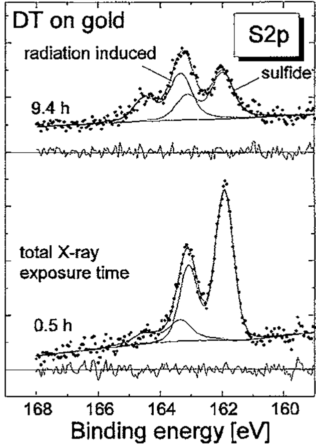


Figure 22. Peak-shape analysis of S2p doublets (representing the sulphide and the radiation-induced species) from DT on gold, including the residuals for exposure times of 0.5 and 9.4 h. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

Hence, we will call this second sulphur doublet component radiation induced. The evolution of the chemical changes in the thiolate layer while continuously irradiated with X-rays is displayed in a quantitative manner by the sulphur component intensity ratios (radiation induced/sulphide) shown in Figure 23a on a longer exposure time scale. This ratio rises

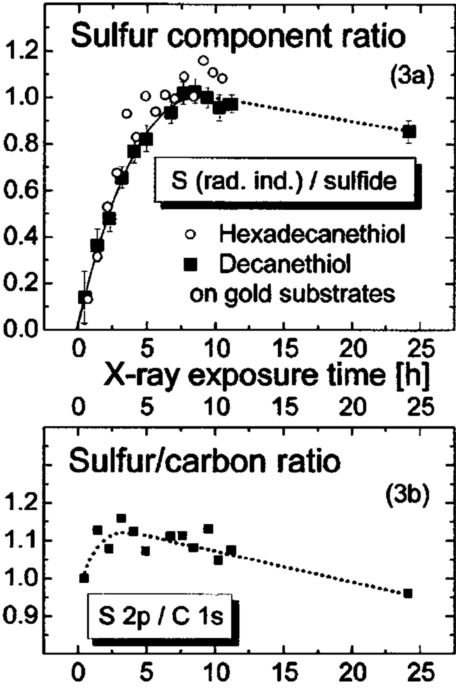


Figure 23. (a) Comparison of the irradiation time dependencies of the relative content of radiation-induced sulphur/sulphide of HDT and DT on gold. Note that extrapolating the data to $t = 0$ suggests a monolayer free of sulphur components other than sulphide from the adsorbed thiol. (b) Comparison of the sulphur/carbon intensity ratios of DT determined on gold. All absolute values have been normalized with respect to the Au4f peak. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

fast during the initial stages of irradiation. Of course, the individual spectra and their intensities represent an average over the necessary recording time of about 30 min for a single high-quality spectrum, and thus the intensity ratio cannot be given for zero irradiation. However, tracing back the evolution of the data toward lower X-ray exposure gives a component intensity identical to or at least very close to zero for irradiation time $t > 0$. This suggests the absence of the second sulphur species at the initial stage of the XPS investigation. After having reached a maximum level at about 8–10 h of irradiation time by monochromatized X-rays, the intensity ratio exhibits a weak trend to decrease very slowly, in particular at the largest duration of X-ray exposures (24 h). It is important to note that the evolution of the sulphur component intensity ratio proceeds very similarly in the case of HDT and DT adsorption layers on gold (see Fig. 23a). This leads to the conclusion that these moderate differences in chain length are of less importance in the case of these radiation-induced chemical changes. But note that the intensity ratios measured for the HDT film exceed those for the DT film by about 5% typically. Normalized intensities of both the total C1*s* and S2*p* signals, as well as of the sulphide and the radiation-induced S2*p* components, are provided in Figure 24. A significant intensity loss of both the total sulphur and the carbon signal of the thiolate film with increasing X-ray exposure can be clearly observed. The decrease of the sulphur signal intensity proceeds fast during the initial stages of the X-ray exposure, but slows down considerably at later stages. The sulphur intensity decreases much slower beyond irradiation times of about 8–10 h, when it has reached about 60% of its initial value. These pronounced intensity changes cannot be explained by a rearrangement of the organic molecules of the adsorbate layer, but suggest a loss of adsorbate from the surface. Interestingly, the C1*s* intensity decreases at about the same rate (Fig. 24) during the whole investigation. There is only a slight increase of the sulphur/carbon ratio (Fig. 23b) by about 10–15% at intermediate irradiation times, which indicates more complex changes in the layer. This observation may indicate that parts of the alkyl chains have been removed from the adsorbed molecules by radiation-induced bond breaking and desorption or that part of the adsorbed organic molecules are more flat lying on the surface than in the well-ordered SAM. Both effects would increase the sulphur/carbon intensity ratio due to reduced sulphur signal attenuation by carbon. Another minor effect to be recently reported is a slight broadening of the C1*s* spectra toward lower BE after prolonged irradiation that has been attributed to chain damage due to double-bond formation possibly accompanying C–H bond breaking [78, 79]. Nevertheless, the major effect regarding the structural layer damage is the intensity decrease of both thiolate signals (C1*s*, S2*p*), which has to be attributed to a large extent to correlated losses of both sulphur and carbon from the surface. Note that all solid curves in Figure 24 have been drawn to guide the eyes only. In fact, the lines have been obtained by fitting the data with exponential functions. The best results have been found by using two exponential functions with characteristic time constants of about 2–3 h and about 100 h, respectively. We emphasize that similar time constants characterize the intensity loss of the total signals (S2*p*, C1*s*), as well as changes of the two individual sulphur component signals (radiation induced, sulphide). This observation strongly supports the assumption that chemical damage and structural damage are somehow correlated on gold substrates. The behavior of different adsorbed thiol molecules has been investigated in the range of alkyl chain lengths from $n = 10$ to $n = 18$ carbon atoms (only even numbers). All investigated layers exhibited the above-described effects. Small deviations in the data (e.g., shown in Fig. 23a) are visible, but not strong enough to be interpreted as significant influences from the deviating chain–chain interactions due to differences in the alkyl chain lengths. We have been able to visualize the structural damage of the adsorbate layers due to the extended X-ray exposure during our studies in SEM pictures, as shown in Figure 25a for the case of an ODT layer on gold.

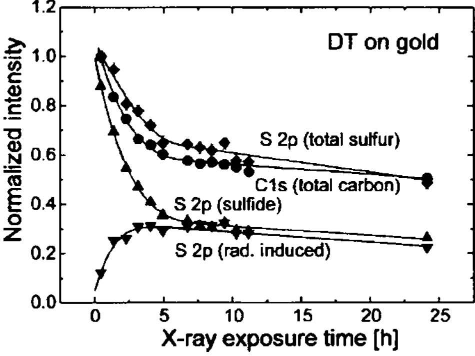
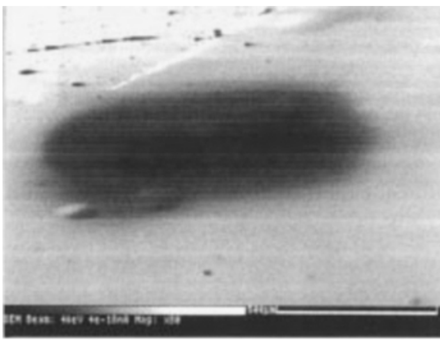


Figure 24. Normalized intensities of the carbon and the sulphur species from DT on gold. The intensities have been calculated from the average peak area of the fitted spectra. Normalizations have been made with respect to the Au4*f* peak of the substrate. This leads to a small overestimation of the loss because the substrate is not covered by the same effective layer thickness at longer exposure times. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

The scanning electron microscopy (SEM) picture of the photoninduced damage of the thiol film was recorded using a VG LEG 1000 electron gun with a primary electron energy of 4 keV and the lowest possible current.

The elliptically shaped dark area represents the spot from the monochromatized X-ray source on the sample. The contrast in the picture is due to different electron reflectivities and secondary electron yields of the irradiated and nonirradiated parts of the thiolate film. The irradiated area can be clearly distinguished from the environment, but within this area, we notice gradual differences in contrast, which may be attributed to an inhomogeneous illumination by X-rays, and thus also inhomogeneous photon density. This renders the application of exact quantitative models for the characterization of the damage evolution difficult.

It should be mentioned that radiation-induced damage proceeds much faster by using nonmonochromatized X-ray sources. This result is quite reasonable, even if the X-ray power is chosen to give equivalent count rates for the photoelectron lines because of the additional photons from the bremsstrahlung and the X-ray satellites, and the much higher background of inelastically scattered and secondary electrons. Radiation-induced damage was also observed



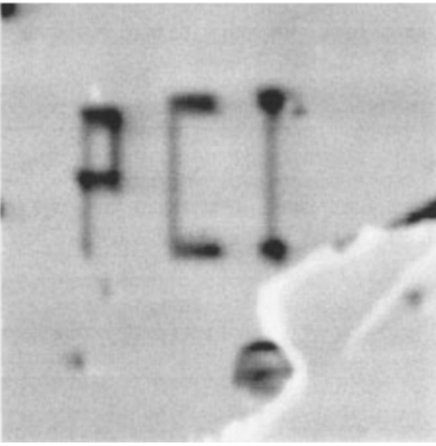


Figure 25. (top) SEM picture of an octadecanethiol monolayer on gold. The dark spot is due to the change of the SE yield by induced damage by the monochromatized X-ray source (1486.6 eV) to the adsorbate. The size of the spot is approximately 1000 μm × 250μm (primary electron energy was 4 keV). (bottom) SEM picture of a dodecanethiol monolayer on InP(110). The pattern "PCI" has been written by electron beam lithography, using a freely programmable electron gun at 10 keV. This picture demonstrates the ease of structuring an alkanethiol layer with electron beams. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

after applying lower energy synchrotron radiation in the range of $h\nu = 50$–320 eV, as was used for X-ray absorption near-edge structure and soft X-ray photoelectron spectroscopy studies. Moreover, damage of SAMs was also found using electrons in the range of 10–14 eV, supplied by a flood gun, and we have further tested that electron beams <500 eV can damage the layer in the same way. These results ensure that low-energy and secondary electrons contribute to the layer damage. We have applied 4–10 keV electron beams of an Auger electron gun for laterally structuring the SAMs.

For example, the pattern "PCI" has been generated in a dodecanethiol monolayer (Fig. 25b). In fact, electron-induced damage has been reported before [80–82], and it has also been used for patterning purposes. To shed more light on the possible importance of substrate–layer interaction regarding the radiation-induced effects, we have also investigated adsorbed thiol monolayers on InP (110). This surface was chosen for comparison to the metallic gold surface because of the quite different surface structure and bonding of this covalent semiconductor with zinc-blende lattice, and finally because of the available experience in the interpretation of sulphur core-level spectra from several adsorption studies of inorganic sulphur species on InP [35–37, 66]. We have studied the composition and structure of such thiolate layers using XPS and XANES. Very pronounced angular dependencies were observed in the σ^*(C–C) and σ^* (C–H)-related structures in the near-edge region of the X-ray absorption spectra on alkanethiolate layers on InP(110) by varying both the polar and azimuth orientation of the sample with respect to the polarization direction of the synchrotron radiation. The data analysis has clearly shown that self-assembly also takes place on this surface, and it results in an excellently ordered organic monolayer. In contrast to gold substrates, where domains with different azimuth orientations of the alkyl chains exist on the surface, only the tilt orientation of chains parallel to the dangling bond directions of the surface atoms were observed on the cleavage face of indium phosphide. However, for both the XANES data and their detailed discussion, we refer to Section 3.

The adsorbed molecules form a dense monolayer on InP(110), as can be concluded likewise from the observed passivation of the semiconductor surface from oxidation by contact to air. This passivating effect is demonstrated by the complete absence of oxygen-related signals in the XPS core-level spectra (Fig. 26). A survey spectrum only shows photoelectron peaks of both the semiconductor substrate and the organic adsorbate on the background of inelastically scattered and secondary electrons [32]. For the purpose of this chapter, we will focus on the S2p spectra and the corresponding intensities as discussed for the gold substrates above. S2p spectra recorded from an HDT layer on InP(110) are provided in Figure 26. Cursory inspection by eye gives only evidence for a single doublet due to the sulphide component, which is present in all of the individual S2p spectra. But careful peak-shape analysis allows identification of a second sulphur component of low intensity at a higher binding energy. This small component also grows with increasing X-ray exposure, similar to the case of thiolate layers on gold. There have been several studies

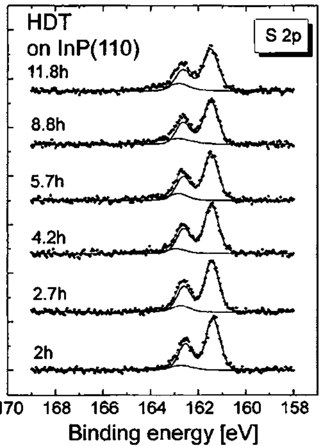


Figure 26. XPS S2p spectra of HDT on InP (110) recorded at different X-ray exposure times. Note the differences compared to alkanethiols on gold. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

of inorganic sulphur species on both InP(001) and InP(110) surfaces [33–35, 37]. In–S bonds have been observed to dominate mostly, with sulphur components between 161 and 162 eV with respect to the Fermi level, depending on the detailed adsorption geometry and Fermi-level pinning. Polysulfidic sulphur was observed with binding energies near 163 eV. Dissociative adsorption has been verified in the case of H2S on InP (110), with P–H and In–SH surface bonds [33, 34]. The evolutions of both the changes in chemistry and composition of the adsorbed monolayer due to the radiation are displayed in a quantitative manner by normalized intensities and intensity ratios, as shown in Figure 27. Looking at the normalized intensities first, we may observe that both sulphur and carbon intensities decrease with increasing X-ray exposure. But this decrease is limited to less than 15% during the period of investigation (12 h). However, the evolution of sulphur and carbon intensities again proceeds about parallel. Furthermore, a small shift toward lower binding energies and a broadening of the C1*s* peak related to the alkyl chain, as was already mentioned for gold substrates, has also been detected on InP (110) in the very last stage of the irradiation process. Thus, there is again sufficient evidence for structural damage of the adsorbed films on indium phosphide. However, the loss of sulphur-containing molecules is strongly reduced under equivalent excitation conditions compared to the gold substrates discussed above.

Second, the sulphur component intensity ratio (radiation induced/sulphide) also increases with continuing irradiation with X-rays, but it increases much slower than that on the gold surface. At the final stages of the present investigations, this ratio does still increase, and there is no evidence for a reversal of this trend yet. In the case of the dataset presented in Figure 27, we have included periods of excitation by nonmonochromatized X-rays to enhance the radiation-induced effects. We have further extended our comparison of alkanethiol layers on gold films and on indium phosphide to the effects of electron beam patterning. It is very interesting to note that patterned structures were consistently smaller on InP substrates by 30–80% than on gold films under identical excitation conditions in the case of applied electron probes of a few 100 nm lateral size and 4–10 keV primary energy as used here. We explain this observation by a larger number of secondaries in the case of gold, and their larger lateral spread due to the higher electron backscattering in gold. We have been able to confirm these effects of higher backscattering of electrons in gold and the higher number of low-energy electrons at the surface compared to indium phosphide quantitatively in our Monte Carlo simulations, which were performed to support separate studies of patterning by electron beams. The Monte Carlo simulations of the electron trajectories were carried out in consideration of relativistic partial-wave calculations of elastic cross sections by solving the Pauli–Dirac equation, combined with Bethe's continuous slowing-down approximation and the Rao–Sahib–Wittrey approach at low energies.

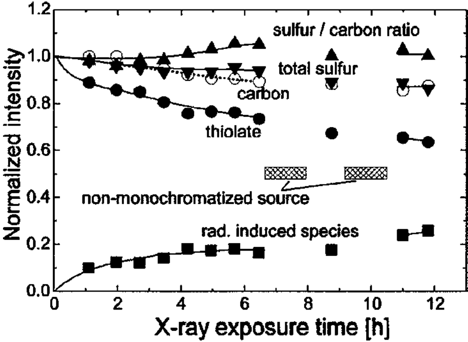


Figure 27. Summary of the carbon and sulphur intensities of HDT on InP (110). All absolute intensities have been normalized with respect to the In3*d* peak of the substrate. The diagonal hatched areas mark the points of time when the nonmonochromatized X-ray source has been used. Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

The higher number of low-energy electrons and the more pronounced backscattering accumulate and generate a high number of low-energy electrons in the surface region, which may introduce damage to the organic films. Thus, both the experimental observation of the substrate-dependent width of patterns in the organic films and the quantitative simulation of electron distribution in the samples provide further supporting evidence for the major importance of secondary electron damage of organic films. It is thus reasonable to suggest that the indirect damage by electrons seems to be very effective compared to direct damage by the X-rays, and thus is likely to dominate the damage, even in the SAMs investigated here. Moreover, care has to be taken in the application of structural probes that use ionizing radiation because of the possibility of damaging the layers by secondaries.

Summarizing this part, we may notice that evidence for both chemical and structural damage of the thiolate layers by extended exposure to X-rays was also observed on InP (110). The striking differences in the quantitative extent of these effects compared to gold substrates provide evidence for the importance of the different substrate–layer interactions and the different substrate properties.

The third part of the data will deal with the analysis of the chemical bonding using core-level spectroscopies. Besides the above-discussed thiolate monolayers on gold and indium phosphide, we have also investigated deposited films of pure dithiol (1,8–octanedithiol) and pure dialkyl or thioether (dioctadecyl–3,3′–thiodipropionate) and HDT as references.

Similar discussions regarding the nature of the adsorbed alkanethiols have been published by Jung et al. on gold [72], and by the author on gold and III–V semiconductors [32]. In the following, we will outline the experiments, main results, and their discussion.

To prepare reference samples for the chemical analysis, the undiluted disulfide DS (poly-octanedithiol, ...–C_8–S–S–C_8–S–S–C_8–..., Aldrich, 97+%; the disulfidic nature was additionally confirmed by Raman spectroscopy) and dialkyl DA (dioctadecyl 3,3′–thiodipropionate, $S[CH_2CH_2CO_2(CH_2)_{17}CH_3]_2$, Aldrich, 99%), as well as HDT were directly applied to a gold surface, and frozen down to liquid nitrogen temperature (LNT) under a protective gas atmosphere to minimize water adsorption. The samples were then inserted into the spectrometer immediately, while they were kept at LNT to prevent degassing of the molecules in UHV. This procedure leads to films which we will call later "thick films." Their thickness is much larger than 100 nm, which is proved

optically and additionally by the fact that the substrate is no longer detectable by XPS measurements. The so-called "thin films" were prepared as follows. The frozen "thick films" were warmed in UHV until they reached room temperature. Thus, most of the film is degassed, and the substrate is detectable by XPS. In the case of long chain reference chemicals, the samples were removed from the UHV environment, and then additionally rinsed with isooctane. After that, the samples were inserted into the spectrometer again. Because of the substrate/adsorbate signal ratio, we estimate the "thin films" to have a thickness approximately up to 5–6 nm. Note that all molecules above the first ML are not supposed to be ordered, but are randomly distributed.

The reference chemicals have been chosen with regard to two aspects: first, because of their sulphur–carbon groups (R–S–S–R or R–S–R, respectively), and second, because of the existence of long alkyl chains which have been used to calibrate the spectra. The X-ray photoelectron spectra were recorded with a VG ESCALAB 220i-XL instrument irradiating the samples with monochromatized Al Kα radiation (22 mA, 10 kV). Spectra were recorded in normal emission, and with an angular averaging mode using an additional magnetic XL lens, if not stated otherwise.

The energy resolution was 0.45 eV (FWHM Ag3d). Binding energies (BE) have been referred to the Fermi energy. The energy scale of the spectrometer was calibrated by adjusting to EB (Au4$f_{7/2}$) 84.00 eV. The spot of the X-ray photons from the monochromatized X-ray source has a size of 1000 μm × 250 μm on the sample, with an approximate elliptic shape. Bremsstrahlung of a nonmonochromatized MgKα source (30 mA, 15 kV) has been used exceptionally to excite the Auger S $KL_{2,3}L_{2,3}$ and S2p spectra, which were used for the chemical state plot. All binding energies in the chemical state plot have been referred to the C1s energy of the alkyl chains, set to 284.60 eV. This allows correct relative adjustment of binding energies from species in highly insulating thick films which exhibit a charging shift of the energy positions.

The deposition of dithiol on gold, in the way described above, leads to the formation of thick films. These are expected to contain sulphur–sulphur bonds such as those in disulphides (DS) mostly, for example, a –C–S–S–C– group, which results in networks of polysulphides. Otherwise, the remarkable stability of these films in an ambient environment can hardly be explained. This assumption has been supported by the detection of specific vibrational modes in Raman spectra, which reveal typical fingerprints of disulphidic species. On the other hand, the dialkyl (DA) films contain –C–S–C– units. In addition to the thick films with, thicknesses in the range of several micrometers, thinner films with thicknesses in the range of only a few nanometers also were prepared by extensive treatment with solvents or partial desorption in UHV. The thickness of the latter films can be crudely estimated from the C1s to Au4f intensity ratios, which were observed roughly in the range of 0.3–3 in these cases. Charging effects were especially observed in the cases of the thick films because of the insulating nature of the adsorbed molecules. The charging-induced shifts in the very thick layers lead to a degeneration of the resolution due to the nonuniformities of the charging. Neither the use of a flood gun nor the change of the X-ray source and the lens modi could completely remove the inhomogeneous charging. However, this does not severely reduce the grade of information obtained. For the comparison of sulphur binding energies from different species and film thicknesses, the energy scales have been calibrated to set the C1s signal of the alkyl carbon of all samples to EB = 284.60 eV to correct any kind of charging-induced shifts. A comparison of sulphur core-level spectra of HDT monolayers on gold and InP (110) and a thick film containing disulphide bonds (DS) is shown in Figure 28. The vertical line indicates the approximate position of the disulphidic S2$p_{3/2}$ components. The faint peak structure at the very high binding energy in the DS spectrum represents some oxidized sulphur. Quick inspection of the spectra by eye reveals that the binding energies of the radiation-induced signals from the monolayer fall quite close to the disulphide signal from the DS. This observation seems to support the hypothesis that the sulphur signal from the monolayers at the higher binding energy may represent disulphidic species, as has been suggested before [78, 79].

Furthermore, we have combined XPS spectra with X-ray-excited Auger electron spectroscopy (XAES) in order to derive Auger parameters, and to apply two-dimensional chemical state plots for chemical analysis of the sulphur bonding. In the "initial state–final state framework," the XPS core-level binding energies and the XAES Auger (kinetic) energies (AE) can be described essentially by the potential energy contribution on the site of the emitting atom due to the electron density distribution in the chemical environment and a relaxation or final state contribution, which characterizes the core-hole screening effects due to the transfer of screening charge or polarization of the environment following the core-hole generation. The Auger parameter α' according to Wagner corresponds to the sum of the BE and the AE [35–37]. The chemical shift $\Delta\alpha'$ of the Auger parameter reflects the changes of the relaxation or core-hole screening due to changes of the chemical bonding situation. Thus, the basic idea of application of Auger parameters and chemical state plots of binding and Auger energies is to use the additional information due to different

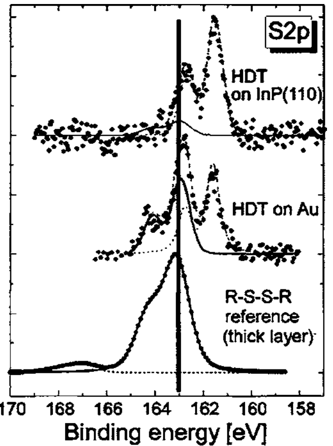


Figure 28. Comparison between the XPS spectra of HDT on InP (110) and gold and a thick layer of a reference disulphide (see text). Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

sensitivities to core-hole screening effects. In some cases, this may permit us to distinguish even between species, which have nearly identical binding energies. Figure 29 shows a chemical state plot, which presents the results from repeated measurements of three different adsorbate systems (HDT, DS, DA) and three different layer thicknesses (monolayer, thin film, thick film). The vertical axis represents the kinetic energy of the S $KL_{2,3}L_{2,3}(^1D)$ Auger peak. Binding energies are shown along the horizontal axis. The diagonal lines with slope $= -1.0$ represent identical values of the Auger parameter α', and thus the same relaxation or core-hole screening energy. There is a striking separation into two different groups of data sets in Figure 29. In the upper right corner, that is at the lowest BE and highest KE and thus the highest Auger parameter values, we observe the data points obtained from the sulphide signal of the HDT monolayer and a signal from the thin HDT films that we attribute to those HDT molecules of the thin film at the interface to gold. All other data points from DA, DS, and including the HDT thick film data, as well as the remaining data points from HDT thin films, may be found in the lower left corner of the chemical state plot. The significantly smaller KE and Auger parameter values of the latter data points suggest a much less effective core-hole screening compared to the sulphide in the HDT monolayer. The lowest KE and Auger parameter values were measured for the thick films of DS, DA, and HDT. This is reasonable in terms of the insulating nature of these films and the much reduced ability to screen the core hole compared, for example, to a metallic environment. Slightly higher Auger parameters are observed for thin films of DS, DA, and HDT. All datasets corresponding to DA films are shifted toward lower BE with respect to comparable film signals of other sulphur species. A shift to lower BE for dialkyl samples by 0.4 eV compared to similar disulphidic samples has already been reported in the early data compilations by Siegbahn [83] and Nefedov [84]. The

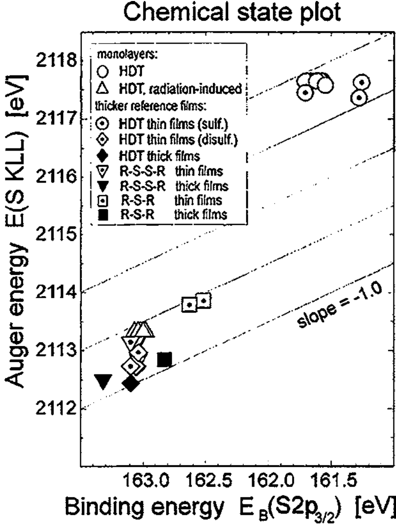


Figure 29. Chemical state plot including binding and Auger energies of three different adsorbates with different layer thicknesses in order to identify the radiation-induced species (see text). Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

relative shifts of about 0.5 eV to disulphides in our investigations are sufficiently large to distinguish DA species from the others. The thin-film HDT results are very similar to the thin-film DS data. We emphasize that the data, which represent the radiation-induced sulphur component of the irradiated SAMs, are very close to the thin DS film data. This observation is another fingerprint pointing to the disulphidic nature of the radiation-induced species in the monolayers. But most importantly, we may conclude from the huge difference in Auger parameters between sulphide and radiation-induced species that the sulphur atoms responsible for the latter signal are not bonded to the gold surface atoms anymore. Otherwise, the metallic substrate should provide a significant contribution to the core-hole screening energy, which is evidently not the case.

In the following, we will discuss the experimental data. The adsorbed monolayers of alkanethiols on metals such as gold are known to exhibit long-term stability. However, it has been shown that extended irradiation by X-rays modifies alkanethiol monolayers on gold films in two ways. First, the chemical bonding of the sulphur atoms at the interface to gold is changed by partial breaking of the sulphur–substrate bonds. Second, the layer is structurally damaged, and loses adsorbate molecules. These results are in good agreement with very recent data by Jäger et al. [79] and Wirde et al. [78], and the appearance of radiation-induced damages of alkanethiolate SAMs on gold films now seems well established.

Besides, we will now focus on the comparison of gold film and InP substrates for the SAMs in order to derive conclusions on the effects contributing to the damage. The experimental observations at hand include the X-ray or radiation-induced changes of the spectra as indicated by the appearance of new spectral components, core-level line broadening and line shifts, total intensity changes, and intensity ratio modifications. These effects have been observed on both investigated substrates, as has been demonstrated in detail in the previous sections. However, there are significant differences regarding the quantitative extent in these effects, and in particular, differences in their evolution with X-ray irradiation time. Most striking is the comparably rapid chemical transformation of the sulphur and the significant loss of material from the SAMs on the gold films in the initial stages of the irradiation (time constant about 2–3 h). This rapid change in the initial stage is not observed on InP. Differences in substrate surface structure and related interface and organic film quality, substrate–layer bonding strength, and substrate-related excitations may contribute to this strikingly different behavior, and call for discussion. Bonding energies of 13.9 and 14 kcal/mol were derived for Au–methanethiol [30] and In–S [85], respectively, suggesting rather similar strengths of sulfur–substrate bonds for the substrates considered here. Note that In–S bonds are, in fact, preferentially formed after the adsorption of sulphur species on indium phosphide [35–37]. In a recent time-of-flight secondary ion mass spectrometry (TOF–SIMS) study of electron-stimulated damage of Langmuir–Blodgett (LB) monolayers and self-assembling monolayers, Rading et al. [86] have concluded that the strength of substrate–film bonding is of less importance for the desorption cross sections by ionizing radiation. Because of these results, we may

reasonably conclude that the quantitative differences of the two substrates considered here should not mainly be related to substrate–film bonding.

The number of surface defects is much larger on evaporated gold films than on the cleavage face of indium phosphide. This was confirmed by STM, as has been mentioned above. Further, due to the existence of several domain types with different azimuthal orientations of the alkyl chains on gold, we may also expect more domains, and thus more adsorbed molecules at domain boundaries, to be present on this substrate. As was mentioned in the brief review of the XANES results, due to the unique azimuthal orientation of the thiolates on InP (110), considerably fewer domains should exist on this surface. But we cannot estimate the relative extent of these effects based on our results. Evidence for reduced order in SAMs on deposited gold films compared to gold single-crystal surfaces has recently been provided in electron energy loss spectroscopy (EELS) studies by Duwez et al. [87]. It has been shown that the stability of SAMs of adsorbed thiols is due to both the sulphur–substrate bonding and the van der Waals interchain interactions to a similar extent for chain lengths of about 10–20 C [28, 29]. Adsorbed molecules near defect sites experience much reduced stabilizing intermolecular interactions compared to those within well-ordered domains on a well-ordered substrate. From defective adsorbate sites, a desorption of rather complete adsorbate molecules after substrate bond breaking seems likely, which may explain that the losses of carbon and sulphur from the surface proceed in a rather correlated fashion. Predominant desorption from defective sites on the gold films may also contribute significantly to the observation of the different time scales that characterize the intensity evolution of the core levels, and thus the loss of organic material from the SAMs on gold in the initial stage (see Fig. 24). The significantly lower rates of material loss, which seem to govern the later stages of the radiation damage on gold films, and evidently the whole range of the irradiation of the SAMs on InP, may then be related to material loss from well-ordered areas. Radiation-induced cross linking of chains is another effect which may stabilize the organic layers against radiation-induced desorption. However, both the observed much faster and larger conversion of sulphur into the radiation-induced sulphur species on gold and the significantly larger increase of the S/C ratio as observed in the initial stage on gold films cannot be sufficiently explained by defect-site-related damage only. The increase of the S/C ratio has indicated partial destruction of the alkyl chains by bond scission, followed by desorption of separated hydrocarbons. Generally, bond breaking is due to vibronic or electronic excitations. In the present case, it is likely to be mostly due to rather localized electronic excitations. In particular, bond breaking in the alkyl chain is an intramolecular process, which is not expected to depend significantly on the substrate. Nevertheless, given the same types of adsorbed alkanethioles and the same photon flux, the larger increase of S/C on gold films indicates an influence of the substrate. We have demonstrated evidence for the importance of the low-energy electrons for the damage of the organic films. Völkel et al. [88] and our group now suggest that the observed difference in the development of S/C ratios may be due to a different number of secondaries originating from the X-ray excitation of the substrate, which may cause severe damage to the organic film. A higher number of secondaries from gold would also be in line with the observed faster chemical transformation of sulphur in the adsorbed alkanethiol layers on gold films, and is expected to contribute to the larger loss of organic material on this type of substrate too. Another important issue is that of the chemical nature and the adsorption geometry of the sulphur atoms of the adsorbed thiols. Our impact on that question remains limited because the XPS method may provide information on the chemical bonding, but it is only indirectly sensitive to structural sites via binding energy shifts. Nevertheless, by focusing on the radiation-induced changes, we have been able to present clear evidence for the disulphidic nature of the radiation-induced sulphur component, as previously suggested by Wirde et al. and Nuzzo et al. [78, 89]. This similarity of the radiation-induced peak to disulphidic or polysulphidic species may also be concluded by comparison to related adsorption studies of thiols and MBT on gold [31], and of several inorganic sulfur molecules on InP too [35–37]. Besides, from the analysis of the Auger data presented here, which included Auger parameters and the chemical state plot (Fig. 29), we have predicted that the radiation-induced sulphur is not strongly bonded to the gold substrate. From our experiments, the interpretation of the radiation-induced sulphur species as a dialkyl seems very unlikely. This problem is, however, linked to the ongoing discussion regarding the chemical nature of the sulphur species at the undamaged interface to gold. In Figure 30, we have sketched a few possible adsorption models, which emphasize the type of adsorbed species, but neglect any details of the local bonding geometries of sulphur on a gold surface. In contrast to the more generally assumed thiolate (RS–Au bonding, Fig. 30a) nature of the adsorbed thiols, there have been papers proposing the disulphidic nature of the adsorbed thiol layers (R–S–S–R bonding to Au via a disulphide group, Fig. 30b,c). These conclusions were derived from analysis of grazing incidence X-ray diffraction (GIXD [90]), X-ray standing wave (XSW [91]), and EELS [87] data

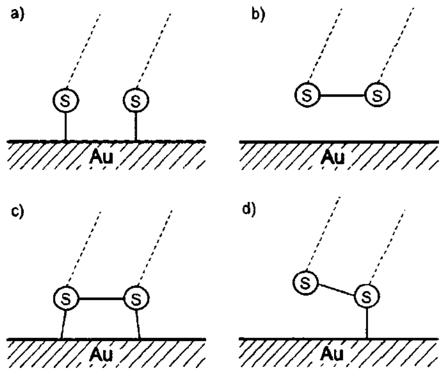


Figure 30. Schematic drawing of different possible bonding configurations of the thiol sulphur head groups: (a) thiolate bonding to the Au surface, (b) disulphide [stabilized above the Au surface by van der Waals forces (not shown)], (c) disulphide with additional dative bondings to the Au surface, (d) disulphide with one single dative bonding to the substrate (note the height offset between the two sulphur atoms). Reprinted with permission from [32], D. Zerulla and T. Chassé, *Langmuir* 15, 5285 (1999). © 1999, American Chemical Society.

obtained on Au (111) single crystals. Fenter et al. suggested that two inequivalent, but spatially close sulphur adsorption sites exist on Au (111), which led them to the conclusion of disulphides being bound on the gold. It has been proposed that different adsorption sites on transition metal surfaces with different coordination numbers will reflect insignificant BE shifts in core-level photoelectron spectra of the sulphur [92]; similarly, a suggestion has been made that inequivalent sites of the disulphide–sulphur atoms on Au (111) should also manifest in such chemical shifts [90]. However, within the limits of our experimental resolution, we cannot find evidence for another sulphur species different from the initial sulphur species (called sulphide), except the radiation-induced one on our gold films after thiol-solution preparation. Note that disulphides have been detected together with thiolate sulphur in photoemission studies of adsorbed alkanethiols on transition metal surfaces, but these observations were mostly limited to low temperatures and low dosages of alkanethiols from the gas phase [93]. To our knowledge, no photoemission study has reported adsorbed disulphides to appear at the binding energies corresponding to our sulphide species. Thus, there is no direct evidence from photoemission which would support disulphide models, and a characterization of the sulphide at the undamaged surface as a thiolate as shown in Figure 30a is most reasonable. Consequently, the disulphide component appears as a result of the radiation-induced effects in the organic film only. Such a prediction for disulphides contradicts neither the long-term stability of the disulphide species observed in our study nor the well-known much higher adsorption probability of disulphides compared to thiols because the strong intermolecular interactions between the neighboring hydrocarbon chains, possibly accompanied by some bending of chains necessary to allow linking of the sulphur atoms in the covalent bonding, may stabilize the disulphide in the film while suppressing retransformation to thiolates. Upon the continuing influence of radiation, even the S–S bonds may be destroyed, and from the radicals, thiolates can be formed again. In this manner, we may explain the fact that the disulphide contribution to the total sulphur signal does not grow further beyond a maximum value (close to 50% in our investigations), and it only changes to a very minor amount at higher irradiation times.

On the other hand, Fenter et al. have ruled out a significant influence of radiation damage on their GIXD and XSW results, which favor disulphides on the gold surface. They estimated a much smaller damage cross section (about 1/100) for the higher energy X-ray photons used in their study compared to photoemission studies using Al K X-rays, based on the energy dependence of the photoionization cross sections of core levels. We have demonstrated evidence for the importance of electrons generated in the substrate, and thus the X-ray energy dependence of the total quantum yields from the substrates rather than a single core-level photoionization cross section, for example, C1*s* of the film, should much better represent the variation of the damage with energy. According to the photon-energy-dependent electron yields for gold determined by Henke et al. [94], the yield at 8–10 keV (GIXD) is only a factor 2–3 smaller than at 1.5 keV (XPS), while at about 3 keV (XSW), it is even larger by a factor 2 than at the energy of the Al K X-rays. We cannot reasonably judge the second argument raised by Fenter regarding the photon flux because the information on photon flux, and in particular on the irradiated area, is incomplete, and the experimental geometries to be compared are quite different. Nevertheless, we feel that the quantitative estimations by Fenter need some reevaluation.

Significant damage of alkanethiol SAMs by X-rays (10 keV) under vacuum conditions was observed on a time scale of 14 days by GIXD [90], which is actually not too far off the time scale of 4 days determined for the long-term damage of the SAMs of gold (and similarly on InP) in our XPS investigations, keeping in mind the differences in damage cross sections (see above) and the different sensitivity to structural damage of the methods. This discussion has demonstrated that, presently, the effects of radiation damage on the sulfur at the interface may not yet be ruled out completely. Further, Jäger et al. [79] also have posed additional questions regarding the possible effects of adsorption-induced gold surface atom redistribution at the interface to the predicted presence of disulphides. Such thiol-induced reconstructions have, in fact, been reported for Cu (111) [95]. On the contrary, our XPS data may not completely rule out all kind of disulphides on gold. A possible adsorbate of this type has been sketched in Figure 30c. Sulphur is known to form a large variety of bonding configurations with metals, for example, in complexes, including disulphide groups and dative bonds [96]. Upon the damaging influence of radiation, such an adsorbed species as assumed in Figure 30c might transform into disulphides with a complete breaking of bonds to the gold substrate as in Figure 30b or a partial breaking of the sulphur–gold bonding with one sulphur atom remaining closely bound to the gold surface, as assumed in Figure 30d. In the latter case, photoemission spectra would represent the two sulphur components at the energies corresponding to disulphides and to the initial sulphide component, as is, in fact, observed after a few hours of irradiation. This latter bonding situation (Fig. 30d) would also fit the stability of the damaged organic layer and the almost similar intensities of the radiation-induced and sulphide component observed after some hours. Again, there is no direct evidence for such an adsorption from photoemission, and these explanations remain speculative and have to be proven. To allow an appropriate comparison of the existing experimental results and the corresponding adsorption models, both substrates (single crystals versus films) and the preparation methods (vacuum deposition versus deposition from solution) have to be chosen consistently, which has not yet been done. There is evidently a need for methods which offer stronger links between the purely structural and the chemical bonding information than is provided by application of GIXD/XSW or XPS/XAES alone. Investigations on gold complex compounds have demonstrated that XANES might be very promising in this context [97].

Summarizing, we have investigated the damage of alkanethiol SAMs on deposited gold films and InP(110) by X-rays using XPS, XAES, and SEM. Both chemical and structural damage of the adsorbed alkanthioles were observed on either of these substrates. But much stronger and faster development of the damage in the organic monolayers was observed on gold films compared to InP. This has been interpreted in terms of a higher defect density and

larger effects of X-ray-generated electrons from the substrate on the organic layers in the case of gold. The importance of low-energy electrons for the damage and patterning of the layers has been emphasized. A radiation-induced new sulphur component has been clearly attributed to have a disulphide character using both binding and Auger energies. During the discussion, we have found more arguments supporting a thiolate-type SAM compared to disulphide models on gold.

In the preceding, we have focused on the molecular mechanisms of ionizing radiation. In order to use X-rays or electrons for structuring, it is not necessary to understand the underlying reactions. Electrons can be used directly to remove small parts of self-assembled monolayers, as shown before in the pattern "PCI." The drawback of this method is the very time-consuming sequential writing of the pattern which makes this method very highly prized and not suitable for mass production. Using X-rays instead requires a partial opaque mask. This mask needs to be transparent to X-rays, where parts of the SAM need to be removed and vice versa. The production of such a mask needs an additional type of lithography, for example, electron beams. The second drawback of both methods is that the irradiated parts are not completely free from covering molecules. This can be seen from our detailed spectroscopic investigation at the beginning of this paragraph. These remaining molecules can be removed by heating the sample in UHV [69] or electrochemical methods ([98] and Section 3.2).

Further related methods use UV light [99–102], near-UV light [103], visible light [104], and ion beams [105] for the photolithography of alkanethiol monolayers.

4.2. Scanning Probes

In the preceding section, we already demonstrated two different techniques for lateral structuring a molecule-covered surface. One of the most obvious ideas to write small structures into a self-assembled monolayer is to use scanning probe techniques like STM, AFM, or SECM.

An STM allows us to write structures by applying voltage pulses to the conducting tip while scanning. If the voltage is chosen high enough, small parts of the self-assembled monolayer are removed [22, 56–61, 88]. The voltage needs to surpass a certain threshold, which is typically in the range of 5–10 V, depending on the substrate–adsorbate combination. In order to demonstrate these technique, we have applied a voltage of 20 V at high currents (10 nA), and therefore near distances, while scanning a dodecanethiol layer on InP(110). The result is depicted in Figure 31. A 15 nm wide straight groove is shown, which is confined by sharp edges. The advantage of the technique is its very high resolution; its disadvantage is the low speed of writing, which is orders of magnitude slower than even *e*-beam lithography. An AFM can be used to literally scratch grooves into a monolayer [61]. Its advantage is that it allows structuring and scanning of topography of even thick nonconducting organic coatings due to the characteristics of its measurement principle. It should be mentioned that, instead of scratching the structures into the films, it could be used parallel to the STM technique if the AFM tip is coated with a metal and is correspondingly connected to a voltage source.

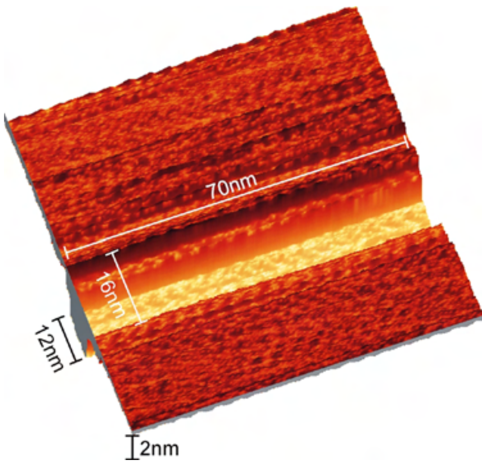


Figure 31. STM-nanostructured dodecanethiol monolayer on InP (110). $U = 20$ V; $I = 10$ nA.

The last microprobe technique which will be mentioned here is the SECM. Contrary to the STM technique, an ultramicroelectrode, a typically on its end flat-polished very thin Au or Pt wire which is coated by a nonconducting glass capillary, poses as a probe. The electrochemical technique is sensitive to redox reactions, and can also be used to deposit organics [98].

4.3. Microcontact Printing and Related Techniques

Microcontact printing (μPC) and related techniques use an old idea which has its roots in the late middle ages. Around the year 1444, Gutenberg invented the idea of using stamps for the purpose of typography. More than 500 years later, Kumar and Whitesides rediscovered this technique (with some changes) in order to transfer microstructured patterns of monolayers on surfaces [106].

The principle of microcontact printing is surprisingly easy. With the help of an elastic stamp of PDMS which is wetted by a self-assembling species, the pattern is transferred to the surface by contact. The technique allows different self-assembling species to be transferred by simply consecutively repeating the procedure with different stamps. Kumar and Whitesides where the first to invent this technique, using the system alkanethiolates on gold [106].

The greatest advantage of this method is its high speed, transferring one whole structure in one single fast stamping.

The time which is needed to obtain a highly ordered SAM from a 10 mM concentrated dodecanethiol in ethanol solution was investigated by Biebuyck et al., and was found to be as short as 0.3 s on Au (111) [21]. The typical time for a 2 mM concentrated HDT solution is between 10–20 s [107–109].

This ability makes the technique especially interesting for industrial applications. Microcontact printing is not the only method of "stamping" the pattern onto a substrate surface. In the following, we will list some further techniques with brief explanations in order to shed some light on the zoo of different techniques.

- Replica molding (REM): This technique deviates from conventional techniques by using a flexible (PDMS) mold. Its elasticity and small surface tension allow for easy removal of the mold after the process is completed The mold itself is produced by high resolving scanning electron lithography [110, 111].
- Microtransfer molding (μTM): The structured surface of a PDMS mold is coated with an appropriate polymer. Afterwards, the excess, which is everything above the mold surface, is removed. The advantage of this method is that nonplane surfaces can be coated, and the consecutive construction of complex three-dimensional structures is possible [112].
- Micromolding in capillaries (MIMIC): Here, the mold comes into contact with the surface first. The polymer (low viscosity) is distributed through the empty channels of the mold via capillary forces [113]. This technique is also very useful for nonplane surfaces, but it cannot produce isolated structures.
- Solvent-assisted micromolding (SAMIM): The last technique briefly described here is SAMIM, which can be seen as a combination of REM and imprinting. A PDMS mold is wetted with a dissolved polymer. After the mold is pressed onto the surface, the solvent dries out, the mold is removed, and the remaining polymer forms the desired structure [114].

This was only a very brief excursion on the topic of stamping. The author has ignored several variations of stamping techniques, as well as other techniques like laser ablation or inkjet printing because the discussion on the topic of lithography is a review of its own. Instead, we refer to the review on soft lithography of Xia and Whitesides [115] and citations therein.

4.4. Consecutive Chemical Reactions

In this last section on structuring, we will investigate the possibilities of creating mixed thiol films, consisting of two structurally and chemically very different thiol species. We will use the aliphatic hexadecanethiol and the aromatic 2–mercaptobenzothiazol (MBT). The competing interaction of both species after successive adsorption on gold is investigated using XPS.

The ability to produce mixed thiol films on surfaces, consisting of molecules with different electronic and chemical properties and combined with structuring in the nanometer scale, however, may lead to additional possibilities of application. Therefore, it is of interest to investigate the behavior of layers of aromatic thiols in comparison to layers formed by alkanethiols. MBT, known to form hydrophobic complexes with many metals, and therefore used as a corrosion inhibitor or a collector for mineral flotation [116, 117], was chosen for the studies presented here. Beside its aromatic character, MBT contains a nitrogen atom, which acts as an additional ligator atom in a chelate-type bonding at the surface, and can be used as tracer atom for identification of the adsorbed species. In connection to the experiments described above, the interaction between alkanethiols and hydrocarbon contamination, produced by sample transfer in air, was investigated. Because of the similarity of the behavior, it seems to be meaningful to discuss them in this context.

The preparation of the samples was described in detail in Section 2. The hexadecanethiol layer (Fluka, 95%) was prepared in a 10^{-3} M solution in isooctane (Merck, Supra-Solv, for organic trace analysis). The adsorption of MBT was carried out in a 10^{-4} M aqueous solution. After an immersion time of 30 min, the samples were rinsed with fresh solvent, dried under a nitrogen stream, and inserted into the spectrometer immediately. The X-ray photoelectron spectra were recorded with a VG ESCALAB 220i XL instrument irradiating the samples with monochromatized AlKa radiation (220 W, 22 mA, 10 kV). Spectra were recorded in normal emission and with angular averaging mode (XL lens), if not stated otherwise. The energy resolution was 0.45 eV. The spectra were fitted by a least squares procedure with a model of the peak shape (Gaussian–Lorentzian convoluted functions) after background subtraction from experimental data using a Shirley-type function. Binding energies are referred to Au4f (84.0 eV).

If a freshly produced gold film is treated with 10^{-4} M MBT solution, an S2p spectrum is recorded, as shown in Figure 32a. Two clearly separated components are found, representing the two differently bound sulphur atoms of the MBT molecule (Fig. 32a). The adsorption of MBT leads to deprotonation of most molecules, confirmed by the main component in the N1s spectrum with the binding energy of 399.0 eV, in comparison to the component of the not deprotonated molecules at 400.6 eV (Fig. 33a). Therefore, the low-energy component in the S2p spectrum (162.3 eV) can be attributed to the exocyclic sulphur of deprotonated and chemisorbed MBT. The second doublet at 163.9 eV is caused by the endocyclic sulphur. A partial oxidation of MBT to products containing sulphur–sulphur bonds like 2,2′–dithiobis(benzothiazole) is responsible for the deviation of the expected intensity ratio of 1:1 [31, 117, 118]. Both types of sulphur atoms in BBTD have nearly the same

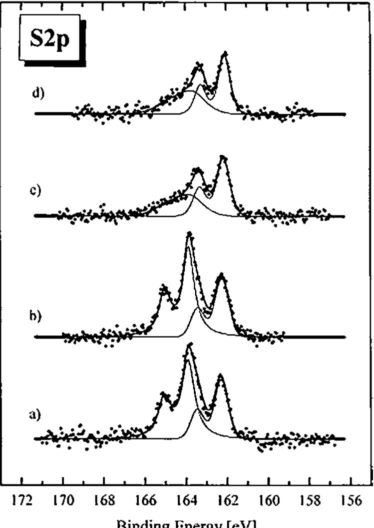


Figure 32. S2p spectra after succsessive treatment of an Au surface with: (a) MBT solution, (b) isooctane, (c) hexadecanethiol solution, and (d) MBT solution. Reprinted with permission from [31], D. Zerulla et al., *Surf. Sci.* 402–404, 604 (1998). © 1998, Elsevier Science.

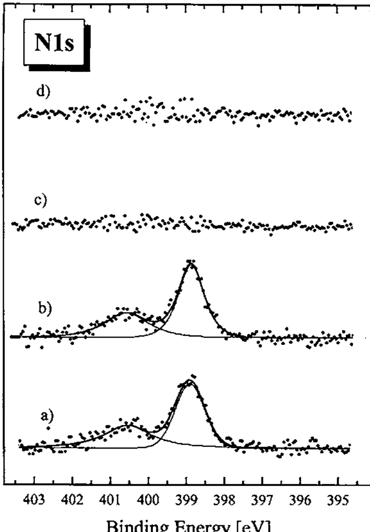


Figure 33. N1*s* spectra after successive treatment of an Au surface with: (a) MBT solution, (b) isooctane, (c) hexadecanethiol solution, and (d) MBT solution. Reprinted with permission from [31], D. Zerulla et al., *Surf. Sci.* 402–404, 604 (1998). © 1998, Elsevier Science.

ground state potential as the endocyclic sulphur in MBT, and cannot be distinguished by their binding energies [118]. After treatment of the MBT-covered sample in pure isooctane for 30 min, no significant changes in the S2*p* and N1*s* spectra were observed (Figs. 32b and 33b), thus excluding any effects caused by this solvent. Only the C1*s* spectrum (Fig. 34b) shows a decrease of the intensity of the signal assigned to aliphatic C–H bonds against the component caused by the aromatic rest, indicating a displacement of hydrocarbon contamination by isooctane. Next, the samples have been treated in an isooctane/hexadecanethiol solution (see Section 3). The changes in the spectra are remarkable. Figure 33c shows the complete absence of nitrogen on the sample surface. This can only be explained by supposing that MBT has been totally removed by the alkanethiol. This conclusion is supported by the S2*p* and C1*s* spectra in Figures 32c and 34c, respectively. These have changed to the typical signals expected for a monolayer of hexadecanethiol on gold. The C1*s* peak shows only one component, belonging to the alkane chain. The S2*p* reveals the typical two doublets for an alkanethiol. The first component at 162.1 eV belongs to the sulphide, which is responsible for the sulphur–gold bonding. The second and less intensive component (163.8 eV) is from disulphide groups. The generation of disulphide in alkanethiol monolayers may arise from air oxidation, but also from the effect of X-rays or electron irradiation. Extended investigations concerning these effects with different energies and intensities on alkanethiol monolayers have been carried out and will be published next. The data in Figures 32d, 33d, and 34d were obtained by exposing the already established alkanethiol film to an aqueous solution of MBT. In contrast to the reversed procedure described above, the surface remains completely covered with hexadecanethiol. Obviously, the adsorbed alkanethiol molecules are resistant against an MBT-containing solution because of the stabilizing effect of the hydrocarbon chain interaction that predominates with respect to the chelating effect of the MBT structure. Consistent with this phenomenon, the MBT adsorbate is completely removed from the gold surface by the hexadecanethiol-containing solution. A similar process to the replacement of the MBT layer by the hexadecanethiol has also been found for the hydrocarbon contamination mentioned at the beginning of this text. In [31], the removal of contamination from a gold surface after the sample has been dipped into a hexadecanethiol solution is demonstrated. The first spectrum shows carbon contamination (C–H and C–O bonds), which was due to exposing the sample to the atmosphere for several days. After a treatment in hexadecanethiol solution for 30 min, additional features of the alkane–carbon appear. The unchanged intensity of the contamination signal gives rise to the assumption that the removal of the carbon does not start until other adsorption sites have been occupied by the thiols. A further treatment for 90 min (total 2 h) removes the contamination and enhances the alkane signal. The last stage shows that an extended treatment (total 16 h) only slightly raises the intensity of the C1*s* signal, indicating a saturation of thiol adsorption with increasing density of the ordered layer. Summarizing, we have confirmed previous results on the ability of thiols to remove organic molecules from gold. In particular, by comparing the bonding of hexadecanethiol and MBT, we have been able to demonstrate that the specific binding of the thiol layers permits stronger bonding to gold, even in relation to other sulphur-containing molecules. This stabilization of a hexadecanethiol layer on a gold surface is believed to be strongly supported by hydrocarbon chain interaction, which is the well-known reason for the self-assembly. This enables the preparation of mixed films.

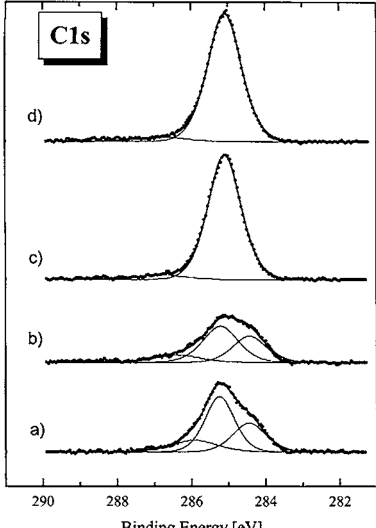


Figure 34. C1*s* spectra after successive treatment of an Au surface with: (a) MBT solution, (b) isooctane, (c) hexadecanethiol solution, and (d) MBT solution. Reprinted with permission from [31], D. Zerulla et al., *Surf. Sci.* 402–404, 604 (1998). © 1998, Elsevier Science.

Starting with a partially covering alkanethiol layer, produced, for example, by stamping the alkanethiol to the surface (see Section 4.3) [21], by different kinds of writing

procedures or with the help of AFM/STM structuring (see Sections 4.1, 4.2) [22, 23], the free parts of the substrate can be covered subsequently with MBT by simply dipping the sample into an MBT solution, leaving the alkanethiol layer structurally undisturbed.

We found that hexadecanethiol displaces the MBT from gold surfaces, similar to the removal of carbon contamination by the alkanethiol. In contrast, MBT cannot replace the adsorbed alkanethiol layer. Therefore, the creation of mixed films has to start with a partial coverage of hexadecanethiol, followed by the MBT, which now can assemble to the uncovered surface without disturbing the alkanethiol.

5. MISCELLANEOUS SELF-ASSEMBLING SYSTEMS

While most of this chapter was focused on alkanethiols as self-assembling species, we will briefly describe further types of molecules which will also self-assemble on semiconductor surfaces.

First, there is the class of functionalized thiols which also consist of a sulphur head group and an alkyl chain, but the terminating group differs from the usual $-CH_3$ group. As long as the terminating group is small in size and does not bind to the semiconductor, we expect the same result as shown in this review. If the terminating group does bind to the substrate, we not only get reversed bounded molecules, but also loops where the head and the terminating group are bounded to the substrate [64]. Instead of exchanging the terminating group, one can change the head group which is responsible for the substrate–molecule binding. Instead of the thiol (–SH) group, sulphonic and phosphorus groups have been used. The choice of the head group should be made with respect to the substrate used. Thiols bind very well on coinage metals (Au, Ag, Cu) and III–V semiconductors, while sulphonic and phosphorus groups bind to oxidized metal surfaces, for example, Al (see Table 1). In the case of the technologically very important Si [42–49], the best choice is to use alkyltricholorosilanes, which give high-quality self-assembled monolayers on this substrate [38–41].

One of the many goals in using SAMs for technological applications is to construct assays for biological and medical sensors. This can be done in two ways. The first is to use well-known self-assembling species as building blocks in the construction of these assays. They may be used as spacers, to specifically bind biomolecules or as a structure-defining substrate themselves [119]. Second, some of the biological species can be used as self-assembling monolayers themselves. One example is the amino acid cysteine, which is able to bind on III–V semiconductor surfaces as well as on coinage metals by its sulphur side group [64]. However, this kind of molecule does not deliver highly ordered monolayers due to two facts. Its chain length is not large enough to have enough van der Waals interactions between neighboring molecules; therefore, one of the most important forces in self-assembly is too weak. Furthermore, it not only can bind with the sulphur, but also with the carboxyl group to most substrates. The result is a nearly unordered monolayer with only weak passivating properties compared to alkanethiols [64].

Table 2. References of semiconductor–SAM combinations in literature.

Substrate	Adsorbate(s)	Author(s)	Ref.
Si(100)	Alkyltrichlorosilane (docosyltrichlorosilane, octyltrichlorosilane)	Nuzzo, Chidsey, Grunze	[38–41, 123]
GaAs	Alkanethiols	Allara	[121–122]
InP(100)	Allkanethiols	Gu Waldeck	[25]
InP(110)	Alkanethiols, mercaptobenzothiazole	Zerulla	[53, 64, 66 67, 75]
GaP(110)	Alkanethiols	Zerulla Chassé	[67]
CdS (1010)	Mercaptobenzothiazole	Mayer Szargan	[75, 124]
PbS, FeS_2	Thiols	Schaufuß, Szargan	[118]

In general, all species that fulfill the following simple rules are likely to form highly ordered self-assembling monolayers:

- a chain length >6 (best: 12–24) [120]
- only one group which binds to the substrate
- the space requirements of the molecules are smaller than the lattice constants of the substrates (more precisely, the distance between binding sites on the substrate).

Of course, these rules ignore complicated molecule–molecule and molecule–substrate interactions, but will suffice in most cases. A compilation of references of semiconductor–SAM combinations in the literature is shown in Table 2. Note that the number of investigations of SAMs on semiconductors is by far lower than on coinage metals. A more detailed treatment, including structure and electronic properties, can be found in Section 4.

6. CONCLUSION

The self-assembly of thiols and related molecules on semiconductors leads to highly ordered systems. Independent of the preparation technique (solution/evaporation), it results in stable, passivating monolayers. For the example of alkanethiols on InP(110), proof is given that the self-assembly leads to single-domain systems with exceptional order. Furthermore, different techniques have been presented which may be used to write structures into the self-assembled monolayers on nanometer scales.

Additionally, even the possibility of constructing monolayers from different self-assembling species was demonstrated. Comparing the results on semiconductor (110) surfaces with those on Au (111), we can conclude that the ease of preparation, passivating properties, and structure are similar. Moreover, it can be stated that the perfection of the adsorbed monolayers is even higher on the semiconductor (110) surfaces than on gold, at least regarding InP (110).

As a future prospect, we suggest using the combination of thiols and a III–V (110) surface in order to construct nanodevices with specifically tailored properties.

GLOSSARY

Chemical vapor deposition (CVD) Technique used to deposit coatings, where chemicals are first vaporized, and then applied using an inert carrier gas such as nitrogen.

Langmuir–Blodgett Name of a nanofabrication technique used to create ultrathin films (monolayers and isolated molecular layers), the end result of which is called a "Langmuir–Blodgett film."

Nanoimprinting Sometimes called soft lithography. A technique that is very simple in concept, and totally analogous to traditional mold- or form-based printing technology, but that uses molds (masters) with nanoscale features. As with the printing press, the potential for mass production is clear. There are two forms of nanoimprinting: one that uses pressure to make indentations in the form of the mold on a surface, and the other, more akin to the printing press, that relies on the application of "ink" applied to the mold to stamp a pattern on a surface. Other techniques such as etching may then follow.

Nanolithography Writing on the nanoscale. From the Greek words nanos—dwarf, lithos—rock, and grapho—to write, this word literally means "small writing on rocks."

OLED or organic LED Not made of semiconductors. It is made from carbon-based molecules. That is the key science factor that leads to potentially eliminating LED's biggest drawback—size. The carbon-based molecules are much smaller.

Self-assembly In chemical solutions, self-assembly (also called Brownian assembly) results from the random motion of molecules and the affinity of their binding sites for one another. Also refers to the joining of complementary surfaces in nanomolecular interaction.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (DFG, Innovationskolleg 24/A1-2313-704-P), BMBF (Laserbiodynamik 13N7515/6), and ELMINOS (AGEF). The work was also supported by the Graduiertenkolleg Physical Chemistry of Interfaces. The author acknowledges stimulating discussions with M. Grunze, T. Chassé, and F. Katzenberg. Finally, the author thanks his wife, N. Zerulla, and his research group (especially G. Isfort and M. Kölbach) for their patience.

REFERENCES

1. A. Pockels, *Nature* 43, 437 (1891).
2. A. Pockels, *Nature* 46, 418 (1892).
3. A. Pockels, *Nature* 48, 152 (1893).
4. A. Pockels, *Nature* 50, 223 (1894).
5. L. Rayleigh, *Philos. Mag.* 48, 321 (1899).
6. W. B. Hardy, *Proc. R. Soc. London A* 86, 610 (1912).
7. H. Devaux, *Smithsonian Inst. Ann. Rep.* 261 (1913).
8. I. Langmuir, *J. Am. Chem. Soc.* 39, 1848 (1917).
9. I. Langmuir, *Trans. Faraday Soc.* 15, 62 (1920).
10. K. Blodgett, *J. Am. Chem. Soc.* 57, 1007 (1935).
11. K. Blodgett, *Phys. Rev.* 51, 964 (1937).
12. W. C. Bigelow, D. L. Pickett, and W. A. Zisman, *J. Colloid Interface Sci.* 1, 513 (1946).
13. L. C. F. Blackman and M. J. S. Dewar, *J. Chem. Soc.* 162, I (1957).
14. G. L. Gaines, Jr., "Insoluble Monolayers at Liquid-Gas Interfaces." Interscience, New York, 1966.
15. A. Ulman, "An Introduction to Ultrathin Organic Films: From Langmuir-Blodgett to Self-Assembly." Academic, New York, 1988.
16. R. G. Nuzzo and D. L. Allara, *J. Am. Chem. Soc.* 105, 4481 (1983).
17. P. G. de Gennes, *Rev. Mod. Phys.* 57, 827 (1985).
18. S. Dietrich, in "Phase Transition and Critical Phenomena," (C. Domb and J. L. Lebowitz, Eds.), Vol. 12. Academic, New York, 1988.
19. C. D. Bain and G. M. Whitesides, *Angew. Chem. Int. Ed. Engl.* 28, 506 (1989).
20. F. Schreiber, *Progr. Surf. Sci.* 65, 151 (2000).
21. H. A. Biebuyck, N. B. Larsen, E. Delamarche, and B. Michel, *IBM J. Res. Dev.* 41, 159 (1997).
22. C. B. Ross, L. Sun, and R. M. Crooks, *Langmuir* 9, 632 (1993).
23. J. K. Schoer and R. M. Crooks, *Langmuir* 13, 2323 (1997).
24. S. R. Lunt, G. N. Ryba, P. G. Santangelo, and N. S. Lewis, *J. Appl. Phys.* 70, 7449 (1991).
25. Y. Gu, Z. Lin, R. A. Butera, V. S. Smentkowski, and D. H. Waldeck, *Langmuir* 11, 1849 (1995).
26. H. Sellers, A. Ulman, Y. Shnidman, and J. E. Eilers, *J. Am. Chem. Soc.* 115, 9389 (1993).
27. J. Kang and P. A. Rowntree, *Langmuir* 12, 2813 (1996).
28. D. L. Allara and R. G. Nuzzo, *Langmuir* 1, 52 (1985).
29. P. E. Laibinis, G. M. Whitesides, D. L. Allara, Y. Tao, A. N. Parikh, and R. G. Nuzzo, *J. Am. Chem. Soc.* 113, 7152 (1991).
30. R. G. Nuzzo, B. R. Zegarski, and L. H. Dubois, *J. Am. Chem. Soc.* 109, 733 (1987).
31. D. Zerulla, I. Uhlig, R. Szargan, and T. Chasse, *Surf. Sci.* 402–404, 604 (1998).
32. D. Zerulla and T. Chasse, *Langmuir* 15, 5285 (1999).
33. E. Dudzik, A. Leslie, E. O'Toole, I. T. McGovern, A. Patchett, and D. R. T. Zahn, *Appl. Surf. Sci.* 104, 101 (1996).
34. T. K. Johal, P. Finetti, V. R. Dhanak, A. W. Robinson, A. Patchett, D. R. T. Zahn, and R. McGrath, *Appl. Surf. Sci.* 104–105, 257 (1996).
35. T. Chassé, A. Chassé, H. Peisert, and P. Streubel, *Appl. Phys. A* 65, 543 (1997).
36. H. Peisert, T. Chassé, P. Streubel, A. Meisel, and R. Szargan, *J. Electron Spectrosc. Related Phenom.* 68, 321 (1994).
37. T. Chassé, H. Peisert, P. Streubel, and R. Szargan, *Surf. Sci.* 331–333, 434 (1995).
38. K. R. Finnie and R. G. Nuzzo, *Langmuir* 17, 1250 (2001).
39. K. R. Finnie, R. Haasch, and R. G. Nuzzo, *Langmuir* 17, 6968 (2001).
40. D. L. Allara, A. N. Parikh, and F. Rondelez, *Langmuir* 11, 2357 (1995).
41. K. Bierbaum and M. Grunze, *Langmuir* 11, 2143 (1995).
42. V. V. Doan, and M. J. Sailor, *Science* 256, 1791 (1992).
43. M. P. Stewart and J. M. Buriak, *Adv Mater.* 12, 859 (2000).
44. O. D. Velev and E. W. Kaler, *Adv Mater.* 12, 531 (2000).
45. S. Schuppler, S. L. Friedman, M. A. Marcus, D. L. Adler, H. Xie, F. M. Ross, Y. J. Chabal, T. D. Harris, L. E. Brus, W. Brown, E. E. Chaban, P. F. Szajowski, S. B. Christman, and P. H. Citrin, *Phys. Rev. B* 52, 4910 (1995).
46. V. Lehmann and U. Gosele, *Appl. Phys. Lett.* 58, 856 (1991).
47. J. Harper and M. J. Sailor, *Anal. Chem.* 68, 3713 (1996).
48. T. F. Harper and M. J. Sailor, *Anal. Chem.* 119, 6943 (1997).
49. K.-P. S. Dancil, D. P. Greiner, and M. J. Sailor, *J. Am. Chem. Soc.* 121, 7925 (1999).
50. [a] W. Schommers and P. von Blanckenhagen, "Structure and Dynamics of Surfaces I," Springer Topics in Current Physics, Vol. 41. Springer, Berlin, New York, Heidelberg, 1985. [b] H. Kalt, in "Optical Properties of Semiconductors," Springer Series in Solid-State Science, Vol. 120. Springer, Heidelberg, 1996.
51. N. Esser, U. Resch-Esser, M. Priestovsek, and W. Richter, *Phys. Rev. B* 53, 13257 (1996).

52. G. E. Poirer and E. D. Pylant, *Science* 272, 1145 (1996).
53. D. Zerulla and T. Chassé, *Langmuir* 18, 5392 (2002).
54. J. P. Song, N. H. Pryds, K. A. Morch, A. R. Thölen, and L. N. Christensen, *Rev. Sci. Instrum.* 64, 900 (1993).
55. K. Dickmann, F. Demming, and J. Jersch, *Rev. Sci. Instrum.* 67, 845 (1997).
56. J. A. Dagata, W. Tseng, J. Bennett, J. Schneir, and H. H. Harary, *J. Appl. Phys.* 70, 3661 (1991).
57. C. R. K. Mariann, F. K. Perkins, S. L. Brandow, T. S. Koloski, E. A. Dobisz, and J. M. Calvert, *Appl. Phys. Lett.* 63, 390 (1994).
58. K. Dickmann, F. Demming, and J. Jersch, *Rev. Sci. Instrum.* 67, 845 (1996).
59. C. A. Widrig, C. A. Alves, and M. D. Porter, *J. Am. Chem. Soc.* 113, 2805 (1991).
60. G. E. Poirer, W. P. Fitts, and J. M. White, *Langmuir* 17, 1176 (2001).
61. A. Ikai, *Surf. Sci. Rep.* 26, 261 (1996).
62. H. Neddermeyer, "Scanning Tunneling Microscopy," 1st ed. Kluwer Academic, 1993.
63. C. Bai, "Scanning Tunneling Microscopy and Its Application," 1st ed. Springer, 1994.
64. D. Zerulla, Ph.D. Dissertation, Universität Leipzig, Wilhelm-Ostwald-Institut, Germany, 1999.
65. S. Schömann and T. Chassé, *J. Vac. Sci. Technol. A* 16, 2990 (1998).
66. D. Zerulla and T. Chassé, *Chem. Phys. Lett.* 311, 8 (1999).
67. T. Chassé, D. Zerulla, and K.-H. Hallmeier, *Surf. Rev. Lett.* 6, 1179 (1999).
68. K. Bierbaum, M. Kinzler, Ch. Wöll, M. Grunze, G. Hähner, S. Heid, and F. Effenberger, *Langmuir* 11, 512 (1995).
69. J. M. Behm, K. R. Lykke, M. J. Pellin, and J. C. Hemminger, *Langmuir* 12, 2121 (1996).
70. R. Gerlach, G. Polanski, and H.-G. Rubahn, *J. Chem. Phys.* (1996).
71. H.-J. Himmel, Ch. Wöll, R. Gerlach, G. Polanski, and H.-G. Rubahn, *Langmuir* 13, 602 (1997).
72. Ch. Jung, O. Danneberger, Y. Xu, M. Buck, and M. Grunze, *Langmuir* 14, 1103 (1998).
73. D. Zerulla, N. Zerulla, and C. Ehlers (to be published).
74. J. Stöhr, NEXAFS. Springer, Heidelberg/New York, 1995.
75. D. Mayer, K.-H. Hallmeier, D. Zerulla, and R. Szargan, in "Topics in Applied Physics," (K. Wandelt, Ed.). Springer, Heidelberg, 2002.
76. M. D. Porter, T. B. Bright, D. L. Allara, and C. E. D. Chidsey, *J. Am. Chem Soc.* 109, 3559 (1987).
77. R. Hesse, T. Chassé, and R. Szargan, *Fresenius J. Anal. Chem.* 365, 48 (1999).
78. M. Wirde, U. Gelius, T. Dunbar, and D. Allara, *Nucl. Instrum. Meth. Phys. Res. B* 131, 245 (1997).
79. B. Jäger, H. Schürmann, H. U. Müller, H.-J. Himmel, M. Neumann, M. Grunze, and Ch. Wöll, *Z. Phys. Chem.* 202, 263 (1997).
80. R. C. Tiberio, H. G. Craighead, M. Lercel, T. Lau, C. W. Sheen, and D. L. Allara, *Appl. Phys Lett.* 62, 468 (1993).
81. M. J. Lercel, G. F. Redinbo, H. G. Craighead, C. W. Sheen, and D. L. Allara, *Appl. Phys Lett.* 65, 974 (1994).
82. M. J. Lercel, H. G. Craighead, A. N. Parikh, K. Seshadri, and D. L. Allara, *Appl. Phys Lett.* 68, 1504 (1996).
83. K. Siegbahn, *Nova Acta Soc. Sci. Upsaliensis* 20, 175 (1967).
84. V. I. Nefedov, "X-ray Photoelectron Spectra of Chemical Compounds; Chemistry." Moscow.
85. D. Gallet and G. Hollinger, *Appl. Phys. Lett.* 62, 982 (1992).
86. D. Rading, V. Liebing, G. Becker, H. Fuchs, and A. Benninghoven, *J. Vac. Sci. Technol. A* 16, 3449 (1998).
87. A. S. Duwez, L. M. Yu, J. Riga, J. J. Pireaux, and J. Delhalle, *Thin Solids Films* 329, 156 (1998).
88. B. Völkel, A. Gölzhäuser, H. U. Müller, C. David, and M. Grunze, *J. Vac. Sci. Technol. B* 15, 2877 (1997).
89. R. G. Nuzzo, L. H. Dubois, and D. L. Allara, *J. Am. Chem Soc.* 112, 558 (1990).
90. P. Fenter, A. Eberhardt, and P. Eisenberger, *Science* 266, 1216 (1994).
91. P. Fenter, F. Schreiber, L. Berman, G. Scoles, P. Eisenberger, and M. J. Bedzyk, *Surf. Sci.* 412–413, 213 (1998).
92. D. R. Mullins, T. Tang, X. Chen, V. Shneerson, D. K. Saldin, and W. T. Tysoe, *Surf. Sci.* 372, 193 (1997).
93. D. R. Mullins, D. R. Huntley, T. Tang, D. K. Saldin, and W. T. Tysoe, *Surf. Sci.* 380, 468 (1997).
94. B. L. Henke, J. P. Knauer, and K. J. Premaratne, *Appl. Phys.* 52, 1509 (1981).
95. M. S. Kariapper, G. F. Grom, G. J. Jackson, C. F. McConville, and D. P. Woodruff, *J. Phys. Cond. Matter* 10, 8661 (1998).
96. N. N. Greenwood and A. Earnshaw, "Chemistry of the Elements." Pergamon, Oxford, 1984.
97. K.-H. Hallmeier, D. Mayer, and R. Szargan, *J. Electron. Spectrosc. Related Phenom.* 96, 245 (1998).
98. R. F. Service, *Science* 265, 316 (1994).
99. J. Huang, D. A. Dahlgren, and J. Hemminger, *Langmuir* 10, 626 (1994).
100. J. Huang and J. Hemminger, *J. Am. Chem. Soc.* 115, 3342 (1993).
101. M. Tarlov, D. R. F. Burgess, and G. Gillen, *J. Am. Chem. Soc.* 115, 5305 (1993).
102. G. Gillen, J. Benett, M. Tarlov, and D. R. F. Burgess, *Anal. Chem.* 66, 2170 (1994).
103. K. C. Chan, T. Kim, J. K. Schoer, and R. M. Crooks, *J. Am. Chem. Soc.* 117, 5875 (1995).
104. C. D. Frisbie, E. W. Wollman, and M. S. Wrighton, *Langmuir* 11, 2653 (1995).
105. P. C. Rieke, B. J. Trarasevich, L. L. Wood, M. H. Engelhard, D. R. Baer, G. E. Fryxwell, C. M. John, D. A. Laken, and M. C. Jaehnig, *Langmuir* 10, 619 (1994).
106. A. Kumar and G. M. Whitesides, *Appl. Phys. Lett.* 63, 2002 (1993).
107. H. A. Biebuyck and G. M. Whitesides, *Langmuir* 10, 4581 (1994).
108. A. Kumar, E. Kim, and G. M. Whitesides, *Adv. Mater.* 6, 600 (1994).
109. A. Kumar, N. L. Abott, E. Kim, H. A. Biebuyck, and G. M. Whitesides, *Acc. Chem. Res.* 28, 219 (1995).
110. Y. Xia, E. Kim, X.-M. Zhao, J. A. Rogers, M. Prentiss, and G. M. Whitesides, *Science* 273, 347 (1996).
111. G. M. Whitesides and Y. Xia, *Photon. Spectra* 31, 90 (1997).
112. X.-M. Zhao, Y. Xia, and G. M. Whitesides, *Adv. Mater.* 8, 837 (1996).
113. E. Kim, Y. Xia, and G. M. Whitesides, *Nature* 376, 581 (1995).
114. E. Kim, Y. Xia, X.-M. Zhao, and G. M. Whitesides, *Adv. Mater.* 9, 651 (1997).
115. Y. Xia and G. M. Whitesides, *Angew. Chem.* 110, 568 (1998).
116. M. D. Fuerstenau, J. D. Miller, and M. C. Kuhn, in "Chemistry of Flotation," p. 27. Am. Inst. Mining Metal. Petr. Eng., New York, 1985.
117. R. Szargan, I. Uhlig, G. Wittstock, and P. Roßbach, *Int. J. Min. Process*, 51, 151 (1997).
118. A. Schaufuß, P. Roßbach, I. Uhlig, and R. Szargan, *Fresenius J. Anal. Chem.* 358, 262 (1997).
119. L. H. Dubois, B. R. Zegarski, and R. G. Nuzzo, *J. Am. Chem. Soc.* 112, 570 (1990).
120. P. Fenter, A. Eberhardt, K. S. Liang, and P. Eisenberger, *J. Chem. Phys.* 106, 1600 (1997).
121. C. W. Sheen, J. X. Shi, J. Martensson, A. N. Parikh, and D. L. Allara, *J. Am. Chem. Soc.* 114, 1514 (1992).
122. O. S. Nakagawa, S. Ashok, C. W. Sheen, J. Martensson, and D. L. Allara, *Jpn. J. Appl. Phys.* 30, 3759 (1991).
123. M. R. Linford and C. E. D. Chidsey, *J. Am. Chem. Soc.* 115, 12631 (1993).
124. D. Mayer, K. H. Hallmeier, T. Chassé, and R. Szargan, *Fresenius J. Anal. Chem.* 361 , 689 (1998).

Self-Assembled Nanobiomaterials

Steve S. Santoso, Shuguang Zhang

Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

CONTENTS

1. Introduction
2. Peptide Systems
3. Nucleic Acid Systems
4. Lipid System
5. Biological Selection of Peptides
6. Summary
Glossary
References

1. INTRODUCTION

Human history has been famously subdivided into distinct ages of materials. Indeed, the progress of scientific sophistication has been tightly linked to the nature of the materials that were predominant in the different eras. The Stone Age, for example, represented the earliest of human ingenuity, when our primitive ancestors created and improved simple tools out of natural rocks. Subsequent Bronze and Iron Ages revealed increased technological savvy as tools and equipments were made from extracted and processed materials. In the "Plastic Age," mankind perfected the science of organic synthesis; and recently, the "Silicon Age" has yielded unprecedented progress in extremely complex electrical devices. In the future, each of these materials will most likely occupy its own niche such that no new material will completely render the previous ones obsolete. However, it can be argued that further progress in particular fields such as regenerative medicine [1], engineering [2, 3], and recently nanotechnology [4] will benefit tremendously from the discovery and development of novel, designed, and integrated materials.

Throughout these ages, scientists have searched for lighter, stronger, more energy-efficient, and in recent times, smaller materials. Materials found in nature in these respects have continued to awe us with their versatility and superiority. The spider silk microfiber, for example, is tougher than steel or Kevlar [5] per unit weight. The male moth antenna can detect a single pheromone molecule secreted by a female over great distances [6], putting current biosensing technology to shame. Breakthroughs in molecular biology and biophysics are starting to give us a glimpse of the molecular basis for these behaviors, and we are starting to learn the basic rules for the organization of biological macromolecules such as proteins, deoxyribonucleic acids (DNA), and ribonucleic acids (RNA) that confer such phenomenal attributes.

Biologists and material scientists are not only interested in the scientific underlying of these natural materials but also in their mimicry and utilization for different applications [7]. Even though spider silk or a moth antenna is currently too difficult to mimic, we have come to realize that biological materials, in general, may have superior qualities over man-made ones for certain applications. Furthermore, any research into biomaterials has benefited from decades of improvements in the chemical synthesis methods of biopolymers, not to mention a wealth of data generated from basic research. Designed peptide molecules, for example, can be synthesized rapidly and analyzed using standard techniques for studying protein structure such as circular dichroism, electron microscopy (EM), and nuclear magnetic resonance (NMR).

A subset of biological materials is the subject of this entry, and all of the members of this set share the same enabling property: they are formed through the self-assembly of particular motifs of biological polymers. Spontaneous associations of diverse molecules into a distinct supramolecular assembly is ubiquitous in biology: for example, the eubacterial ribosome consists of more than 50 proteins and three ribosomal ribonucleic acid molecules that self-organize into a molecular machinery which can perform complex reactions that synthesize proteins [8]. Such nanoscale organizations occur without human instructions or interference, resulting in the formation of structures with little energy expenditure [9].

We will first present the case of self-assembling peptides and describe how researchers have used these systems to create different materials for applications such as mineralization, formation of nanowires, anti-microbial agents, and growth of tissues. We will then focus on DNA molecules that have been utilized as templates for metallization and

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (459–471)

engineered to form crystalline structures. Last, we will summarize the use of larger biological entities: bacterial viruses that allow inorganic metal ions to self-assemble on their surfaces. While practically all of these materials are still in their infancy, the promise of biologically inspired materials has galvanized some researchers to study and improve them vigorously. It is our hope that this entry will show the vast promise that biologically inspired materials have on the advancement of many nano-engineering and medical disciplines.

2. PEPTIDE SYSTEMS

2.1. Introduction to Peptides

Peptides and proteins are biopolymers that are the workhorses of the cell. Their functions are as various as their structures, most playing the roles of enzymes that accelerate reaction rates, membrane proteins that act as channels and molecular transporters, while others act as the structural components which enable a cell to hold its shape and to move about. Each is a polymer that is made up of a linear combination of amino acids (Figures 1A and 1B) and can vary in length between a few to tens of amino acids (also called "residues" in the vernacular of biology) in the case of signaling peptides to ten thousand long in a membrane protein of a human cell [10]. This structural versatility of proteins is afforded by the vastness of the amino acid alphabet.

There are 20 naturally occurring amino acids to choose from so that for a 10-amino-acid-long peptide, nature has the option of making 20^{10} (approximately 10 trillion) different species. The peptides and proteins that are found in nature, however, have been selected through billions of years of evolution so that what is left is a miniscule subset of all the possible sequences. These are the proteins whose amino acid sequences give distinct three-dimensional structures or folds so that they can perform tasks in an extremely specific and efficient manner. Indeed, the structure of a protein is the chief reason for its function, and it has been the holy grail of structural biology to predict the three-dimensional structure of a protein from its amino acid sequence.

In essence, biomaterial and nanoscale scientists want to do just the reverse: knowing a desired structure, can he or she design a sequence that can self-assemble to form it? Even though our knowledge of how linear protein molecules fold into higher order structures is still rather rudimentary, the basic principles that we have learned so far have enabled researchers in biological materials to design peptide systems which have already shown considerable promise for a broad range of applications.

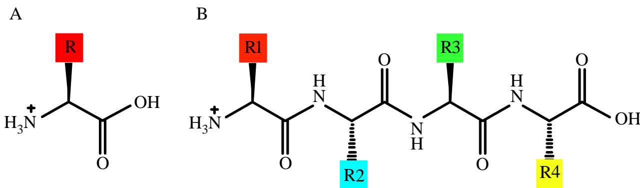


Figure 1. (A) Chemical structure of an amino acid. The side chain, labeled "R" here, can be any one of the 20 naturally occurring amino acids or their derivatives; (B) a four residue (amino acid) peptide. Each peptide is covalently linked to the adjacent one through a peptide amide bond.

2.2. Peptides That Self-Assemble Into Fibers for Biomineralization

Hartgerink et al. reported the rational design [11] and application of a peptide-amphiphilic molecule in a biomineralization process that is akin to the collagen fibers in bone tissue [12]. This lowest structural level in bone is already rather complex, since it involves the integration of an organic protein component with an inorganic, hydroxyapatite (HA) crystal component. Furthermore, the crystals grow in a particular orientation within the collagen fibrils, with their unit cell *c* axis parallel to the long axis of the fibrils [13]. Certainly, any biomaterial that has the potential to replace bone tissue at this level must have similar properties.

Hartgerink et al. synthesized a peptide molecule with a 16-carbon alkyl tail (Figure 2A). The hydrophobicity of the tail and the conical shape of the polymer would drive the self-assembly into cylindrical micelles [14] (Figure 2B). The peptide itself contained amino acids that were deemed to be crucial for particular functions: a string of cysteine residues following the alkyl chain would form intermolecular disulfide bridges, which would confer structural integrity for the self-assembled nanofibers; a linker group made up of glycine residues provided structural flexibility; a phosphoserine residue promoted the nucleation of HA crystals; and an

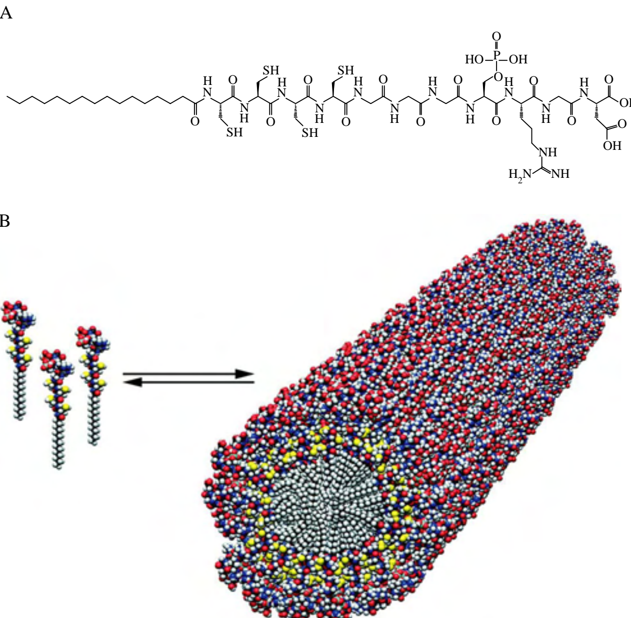


Figure 2. (A) Peptide amphiphile synthesized by Hartgerink et al. for use in biomineralization of hydroxyapatite crystals [11, 12]; (B) the self-assembly of this peptide amphiphile molecule into a cylindrical micelle. The peptide is diagrammed as a van der Waals space-filling model. Reprinted with permission from [12], J. D. Hartgerink et al., *Science* 294, 1684 (2001). (©) 2001, American Association for the Advancement of Science.

RGD sequence motif, which capped the peptide, promoted cell adhesion.

They discovered the assembly of cylindrical micelles under acidic conditions and disassembly in alkaline ones. Cross-linking of the cysteine residues through air oxidation further stabilized the assembled nanostructures. The ability to assemble or disassemble supramolecular assemblies selectively under controlled conditions is beneficial when making new materials since it allows for reversibility of any undesired structures. Further experiments also revealed that the nanofibers have the potential to nucleate HA crystals in the correct orientation. This promising result highlighted the possibility of using biomolecules or their derivatives in making novel materials for medicine and tissue repair and regeneration.

2.3. Peptides That Form Tubes and Vesicles

2.3.1. Short Amphiphilic Peptides

Ideally, the biopolymer constituents used for self-assembly into materials should be rather inexpensive to synthesize in large quantities, easy to handle, and chemically uncomplicated so analysis and modeling can be performed with a reasonable amount of effort. Our laboratory has designed a simple peptide system having those properties [15, 16]. We made short peptides of around six to seven amino acids that had the properties of surfactant molecules in that each monomer contained a polar and a nonpolar region. For example, a peptide called A_6D had six hydrophobic alanine amino acids followed by one polar, negatively charged aspartic acid (Figure 3A). In essence, the molecule looked like a biological phospholipid in that it had a polar head group and a nonpolar tail.

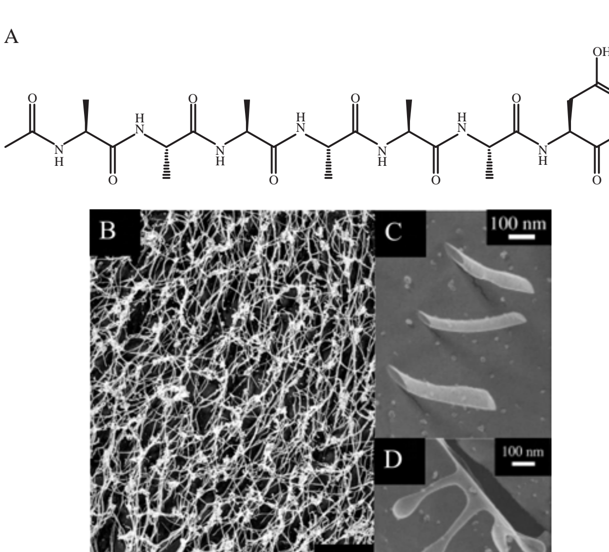


Figure 3. (A) Short surfactant-like peptide A_6D, consisting of six hydrophobic alanine amino acids followed by one negatively charged aspartic acid. (B) The nanotubes formed when this peptide was resuspended in water at pH 7. (C) Picture of V_6D peptide solution showing the presence of tubular structures. (D) V_6D solution showing vesicles budding out.

Dynamic light-scattering experiments on such small peptides in aqueous solution revealed the presence of a uniform assembly that was substantially larger than the size of individual monomers. The homogeneity and size of the supramolecular assembly were sequence-sensitive: peptides of the same length behaved differently when they had different polar head or hydrophobic tail sequences. Such phenomena have been described theoretically and experimentally in other amphiphilic systems. The shape and size of the assemblies are ultimately dependent on the size and geometry of their constituents [14].

In order to visualize the structures in solution, we utilized the transmission electron microscope (TEM). We used the quick-freeze/deep-etch method for sample preparation [17] to preserve the structures that formed in solution for electron microscopy. Each peptide solution was flash-frozen in liquid propane to prevent the formation of ice crystals. One hundred nanometers of vitreous ice were then sublimed, and the sample was coated with a 1-nm-thick platinum layer and a 20-nm-thick carbon coating and visualized using the TEM.

We observed discrete nanotubes and vesicles in the samples that gave homogeneous size distribution in the dynamic light scattering experiment (Figures 3B and 3C). Those samples that were polydispersed tended to give irregular membranous layers. The nanotubes that formed had an average diameter of around 30 nanometers as examined by TEM, consistent with results obtained from the dynamic light scattering. We also observed vesicles budding out of or fusing into a nanotube, suggesting the existence of a dynamic behavior between the different structures in the surfactant peptide system (Figure 3D).

These nanotubes have the potential to act as templates for metallization and formation of nanowires. Furthermore, the nanovesicles may be useful as an encapsulating system for drug delivery. Chemical modification of the peptide monomer may expand the function of these structures: for example, a specific cell-surface ligand can be directly incorporated into a vesicle for targeted delivery of insoluble drugs to particular cells. Current research focuses on these applications, along with more detailed structural studies and modeling of the nanostructures.

2.3.2. Cyclical Peptides with Alternating D- and L-Amino Acids

A prime example of how researchers have applied structural information of biomolecules into designing and characterizing a novel nanobiomaterial is the recent development of nanotube arrays made from cyclic peptides. The concept was first proposed by Hassal in 1972 [18] but has only recently been realized. Through several elegant experiments, Ghadiri et al. have shown that cyclic hexa- and octa-peptides with alternating D- and L-enantiomers of amino acids can regularly stack in an anti-parallel fashion to form very regular nanotube arrays, similar to hollow disks that form long cylinders [19] (Figures 4A and 4B). More importantly, the nanotubes were formed by self-assembly of the peptide disks under particular conditions, which depended on the sequences of the peptides. For example, the initial experiment involved the octa-peptide *cyclo*[-(D-Ala-Glu-D-Ala-Gln)$_2$-], a ring that contained the

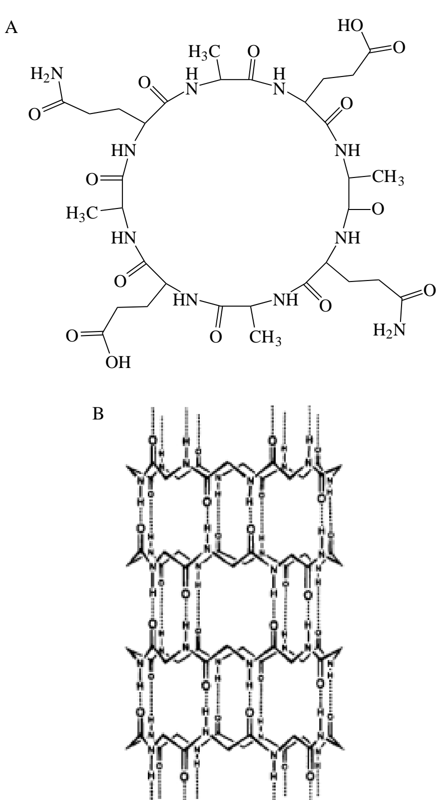


Figure 4. (A) Top-view of a *cyclo*[-(D-Ala-Glu-D-Ala-Gln)$_2$-] monomer, a cyclic peptide with alternating D- and L-amino acids [19]. (B) The structure of the nanotube formed by the antiparallel stacking of the cyclic peptide disks [30]. The dotted lines depict hydrogen bonding between the amide nitrogen of one disk to the amide oxygen of an adjacent disk. Reprinted with permission from [30], M. Engels et al., *J. Am. Chem. Soc.* 117, 9151 (1995). © 1995, American Chemical Society.

hydrophobic D-amino acid alanine (Ala), interspaced with either a negatively charged glutamic acid (Glu) or the polar glutamine (Gln) (Figure 4A). At neutral or alkaline pH, the glutamic acids were effectively deprotonated, and self-assembly was inhibited by electrostatic repulsion of these side chains. At lower pH values, however, the glutamic acids became protonated, the negative charges disappeared, and self-assembly occurred. Such selectivity in initiating assembly adds an important level of control for handling these nanostructures.

This cyclic peptide system has several other important properties: first, the diameter of the pore of the nanotube has the potential to be easily adjustable by changing the diameter of the peptide disk, although eight amino acid-residue peptides may be energetically preferred [20]. Second, due to the geometrical constraints imposed by the alternating amino acid enantiomers, the variable side chains of the amino acids project away from the side of the nanotubes. This is important for added functionality of the system and also for incorporation of the nanotubes into a lipid layer or a hydrophobic self-assembled monolayer (SAM).

One aspect that Ghadiri et al. investigated was the ability of these nanotubes to allow ions [21], small molecules such as glucose [22] and glutamate [23], and metal complexes of various sizes [24] to traverse passively. With an expected pore size of 7.5 angstroms for the eight-residue peptide, ions, small organic molecules, and small metal complexes such as $K_3[Fe(CN)_6]$ and $[Ru(NH_3)_6]Cl_3$ should diffuse through, while the passage of the larger $K_4[Mo(CN)_8]$ should be hindered. For these studies, they incorporated the peptide nanotubes inside a nonpolar lipid environment [25], either in a unilamellar vesicle system or an organosulfur monolayer on gold substrates. The system allowed for a more rapid diffusion of ions than some natural systems such as gramicidin A and amphotericin B. Cyclic voltammogram experiments also revealed the size selectivity of the nanotubes. These results showed that the nanobiomaterial can serve in many applications as synthetic ion channels or as sensors for chemicals of a particular size [26, 27].

By systematically changing the sequence of the residues of the octa-peptide, Fernandez-Lopez et al. have also identified peptides that show antibacterial activity [28]. The peptide nanotube presumably increased the permeability and disrupted the integrity of the bacterial cell membrane. The sequence of the peptide affected the efficiency at which it killed a particular strain of bacteria, most likely because of specific interactions between the peptide side-chains with unique components that exist on the bacterial cell surface. This property is beneficial since inert or useful bacteria inside the body will be spared while pathogenic ones can be specifically targeted. Moreover, the peptides exerted no detrimental effects on mammalian cells, increasing their prospects to serve as antibacterial drugs.

Clark et al., in Ghadiri's laboratory, expanded the system to include nonnatural, β^3-amino acids [29]. Through molecular modeling, the researchers proposed that these cyclic monomers should hydrogen-bond in a parallel manner to form stacked rings similar to the D, L-amino acid system. Furthermore, nanotubes made from cyclical β^3-amino acids are expected to have a dipole moment along their lengths, which may facilitate transport of charged species.

The distinct physical parameters of the nanotubes enabled theorists and computer scientists to perform molecular dynamic studies and *ab initio* calculations of their electronic and molecular structures [30–32]. All calculations have benefited tremendously not only from previous measurements of bond angles and bond lengths of peptide molecules but also from the development of molecular dynamic software for biomacromolecules such as CHARMM [33]. Designing biologically inspired materials is facilitated by these rapid developments in molecular biology and biophysics.

2.3.3. Bolaamphiphilic Peptides

Bolaamphiphiles are amphiphilic molecules that have two hydrophilic ends joined by a hydrophobic segment. Matsui et al. investigated a peptide bolaamphiphile molecule that assembled into a nanotube at acidic pH [34]. The polymer, a bolaamphiphile that contained two glycine residues at each end and a heptane moiety at the center (Figure 5A), self-assembled into a helical ribbon at alkaline pH (Figure 5B), presumably due to the lengthening of hydrogen bonds between the carboxylic acid groups [35].

The nanotube formed was then modified in various ways. Reduction of the nanotube in nickel and copper baths yielded metal-amide complexes that may be used as nanowires in

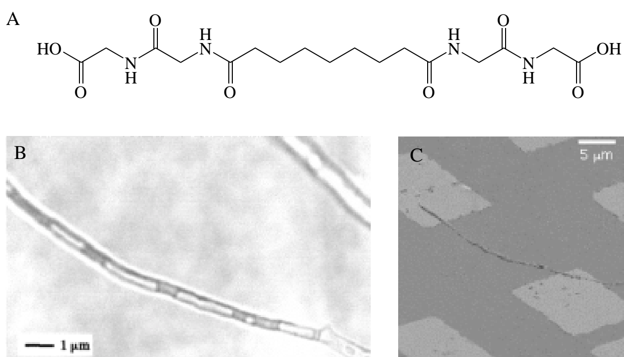


Figure 5. (A) Glycylglycine bolaamphiphile containing two glycine residues at both ends and a hydrophobic heptane group joining them [35]. (B) Tubule formed by the bolaamphiphilic molecule at pH 4. The average diameter of the tube is 500 nm. Reprinted with permission from [35], H. Matsui and B. Gologan, *J. Phys. Chem. B* 104, 3383 (2000). © 2000, American Chemical Society. (C) A scanning electron micrograph (SEM) of a protein tubule immobilized onto biotin-SAM/Au surfaces [39]. Reproduced with permission from [39], H. Matsui et al., *Nano. Lett.* 1, 461 (2001). © 2001, American Chemical Society.

electrical circuits [36]. Interestingly, no metal-amide complex formed when the self-assembled structure was in the helical form, where all amide groups were hydrogen-bonded to one another. This suggested that amide hydrogen bonds are required for the formation of the complex, and that metallization will be unique to the nanotube state of this peptide system.

Nanotube immobilization on surfaces had also been achieved under two different conditions. The first condition was via hydrogen bonding to a SAM on gold surfaces [37]. The nanotube was then metallized using the aforementioned technique. The second immobilization was through specific avidin-biotin interactions [38, 39]. The nanotube was initially coated with the protein avidin, and the complex was immobilized on biotinylated SAMs. A scanning electron micrograph showed that the peptide nanotube was immobilized between two gold contacts (Figure 5C).

The experiments mentioned above are preliminary steps toward the use of biological nanotubes as construction scaffolds for nano-electronics. However, many issues remain to be resolved, such as sample heterogeneity and the effective conductivity of the metal-nanotube complex.

2.4. Peptides That Form 3D Scaffold Hydrogels

One of the self-assembling peptide types, also called "molecular Lego," form very stable beta-sheet structures in aqueous solution because they contain two distinct surfaces—one hydrophilic, the other hydrophobic. Like LEGO® bricks that have pegs and holes and can be assembled into particular structures at centimeter and meter scale according to a program, these peptides can do so at the nanoscale level without external instructions. The unique structural feature of these peptides is that they form complementary ionic bonds with regular repeats on the hydrophilic surface (Figure 6A). The complementary ionic sides have been classified into several

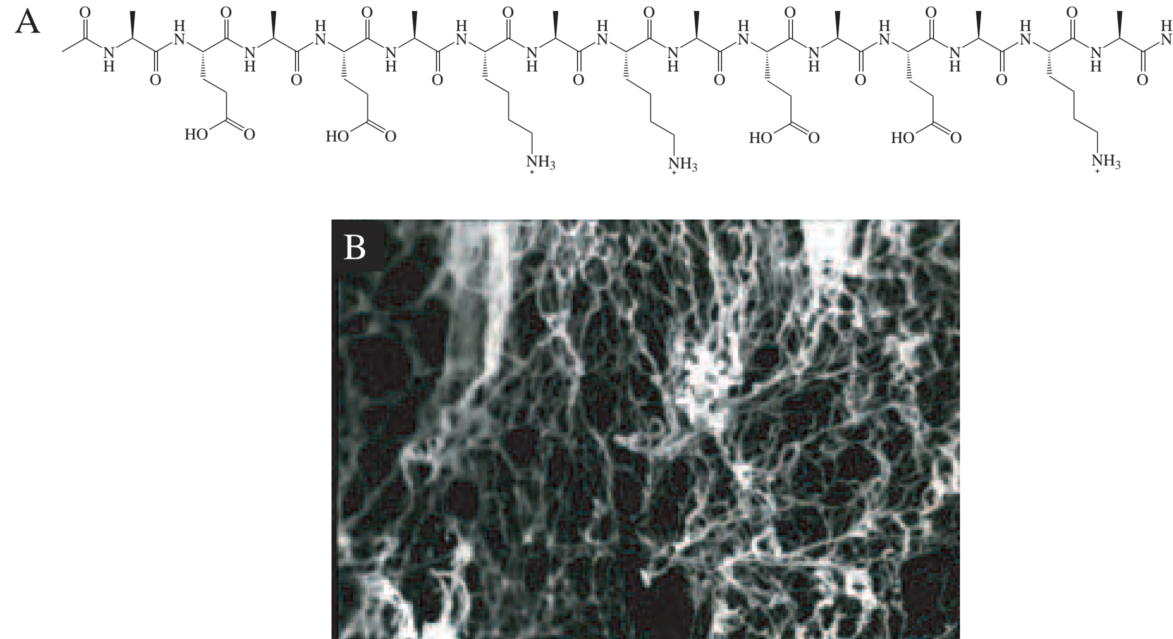


Figure 6. Self-assembling peptide to form scaffold hydrogel. (A) The sequence of the first member of the self-assembling peptides, EAK16, AEAEAKAKAEAEAKAK, serendipitously discovered in a yeast left-handed Z-DNA binding protein [74]. (B) The SEM structure revealed that the peptide self-assemble to form individual interwoven nanofibers with 50–200 nm pores. The diameter of the fiber is about 10–20 nm. Under high resolution by AFM, the filaments are revealed to be a twisted helix with regular helical repeats at early stage of self-assembly. They form scaffold hydrogel with extremely high water content (99.5–99.9%). It is likely that the nanofibers may organize water molecules.

moduli, that is, modulus I, II, III, IV, etc., and mixed moduli. This classification is based on the hydrophilic surface of the molecules that have alternating positively (+) and negatively (−) charged amino acid residues, alternating by 1, 2, 3, 4, and so on. For example, molecules of modulus I have − + − + − + − +, modulus II, − − + + − − + +, modulus IV, − − − − + + + +. These well-defined sequences allow them to undergo ordered self-assembly, resembling situation found in well-studied polymer systems.

Upon the addition of monovalent cations or the introduction of the peptide solutions into physiological media, these peptides spontaneously assemble to form macroscopic structures which can be fabricated into various geometric shapes [40, 41]. Scanning electron microscopy (SEM) and atomic force microscopy (AFM) reveal that the matrices are made of interwoven nanofibers having 10–20 nm in diameter and pores about 50–200 nm in diameter [40–45] (Figure 6B).

Atomic force microscopy and TEM experiments, as well as molecular simulation of the FKFEFKFE peptides, showed that billions of individual peptides self-assemble into nanofibers. Molecular simulation using CHARMM suggests that these individual peptides coalesce together to form a left-handed double helix with hydrophobic phenyalanines on the inside away from water, and lysines and glutamates on the outside to interact with water.

The self-assembly process is a function of time, progressing from mostly short, left-handed helical segments in the first few minutes to long nanofibers after a few hours [45]. This structure represents an example of this class of self-assembling beta-sheet peptides that spontaneously undergo association under water and physiological conditions. If the charged residues are substituted, that is, the positively charged lysines are replaced by positively charged arginines and the negatively charged glutamates are replaced by negatively charged aspartates, there are essentially no drastic effects on the self-assembly process. However, if the positively charged residues, Lys and Arg, are replaced by negatively charged residues, Asp and Glu, the peptide can no longer undergo self-assembly to form macroscopic materials, although they can still form beta-sheet structures in the presence of salt. If the alanines are changed to more hydrophobic residues, such as Leu, Ile, Phe, or Tyr, the molecules have a greater tendency to self-assemble and form peptide matrices with enhanced strength [43–45].

Several peptide materials have been tested for their ability to support cell proliferation and differentiation. A number of mammalian cells have also been tested and all have been found to be able to form stable attachments to the peptide scaffolds [40]. These results suggested that the peptide scaffolds cannot only support various types of cell attachments, but can also allow the attached cells to proliferate and differentiate. For example, once rat PC12 cells on peptide matrices were exposed to neural growth factor (NGF), they underwent differentiation and exhibited extensive neurite outgrowth. In addition, when primary mouse neuron cells were allowed to attach the peptide materials, the neuron cells projected lengthy axons that followed the specific contours of the self-assembled peptide surface and made active and functional connections [41]. Furthermore, when chondrocytes were encapsulated in the scaffolds, the cell exhibited fully functional cartilage properties. These cells not only underwent limited proliferation, but also produced large amounts of type II and XI collagens and glycosaminoglycan, which are typical cartilage cell products [46]. This peptide scaffold is now being developed as a general three-dimensional culture system not only for a broad range of tissue cell cultures but also for tissue repair in regenerative medicine.

2.5. Peptides That Coat Surfaces

One class of peptides has been designed to self-assemble into a monolayer on surfaces and allow cells to adhere to them [47]. These peptides have three general regions along their lengths: a ligand for specific cell recognition and attachment, a linker for physical separation from the surface, and an anchor for covalent attachment to the surface (Figure 7A). The ligand may be of the RGD sequence motif that has been known to promote cell adhesion; the linker is usually a string of hydrophobic amino acids such as alanine or valine; and the anchor can be a cysteine residue for gold surfaces. Zhang et al. have used this technology in conjunction with SAMs prepared through micro-contact printing to place cells into complex patterns (Figures 7B and 7C). This approach may facilitate research into cell-cell communication.

3. NUCLEIC ACID SYSTEMS

3.1. Introduction to Nucleic Acids

Unlike proteins, nucleic acids predominantly play more inert roles in the cell. Deoxyribonucleic acid, for example, is the biopolymer that contains genetic information in its sequence and is processed and replicated by protein machineries. Even though some natural RNAs do perform enzymatic reactions and act as structural scaffolds, they are most renowned as the passive carriers of genetic information that is to be translated into proteins. The simplicity of the roles of DNA and RNA, relative to proteins, is reflected in the limited number of the letters in their alphabet—four as opposed to 20 in the case of proteins. The four letters, called nucleotides, are adenine, thymine or uracil, guanine, and cytosine. Adenine is geometrically compatible with thymine, and guanine with cytosine, such that in solution, one will try to hydrogen bond to the other (Figure 8). This is called complementarity. The most ubiquitous fold that a nucleic acid can form is the complementary right-handed double helix, where two strands of DNA with sequences that complement each other spontaneously form an intertwining double helical structure. For example, a poly-guanine will form a double helix with a poly-cytosine. Unlike most protein folds, this structural body is utterly predictable, as one only needs to know the sequences of the two strands to predict whether a duplex will form. This ease in knowing the conformation of nucleic acids in solution has led to many ingenious experiments that manipulate DNA to perform unnatural functions, the most recent of which are reviewed below.

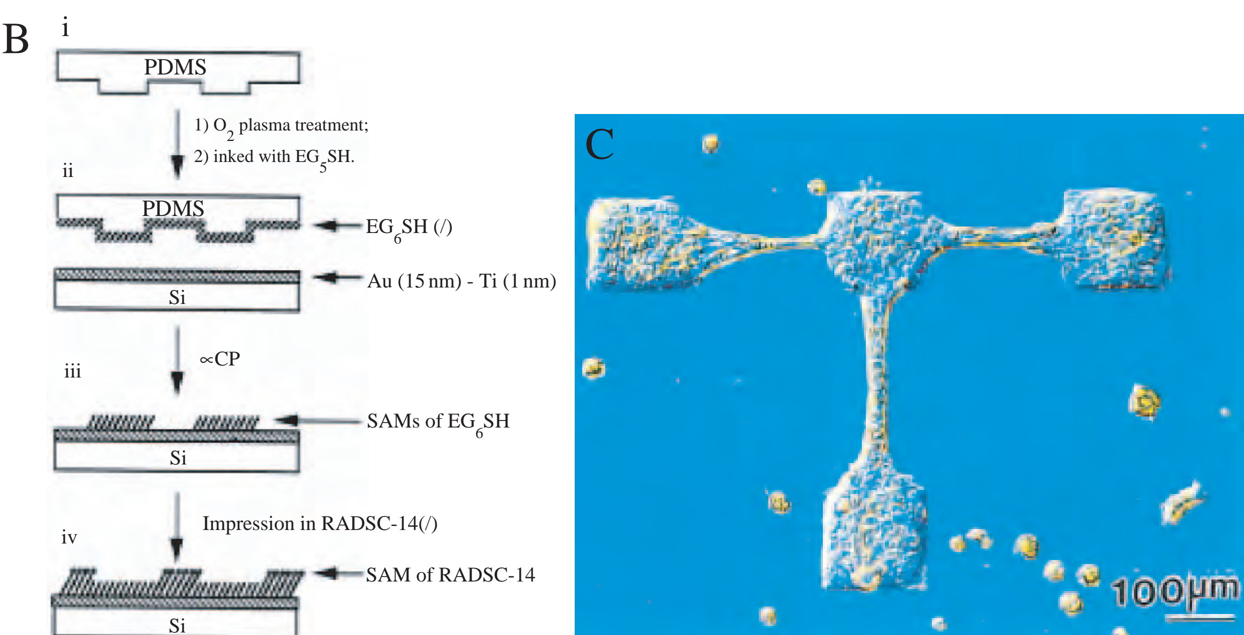


Figure 7. (A) The chemical structure of a surface-coating peptide RADSC-14 [47]; (B) a schematic diagram of micro-contact printing using RADSC-14 peptide and EG_6SH to generate a surface for cell immobilization; (C) human epidermal carcinoma cells grown on an imprinted surface. Reprinted with permission from [47], S. Zhang et al., *Biomaterials* 20, 1213 (1999). © 1999, Elsevier Science.

3.2. Template for Nanowire

The DNA molecule has been suggested as a template for making nanoscale wires for the emergent field of nanoelectronics. This is due to the regularity of the width of the DNA double helix and its robust mechanical properties. Several groups have succeeded in coating DNA molecules with metallic particles and have shown data on the conductive properties of these biotemplated materials.

Braun et al. noncovalently bound a stretch (16 μm) of bacteriophage λ-DNA between two gold electrodes by allowing it to hybridize with short DNA fragments that had been covalently attached to those surfaces [48]. A solution of silver ions was flowed through the self-assembled complex, allowing for the positively charged silver ions to bind to the negatively charged DNA backbone. Reduction of the silver DNA complex resulted in enlarged silver metal particles effectively coating the DNA molecule, which was confirmed using AFM.

Electrical measurements indicated that the wires were nonconducting at low voltage bias, with resistances greater than the experimentally measurable 10^{13} Ω. Furthermore, the shape of the *I-V* curve obtained was dependent on the voltage scan direction. Increasing the silver deposition reduced the severity of some of these issues, implying that modification of certain conditions may yield low-resistance metal wires which have ohmic behavior over a reasonable range of voltages.

Richter et al. employed a similar strategy to produce DNA-templated nanowires that showed relatively low resistances under low-voltage bias [49] (Figures 9A and 9B). They reduced palladium on λ-DNA and immobilized the nanowire on gold electrodes. Electron-beam-induced carbon lines were then formed between the electrodes and the ends of the nanowire, lowering the contact resistances. The resistances obtained were lower than 1 kΩ, with the specific conductivity approximately one order of magnitude lower than bulk palladium.

The resistances of these palladium nanowires were subsequently studied at low temperatures [50]. The study was performed to determine whether the biology of the material affected its behavior as a normal metal system. Theory and experiments on disordered metals showed a trend of increasing resistance with decreasing temperature at sufficiently low temperatures due to weak localization and/or enhanced electron-electron interactions. It was discovered

Figure 8. Chemical structure of the bases in DNA and their complementary partners. Adenine (A) form two hydrogen bonds with thymine (T), and guanine (G) form three hydrogen bonds with cytosine (C). Not pictured is the phosphate backbone of the DNA molecule that connects and arranges these bases in a spiral, ladder-like configuration.

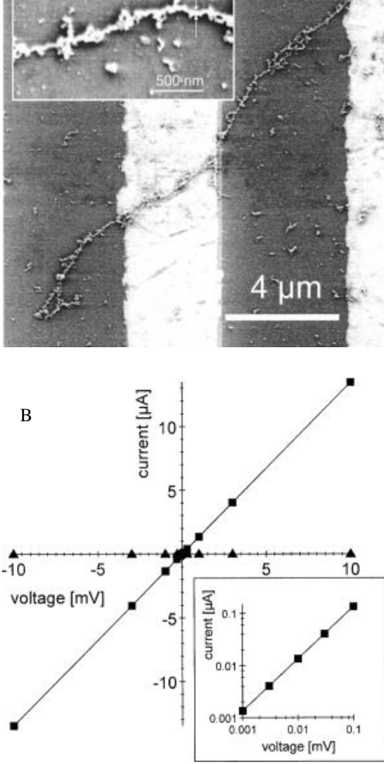


Figure 9. (A) Scanning electron micrograph of a palladium-coated λ-DNA [49]. The templated wire was laid on two gold electrodes over a SiO_2 substrate. (B) *I-V* curve of the wire in figure A. The inset shows the ohmic behavior of the wire down to 1 μV. Reprinted with permission from [49], J. Richter et al., *Appl. Phys. Lett.* 78, 536 (2001). © 2001, American Institute of Physics.

that the palladium metals reduced on a DNA template showed the expected quantum mechanical behavior, with their resistances increasing at low temperatures. This behavior is similar to that of thin palladium films and shows that wires templated with DNA molecules behave normally.

By using first-principle molecular dynamics (FPMD), Mertig et al. discovered conditions in which fine and regular platinum clusters formed on DNA molecules [51]. They hypothesized that the rate of metal formation upon reduction of the DNA-metal ion complex depends on the number of metal nucleation events. Furthermore, this rate-limiting nucleation was controlled by Pt(II)-DNA formation during an activation step when platinum ions bound electrostatically to DNA molecules. Lengthening the activation step resulted in higher occupancy of platinum molecules in the DNA prior to reduction, and ultimately yielded a faster rate of growth and finer metal clusters on the template.

Another metal that has been investigated for surface templating of DNA is gold. Harnack et al. investigated the binding and reduction of tris(hydroxymethyl)phosphine-derivitized gold particles on calf-thymus DNA [52]. The rapidly formed nanowires show electrical conductivities about 1/1000th that of gold, which the authors attributed to the graininess of the material.

Patolsky et al. modified *N*-hydroxysuccinimide-gold nanoparticles with a nucleic-acid intercalating agent, amino psoralen [53]. Addition of this complex with poly-adenine/poly-thymine DNA double-strand resulted in the interacalation of gold nanoparticles to the DNA. Subsequent UV-irradiation covalently cross-linked the two species. Interestingly, deposition of the nanowires onto a mica surface resulted in their structural alignment. Even though electrical studies of these complexes were not performed, the authors showed how they utilized chemical knowledge of a biological system to engineer a desired nanostructure.

3.3. Self-Assembly Into Geometrical Objects

Designing higher-ordered structures such as a polyhedron and other geometrical objects from defined biological building blocks requires the perfect understanding of the behavior of those building blocks. Nucleic acid, in this sense, is the perfect candidate since, as mentioned previously, designing and predicting the complementarity of two nucleic acid strands is trivial. Seeman's group has intelligently designed a quadrilateral [54], a cube [55], a truncated octahedron [56], and Borromean rings [57] made of DNA by using a solid support system [58] (Figures 10A, 10B, and 10C, respectively). Not only did they use the knowledge they have of nucleic acids, but they also employed commonly used nucleic acid modification enzymes to build and analyze the supramolecular objects they designed.

3.4. Formation of DNA Crystal Tiles

Winfree et al. designed a set of Wang tiles made of double-crossover DNA complexes that self-assembled in solution to form macroscopic crystals [59] (Figures 11A, 11B, and 11C). The anti-parallel, double-crossover motif provides structural rigidity and predictability and was inspired by the four-way

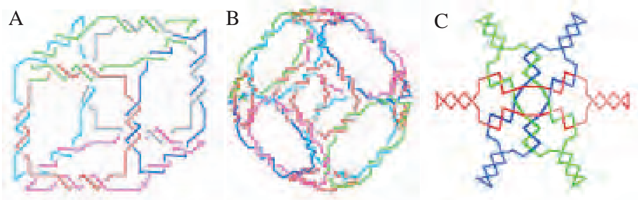


Figure 10. Geometrical objects built with DNA [75]. Pictured are schematics of (A) a cube [55]; (B) a truncated octahedron [56]; (C) a Borromean ring [57]. Reprinted with permission from [75], N. C. Seeman, *Trends in Biotechnology* 17, 437 (1999). © 1999, Elsevier Science.

Holliday junction structure found during meiosis, a biological process that produces the specific chromosome count in eggs and sperm. These mesoscopic tile structures grew to be as large as 2×8 μm in size, with thickness that of a single tile of DNA. Each tile has the approximate size of $2 \times 4 \times 13$ nm and can be associated with one another through "sticky-end" ligation (hybridization and subsequent covalent linkage of complementary sequences between the DNA tiles).

Atomic force microscope images and Fourier analysis showed the presence of two-dimensional sheets with numerous column widths of 13 nm (Figure 11D), consistent with the width of a single DNA tile. Decorating one set of tiles in the two-tile system with two DNA hairpin sequences revealed this columnar structure further in the AFM, since the two hairpins protruded out and provided contrasting height with the rest of the sheet (Figure 11E). The

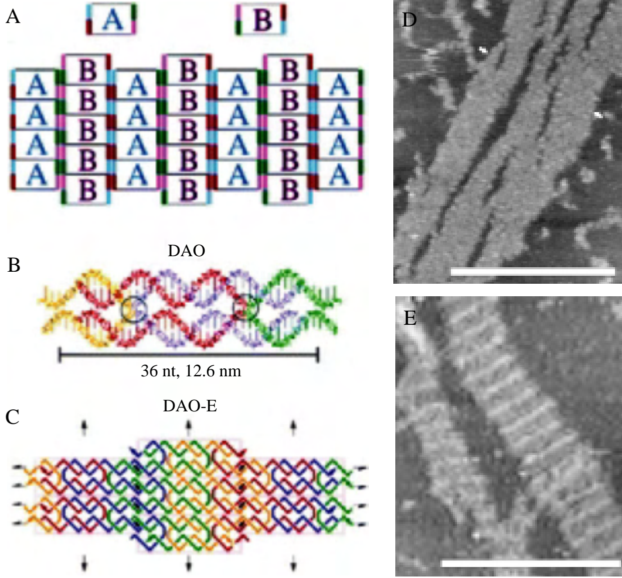


Figure 11. (A) A 2D DNA lattice made up of two Wang tiles, labeled A and B [59]. One side of the A DNA tile had sequences that matched one side of the B tile; (B) a molecular model of a tile. Each colored piece corresponds to one DNA strand; (C) the lattice topology of the assembled tiles. Black arrows indicate dyad symmetry axes; (D) AFM images of two-unit lattice as above. Scale bar corresponds to 300 nm; (E) lattice that incorporated a tile with a protruding DNA hairpin. Reprinted with permission from [59], E. Winfree et al., *Nature* 394, 539 (1998). © 1998, Macmillan Magazines Ltd.

authors also decorated the sheet with nanogold-streptavidin by incorporating a biotin moiety to the 5' end of one set of the double-crossover DNA tile.

3.5. DNA Molecular Machine

Nanometer-sized molecular "machines" based on DNA, which can cycle through different states, have also been designed. Yurke et al. designed a system, which consisted of three DNA molecules that hybridized to form a complex that could accept an incoming "fuel strand" made of another DNA molecule [60] (Figure 12A). The fluorescent quenching between two dyes on one strand (strand A in the figure) was monitored and served as an indication of the state of the system. Briefly, in an open state the two fluorescent dyes were separated such that quenching was lessened. Upon addition of a fuel strand (strand F), the system became "closed" and the two dyes were located in proximity with one another, causing quenching. Addition of a "removal strand," $\overline{F}$, which was complementary to the fuel strand, reset the system back to the open state and generated a double-stranded DNA waste product. The cycle could then be repeated (Figure 12B).

Another system built by Yan et al. consisted of four-stranded DNA molecules that formed two parallel double helices, joined together by the crossing over of the strands [61]. This complex could adopt two stable conformations—a paranemic crossover and its topoisomer. Each was related to one another by a 180° rotation of one strand end (Figure 13A). In their scheme, Yan et al. broke apart two of the strands to make three strands (Figure 13B), the middle one called the "set" strand, and played a role in the switch between the two conformations. Addition of a "fuel" strand, which is perfectly complementary to the "set" strand, left the complex in an intermediate state. To switch the DNA machine into the other state, a different set of set strands was added which would stabilize the other conformer. Since each of these steps involved only noncovalent, hydrogen

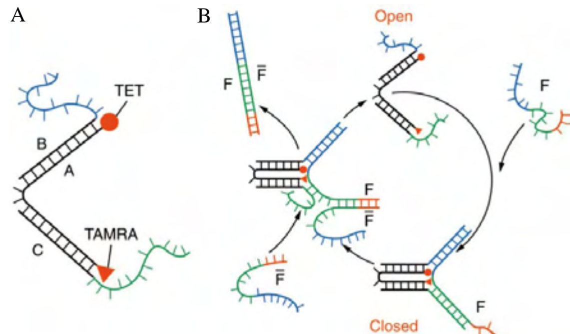


Figure 12. (A) Three strands of DNA that assembled to form the open state of a molecular tweezer [60]. TET and TAMRA are two fluorescent dyes that were attached to the ends of strand A and act as a signal of the state of the system. (B) A schematic showing the operation of the DNA-fueled molecular machine. Addition of fuel strand F closed the system and placed the two fluorescent dyes in promixity to one another, causing fluorescent quenching. The system was reset by adding the release strand, $\overline{F}$, which hybridized to the fuel strand. Reprinted with permission from [60], B. Yurke et al., *Nature* 406, 605 (2000). © 2000, Macmillan Magazines Ltd.

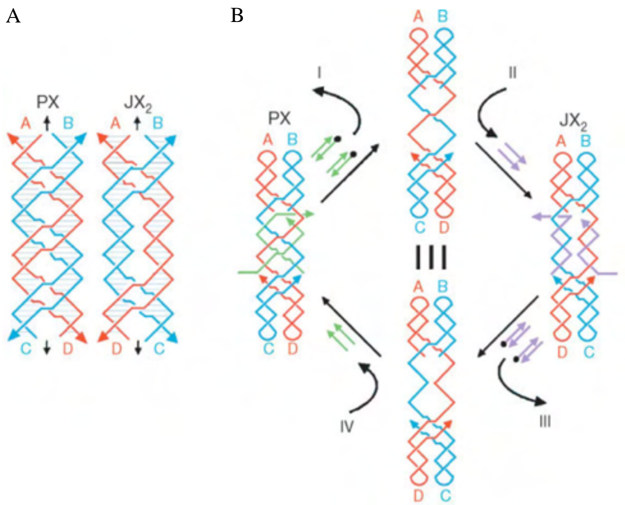


Figure 13. (A) Schematic diagram of the two topological states of a DNA machine designed by Yan et al. [61]. The black arrows indicate a central dyad axis that separates the two helical domains. The arrowheads at the end of the strands indicate the 3'-end of the DNA molecules. (B) Operation of the device. A green set strand in the PX molecule could be removed with the addition of a biotinylated complementary fuel strand, resulting in an unstructured intermediate. The intermediate can be converted into the JX_2 or back to the PX molecule by adding the appropriate set strand. Reprinted with permission from [61], H. Yan et al., *Nature* 415, 62 (2002). © 2002, Macmillan Magazines Ltd.

bonding between the DNA single strands, the cycle could be repeated indefinitely.

4. LIPID SYSTEM

Lipids have long been known to undergo self-assembly to form a variety of structures and colloidal materials. The traditional lipid surfactants that form nano- and micro-structural materials have penetrated into a diversity of applications, from lubricate, detergents, surface-coating to encapsulation for deliveries. The size of individual phospholipid molecules is approximately 2.5 nm in length, but they can self-assemble into millimeter-size lipid tubules with defined helical twists, many million times larger.

Schnur et al. have not only developed an elegant theory to explain the self-assembly of the lipid tubular structures [62], but have also developed a number of applications [63, 64]. They found that the chirality of the monomers plays a key role in the tubular structural formation. Furthermore, they can finely tune the structures at the molecular level both through synthesis of derivatives and using various solvents, especially mixing different types and concentration of alcohols [65]. They coated the metal nanocrystals onto the well-formed, left-handed microhelical tubules. These crystals metal-aligned along the helical ridge on the tubules [66] (Figures 14A and 14B). They also used the metal-coated lipid tubules to coat surfaces so that the properties of the coated surfaces can be completely altered. Their pioneering and innovative research activities have inspired more and more scientists and engineers to conduct multidisciplinary research, namely, to use biological scaffold to construct nanodevices [63].

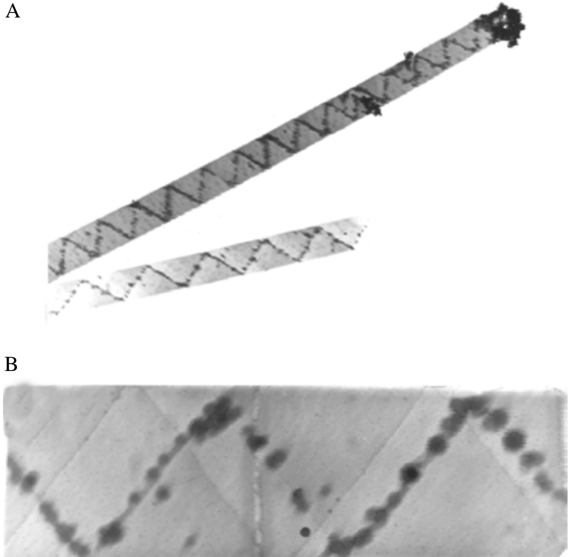


Figure 14. (A) Alignment of nanoparticles along helical ridges inside the 1 μm lipid tubules [66]. (A) magnification at 5000 ×; (B) part of the tubule at 50,000 × magnification. Reprinted with permission from [66], Y. M. Lvov et al., *Langmuir* 16, 5932 (2000). © 2000, American Chemical Society.

5. BIOLOGICAL SELECTION OF PEPTIDES

5.1. Phage Display for Generating Peptides with Novel Properties

5.1.1. Peptides that Specifically Bind to Inorganic Surfaces

For particular applications that have no known analog in biology, molecular design may not be an efficient route to pursue. Even though one can potentially test many different biomolecular species to perform a particular function, the sheer number of samples that must be screened makes such an endeavor prohibitive in cost. This has led Belcher's research group to modify an established method called phage display to generate new peptides that can selectively bind to semiconductors.

Whaley et al. started with a random library of 12 amino acid peptides that were fused to the coat protein of M13 bacteriophage (bacterial virus), resulting in one phage displaying one particular sequence of peptide [67]. The DNA that encoded for the peptide was fused with the DNA of the coat protein; sequence determination of the peptide could then be carried out by standard nucleotide sequencing of this stretch of DNA. They proceeded to allow some 10^9 phages bind to different crystalline semiconductor surfaces such as GaAs (100), GaAs (111), InP (100), and Si (100). Any phage that did not bind to the surfaces was washed away, and those that did were eluted, amplified, and rereacted under more stringent binding conditions. The process was repeated five times, yielding an exponential enrichment of those phages that displayed peptides which enabled them to bind to these inorganic surfaces.

For binding to GaAs (100), peptides with a higher number of uncharged polar and Lewis-base side-chains became more predominant with successive rounds of selection. This could be attributed to the interaction of these functional groups to the Lewis-acid sites of the GaAs surface. Furthermore, phages that bound to one particular surface showed poor binding to other surfaces, indicating that the binding was sequence-specific (Figure 15B).

5.1.2. Peptides that Bind to ZnS/Ordering of Quantum Dots

Using a similar selection strategy, Lee et al. identified a bacteriophage that had the propensity to bind to ZnS crystal surfaces (Figure 15A) [68]. These phages were then mixed with ZnS quantum dots, forming a liquid crystalline suspension of the complex. Differential interference contrast imaging and AFM studies of this suspension revealed the ordering of the bacteriophages, most likely due to the geometrical constraints of the phage body, not the attached quantum dot.

Transparent films from a suspension of viral-quantum dot assembly were made and their properties analyzed. It was discovered that the system self-assembled into a smectic-like lamellar layer, with the ZnS quantum dots localized in the region where lamellar layers met (Figures 15C and 15D). Understanding how this self-assembly occurs when the suspension is dried into a film will enable researchers to design various three-dimensional arrangements of inorganic crystals. This will push forward the areas of nano-electronic, optical, and magnetic sciences and engineering.

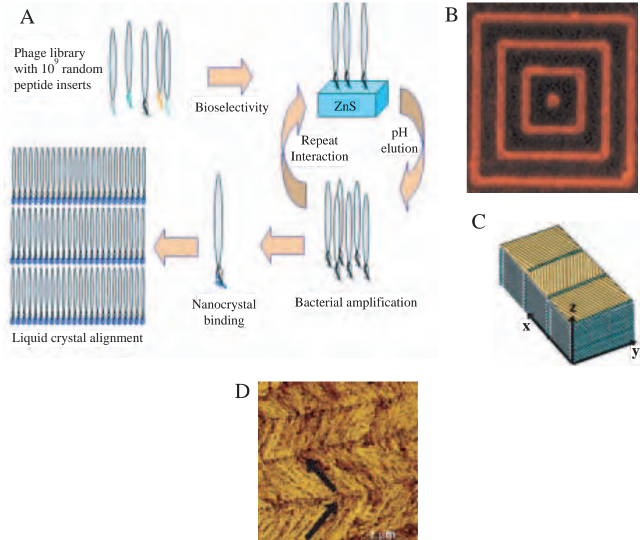


Figure 15. (A) The process employed to select phage presenting a peptide that can selectively bind to ZnS surface [68]. Reprinted with permission from [68], S. W. Lee et al., *Science* 296, 892 (2002). © 2002, American Association for the Advancement of Science. (B) Fluorescently labeled phage that specifically bids to GaAs. The surface consisted of 1 μm GaAs lines and 4 μm SiO_2 spaces [67]. Reprinted with permission from [67], S. R. Whaley et al., *Nature* 405, 665 (2000). © 2000, Macmillan Magazines Ltd. (C) Schematic drawing of the assembly of phage-bound ZnS quantum dots into a film. Reprinted with permission from [68], S. W. Lee et al., *Science* 296, 892 (2002). © 2002, American Association for the Advancement of Science. (D) AFM image of the free surface of the phage-ZnS film. Reprinted with permission from [68], S. W. Lee et al., *Science* 296, 892 (2002). © 2002, American Association for the Advancement of Science.

5.2. Artificial Peptide and Protein Libraries

Artificial peptide and protein libraries have been constructed for selection of novel proteins and peptide motifs that nature never made [69–73]. Many investigators completely designed the peptide and protein libraries *de novo*, without pre-existing protein basis. The idea of constructing protein libraries is for several reasons:

1. to further explore the enormous diversity of protein species,
2. to select for a particular interesting species for a defined purpose,
3. to study protein-folding and interactions,
4. to expand knowledge of biochemistry and life forms,
5. to go beyond biology into materials sciences, molecular engineering, nanotechnology, and uncharted frontiers.

Although nature has selected and evolved many diverse proteins for all sorts of functions that support life, it has not ventured into the functions outside of life. The protein universe is enormous, in comparison with what we know today. There are, undoubtedly, a great number of more proteins that can exist beyond what has been founded in living systems.

Numerous new proteins and peptides with desired and novel properties have been selected for a particular application. This strategy permits us to purposely select and rapidly evolve nonnatural materials, nano-scaffolds, and nano-construction motifs for a growing demand in nanotechnology. The numbers of these biologically based scaffolds are limitless and they will likely play an increasingly important role for design molecular machines, nanodevices, and countless other novel, unanticipated new tools and applications.

6. SUMMARY

In this article, we summarized how different research groups have utilized the self-assembling properties of some biological molecules to form nano-materials for different applications. We categorized the systems into three broad areas, following with the identity of the biopolymer and methodology of the research—designed peptide systems, designed nucleic acid systems, and phage display (*in vivo* evolution). Each system carries distinct advantages over the others. Nucleic acids, for example, exhibit predictability in simple molecular associations. While this increases the ease of designing particular types of structures (such as sheets), it may prohibit the formation of others (such as three-dimensional gels or large nanotubes). Peptides form different nanostructures, but their self-assembling behavior is more complex, has larger degrees of freedom, and is not so well understood. Nevertheless, the different research groups have made headway in assembling different biopolymers into nano- and meso-scale structures and are starting to optimize and make modifications to them.

It is likely that these designed molecular construction motifs and nanoscale materials will become an integral part of future technology for both anticipated applications and even more importantly, unanticipated discoveries.

GLOSSARY

Bacteriophage A virus that infects bacteria.

Dynamic light scattering A scattering technique to determine the hydrodynamic radius and diffusion constant of particles in solution.

Hydrogel A gelatinous substance composed predominantly of water molecules.

Nanobiomaterial Material made from the self-organization of designed biological polymers or their derivatives.

Peptide Short polymers made from amino acids. Peptides are usually shorter than naturally occurring proteins.

Quick-freeze/deep-etch method A method to prepare samples for electron microscopy. Aqueous samples are flash frozen in liquid propane and coated with platinum and carbon. The flash freezing preserves structures in the solution.

Surfactant A compound that reduces the surface tension of water. Usually contains hydrophilic and hydrophobic groups.

ACKNOWLEDGMENTS

We gratefully acknowledge the grant support from the U.S. Army Research Office, Office of Naval Research, Defense Advanced Research Project Agency/Naval Research Laboratories, and NSF CCR-0122419 to Massachusetts Institute of Technology Media Lab's Center for Bits & Atoms, National Institute of Health, Du Pont-MIT Alliance, and the Whitaker Foundation. We also gratefully acknowledge the generous educational donation of a computer cluster from the Intel Corporation.

REFERENCES

1. E. L. Chaikof, H. Matthew, J. Kohn, A. G. Mikos, G. D. Prestwich, and C. M. Yip, *Ann. N. Y. Acad. Sci.* 961, 96 (2002).
2. D. W. Hutmacher, *J. Biomater. Sci. Polym. Ed.* 12, 107 (2001).
3. L. G. Griffith and G. Naughton, *Science* 295, 1009 (2002).
4. K. E. Drexler, "Nanosystems: Molecular Machinery, Manufacturing, and Computation." Wiley, New York, 1992.
5. J. M. Gosline, P. A. Guerette, C. S. Ortlepp, and K. N. Savage, *J. Exp. Biol.* 202 Pt 23, 3295 (1999).
6. K. E. Kaissling, *Chem. Senses.* 21, 257 (1996).
7. J. M. Benyus, "Biomimicry: Innovation Inspired by Nature," 1st ed. Morrow, New York, 1997.
8. M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller, *Science* 292, 883 (2001).
9. G. M. Whitesides and B. Grzybowski, *Science* 295, 2418 (2002).
10. C. C. Lee, J. A. Pearlman, J. S. Chamberlain, and C. T. Caskey, *Nature* 349, 334 (1991).
11. J. D. Hartgerink, E. Beniash, and S. I. Stupp, *Proc. Natl. Acad. Sci. U.S.A.* 99, 5133 (2002).
12. J. D. Hartgerink, E. Beniash, and S. I. Stupp, *Science* 294, 1684 (2001).
13. S. Weiner and H. D. Wagner, *Ann. Rev. Mater. Sci.* 28, 271 (1998).
14. J. N. Israelachvili, "Intermolecular and Surface Forces," 2nd ed. Academic Press, San Diego, 1991.
15. S. Vauthey, S. Santoso, H. Gong, N. Watson, and S. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* 99, 5355 (2002).
16. S. Santoso, W. Hwang, H. Hartman, and S. G. Zhang, *Nano. Lett.* 2, 687 (2002).
17. J. Heuser, *Meth. Cell. Biol.* 22, 97 (1981).
18. C. H. Hassall (Ed.). "Proceedings of the Third American Peptide Symposium," Ann Arbor Science, 1972.
19. M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee, and N. Khazanovich, *Nature* 366, 324 (1993).
20. J. P. Lewis, N. H. Pawley, and O. F. Sankey, *J. Phys. Chem. B* 101, 10576 (1997).
21. M. R. Ghadiri, J. R. Granja, and L. K. Buehler, *Nature* 369, 301 (1994).
22. J. R. Granja and M. R. Ghadiri, *J. Am. Chem. Soc.* 116, 10785 (1994).
23. J. Sanchez-Quesada, H. S. Kim, and M. R. Ghadiri, *Angew. Chem. Internat. Ed.* 40, 2503 (2001).
24. K. Motesharei and M. R. Ghadiri, *J. Am. Chem. Soc.* 120, 1347 (1998).
25. H. S. Kim, J. D. Hartgerink, and M. R. Ghadiri, *J. Am. Chem. Soc.* 120, 4417 (1998).
26. K. Motesharei and M. R. Ghadiri, "Abstracts of Papers of the American Chemical Society" 213, 375 (1997).
27. J. Sanchez-Ouesada, M. P. Isler, and M. R. Ghadiri, *J. Am. Chem. Soc.* 124, 10004 (2002).
28. S. Fernandez-Lopez, H. S. Kim, E. C. Choi, M. Delgado, J. R. Granja, A. Khasanov, K. Kraehenbuehl, G. Long, D. A. Weinberger, K. M. Wilcoxen, and M. R. Ghadiri, *Nature* 412, 452 (2001).
29. T. D. Clark, L. K. Buehler, and M. R. Ghadiri, *J. Am. Chem. Soc.* 120, 651 (1998).
30. M. Engels, D. Bashford, and M. R. Ghadiri, *J. Am. Chem. Soc.* 117, 9151 (1995).
31. M. R. Ghadiri, K. Kobayashi, J. R. Granja, R. K. Chadha, and D. E. McRee, *Angew. Chem. Internat. Ed.* 34, 93 (1995).
32. D. Bashford, M. Engels, and M. R. Ghadiri, "Abstracts of Papers of the American Chemical Society" 212, 323-PHYS (1996).
33. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* 4, 187 (1983).
34. M. Kogiso, S. Ohnishi, K. Yase, M. Masuda, and T. Shimizu, *Langmuir* 14, 4978 (1998).
35. H. Matsui and B. Gologan, *J. Phys. Chem. B* 104, 3383 (2000).
36. H. Matsui, S. Pan, B. Gologan, and S. H. Jonas, *J. Phys. Chem. B* 104, 9576 (2000).
37. H. Matsui, B. Gologan, S. Pan, and G. E. Douberly, *Eur. Phys. J. D* 16, 403 (2001).
38. G. E. Douberly, S. Pan, D. Walters, and H. Matsui, *J. Phys. Chem. B* 105, 7612 (2001).
39. H. Matsui, P. Porrata, and G. E. Douberly, *Nano. Lett.* 1, 461 (2001).
40. S. Zhang, T. C. Holmes, C. M. DiPersio, R. O. Hynes, X. Su, and A. Rich, *Biomaterials* 16, 1385 (1995).
41. T. C. Holmes, S. de Lacalle, X. Su, G. Liu, A. Rich, and S. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* 97, 6728 (2000).
42. S. Zhang, T. Holmes, C. Lockshin, and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* 90, 3334 (1993).
43. M. R. Caplan, P. N. Moore, S. Zhang, R. D. Kamm, and D. A. Lauffenburger, *Biomacromolecules* 1, 627 (2000).
44. E. J. Leon, N. Verma, S. G. Zhang, D. A. Lauffenburger, and R. D. Kamm, *J. Biomater. Sci.-Polymer Ed.* 9, 297 (1998).
45. D. M. Marini, W. Hwang, D. A. Lauffenburger, S. G. Zhang, and R. D. Kamm, *Nano. Lett.* 2, 295 (2002).
46. J. Kisiday, M. Jin, B. Kurz, H. Hung, C. Semino, S. Zhang, and A. J. Grodzinsky, *Proc. Natl. Acad. Sci. U.S.A.* 99, 9996 (2002).
47. S. Zhang, L. Yan, M. Altman, M. Lassle, H. Nugent, F. Frankel, D. A. Lauffenburger, G. M. Whitesides, and A. Rich, *Biomaterials* 20, 1213 (1999).

48. E. Braun, Y. Eichen, U. Sivan, and G. Ben-Yoseph, *Nature* 391, 775 (1998).
49. J. Richter, M. Mertig, W. Pompe, I. Monch, and H. K. Schackert, *Appl. Phys. Lett.* 78, 536 (2001).
50. J. Richter, M. Mertig, W. Pompe, and H. Vinzelberg, *Appl. Phys. A Mater. Sci. Proc.* 74, 725 (2002).
51. M. Mertig, L. C. Ciacchi, R. Seidel, W. Pompe, and A. De Vita, *Nano. Lett.* 2, 841 (2002).
52. O. Harnack, W. E. Ford, A. Yasuda, and J. M. Wessels, *Nano. Lett.* 2, 919 (2002).
53. F. Patolsky, Y. Weizmann, O. Lioubashevski, and I. Willner, *Angew. Chem. Internat. Ed.* 41, 2323 (2002).
54. J. H. Chen, N. R. Kallenbach, and N. C. Seeman, *J. Am. Chem. Soc.* 111, 6402 (1989).
55. J. H. Chen and N. C. Seeman, *Nature* 350, 631 (1991).
56. Y. W. Zhang and N. C. Seeman, *J. Am. Chem. Soc.* 116, 1661 (1994).
57. C. Mao, W. Sun, and N. C. Seeman, *Nature* 386, 137 (1997).
58. Y. W. Zhang and N. C. Seeman, *J. Am. Chem. Soc.* 114, 2656 (1992).
59. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature* 394, 539 (1998).
60. B. Yurke, A. J. Turberfield, A. P. Mills, Jr., F. C. Simmel, and J. L. Neumann, *Nature* 406, 605 (2000).
61. H. Yan, X. Zhang, Z. Shen, and N. C. Seeman, *Nature* 415, 62 (2002).
62. J. V. Selinger and J. M. Schnur, *Phys. Rev. Lett.* 71, 4091 (1993).
63. J. M. Schnur, *Science* 262, 1669 (1993).
64. M. S. Spector, K. R. K. Easwaran, G. Jyothi, J. V. Selinger, A. Singh, and J. M. Schnur, *Proc. Nat. Acad. Sci. USA* 93, 12943 (1996).
65. M. S. Spector, J. V. Selinger, and J. M. Schnur, *J. Am. Chem. Soc.* 119, 8533 (1997).
66. Y. M. Lvov, R. R. Price, J. V. Selinger, A. Singh, M. S. Spector, and J. M. Schnur, *Langmuir* 16, 5932 (2000).
67. S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, and A. M. Belcher, *Nature* 405, 665 (2000).
68. S. W. Lee, C. Mao, C. E. Flynn, and A. M. Belcher, *Science* 296, 892 (2002).
69. A. A. Furka, *Drug Discov. Today* 7, 1 (2002).
70. S. E. Blondelle and R. A. Houghten, *Trends Biotechnol*. 14, 60 (1996).
71. D. A. Moffet and M. H. Hecht, *Chem. Rev.* 101, 3191 (2001).
72. Y. Wei, T. Liu, S. L. Sazinsky, D. A. Moffet, I. Pelczer, and M. H. Hecht, *Protein Sci.* 12, 92 (2003).
73. M. Takahashi, A. Ueno, and H. Mihara, *Chemistry* 6, 3196 (2000).
74. S. Zhang, C. Lockshin, A. Herbert, E. Winter, and A. Rich, *Embo. J.* 11, 3787 (1992).
75. N. C. Seeman, *Trends in Biotechnology* 17, 437 (1999).

www.aspbs.com/enn

Self-Assembled Nanofibers

Hirotaka Ihara, Makoto Takafuji, Toshihiko Sakurai

Kumamoto University, Kumamoto, Japan

CONTENTS

1. Introduction
2. Nanofibers from Covalent Systems
3. Nanofibers from Noncovalent Systems
4. Summary
Glossary
References

1. INTRODUCTION

This article introduces nanofibrillar structures created by self-assembling. The term "self-assembled nanofibers" refers to strands with various shapes such as helices, ribbons, and tubes. Biological organisms are constructed by molecular building blocks and these molecules are assembled spontaneously through various intermolecular interactions [1–6]. Life itself is hierarchically composed of self-assembling molecular building blocks.

There are many kinds of assembled nanofibers in nature. DNA is a double-stranded molecule [7–11]. Each spiraling strand, comprised of a sugar-phosphate backbone and attached bases, is connected to a complementary strand by noncovalent hydrogen bondings between paired bases. Actin forms twisted, rope-like filaments known as F-actins, which are made up of identical building blocks called G-actin subunits [12–14]. Tobacco mosaic virus, for example, forms 300 nm in length × 18 nm in diameter rod-like assemblies composed of 2130 identical protein-armored cylindrical RNAs [15–19]. Each protein subunit contains 158 amino acids.

On the other hand, many artificially self-assembled nanofibers have also been reported in the last two decades [20–32]. Various aggregation morphologies such as helices, tubes, fibers, and ribbon-like and rod-like morphologies have been identified by optical, electron, and probe microscopic observations. Most of their root compounds were serendipitously discovered to form self-assembled nanofibers [33]. The morphologies and properties of these molecular assemblies have attracted the interest of many researchers in a wide range of research fields.

The self-assembled nanofibers themselves can be obtained by simple preparation methods—for instance, by dispersing compounds into media by heat or sonication and then allowing them to stand at a given temperature for a few minutes to several days. These self-assembled nanofibers are interesting due to their numerous potential applications and for understanding biological systems.

Now it is possible to select from a large number of molecular models because many synthetic compounds have been found to form self-assembling nanofibers as well as biomolecules. In this article, the authors would like to introduce self-assembled nanofibrillar aggregates, and specific attention will be paid to aqueous bilayer membrane systems and organogel systems. These self-assemblies show not only unique morphologies but also high molecular orientation toward special functions, for which molecular chirality is an especially important factor. Mirror images of morphologies are formed from enantiomer, and racemates will often destroy developed aggregates [34, 35]. This suggests that chirality is not unrelated with the evolution of life. Well-designed, self-assembled nanofibers can support nanoscopic technologies and their applications.

2. NANOFIBERS FROM COVALENT SYSTEMS

Carbon nanotubes were first discovered in 1991 by S. Iijima et al. as a by-product, while they were examining the generation mechanism of fullerene [36, 37]. These carbon nanotubes form a graphite structure made of a net-like carbon surface in the shape of a cylinder, and both single-wall carbon nanotubes (SWNT) (Fig. 1a) [38, 39], consisting of a single graphitic carbon sheet, and multi-wall carbon nanotubes (MWNT) (Fig. 1b) [37] with a multi-wall structure have been confirmed. The diameter of the single-wall nanotubes is 3 nm and its central cavity is 1–2 nm. On the other hand, the diameter of multi-wall nanotubes ranges from 5–50 nm and the outer diameter of the central cavity can be from 3–10 nm. Characteristically, their lengths exceed 10 μm and both types have a high aspect ratio in every direction. Until now, they were considered expensive solid

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (473–495)

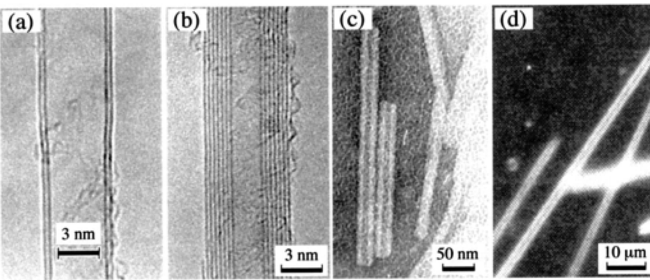


Figure 1. (a) Single-wall and (b) multiwall carbon nanotubes [36–38], and (c) single-wall organic nanotube [21] and (d) organic microtube from synthetic lipids [43]. Reprinted with permission from [36], S. Iijima, *Nature* 354, 56 (1991). © 1991, from [21], K. Yamada et al., *Chem. Lett.* 1713 (1984). © 1984, The Chemical Society of Japan; and from [43], N. Nakashima et al., *J. Am. Chem. Soc.* 107, 509 (1985). © 1985, American Chemical Society.

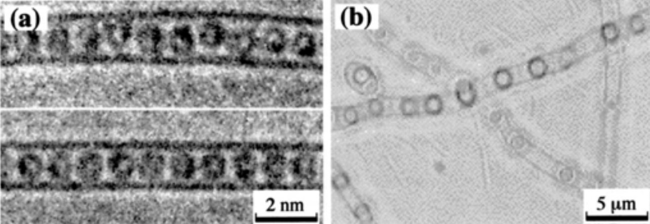


Figure 2. TEM images of (a) carbon nanotube encapsulated Ga@C82 [57] and (b) organic nanotubes encapsulated vesicles [58]. Reprinted with permission from [57], S. Iijima, *Physica* B 323, 1 (2002). © 2002, Elsevier Science; and from [58], T. Shimizu, *Macromol. Rapid Commun.* 23, 311 (2002). © 2002, Wiley-VCH.

carbon compounds as their yield was low. Presently, however, these materials can be created relatively cheap using new industrial methods [40, 41] including discharge, laser evaporation, and catalyst chemical vapor deposition methods, and their application in various fields is being actively explored.

On the other hand, the history of organonanotubes produced from organic matter is much older. In 1984, one of the present authors confirmed that one polypeptide lipid formed aggregates in the shape of a tube with an inner diameter of 4–10 nm and a length of 50–200 nm by self-assembly in an aqueous solution (Fig. 1c) [21]. Since the width of the wall was approximately 5 nm, it is presumed to have been formed as a single-wall bilayer. Nanotubes from longer single-wall bilayers have also been reported in subsequent research [24, 31, 32]. Nakashima et al. discussed microtubes (Fig. 1d) based on bilayer structures [42, 43].

In 2001, *Science* chose carbon nanotubes as one of the top 10 research domains, and growing interest in nanotubes from the field of molecular device physics indicates their application potential in the electronics industry of the future. Especially, SWNTs of approximately 1 nm in diameter have been shown theoretically [44–49] and experimentally [50–53] to become both metal-like and semiconductor-like, and the production and operation of transistors using carbon nanotubes has been reported [54].

Research has already begun on the electronic characteristics of hybrid structures created by inserting different molecules into the space in the center of a nanotube. Iijima and Smith et al. have reported one-dimensional crystallization of C60 molecules by Van der Waals interaction formed by introducing fullerene into nanospace through the deficit part of SWNT and annealing it [55–57]. Further tests have been performed by inserting Ga encapsulated fullerene (Ga@C82) into the SWNT instead of C60, and these results also showed the formation of structures crystallized in one dimension by regular intervals like C60 (Fig. 2a) [57]. On the other hand, similar encapsulation behavior can be attained by organic nanotube systems. Shimizu has excellent photographs of vesicle-encapsulated nanotubes (Fig. 2b) [58].

In contrast to organonanotubes, carbon nanotubes have an essential characteristic property of high electroconductivity, as well as a stable chemistry structure and a morphology that can be kept stable as well. For these reasons, these nanomaterials are certain to be very important in future nanotechnology [54, 59–62]. However, conversely, only limited chemical modification can be performed and addition or conversion of functions is not easy. Self-assembling organotubes, on the other hand, are admittedly unstable both physically and chemically, but the molecules that form them can be exchanged by taking advantage of their aggregate feature, allowing them to take on many different kinds of functions. Moreover, morphologically, they do not necessarily generate tubes alone—various nanofibrillar aggregate morphologies are possible, among which the tube-like aggregate is simply comparatively one of the most stable. This section will end with a brief description of the names of various morphologies that have been produced by self-assembly until now (Fig. 3).

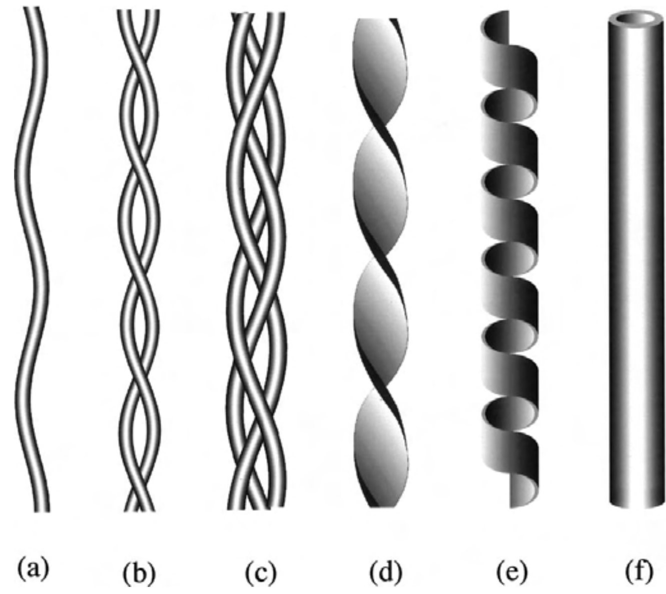


Figure 3. Classification of nanofibrillar aggregates. (a) rope-like fibril, (b–c) multistrand fibril, (d) twisted ribbon-like fibril, (e) helical ribbon-like fibril, and (f) tubular fibril.

3. NANOFIBERS FROM NONCOVALENT SYSTEMS

3.1. Characterization

Self-assembled nanofibers are generally prepared by dispersing into aqueous or nonaqueous (organic solvents) media with heating and sonication procedures, and then allow to stand the solution at designated temperature for few minutes to several days. Some cases are accompanied by macroscopic transformations through formation of well-developed aggregates. A typical example is a viscosity increase and, thus, the judgment of gelation of organic solvents are conventionally carried out by an inversion fluid method [63].

Microscopic techniques are useful for observation of aggregation morphologies—optical microscopes and scanning and transmission electron microscopes (SEM and TEM). Typical examples will be shown in Figures 1, 2, 5, 8, and 13. Scanning probe microscopes such as atomic force microscope (AFM) have also been used to obtain detailed information on the self-assembled morphologies. Figure 4 shows typical AFM images of the lithosatahine protofibriles [64]. A large number of photographs and images of self-assembled aggregates show various fiber-like morphologies such as rods, tubes, helices, ribbons, tapes, and twisted multiple strands. The freeze-fracture and freeze-drying techniques can be combined with these microscopic observations. Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) are important techniques to obtain the information of practical quantities such as diameter, thickness, and length of aggregates. Figure 5 shows an example of the powder X-ray diffraction (XRD) spectra of zerogels prepared by freeze-drying of fibrillar aggregate-containing organogels. The authors discussed the molecular packing mode [65]. Thermodynamic properties are, in most cases, performed by differential scanning calorimeter (DSC) [66–73]. Most self-assembled aggregates show phase-transition phenomena such as gel (crystal)-to-liquid crystal and gel (crystal)-to-sol, and thus physicochemical properties of the aggregates drastically change at their transition temperatures. Spectroscopic observations provide information on molecular orientations, packing states, and lateral diffusion behavior. UV-visible [32, 69, 73–76], Fourier transform infrared (FT-IR), circular dichroism (CD) [21, 22, 32, 73, 76–78], fluorescence [79, 80], and nuclear magnetic resonance (NMR) [81] spectrometers have been widely used for investigation and analysis on the molecular orientation. Typical investigations are found in chromophore-containing lipid systems. Shimomura et al. discussed on H- or J-aggregations with λ_{max}-shift of a UV-visible spectra [82]. Ihara et al. discussed chiral-stacking behavior among the sorbyl groups of lipids with both UV-visible and CD spectra [32]. They also discussed the photo-induced polymerization process by following the spectral changes. Hachisako et al. discussed the critical aggregation concentration with visible spectral change induced by the isomerization of the spiropiran-containing lipid [83]. Schnur et al. discussed the relationship between the molecular chirality of lipids and the helicity of aggregation morphology with CD spectra [84]. These suprastructural aggregates often provide specific binding behavior for guest molecules such as dyes. These phenomena can also be detected by UV-visible and CD spectra [76, 85–87]. Fluorescence spectra are helpful for knowing the microenvironment around lipids. Sagawa et al. reported excimer formation when a pyrenyl group-containing lipid forms highly ordered aggregates in organic solvents [80]. These typical spectral data are shown in Figures 6 and 12.

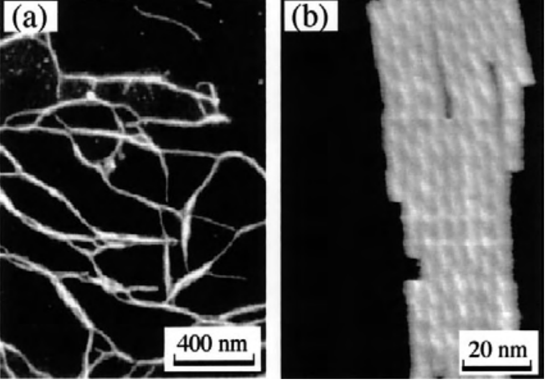


Figure 4. Morphological characterization of lithostathine protofibrils by tapping mode AFM in air (a), and in solution (b).

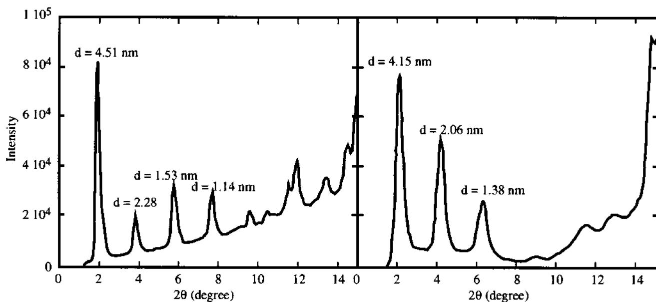


Figure 5. Powder XRD spectra of xerogels prepared from neat gels with the azobenzene-containing cholesterol derivatives [65]. Reprinted with permission from [65], J. H. Jung and S. Shinkai, *Journal of Inclusion Phenomena and Macrocyclic Chemistry* 41, 53 (2001). © 2001, Kluwer Academic Publishers.

3.2. Nanofibers in Aqueous Systems

Bio-membranes are spontaneously organized from many kinds of molecules such as phospholipids, proteins and polysaccharides. Some phospholipids (**N-1** ~ **N-4**) listed in Figure 7, which are representative of amphiphiles, form bilayer membrane structures spontaneously in water and their hydrophobicity is a major driving force in the aggregation and maintenance of the bilayer structures. These lipids usually form small globules and vesicles in water. A typical example is the vesicular structures observed when soybean-derived phosphatidyl choline is dispersed in water, which was reported in 1965 [88].

In 1977, Kunitake et al. reported in a landmark study that didodecyldimetylammonium bromide, as a totally synthetic lipid, could form bilayer structures in water [33]. Since this turning point, a large number of double-chain alkyl amphiphiles have been synthesized and characterized by many researchers. These findings led to the next step of lipid chemistry. Nobody doubts that the number of alkyl

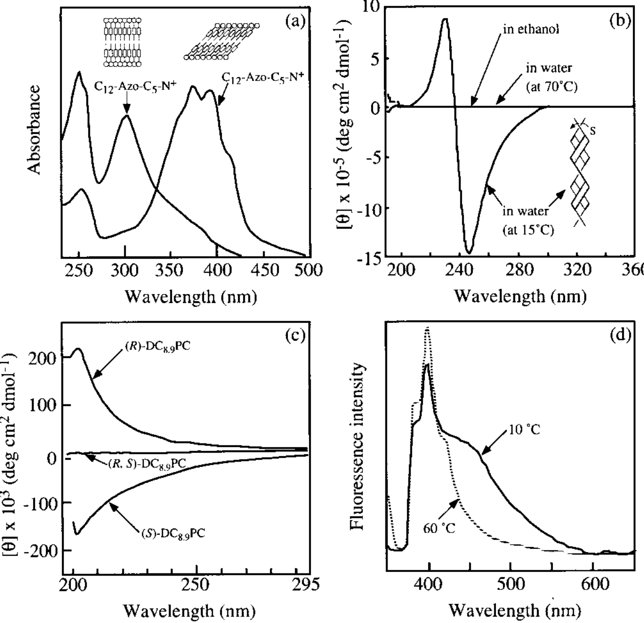


Figure 6. Typical spectral data for self-assembled nanofibers: (a) UV-visible spectra of the azobenzene-containing single alkyl-chain lipid in aqueous solution systems [82]; (b) CD spectra of the sorbyl group containing double chain-alkyl lipid in aqueous solution systems [32]; (c) CD spectra of 1,2-bis(tricosa-10,12-diynoyl)-sn-glycero-3-phosphocholines in aqueous solution systems [84]; (d) fluorescence spectra of the pyrene-containing double alkyl-chain lipid in organic solvent systems [80]. Reprinted with permission from [82], M. Shimomura et al., *Ber. Bunsenges. Phys. Chem.* 87, 1134 (1983). © 1983, Wiley VCH; from [32], H. Ihara et al., *Langmuir* 8, 1548 (1992). © 1992, American Chemical Society; from [84], J. M. Schnur et al., *Science* 264, 945 (1994). © 1994, American Association for the Advancement of Science; and from [80]. T. Sagawa et al., *Langmuir* 18, 7223 (2002). © 2002, American Chemical Society.

chains in a hydrophobic part is not directly related whether or not a lipid can form bilayer structures but its molecular shape and intermolecular interaction are rather important.

It has been recognized that bilayer membrane structures can be formed from single chains [20, 89], triple chains [90], and others [29]. Also a hydrophilic part is not within the specified structure. It has been reported forming bilayer membrane structures from amphiphiles with anionic and nonionic groups, as well as twitter ionic and cationic groups.

Through these investigations, in 1984, one of the authors found that special synthetic lipids (**B-1a**) with hydrophilic oligopeptide head groups can form helical or tubular structures in water (Figs. 8a and 8e) [21]. This finding is significant in light of the fact that the thickness of the aggregates corresponds to that of single-walled bilayer structures and that the tubular and helical forms are closely related. At the same time this report on similar tubes through lipid aggregation was published, Nakashima et al. reported that helical ribbon-like aggregates could be produced from L-glutamate-derived lipids but that these were much larger—in microsize rather than nanosize [42, 43]. Further investigation made it clear that the helical form was rather an intermediate to the tubular form and that the formation of fibrillar structures with helices and tubules is deeply related to their chiral properties [22, 24, 31].

Since 1984, many researchers have designed and synthesized chiral lipids that produce nanofibrillar structures in water. They can be roughly classified into two categories: (1) micellar-based aggregates, and (2) bilayer sheet-based aggregates. Figure 7 includes the chemical structures of synthetic lipids that can form nanofibrillar aggregates in water.

3.2.1. Micellar-Based Fibrillar Aggregates

It may be difficult to define a micellar system, but one well-known case of micellar-based aggregates from N-alkyl aldonamides (**B-6**) and a series of diastereomeric and enantiomeric N-octylaldonamides has been studied by Fuhrhop et al. since 1987 [91]. Their TEM microscopic observations indicated that the solubilities and aggregation morphologies of aldonamides depended directly on the stereochemistry of their polyol head groups (Fig. 8g). All antichain head groups, like D-Man for instance, of N-octylaldonamides-induced formation of sheet-like aggregates based on bilayer structures, and highly irregular chain (i.e., twisted chain) conformations, like D-Alt and D-Ido for instance, induced high water solubility with no consequent aggregation. On the contrary, cylindrical micellar aggregates were obtained from N-octylaldonamide with L-Glu, D-Glu, and D-Tal head groups. It was estimated that a moderate molecular bending at the head group induced high curvature in the aggregates. They proposed a model of fibrillar (cylindrical) micelles and then this arrangement was supported by ^{1}H-NMR spectroscopic measurements [92]. Furthermore, it was confirmed by detailed TEM observations and image-processing that micellar fibrils often produced multihelices by twisting themselves. N-Octyl and N-dodecyl-D-gluconamides formed self-organized quadruple helices with magic angles whose pitch is equal to $2\pi\times$ molecular bilayer diameter. Figure 8h shows its TEM image and a contour line diagram obtained by image analysis [93]. Here, each micellar fibril preserved its independence without fusion. A computer model of quadruple micellar fibrils is also shown in Figure 9.

In 1992, Imae et al. [94] reported that N-acyl amino acids formed fibrillar, distorted fibrillar, or twisted ribbon-like aggregates (Fig. 8i), and their aggregation morphologies were dependant upon the types of amino acids and pH of aqueous solutions. Transmission electron microscopic observations showed that N-dodecyl-L-aspartic acid and N-dodecyl-L-alanine formed distorted fibers (pH = 3.6–5.5, temperature < 12 °C) and cylindrical structures without distortion (pH < 6.0, ambient temperature), respectively. Since N-dodecyl-L-glutamic acid formed globular aggregates, it was estimated that the bulkiness of the head group may have influenced the aggregation morphologies. Similar micellar formation was observed in poly(amino acid)s with terminal dialkyl groups. Their aggregation morphology is influenced by the pH-dependant secondary structures of poly(amino acid)s as head groups (**B-1, 2, 3**). [22, 30, 31] The complex **B-9** is known as bolaamphiphiles, which form spherical micelles or vesicles in aqueous solutions as well as amphiphiles. **B-9** is an asymmetric bolaamphiphile with one amino acid head group and one ammonium chloride head group. The diameter of their fibers is 2.5 nm, which corresponds exactly to that of micellar monolayers. (Fig. 8j) [58].

a) Bilayer-forming natural lipids

N-1

N-2

N-3

N-4

b) Synthetic lipids forming fibriller aggregates

B-1a

B-1b

B-1c

B-2

B-3

B-4

B-5

B-6a (D-Glu)

B-6b (D-Man)

B-6c (D-Ido)

B-7

B-8

B-9

B-10 D-tartrate L-tartrate

B-11

B-12

B-13

B-14

c) Polymerizable bilayer-forming lipids

B-15

B-16

B-17

B-18

Figure 7. Chemical structures of aqueous bilayer membrane-forming chiral lipids (Ref.: **B-1a** [21, 24], **B-1b** [22, 24], **B-1c** [31], **B-2** [24], **B-3** [30], **B-4** [76], **B-5** [170], **B-6** [91, 92], **B-7** [94], **B-8** [58], **B-9** [94], **B-10** [95, 96], **B-11** [74], **B-12** [29], **B-13** [26], **B-14** [28], **B-15** [25], **B-16** [103], **B-17** [103], **B-18** [32]).

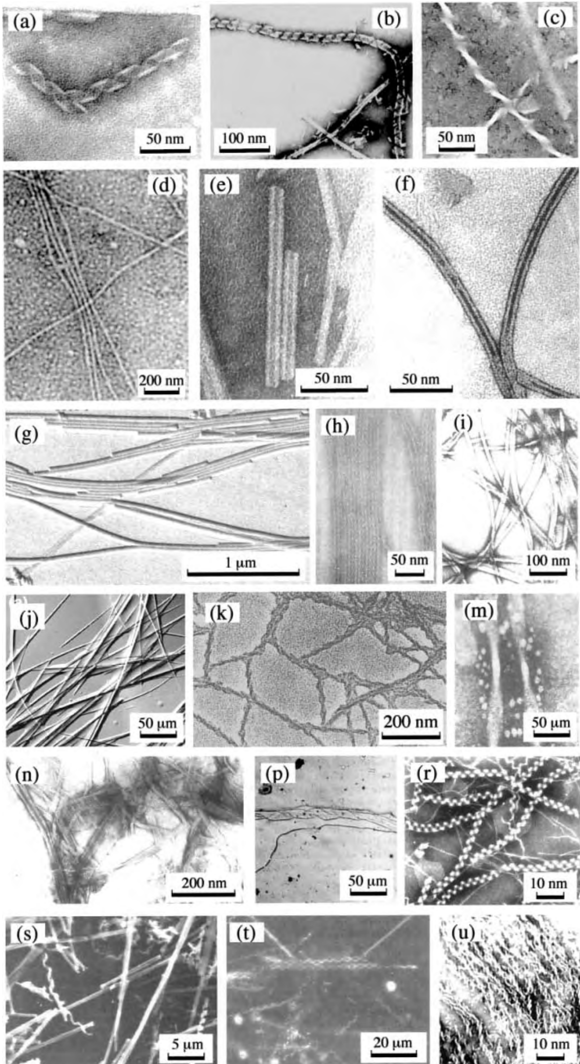


Figure 8. Various nanofibrillar aggregates from aqueous lipid membrane systems. (a) **B-1a** [21], (b) **B-3**, (c) **B-3** [30], (d) **B-1b** [22], (e) **B-1a** [21], (f) **B-3** [30], (g) D-talonamide derivative [27], (h) *N*-octyl-D-gluconamide [93], (i) *N*-dodecanoyl-β-alanine [94], (j) bolaamphiphile [58], (k) **B-10** [95], (m) **B-10** [96], (n) **B-12** [29], (p) **B-13** [26], (r) **B-14** [28], (s) **B-15** [25], (t) **B-16** [103], (u) **B-18** [32]). Reprinted with permission from [21], K. Yamada et al., *Chem. Lett.* 1713 (1984). © 1984, The Chemical Society of Japan; from [30], H. Ihara et al., *J. Chem. Soc. Jpn.* 1047 (1990). © 1990, The Chemical Society of Japan; from [27], J.-H. Fuhrhop et al., *J. Am. Chem. Soc.* 110, 2861 (1988). © 1988, American Chemical Society; from [93], J. Koning et al., *J. Am. Chem. Soc.* 115, 693 (1993). © 1993, American Chemical Society; from [94], T. Imae et al., *J. Am. Chem. Soc.* 114, 3414 (1992). © 1992, American Chemical Society; from [58], T. Shimizu, *Macromol. Rapid Commun.* 23, 311 (2002). © 2002, Wiley-VCH; from [95], R. Oda et al., *Nature* 399, 566 (1999). © 1999, Macmillan Magazines Ltd; from [96], R. Oda et al., *Angew. Chem. Int. Ed.* 37, 2689 (1998). © 1998, Wiley-VCH; from [29], N. Kimizuka et al., *Chem. Lett.* 29, (1990). © 1990, The Chemical Society of Japan; from [26], I. Cho and J. G. Park, *Chem. Lett.* 977 (1987). © 1987, The Chemical Society of Japan; from [28], H. Yanagawa et al., *J. Am. Chem. Soc.* 111, 4567 (1989). © 1989, American Chemical Society; from [25], J. H. Georger et al., *J. Am. Chem. Soc.* 109, 6169 (1987). © 1987, American Chemical Society; from [103], T. Kunitake et al., *Macromolecules* 22, 3544 (1989). © 1989, American Chemical Society; and from [32], H. Ihara et al., *Langmuir* 8, 1548 (1992). © 1992, American Chemical Society.

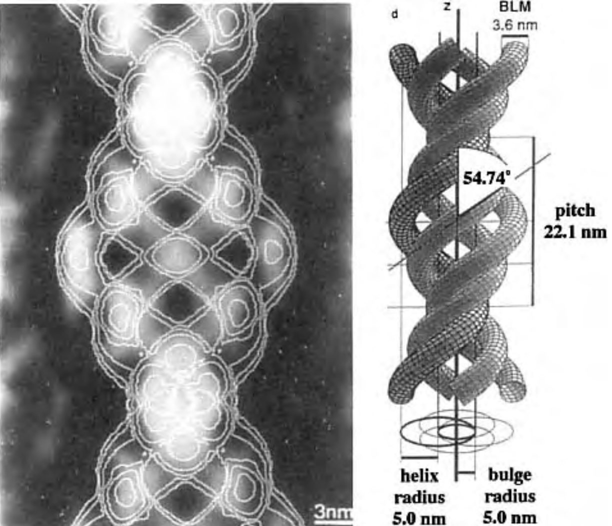


Figure 9. A computer model of quadruple micellar fibrils [93]. Reprinted with permission from [93], J. Koning et al., *J. Am. Chem. Soc.* 115, 693 (1993). © 1993, American Chemical Society.

3.2.2. Bilayer Sheet-Based Fibrillar Aggregates

So far, a large number of bilayer-forming lipids have been reported, including some that can form helical and twisted ribbon-like aggregates. Their assemblies are based on distorted sheet-like bilayer membranes with large curvature and are morphologically divided into helical and tubular structures with cylindrical curvature and twisted ribbons with Gaussian, saddle-like curvature.

The first observation was reported in 1984. Didodecyl L-glutamide derivatives with oligo(L-glutamic acid) head groups (**B-1a**) formed helical or tubular aggregates [21]. Transmission electron microscopic observations showed that the **B-1a** formed helical aggregates (Fig. 8a) immediately after dispersion and then grew to tubular aggregates (Fig. 8e). It was confirmed that the tubular aggregates were 5–6 nm thick, which corresponds to the thickness of single-layered bilayer membranes (two molecular lengths) of lipid **B-1a**, and their diameters were 20–25 nm. On the other hand, when the head group is composed of the corresponding DL-glutamic acid, only large sheets without curvature were produced [24].

In addition, when oligo(L-aspartic acid) was used instead of oligo(L-glutamic acid) as the head group of **B-1b**, fibrillar and twisted ribbon-like aggregates were formed (Fig. 8d) [22]. Through these studies, we were also able to reveal the process of morphogenesis of these suprastructural aggregates by TEM observation and light scattering measurement:

1. **B-1b** formed globular aggregates with the largest curvature at the initial step;
2. these globules grew to fibrillar aggregates;
3. double or multiple strands were produced and then fused themselves to make ribbon-like aggregates.

These findings indicate that even a very small difference between L-glutamic and L-aspartic acid in the head group strongly influences the morphology of the aggregates. This

means that aggregation morphology is also very sensitive to temperature and pH factors because these can drastically change the secondary structures of the head groups.

Oda et al. described the difference between helical (and tubular) and ribbon-like aggregates formed from charged gemini surfactants **B-10** with chiral counterions (Figs. 8k and 8m) [95, 96]. In these studies, helical aggregates were formed from long-chain (C_{18}) surfactants, whereas twisted ribbon-like aggregates were formed from short-chain (C_{14} or C_{16}) surfactants. The degree of twist and the pitch of the ribbons could be controlled by the introduction of opposite-handed chiral counterions in various proportions.

In 1985, Helfrich and Harbich [97] attempted to define a theoretical approach for the morphologies of helical and twisted aggregates. An explanation of these phenomena, Oda et al. demonstrated their own theoretical approach using the extended Helfrich model [98–100]. Other researchers have endeavored to solve a riddle of being between geometrical morphologies and chemical structures [101, 102]. So far, no consensus has been established, because supramolecular assembly-forming compounds include many different chemical structures. Several questions still remain; for instance, the nature of the edge of the bilayers has not been clarified. Future work in both experimental and theoretical studies will develop technologies for order-made self-assembling nanofibers.

Two common features necessary for the chemical structure of molecules that form supramolecular assemblies are a chiral carbon atom and moieties feasible for intermolecular interactions. Artificial chiral lipids that can form supramolecular assemblies such as helical, tubular, and ribbon-like aggregates based on bilayer membrane structures are listed in Figure 7. This includes single chain (**B-11**) [74], quadriplex chain (**B-12**) [29], and cholesterol group (**B-13**) [26] as hydrophobic part, and nucleotide group (**B-14**) [28] as a hydrophilic part. Moreover, polymerizable moieties such as diacetyrene group (**B-15**) [25], acrylate group (**B-16, 17**) [103], methacrylate (**B-13**) [26], and sorbate group (**B-18**)-containing lipids [32] have been also reported. Figure 8 includes their typical TEM images.

3.2.3. Non-Lipid Based-Nanofibers

The preceding section discussed special nanofiber structures formed by hydrophobic aggregation and by intermolecular interaction with polar groups. Here, is described the formation of nanofibrillar structures through multiple hydrogen bonding and, consequently, molecular orientation (Figure 10).

On the other hand, the construction and structural analyses of aggregates caused by β-structures are also performed from the viewpoint of protein chemistry [104–108]. For example, prion proteins from patients with Alzheimer's disease or Bovine Spongiform Encephalopathy (BSE) form a different fiber-like aggregate [109–111] than prion proteins from a healthy body, and these proteins include β-structure [112–114].

Although Scrocchi et al. did not aim to produce nanofibers, they chose an arrangement of 6 out of the 37 residues in the human islet amyloid polypeptide (hIAPP) amino acids sequence (**P-1**), and specified the sequence that

P-1 (hIAPP)
1 10 20 30 37
KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY
SNNFGA

P-2 **P-3** **P-4** **P-5** **P-6** **P-7**

Figure 10. Chemical structures of peptides and other polymers that can form nanofibrillar aggregates (Ref.: **P-1** [109], **P-2** [115], **P-3** [111], **P-4** [192], **P-5** [191], **P-6** [187], **P-7** [198]).

makes hIAPP β-formed [109]. Consequently, it was found that the hexapeptides of hIAPP (residues 20–25) promote β-transition of IAPP most strongly, and were seen to produce nanofibers (Fig. 11a). This was especially interesting as it showed that an amino acid sequence of only six residues could serve as a trigger for proteinic β-form induction to

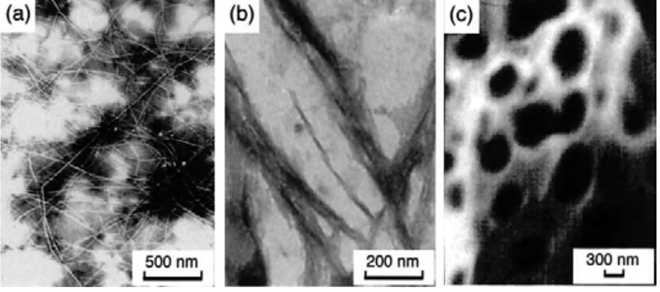


Figure 11. TEM images of (a) **P-1** fibrils after incubation with SNNFGA [109], (b) **P-2** fibrils obtained from an aqueous solution with 200 eq [115]. NaCl and (c) **P-3** aggregates formed from a chloroform solution [111]. Reprinted with permission from [109], L. A. Scrocchi et al., *J. Mol. Biol.* 318, 697 (2002), © 2002, Academic Press; from [115], T. Sakurai et al., *Chem. Lett.* (in press). ©—, The Chemical Society of Japan; and from [111], R. Jayakumar et al., *Langmuir* 16, 1489 (2000). © 2000, American Chemical Society.

form a nanofiber. However, nanofibers of β-structure origin have extremely low solubility and do not dissolve easily. Therefore, they are not suitable for use as nanostructure materials.

To overcome this hurdle, construction of water-soluble nanofibers from β-structures is now being studied [115]. We have noticed that when sequential oligopeptides with alternating hydrophilic and hydrophobic residues form β-structures, they may also be able to form amphiphilic structures. As an example, a Lys-Leu alternating sequential oligopeptide (**P-2**) was synthesized, and it was confirmed that helical tape-like aggregates were formed in the solution. Transmission electron microscopic images showed that the width of the thinnest portion was approximately 12 nm (Fig. 11b). Moreover, when fluorescence spectra of pyrenyl chromophores at the terminal end of the main chain were investigated, we found excimer formation that demonstrates orientation between the pyrenyl chromophores at the time of nanofiber formation. On the other hand, neither nanofiber formation nor excimer generation was observed in random copolypeptide.

Similar formation of aggregates from β-structures of oligopeptides (**P-3**) was reported by Jayakumar et al., who observed that L-Ileu pentapeptides form aggregates in toluene and DMF (Fig. 11c) [111]. In this case, however, they were insoluble and it was not specified whether nanofiber structure was formed.

3.2.4. Chiral Properties of Suprastructural Bilayer Membranes

Most nanofibrillar assemblies based on micells and bilayer membranes possess chiral centers in their structures. Of course, the chirality of molecules plays an important role in the production of distorted supramolecular assemblies. We have observed that oligo(L-glutamic acid)-containing lipids (**B-1a**) produce nanofibrillar aggregates including helices and tubules but that corresponding DL-derivatives produce only large sheet aggregates [24]. Similar observations have been made for other lipids (**B-6** and **10**): enantiomeric compounds form right-handed and left-handed helical aggregates, and equimolar mixtures (= racemates) often induce drastic morphological change [27, 95]. It is considered that the molecules in the distorted aggregates pack with slight angle with their neighbors, and this feature of molecular orientation produces some unique properties. **B-3** [30], in which dialkyl chains and pyridinium moieties are connected by L-glutamide through three amide groups, formed helical and tubular aggregates. Here, DSC measurement showed two endothermic peaks (with top temperatures of 33 °C (T_1) and 45 °C (T_2), respectively) in the heating process [30]. These aggregates showed extremely strong exciton coupling at wavelengths near the absorption of amide groups and pyridinium groups. The optical activity of amide groups changed from negative to positive around T_1 and optical activities of both groups disappeared above T_2 (Fig. 12a). Neither formation of tubular aggregates nor enhanced CD spectra were obtained from the corresponding ester-type compound at which dialkyl chains were introduced through ester bonds. These results indicate that the three amide bonds around the chiral carbon play an important role in the formation of supramolecular assemblies. The reversal of chirality is probably due to change of the molecular orientation, for instance, from S-chiral to R-chiral. Similar induced CD spectra based on chirally ordered structure have been observed in bilayer aggeregates formed by chromophoric group-containing lipids (**B-11** and **B-18**).

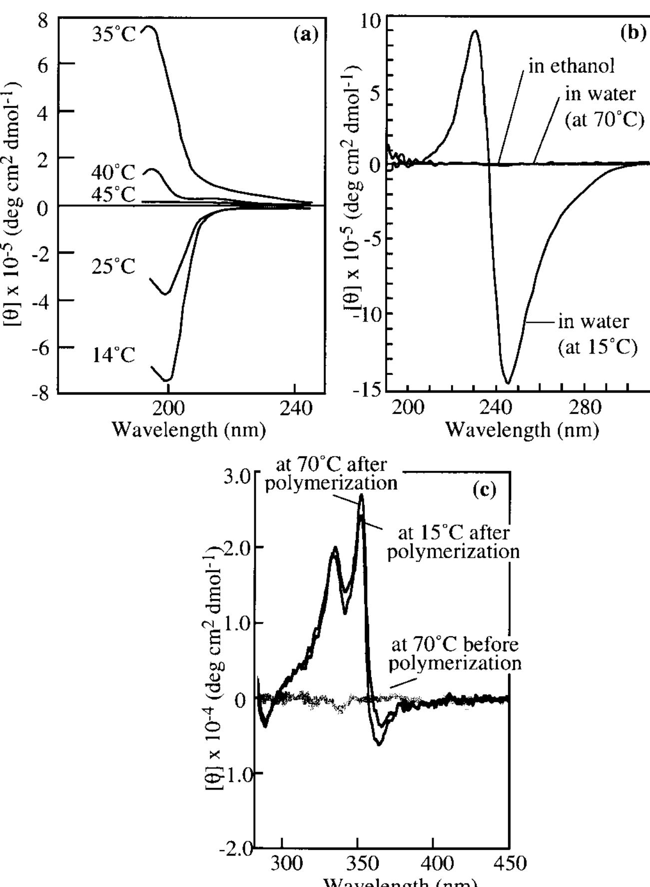


Figure 12. Specific optical activities in (a) **B-3** [30] and (b) **B-18** aggregates in aqueous solutions [32], and (c) **G-11** organogels [163]. (a) The optical activity changes reversibly depending on temperature. (b) The CD pattern indicates that the sorbyl groups as polymerizable moiety are in chirally stacking states among them at a temperature below T_c. (c) Polymerization of methylmethacrylate as a bulk solution promotes stabilization of the highly oriented structure of **G-11** aggregates [163]. Reprinted with permission from [30], H. Ihara et al., *J. Chem. Soc. Jpn.* 1047 (1990). © 1990, The Chemical Society of Japan; and from [32], H. Ihara et al., *Langmuir* 8, 1548 (1992). © 1992, American Chemical Society.

Mirror-image of CD spectrum was obtained in bilayer membrane aggregates from D- and L-form lipids, respectively [84]. Specter et al. [116] reported on the CD spectra and TEM observations of diacetylenic phospholipid in various conditions. The CD spectra of R- and S-enantiomers showed similar patterns but opposite signs and the CD signal of racemic mixture showed approximately nothing. Moreover, TEM observations showed that the R- and S-enantiomers formed tubular aggregates with the same dimensions, and their racemates also formed tubular aggregates without dimensional changes. Since the tubular aggregates of pure enantiomers were grown from

left- and right-handed helical aggregates, respectively, it was anticipated that racemic mixture almost phase-separated.

It is well known that achiral dye molecules bound to chiral poly(α-amino acid) show CD spectra (induced circular dichroism (ICD)) around their absorption bands [117–119]. Similar phenomena have been observed on chiral bilayer assemblies. When an achiral dye molecule, such as methyl orange, was added to **B-4** aggregates in an aqueous solution, it combined with the cationic membranes to give three dispersion species, belonging to S-chirally oriented dimmers on the surface, monomers on the surface, and monomers incorporated into the bilayers, respectively [76]. Many chiral lipids possess similar functions through formation of highly oriented structures [85–87].

3.2.5. Stabilization of Fibrillar Assemblies

Compared with nano assemblies from lipids, carbon nanotubes are useful in light of their physical and chemical stability, but this stability causes limitations on modification. The assembling systems, however, provide unlimited opportunities for modification. For example, the mixing of L- and D-isomers of **B-1a** causes their racemates to show absolutely different morphologies from their original forms. **B-1b** ($n = 4$) can form only fragmented fibrillar aggregates spontaneously, but this morphology can be transformed to well-developed twisted ribbon-like aggregates by mixing with a triple chain-alkylated, lipid-forming tubular aggregates [32]. However, these morphological transformations mean to be so weak for their industrial use. Therefore, it must be valuable to stabilize their aggregates. This attempt has been carried out by physical and chemical methods. The chemical methods are roughly classified into three types as summarized in Figure 13. However, most of these methods include polymerization, and, thus, the resulting bilayer membranes lose some functions such as molecular fluidity instead of obtaining stability. Of course, stabilization is required in some cases to expand their applications.

Polymerizable lecithins (**B-15**) with diacetylenic alkyl chains such as 1,2-bis(10,12-tricosadi-ynoyl)-sn-glycero-3-phosphocholine ($DC_{8,9}PC$) are known to form tubular microstructures when their liposomes are cooled below the chain-melting transition [25]. Tubular and helical structures from **B-15** measured from 0.3 to 3 μm in diameter and from

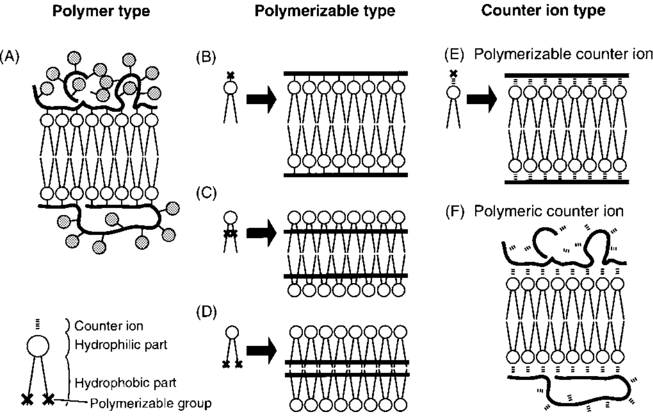


Figure 13. Stabilization of bilayer aggregates by polymerization.

5 to over 1000 μm in length. Both dimensions depended on the solvent system (ratio of water-to-alcohol). Irradiation with UV rays or γ-rays induced photopolymerization of diacetylenic units, but the helical or tubular structures were lost. A series of monoacryl and diacryl double-chain ammonium amphiphiles (**B-16** and **17**) were synthesized for investigation of their dispersion characteristics before and after photopolymerization [103]. **B-16** formed vesicular aggregates after dispersion into water and transformed into helices when allowed to stand at room temperature for 1 month. Polymerization of acryloyl groups leads to destruction of aggregation morphologies.

On the other hand, we reported the dispersion states and effect of polymerization on the aggregation morphology from polymerizable lipid (**B-18**) [32]. The introduction of polymerizable sorbyl groups at the end of the hydrophobic alkyl chains is far enough from the chiral carbon in order to prevent obstruction of the chiral interactions between lipids. **B-18** formed single-wall helical bilayer structures (Fig. 8u) with large chirality (Fig. 12b) in aqueous solutions below T_c (= 51 °C). Ultraviolet irradiation of the **B-18** aggregates caused photoreaction of the sorbyl groups at a rate 25 times faster in the stacked species (below T_c) than in the nonstacked species. This photoreaction was accompanied by morphological transitions from helical aggregates to tubular aggregates below T_c and to twisted fibrillar aggregates above T_c. Almost no morphological change was observed as a result of this polymerization over a wide range of temperatures.

3.3. Nanofibers in Organic Media

Hydrogels can often be formed from aqueous solutions of hydrophilic polymers, biomolecules such as proteins, and inorganic materials such as silicates. They have been studied in detail, and related books [120, 121] and reviews [122] have been published by many researchers. Some of these gels and gelators are widely used in industries, food science, and cosmetic science. Recently, it has been found that some specia, low molecular weight compounds formed gels in organic media [123]. These can be referred to as organogels (organic gels). Organogels are very unique, not only in that the gelation is induced by three-dimensional network formation with well-developed fibrous aggregates, but also in that these aggregates are on the basis of highly oriented structures like aqueous lipid bilayer membranes. Therefore, they attract our interest in spite of their instability, and we further label them "self-assembled organogels" to distinguish them from conventional gels. Self-assembled organogels include various fibrillar aggregates such as rods, helices, and sheets, and the challenge to stabilize their morphologies and molecular orientation to widen their applications is now being met. In this section, we focus on nanofibrillar aggregates formed in organic solvents.

Low molecular weight compounds that can produce gels from organic solvents have been known for 50 years. However, the gelation of organic solvents and oils as macroscopic phenomena has been a main subject of interest for several decades thereafter. From the late 1980's to the early 1990's, several kinds of organogels were discovered simultaneously. In this period, many researchers joined the research field

of nano-sized molecular architectures to develop molecular devices for supersensitive sensors, high-density memory storage, and so on. It appears that a point-of-view toward organogels has been shifted into explication and control of molecular buildings formed from low molecular weight organogelators, spontaneously. The original root compounds of organogelators (Fig. 14) have been discovered serendipitously by researchers who were working in various research fields [70, 124–127]. Therefore, the chemical structures of most organogelators are derived from intermediate molecules designed for specific functions.

3.3.1. On Driving Forces and Chemical Structures

A hydrophobic effect is the most essential driving force for molecular aggregates in aqueous solution systems but almost disappears in organic media. More positive intermolecular interactions play an important role in molecular aggregation in organic media. Hydrogen-bonding interaction is especially effective and many organogelators are classified into this category. 12-Hydroxystearic acid (12HSA, **G-6**) and its related salts were early examples of organogelators [128–130] and the first report of 12HSA was brought by Tachibana et al.

G-1 Cholesterol-based compound

G-2 PAH-based compound

G-3 Perfluoroalkyl-based compound

G-4 Peptide-based compound

G-5 Saccharide-based compound

Figure 14. Root organogelators developed in the late 1980's to the early 1990's **G-1** [124], **G-2** [125], **G-3** [126], **G-4** [70], **G-5** [127].

in 1965. Lithium 12HSA forms twisted fibrous aggregates, whose direction of the twist depends on the optical isomerization of lithium 12HSA. Transmission electron microscopic observations show that the direction of the twist is left-handed for the L-form, but right-handed for the D-form, and that no twisted fibers are obtained for the DL-form. Furthermore, the helical structures formed from 12HSA are sensitive to counter metal ions. 12HSA forms developed ribbon-like aggregates with diameters of 10 to 100 nm in soybean oil. However, the number of cross-sections is relatively few according to SEM observations. Addition of sodium salt of 12HSA to these organogels contributes to improvement of gelation ability with metamorphoses from flexible ribbon-like aggregates to bundled rigid fibers with equal diameters. Terech et al. elucidated the oriented structure of 12HSA organogels at the molecular level by a combination of several techniques [131]. Small-angle neutron scattering (SANS), small-angle X-ray scattering (SAXS), and wide-angle X-ray scattering (WAXS) investigations revealed the structural model of the 12HSA fibrous or twisted, strand-like aggregates in organic media. The 12HSA molecules form head-to-head aggregates in fibrous networks and the growth of fibrous aggregate in the fiber axis direction is promoted by hydrogen bonds between the hydroxyl groups of 12HSA. A polarity of solvents influences the diameter of fibrous aggregates since the laminating of fibrous aggregates in the vertical direction is brought by dipole-dipole interactions between carbonyl groups of 12HSA. Microscopic morphological arrangements of fibrous aggregates inevitably lead to structural changes of cross-section and/or junction zones, and finally exterior transformations such as transparency (turbidity) and other physical properties will appear.

Peptide-based derivatives will be useful as organogelators because their amide bonds work as a stronger driving force for molecular aggregation. Peptide-based organogelators have a plural number of hydrogen bondable moieties. **G-4** [70, 73] as a typical example, possesses three amide bonds around an L-glutamic moiety, which works as a good organogelator. Transmission electron microscopic and SEM observations showed a three-dimensional network with fibrillar aggregates in its organogel and xerogel (Figs. 15a and 15b, respectively). The minimum diameter of the aggregates in the picture is 20 nm, which is 2–3 times larger than the molecular length estimated by SAXS [73]. However, if two of the three amide bonds are replaced by the ester bonds, no gelation is observed even when their concentration is 10 times higher than the former. It was also confirmed that addition of trifluoroacetic acid as an inhibitor for hydrogen bonding causes gel-to-sol transition. MOPAC calculation indicated that the three amide bonds around the L-glutamic acid moiety provided a proper conformation for intermolecular interaction.

K. Hanabusa et al. investigated the gelation abilities of single-chain alkylated mono- or dipeptides toward general organic solvents [132–134]. Peptides containing L-isoleucine and L-valine (**G-7**) work as distinguished gelators for various organic solvents such as alcohols, ketones, esters, dimethyl formamide (DMF), dimethyl sulfoxide (DMSO), and aromatic compounds. Both the length of the alkyl chain and the side chain structure of amino acids are important in the

Figure 15. Chemical structures of various low-molecular mass organogelators (Ref.: **G-6** [128], **G-7** [134], **G-8** [157], **G-9** [169], **G-10** [213], **G-11** [80], **G-12** [172], **G-13** [83], **G-14** [135], **G-15a** [63], **G-15b** [63], **G-16** [138], **G-17**[215], **G-18** [78], **G-19** [147], **G-20** [123], **G-21a** [216], **G-21b** [176], **G-21c** [173], **G-22a** [177], **G-22b** [166], **G-22c** [166], **G-23** [214], **G-24** [151], **G-25** [158], **G-26** [159]).

production of fibrous aggregates. If a hexyl chain is introduced instead of an octadecyl chain, the compound neither produces organogels in solvent nor forms fibrous aggregates. The possible structure for the long chain-alkylated dipeptides was proposed to be an antiparallel β-sheet through intermolecular hydrogen bonds between the amide and urethane groups.

Cyclohexane diamide derivatives have hydrogen bonding sources [135, 136]. Although the two isomers of *trans*-1,2-cyclohexane diamide derivatives (**G-14**), which are (1R, 2R) and (1S, 2S), can produce a gel for many organic solvents, *cis*-isomer derivatives have no gelation ability. Furthermore, dialkylated 1,3- and 1,4-cyclohexane diamide derivatives did not produce organogel. It was considered from their aggregation morphologies that one-dimensional growth of molecular assemblies through intermolecular interactions was indispensable to the formation of fibrous aggregates. The alkyl chains connected to the *cis*-isomer are spread spatially since they are located at axial and equatrial positions, respectively. The molecular packing formation in the aggregates is also important for one-dimensional growth and this can also apply to 1,3-isomers and 1,4-isomers.

A two-component gelation system revealed essential information about the gelation mechanism. A pair of compounds, **G-15a** and **G-15b**, was designed as a two-component system because a dumbbell-type cocrystal was produced from a barbiturate derivative and triaminopirimidine through the intermolecular hydrogen bonds [63]. The heating process normally carried out to dissolve an organogelator into the solvents is unnecessary to produce organogels in chloroform with the combination system of **G-15a** and **G-15b**. Molecular orientation of the two-component system was supported by FT-IR and SAXS measurements. Transmission electron microscopic observations show the twisted fibrous aggregates in organic solvents. Since each molecule cannot form an organogel individually, it is understandable that the arrangement of the hydrogen bond-forming moiety and bulkiness of the molecule are important in inducing self-assembled organogels. This argument is supported by the results of dendritic two-component gels. Dendrimetric peptides with carboxylic acid form organogels with a linear aliphatic diamine in nonhydrogen bonding solvents [137].

Saccharide-based lipophilic derivatives (**G-5** and **G-16**) can be organogelators. Hafkamp et al. reported that gluconamide derivatives (**G-16**) form stable organogels in a number of solvents such as *o*-xylene, chloroform, and ethyl acetate [138]. Transmission electron microscopic observations showed that **G-16** formed fibers or bundled fibers. Studies of gluconamide derivatives were carried out to explicate gelation mechanisms. **G-16**, whose free hydroxyl groups were protected by bismethylene, was synthesized to confirm the effects of free hydroxy groups on the formation of organogels. Since these compounds did not form organogels in any solvents, it was clarified that the hydroxyl groups played an important role in the gelation.

Since saccharides have many hydrogen-bonding sources, they are, like amino acids, invaluable materials for design of organogelator molecules. Friggeri and Gronwald et al. have reported 11 methyl 4,6-benzylidene derivatives of the monosaccharides D-glucose, D-mannose, D-allose, D-altrose, D-galactose, and a-D-isose (**G-18**) [139–141]. The driving force of aggregation was estimated as intermolecular hydrogen bondings by FT-IR and ^{1}H-NMR spectroscopies. Gelation ability was evaluated in a large number of organic solvents, and it was found that insignificant differences among the chemical structures are effective decisively on their gelation ability. Furthermore, a remarkable change in molecular mobility around T_{gel} was observed by the line-broadening effect of ^{1}H-NMR. Structural details in fiber network in the solid state and gel state were deduced by ^{1}H-NMR, FT-IR, and SAXS. The series of these sugar-based derivatives is useful to construct architectures of molecular assemblies as base compounds.

There are some derivatives that can produce organogels without intermolecular hydrogen-bonding interaction. A typical example is the steroid derivative with polyaromatic group (**G-1**), first reported in 1987 by Weiss et al., who investigated the gelation ability of the isoandrosterone derivatives [124, 142–144]. The kinetics were investigated in detail by using electron paramagnetic resonance (EPR), SANS, infrared (IR), and CD spectroscopies. After this finding, more than 40 derivatives, including steroid and aromatic groups, have been reported. They are sometimes classified by the abbreviations ALS, where A, L, and S correspond to aromatic (A), linking (L), and steroid (S) groups, respectively. Effects of chemical structures of each part on the gelation were investigated in detail; for instance

1. stereo-chemistry at C-3 and the nature of the chain at C-17 of the steroidal part,
2. various aromatic groups such as 9,10-anthraquinones, cinnamate, 2-naphtyl, 1-pyrenyl, phenyl, and their substituted compounds,
3. the length and functionality of the linking groups.

Several important considerations for organogelators from aromatic compound-linked steroid were concluded as follows [145]:

1. H-bonding, even when possible, may be absent in low molecular-mass organic gelator (LMOG) assemblies when other packing contributions (e.g., π-π interaction and London dispersion forces) dominate [146].
2. Charge-transfer interactions within gelator strands can stabilize gels [147].
3. Thixotropy can be induced by adding a small concentration of a second (nongelling) ALS molecule whose size and shape are similar to those of a good ALS gelator.
4. The fraction of ALS gelator within the solid network is dependant on temperature and on the solubility of the gelator in the liquid component [143, 144].
5. The bulk properties of a liquid mixture, rather than the properties of the individual components, determine the dimensions and shape of the gelator assemblies [142], as well as the relative ordering of T_{gel} values within a series of ALS gelators.
6. Subtle changes in molecular shape can alter profoundly the ability of an ALS to gel organic liquids.

Inductions of functional groups such as crown ether and azo-benzene into cholesterol were performed by Shinkai et al. [175–177]. Application of these organogelators are described in Section 3.3.4.

We estimate that a steroid group will give limited solubility to organic solvents compared to a long chain alkyl group and thus may work as a sorbophobic moiety. Positive interaction can probably be induced by polyaromatic groups. Supporting this, anthryl derivatives connected to dialkyl chains have been studied as organogelators [125, 148–150] 2,3-Bis-*n*-decyloxyanthracene, **G-23**, produced organogels in many organic solvents [125]. Since 2,3-dialkoxynaphthalene showed no ability of gelation for any organic solvents, increasing aromaticity promoted gelation ability. Even if the anthracene moiety was replaced by anthrax quinone and phenazine, effective gelation was observed. The number and length of alkyl chains were also sensitive to gelation abilities. Freeze-fracture electron micrographs of **G-23**-propanol gels indicated a three-dimensional network of fibrous rigid bundles with 60–70 nm diameters.

Some organometallic compounds can also be organogelators, in which case the essential interaction source is surely derived from the chelation behavior. Terech et al. reported that mononuclear copper β-diketonate derivatives **G-24** formed both organogels [151] and thermotropic disc-like mesophases in cyclohexane [152, 153]. Some binuclear copper tetracarboxylates [154–156] form neat thermotropic columnar mesophases and organogels at less than 1 wt% concentrations in hydrocarbons. These organogels form disc-like molecules including semi-rigid, rod-like threads whose diameter was 1.7 nm, which corresponds to the diameter of disc-like molecules.

At the end of this section, chemical structures and aggregation forms of low-molecular organogelators are summarized in Fig. 16 and Table 1.

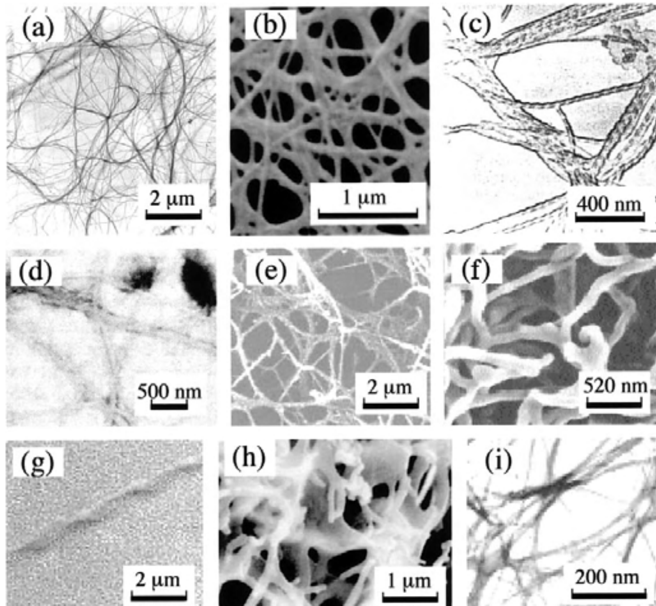


Figure 16. Fibrillar networks formed from various low-molecular mass organogelators (a) **G-4** [70], (b) **G-4** [73], (c) **G-20** [123], (d) **G-7** [134], (e) **G-25** [158], (f) **G-17** [215], (g) **G-5** [127], (h) **G-21c** [173], (i) **G-26** [159]. Reprinted with permission from [123], P. Terech and R. Weiss, *Chem. Rev.* 97, 3133 (1997). © 1997, American Chemical Society; from [134], K. Hanabusa et al., *J. Chem. Soc. Chem. Commun.* 390 (1993). © 1993, Royal Society of Chemistry; from [158], M. de Loos et al., *J. Am. Chem. Soc.* 119, 12675 (1997). © 1997, American Chemical Society; from [215], K. Yoza et al., *J. Chem. Soc., Chem. Commun.* 907 (1998). © 1998, Royal Society of Chemistry; from [173], T. Ishi-i et al., *J. Mater. Chem.* 10, 2238 (2000). © 2000, Royal Society of Chemistry; and from [159], M. Masuda et al., *Macromolecules* 31, 9403 (1998). © 1998, American Chemical Society.

3.3.2. Chemical Stabilization of Organogels

Some lipophilic peptides have been investigated not only as organogelators but also with respect to their self-assembling behaviors. This latter property gives them distinct advantages compared with conventional gel systems. The fibrillar aggregates are based on highly ordered structures, and thus show aqueous lipid membrane-like behaviors such as phase transition, phase separation, and chirality enhancement through molecular orientation. Although these features are advantageous for extended applications, it is also clear that their thermal and mechanical instabilities are a disadvantage in some application fields.

Some approaches have been proposed for stabilization of organogels. Introducing polymerizable group into a gelator is a reasonable method for this purpose. The first example of polymerizable organogelator was a sorbyl group-introduced peptide (**G-8**) in 1995. In this case, significant stabilization was not observed by photo-induced cross-linking among the peptides because oligomerization was a predominant reaction in the process [157]. On the other hand, de Loos et al. reported a bis(ureido)cyclohexane derivative containing a methacrylate moiety as a polymerizable organogelator (**G-25**) [158]. This compound produces organogels with developed fibrous aggregates in various organic solvents. Gel formation was maintained after polymerization by UV irradiation in the presence of a photo-initiator and the resultant gel showed highly thermal stability up to temperatures above the boiling point of the solvents. Polymerization of photo-induced polymerizable groups containing organogelators (**G-26**) was demonstrated by Masuda et al [159, 160]. Diacetylene containing organogelator (bolaamphiphile) was used for the purpose of stabilizing. Polymerization could be monitored by UV spectroscopic observations, and was induced by photo or γ-ray irradiation. In each case, stability of the organogel preserving fibrous aggregates rose in several samples after polymerization. However, reversibility of sol-to-gel transition and most properties based on molecular fluidity disappeared.

Sometimes, metal ions increase the mechanical strength of organogels. 1-O-(p-Aminophenyl)-4,6-O-benzylidene-α-D-glucopyranoside (**G-18**) behaves as a good gelator for various organic solvents [78, 139]. The T_{gel} values for ethanol gel were markedly improved by addition of $AgNO_3$, $CoCl_2$, or $CdCl_2$. The T_{gel}s for 1 wt/vol% of organogel in ethanol are -10 °C and 71 °C, respectively, in the absence and presence of equimolar $CoCl_2$. This remarkable change is due to cross-linking of **G-18** molecules by Co(II)-amino group interaction. It seems that hydrogen bonds and coordination bonds work cooperatively for reinforcement of organogels. Similar observations were obtained using a diketonate ligand-containing organogelator [161]. Maitra et al. reported that the donor-acceptor interaction promoted

Table 1. Classification of self-assembling organogelators.

Ref.	Compound	Year	Structual unit										Driving force for aggregation	Aggregates		Notes
			Nonpolar moiety					Polar moiety								
			sac	dac	mac	cho	pah	Acidic	Basic	Saccharide	Peptide	Amide urea		Diameter (nm)	Form	
[128]	L, D	1965	*					*					hb	10–100	tr	**G-6**, Figure 15
[63]	1 + 4	1993	*	*					*				hb	80	msf	**G-15a** + **G-15b**, Figure 15
[134]	6	1993	*								*		hb	100–300	tr	**G-7**, Figures 15 and 16d
[217]	6	1994	*				*						hb, sp	8–10	ssf, msf	**G-3**, perfluoroalkyl-containing, Figure 14
[138]	4	1997	*							*			hb, mc	ca. 40	hr	**G-16**, Figure 17a
[159]	1	1998	*							*			hb	3–20	ssf, msf	**G-26**, polymerizable, Figure 15 and 16i
[218]	3	1999	*				*						hb	15–40	hr	bolaamphiphiles
[219]	1	1999	*								*		hb	—	ssf	
[220]	1	2000	*									*	hb	30–100	msf	
[221]	1	2000	*				*						sp, pp	6–25	ssf, msf	stilbene-containing
[222]	2	2001	*								*		hb	30–100	msf	
[223]	1	2001	*					*			*		hb	0.5–15μm	ssf	
[224]	9b	2002	*										hb, sp	10–12	ssf	cholate-containing
[225]	6	2002	*				*	*				*	hb, pp	50–70	ssf	
[125]	DDOA	1991		*			*						pp	60–70	msf	**G-2**, anthracene-containing, Figure 14
[70]	2	1992		*			*				*		hb	—	—	azobenzene-containing
[226]	5a	1995		*							*		hb	3	ssf, msf	**G-8**, polymerizable, Figure 15
[136]	4, 5	1996		*								*	hb, pp	20–80	msf	
[135]	1, 2	1996		*								*	hb	40–70	tr	**G-14**, Figure 15
[83]	2	1997		*			*				*		hb	—	ssf, msf	**G-13**, piran-containing, Figure 15
[227]	2	1997		*								*	hb	50	msf	**G-25**, polymerizable, Figures 15 and 16e
[213]	2	1999		*			*				*		hb, pp	—	—	**G-10**, biphenyl-containing, Figure 15
[228]	D-1	1999		*				*			*		hb	—	hr	**G-9**, Figure 15
[229]	1a	1999		*							*		hb	40–100	ssf, msf	**G-4**, Figure 14 and 16a
[230]	1	1999		*			*					*	hb, pp	2–10	tr	
[231]	1	2000		*					*		*	*	hb	20	ssf, msf	
[232]	2	2000		*			*					*	hb	20–70	ssf	porphyrin-containing
[214]	1	2000		*			*						pp	—	ssf	**G-23**, anthracene-containing, Figure 15
[233]	2	2001		*			*				*		se, hb	30	ssf, msf	bipyridinium-containing
[234]	2	2001		*								*	hb	—	msf	
[235]	2	2001		*			*					*	hb, pp	—	—	azobenzene-containing
[236]	3	2002		*								*	hb	—	ssf, msf	polymerizable
[237]	1, 2	2002		*			*					*	hb, pp	10–20	ssf	azobenzene-containing
[80]	2	2002		*			*				*		hb, pp	—	ssf	**G-11**, pyrene-containing, Figure 15

[172]	1	2002	*				*		*		hb	50–100	msf	**G-12**, Figure 15
[151]		1987		*							sp, mc	—	—	**G-24**, Figure 15
[238]	1a	1995		*		*					pp	6	ssf, msf	crown ether- and phthalocyanine-containing
[127]	CB 10	1995		*				*			sp	—	—	**G-5**, Figures 14 and 16g
[162]	TOAB	1997		*			*				sp	240–1400	ssf, msf	
[239]	2	2000		*							pp, mc	3.5–3.4	tu	olefine- and crown ether-containing
[80]	1	2002		*		*			*		hb, pp	—	ssf	porphyrin-containing
[240]	2	2002		*						*	sp	50–250	tu	perfluoroalkyl-containing
[241]	CAB	1987			*	*					sp, pp	16–19.2	msf	**G-1**, Figure 14
[175]	1, 2, 3	1991			*	*					pp	100	msf	**G-22**, azobenzene- and crown ether-containing, Figures 15 and 18
[176]	3R′	1994			*	*	*				sp, pp	—	hr	**G-21b**, azobenzene-containing, Figure 15
[177]	13Me	1998			*	*	*				sp	—	ssf	azobenzene-containing
[242]	S, R	1998			*	*					sp, pp	10–50	ssf	porphirin-containing
[147]	4 + 6	1999			*	*					sp, ct	2 mm	ssf	**G-19**, pyrene-containing, Figure 15
[243]	1	2002			*	*					se, pp	500	tu, hr	crown ether-containing
[244]	PFS-b-PDMS	1998				*					ct	20	tu	ferrocene-containing, polymer
[245]	4b	2001				*					pp	100–150	—	
[246]	PBZT	2002				*					pp	10	—	
[247]	Ib	1999						*			hb	20–120	msf	**G-18**, Figure 15
[248]	Z-1	1999							*		hb	—	tu	azobenzene-containing
[249]	9	2000							*		hb	20–50	—	cyclic dipeptide
[250]	1	2001							*	*	hb	20–50	tu	
[251]	P_{11}	2001							*		hb	—	tr	origopeptide
[252]	6	1997								*	hb	—	—	
[253]	1 + 3	1999								*	hb	20–50	—	
[254]	1	2000								*	hb	30–300	msf	
[255]	3, 4	2000								*	hb	20	msf	
[256]	4b	2001								*	hb	50	—	cholate-containing
[257]	1	2001								*	hb	15	msf	
[123]	STNH	1997									sp	ca. 40	tr	**G-20**, Figures 15 and 16c
[215]	1	1998									sp	30	ssf	**G-17**, Figures 15 and 16f

Note: sac: single-chain alkyl, dac: double-chain alkyl, mac: multiple-chain alkyl, cho:cholesterol-containing, pah: polycyclic aromatic hydrocarbons, ssf: single-strand fiber, msf: multi strand fiber, tr: twisted ribbon, hr: helical ribbon, tu: tubles, hb: hydrogen bonding, sp: solvophobic effect, mc: metal-coordinate interaction, pp: π-π interaction, ct: charge transfer interaction.

organogels (**G-19**) [147]. Pyrene-containing bile acid derivatives formed highly stable organogels with needle-like aggregates in various solvents even with less than 1% of the gelator. The T_{gel} increased with increasing amounts of trinitrofluorenone as an electron acceptor that could not form a gel in solvents. In an organogel system, several intermolecular interactions participate to maintain the gel formation. Control of these interactions is useful not only for understanding the gelation mechanism but also for developing functional organogels.

3.3.3. Approach for Nanostructured Materials

Organogel systems can be applied as a liquid organic media. Gu et al. [162] and Hafkamp et al. [138] described morphological imprinting of fibrous aggregates using tetraoctadecylammonium bromide and gluconamide derivatives coordinated with metal ion, respectively. Each organogelator can produce organogels with fibrous aggregates in styrene and methyl methacrylate. Polymerization of the solvents was carried out with UV light in the presence of a photoinitiator. Fibrous aggregates with similar diameter were observed before and after photo-polymerization, and the gelator molecules could be removed from the resultant polymer matrix by the solvent extraction method (Fig. 17) [138]. According to optical and electron microscopic measurements, the diameters of the strand-like pores were bigger than those of the original fibrous aggregates. It was expected that the monomers that exist near the surface of fibrous aggregates could not react since their mobility and fluidity were restricted. Similar observations were obtained with organogels from a pyrenyl group-containing peptide lipid (**G-11**) in styrene and methylmethacrylate. In this study, we obtained significant information on the molecular orientation states before and after polymerization of the bulk solution. Enhanced CD spectra around the pyrenyl group was observed after polymerization and maintained even at 70 °C, a temperature was even higher than its T_{gel}. These results indicate that highly oriented structures can be stabilized by polymerization of a bulk solution [163].

Ono et al. have reported that hollow, fibered silicas are prepared by transcription of various suprastructures formed in organogels [164, 165]. Three cholesterol-based gelators with monobenzo-18-crown-6 (**G-22a**), monoaza-18-crown-6 (**G-22b**), and 1,10-diaza-18-crown-6 (**G-22c**), respectively, were synthesized. These compounds produced organogels. Scanning electron microscopic observations of the xerogels showed that their organogelators assemble into a fibrous network structure, a curved lamellar structure, and a cylindrical tubular structure, respectively, in cyclohexane. Sol–gel polymerization of tetraethoxysilane was carried out in each of the gel systems. The silica obtained from **G-22a** showed a granular structure and a hollow fiber structure featuring the rough surface and the thick tube wall, respectively, in the absence and the presence of metal. This structure was created by adsorption of the anion-charged silica particles onto the cation-charged organogel fibers. On the other hand, the silica obtained from **G-22b** and **G-22c** had a hollow fiber structure featuring a smooth surface and thin tube wall both in the absence and the presence of metal salt. In the absence of metal salt, it was considered that the cationic charge generated by protonation of azacrown ethers plays a crucial role in the creation of such hollow fiber structures. In the presence of $KClO_4$, sol–gel polymerization resulted in tubular silica with a multilayered structure like a roll of paper (Fig. 18) [166]. The findings suggested that sol–gel polymerization proceeded along the surface of the curved lamellar stucture of **G-22b** or **G-22c** and the silica eventually grew

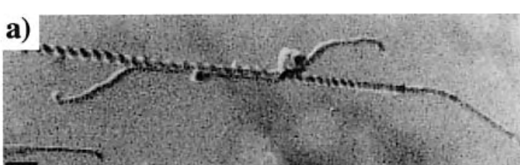
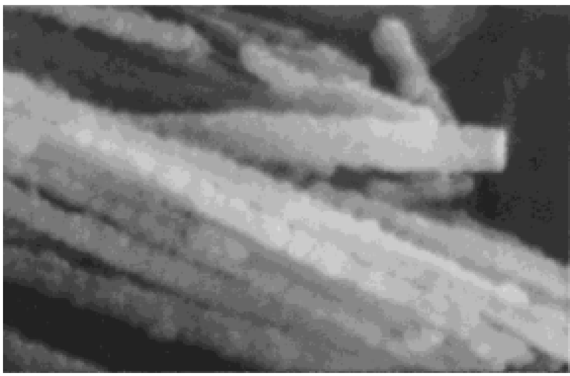


Figure 17. TEM images (Pt shadowing, bar 110 nm) of a dried gel (a) of **G-16** [138] in ethyl acetate and imprinted pores (b) of **G-16** [138] (bar 1.35 μm). Reprinted with permission from [138], R. J. H. Hafkamp et al., *Chem. Commun.* 545 (1997). © 1997, Royal Society of Chemistry.



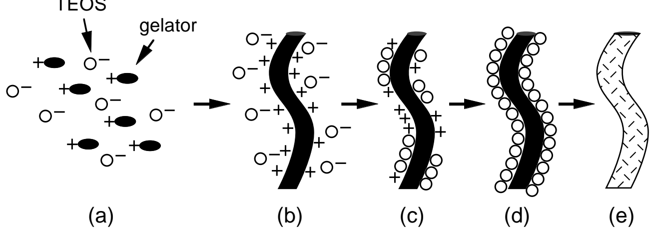


Figure 18. Schematic representation for the creation of hollow fiber silica by sol–gel polymerization of TEOS in the organogel state of **G-22** [166]: (a) mixture of gelators and TEOS; (b) gelation; (c) sol–gel polymerization of TEOS and adsorption onto the cationic gelator fibrils; (d) before calcination; (e) the hollowfiber silica formation after calcination. Reprinted with permission from [166], J. H. Jung et al., *Langmuir* 16, 1643 (2000). © 2002, American Chemical Society.

into a tubular structure. Vesicular and helical silica structures with several tens to several hundreds nm in diameter were obtained by using similar methods [166–168]. Future studies will certainly realize the production of more complicated and uniform nano-transcribed materials with desired fashion.

3.3.4. Control of Highly Ordered States

Some organogelators form highly ordered structures like those in aqueous bilayer membranes. Therefore, it is a convenient method to arrange the functional groups in the molecular assemblies. The arrangement and control of functional groups will certainly be an invaluable technique in many application fields such as sensors, molecular devices, and nano fabrication.

Lipophilic peptide-derived organogelators such as **G-10** and **G-11** show specific chirality, which is detectable by CD spectroscopic measurement around the chromophore groups. We have made some examinations to confirm whether or not specific properties observed in aqueous bilayer membranes can be reproduced in organic solvents. Phase separation behavior was observed in a mixture of **G-4** and an azobenzene-attached **B-4** derivative in benzene. Distinct CD spectra based on the azobenzene moiety were observed below T_{gel} [69]. The cotton effect of this mixture disappeared above T_{gel} and thermal reversibility was observed. On the other hand, a carboxylic acid-containing L-glutamate derivative (**G-9**) produced organogel with developed fibrous aggregates in several organic solvents [169].

These specific chiral properties stimulated us to implement optical resolution with chiral organogels. We reported the first example in 1995. When danoyl phenylalanine as a chiral guest molecule was dissolved in organogels from **G-4** and **B-3**, distinct enantioselective elution to an alkaline solution was observed. Interestingly, CD and DSC measurements showed that the best result was obtained through **B-3** domain formation phase separated from G-4 aggregates [170].

This finding encouraged us to control the chirality of organogels. For this purpose, we synthesized a double-chain alkyl L-glutamide derivative with an isoquinoline-head group (**G-12**) [171, 172]. This behaves as an organogelator with developed twisted fibers. It was confirmed that chelation with metal chloride in cyclohexane-ethanol (100:1) solution remarkably perturbed the chirality and morphology of the aggregates. Addition of copper ion ($CuCl_2$), which can form a square planar coordination, induced the chirality enhancement with morphological change from twisted fibrous aggregates to ribbon-like aggregates. On the contrary, cobalt ion ($CoCl_2$) and zinc ion ($ZnCl_2$), which can form an octahedral coordination state, caused serious morphological change with remarkable decrease of the chirality [171, 172].

On the other hand, Ishi-i et al. reported that a Zn(II)-porphyrin appended-cholesterol organogelator (**G-21c**) interacts with [60]fullerene to form a 2:1 Zn(II) porphyrin/[60]fullerene sandwich complex [173, 174]. The distinct bathochromic shift of the soret absorption band, which was found in the organogel with [60]fullerene in toluene, indicates that intermolecular electronic interaction exists between the Zn(II) porphyrin moiety and the [60]fullerene in the gel phase. Circular dichroism measurements suggested that the Zn(II) porphyrin moieties are strengthened to orient chirally in the gel phase by the interaction with [60]fullerene. The formation of a Zn(II) porphyrin/[60]fullerene sandwich complex would induce a morphological change-in-pitch length of twisted fibrous aggregates.

Light and metal ion-responsive organogelators were demonstrated by Murata and Shinkai et al. [175–177]. These gelators appended a crown ether moiety (**G-22a**), azacrown ether moiety (**G-22b** and **22c**), and azobenzenyl moiety (**G-21b** and **G-22**) to cholesterol, respectively. The azobenzenyl appended-cholesterol derivatives with a natural (S)-configuration at C-3 and those with an inverted (R)-configuration at C-3 formed organogels, but their solubilities and gelation abilities toward each solvent were quite different. LD, CD, and UV spectroscopic measurements showed that the azobenzene moiety was oriented in clockwise (in (R) chirality) or anticlockwise (in (S) chirality) direction when they interacted in the excited state. Furthermore, a sol–gel phase transition was induced by photoresponsive *cis*-trans isomerism of the azobenzene moiety. The molecular arrangement of crown ether appended-cholesterol derivatives was altered by the addition of metal ions and ammonium ion. The T_{gel} of this organogelator in methylcyclohexane/benzene (4:1) increased with increasing concentration of Li^+, Na^+, K^+, Rb^+, and NH_4^+, but decreased with increasing concentration of Cs^+. It was presumed that the 1:2 metal/crown complex with Cs^+ resulted in a disordered structure of organogels, but contrarily that 1:2 metal/crown complex with other metal ion and ammonium ion promoted their stability.

3.4. Nanofibrillar Structures from Polymers

3.4.1. Phase Separation Method

It is well known that block copolymers consisting of two or more compositions will form cube-like, rod-like, vesicle-like, and even lamella-like shaped microphase separation structures depending on the membrane casting conditions [178–185]. Therefore, it is also possible to produce a nanofibrillous microphase separation structure by choosing the chemical composition and varying the membrane casting conditions [186–193].

Liu et al. and Yang et al., for example, succeeded in producing a nanofibrillar-like microphase separation structure from A-B-C block polymers resulting from the combination of polystyrene (A), poly(2-cinnamoylethyl methacrylate) (B), and poly(tert-butyl acrylate) (C) (**P-4**) (Fig. 19a) [190]. To briefly introduce the method of production, a toluene solution of polymer was put into a polyethylene capsule, and a solid was obtained by evaporation and annealing. Transmission electron microscopic observation of the section of this solid showed a very beautiful nanofibrillar-like microphase separation structure with a diameter of about 25 nm as seen in Figure 19b [192].

On the other hand, Fujiwara and Kimura succeeded in producing a nanofiber from A-B or A-B-A block copolymer by the cast method from solution [191]. Specifically, it was obtained from a 0.2 wt% solution of block copolymer from poly(L-lactide) (A) and poly(oxyethylene) (B) which

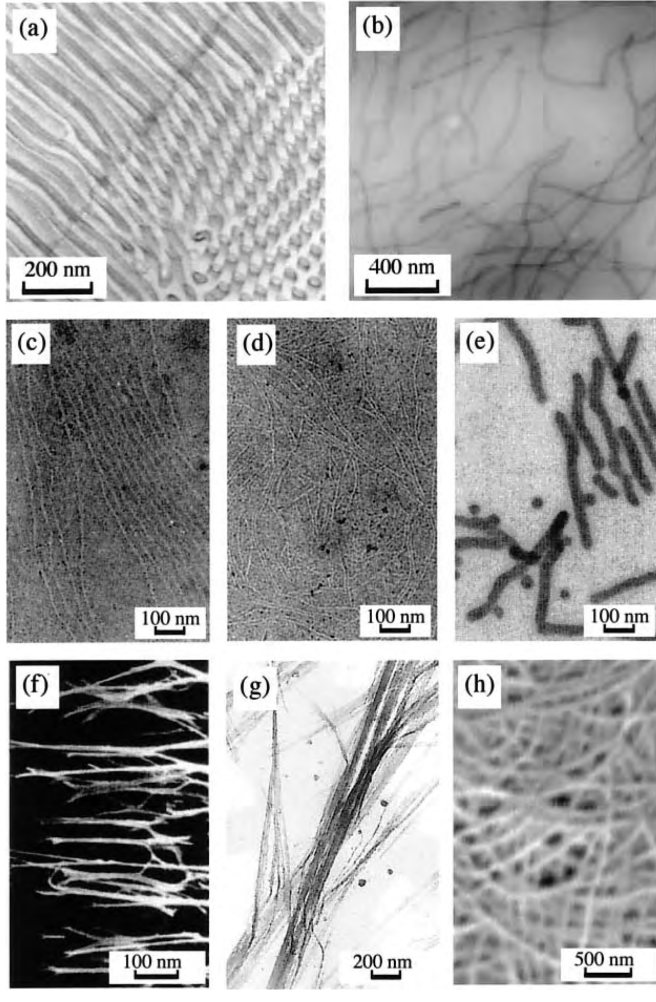


Figure 19. TEM images of (a) a 50-nm-thick section of a PS-b-PCEMA-b-PtBA/hPS film and (b) triblock/Fe_2O_3 hybrid nanofibers [192]. TEM images of the copolymer samples exfoliated from a silicon wafer surface after annealing at 60 °C for 10 h: (c) PLLA-PEG [191], (d) PLLA-PEG-PLLA [191], and (e) P4VP-(20.7)-b-PS(21.4k) diblock copolymers [187]. SEM images of (f) the self-synthesized PAN nanofibers [199], (g) the PPP nanofibers [198], and (h) electrospun fibers of polystyrene [197]. Reprinted with permission from [192], X. Yang et al., *Angew. Chem. Int. Ed.* 40, 3593 (2001). © 2001, Wiley-VCH; from [191], T. Fujiwara and Y. Kimura, *Macromol. Biosci.* 2, 11 (2002). © 2002, Wiley-VCH; from [187], K. de Moel et al., *Chem. Mater.* 13, 4580 (2001). © 2001, American Chemical Society; from [199], L. Feng et al., *Angew. Chem. Int. Ed.* 41, 1221 (2002). © 2002, Wiley-VCH; from [198], B. Yu and H. Li, *Mater. Science Eng.* A325, 215 (2001); © 2001, Elsevier; and from [197], A. G. MacDiarmid et al., *Synthetic Metals* 119, 27 (2001). © 2001, Elsevier Science.

was prepared (**P-5**), cast on mica, and afterwards annealed for 2 hs at 60 °C. From TEM observation, the diameter of the nanofiber formed from the A-B block copolymer was about 20 nm (Fig. 19c) [191], and that from the A-B-A block copolymer was about 6 nm (Fig. 19d) [191]. The authors hypothesize that the core is made from lactide (A) and that this part has a crystal structure.

de Moel et al. have proposed the crew-cut method as a new technique of attaining nanofibrillar microphase separation structures from block polymers like polystyrene (A) and poly(4-vinylpyridine) (B) (**P-6**) which are difficult to separate [187]. Specifically, a 1:1 complex was produced by adding equimolar amounts of pyradine and 3-pentadecylphenol (PDP) to a chloroform solution of A-B block polymer. This was then evaporated very slowly by drying at 50 °C for at least 12 hs. Next, a 0.5 g sample was put into a dialysis tube filled with ethanol, and most of the PDP(s) were removed by dialyzing for 2 weeks. Decompression dryness was carried out at 50 °C after washing by ethanol overnight, and a sample was obtained. Transmission electron microscopic observation showed that a nanofiber with a diameter of about 30 nm was formed from (A) in the core part (Fig. 19e) [187].

3.4.2. Injection Molding Method

When using a high polymer, it is possible to produce nanofibers by adapting an injection molding method. Although production of the nanofiber from a high polymer is somewhat forced, two techniques may be used: either injecting high polymer solution into nanosize pores [194–197] or polymerization and pushing out from nanofiber-like capillaries [198].

Feng et al. reported the production of nanofibers using the former technique by injecting poly(acrylonitrile) as a template into porous anodic aluminum oxide (Fig. 19f) [199]. Yu and Li adapted the latter technique with porous metal using poly(*p*-phenylene) (**P-7**) as a template to produce nanofibers (Fig. 19g) [198]. Features such as super-hydrophobicity [200–203] and erection luminescence [204–206] were obtained using both injection methods, but the nanofibers produced also had disadvantages such as low diameter uniformity and a tendency to become condensed.

In order to solve these problems, a technique called the electrospinning method has come into practical use. This technique was patented in 1934 by Formhals [207] and was little known at first, but it has been attracting attention for the past several decades as a nanomanipulation technique [208–211]. The electrospinning method is the technique of putting polymer solution and metal electrodes into a poor solvent such as water, and blowing off the polymer solution by applying high voltage between it and the metal electrodes, and producing nanofibers. In this case, the diameter and length of fibers formed depends on the concentration of polymer solution, the distance between the polymer solution and metal electrodes, and the voltage applied. MacDiarmid et al. have reported nanofibers formed using this technique with an average diameter of about 50 nm, by injection using a 2 wt% poly(aniline) chloroform solution and applying a voltage of 25,000 V from a distance of 25 cm [197]. Polystyrene nanofibers of about 30 nm in diameter were also prepared using the same technique, meaning that micron-order fibers can be prepared from nanofibers cheaply and easily (Fig. 19h).

4. SUMMARY

"Self-assembled nanofibers" are formed from miscellaneous synthetic compounds and show various charming shapes as described herein. Although it is hard to completely understand the relationships between their morphologies and chemical structures, experimental and theoretical

approaches indicate several important factors such as:

a) moderate solubility into media,
b) intermolecular interaction moieties and their sterical position,
c) molecular shapes for highly ordered molecular packing,
d) molecular chirality.

Many reports also indicate a wide range of possible applications for these self-assembled nanofibers, from morphological applications such as preparation and utilization of replica of them to molecular level applications of specific molecular orientation. In this article, we introduced self-assembled nanofibers constructed mainly of relatively simple compounds, but oligomers and polymers such as cyclic polypeptide, cyclodextrin, and cyclic polysaccharide are also known to form self-assembled nanofibers [212].

It is prospected that self-assembling nano architectures including fibrillar aggregates have many potential applications because nature, especially human, is an excellent example of a hierarchical product of self-assembling molecules—there are so many excellent examples of hierarchical products of self-assembling molecules in nature itself, not the least of which is the human body. We sincerely hope that research on self-assembled nano architectures will be helpful in the development of molecularly precise materials and devices. The diversity of self-assembled nanofiber systems, including a large number of molecules, provides much opportunity to modify the chemical structures of self-assembling molecules. With the present rapid progress and expansion of this field, we feel a premonition that many applications will be found for nanofibers in the near future, including using combinations of morphologies and functions to provide suggestions for the origin of life.

GLOSSARY

α-helix One of the secondary structures of peptides, rod-like shape caused by intermolecular hydrogen bonding.

Amiloid fiber Fiber-like aggregate structures formed β-structure protein. These structures cause prion protein structure such as Alzheimer's disease or Bovine Spongiform Encephalopathy (BSE).

Atomic force microscope (AFM) Provide pictures of atoms on or in surfaces to measure the forces (at the atomic level) between a sharp probing tip (which is attached to a cantilever spring) and a sample surface.

β-sheet One of the secondary structures of peptides, its formed sheet-like shape is caused by intramolecular hydrogen bonding (antiparallel and parallel structure).

Bilayer membranes Amphiphiles self-organize in water to form molecular bilayers which assume the shape of globule, vesicle, lamella, fiber, and so on.

Carbon nanotube Fullerene-related structures that consist of graphene cylinders closed at either end with caps containing pentagonal rings.

Chiral/chirality Nonsuperimposable upon their mirror images and chirality is particularly important for molecules interacting within a biological environment, where many of the native molecules are chiral.

Circular dichroism (CD) The difference in absorption between left- and right-handed circulary polarized light.

Differencial scanning calorimetry (DSC) DSC measures the amount of energy (heat) absorbed or released by a sample as it is heated, cooled, or held at constant temperature. DSC also performs precise temperature measurements.

Hydrogen bonding Occurs where the partially positively charged hydrogen atom lies between partially negatively charged oxygen and nitrogen atoms.

Organogel Self-assembly of low molecular weight gelators to fiber-like structures, which entangle to form a three-dimensional network, immobilize organic fluids.

Peptide Two or more amino acids can be linked together by a dehydration synthesis to form a peptide. The characteristic chemical bond is called a peptide bond.

Supramolecular assemblies Infinite repeating structure comprised of several chemical species held together by intermolecular interactions and showing transcendent properties which are not observed in its constituent molecule alone.

Transmission electron microscopy (TEM) TEM allows to determine the internal structure of materials, either or biological or nonbiological origin used by a beam of electrons, to illuminate the sample.

ACKNOWLEDGMENTS

The authors deeply thank Ms. Yu Hamasaki, Mr. Tomohiro Shirosaki, and Mr. Takao Satoh for their capable assistance.

REFERENCES

1. J. Inglese, J. F. Glickman, W. Lorenz, M. G. Caron, and R. J. Lefkowitz, *J. Biol. Chem.* 267, 1422 (1992).
2. J. E. Tousa, W. Baehr, R. L. Martin, J. Hirsh, and W. L. Pak, *Cell* 40, 839 (1985).
3. A. F. Cowman, C. S. Zuker, and G. M. Rubin, *Cell* 44, 705 (1986).
4. C. S. Zuker, A. F. Cowman, and G. M. Rubin, *Cell* 40, 851 (1985).
5. S. S. Karnik, T. P. Sakmar, H. Chen, and H. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* 85, 8459 (1988).
6. I. Nishie, K. Anzai, T. Yamamoto, and Y. Kirino, *J. Biol. Chem.* 265, 2488 (1990).
7. R. E. Dickerson, *Sci. Am.* 249, 94 (1983).
8. J. Kypr and M. Vorlickoba, *Gen. Phys. Biophys.* 4, 471 (1985).
9. E. Palecek, *Biol. Listy.* 53, 15 (1988).
10. M. Tsuboi, *Gendai Kagaku* 36, 360 (2001).
11. N. Sugimoto and S. Nakano, *Kagaku to Seibutsu* 39, 714 (2001).
12. E. H. Egelman, N. Francis, and D. J. DeRosier, *Nature* 298, 131 (1982).
13. M.-F. Carlier, D. Didry, I. Erk, J. Lepault, M. L. V. Troys, J. Vandekerckhove, I. Perelroizen, H. Yin, Y. Doi, and D. Pantaloni, *J. Biol. Chem.* 271, 9231 (1996).
14. B. J. Pope, S. M. Gonsior, S. Yeoh, A. McGough, and A. G. Weeds, *J. Mol. Biol.* 298, 649 (2000).
15. T. Ohno, *Saibo* 8, 170 (1976).
16. P. J. Butler, *J. Mol. Biol.* 72, 25 (1972).
17. P. J. Butler and J. T. Finch, *J. Mol. Biol.* 78, 637 (1973).
18. T. Ohno, T. Okada, Y. Nonomura, and H. Inoue, *J. Biochem.* 77, 313 (1975).
19. K. Raghavendra, J. A. Kelly, L. Khairallah, and T. M. Schuster, *Biochemistry* 27, 7583 (1988).
20. T. Kunitake, Y. Okahata, S. Shimomura, S. Yasunami, and K. Takarabe, *J. Am. Chem. Soc.* 103, 5401 (1981).

21. K. Yamada, H. Ihara, T. Ide, T. Fukumoto, and C. Hirayama, *Chem. Lett.* 1713 (1984).
22. H. Ihara, T. Fukumoto, C. Hirayama, and K. Yamada, *Polym. Commun.* 27, 282 (1986).
23. T. Kunitake and N. Yamada, *J. Chem. Soc., Chem. Commun.* 655 (1986).
24. H. Ihara, T. Fukumoto, C. Hirayama, and K. Yamada, *J. Chem. Soc. Jpn.* 543 (1987).
25. J. H. Georger, A. Shingh, R. R. Price, J. M. Schnur, P. Yager, and P. E. Schoen, *J. Am. Chem. Soc.* 109, 6169 (1987).
26. I. Cho and J. G. Park, *Chem. Lett.* 977 (1987).
27. J.-H. Fuhrhop, P. Schnieder, E. Boekema, and W. Helfrich, *J. Am. Chem. Soc.* 110, 2861 (1988).
28. H. Yanagawa, Y. Ogawa, H. Furuta, and K. Tsuno, *J. Am. Chem. Soc.* 111, 4567 (1989).
29. N. Kimizuka, H. Ohira, M. Tanaka, and T. Kunitake, *Chem. Lett.* 29 (1990).
30. H. Ihara, M. Yamaguchi, M. Takafuji, H. Hachisako, C. Hirayama, and K. Yamada, *J. Chem. Soc. Jpn.* 1047 (1990).
31. H. Ihara, K. Yoshikai, M. Takafuji, C. Hirayama, and K. Yamada, *Jpn. J. Polym. Sci. Tech.* 48, 377 (1991).
32. H. Ihara, M. Takafuji, C. Hirayama, and D. F. O'Brien, *Langmuir* 8, 1548 (1992).
33. T. Kunitake and Y. Okahata, *J. Am. Chem. Soc.* 99, 3860 (1977).
34. R. Jetter and M. Riederer, *Planta* 195, 257 (1994).
35. M. Lohmeyer and P. Workman, *Biochem. Pharmacol.* 819, 44 (1992).
36. S. Iijima, *Nature* 354, 56 (1991).
37. S. Iijima, T. Ichihara, and Y. Ando, *Nature* 356, 774 (1992).
38. J. Han, M. P. Anantram, R. F. Jaffe, J. Kong, and H. Dai, *Phys. Rev. B* 57, 14983 (1998).
39. D. S. Bethune, *Physica B* 323, 90 (2002).
40. S. Iijima and T. Ichihashi, *Nature* 363, 603 (1993).
41. D. S. Bethune, C. H. Kiang, M. S. de Vries, G. Gorman, R. Savoy, J. Vanquez, and R. Beyers, *Nature* 363, 605 (1993).
42. N. Nakashima, A. Asakuma, J.-M. Kim, and T. Kunitake, *Chem. Lett.* 1709 (1984).
43. N. Nakashima, A. Asakuma, and T. Kunitake, *J. Am. Chem. Soc.* 107, 509 (1985).
44. N. Hamada, S. Sawada, and A. Oshiyama, *Phys. Rev. Lett.* 68, 1579 (1992).
45. J. W. Mintmire, B. I. Dunlap, and C. T. White, *Phys. Rev. Lett.* 68, 631 (1992).
46. R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 46, 1804 (1992).
47. R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Appl. Phys. Lett.* 60, 2204 (1992).
48. K. Tanaka, K. Okahara, M. Okada, and T. Yamabe, *Chem. Phys. Lett.* 191, 469 (1992).
49. H. Ajiki and T. Ando, *J. Phys. Soc. Jpn.* 62, 1255 (1993).
50. J. E. Fischer, H. Dai, A. Thess, R. Ree, N. M. Hanjani, D. L. Dehaas, and R. E. Smalley, *Phys. Rev. B* 55, R 4921 (1997).
51. M. Bockrath, D. H. Cobden, P. L. McEuen, N. G. Chopra, A. Zettl, A. Thess, and R. E. Smalley, *Science* 275, 1922 (1997).
52. S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerlings, and C. Dekker, *Nature* 386, 474 (1997).
53. D. H. Cobden, M. Bockrath, P. L. McEuen, A. G. Rinzler, and R. E. Smalley, *Phys. Rev. Lett.* 81, 681 (1998).
54. S. J. Tans, A. R. M. Verschueren, and C. Dekker, *Nature* 393, 49 (1998).
55. B. W Smith and D. E. Luzzi, *Chem. Phys. Lett.* 321, 169 (2000).
56. B. W Smith, M. Monthioux, and D. E. Luzzi, *Chem. Phys. Lett.* 315, 31 (1999).
57. S. Iijima, *Physica B* 323, 1 (2002).
58. T. Shimizu, *Macromol. Rapid Commun.* 23, 311 (2002).
59. A. C. Dillon, K. M. Jones, T. A. Bekkedahi, C. H. Kiang, D. S. Bethune, and M. J. Heben, *Nature* 386, 377 (1997).
60. P. Chen, X. Wu, J. Lin, and K. L. Tan, *Science* 285, 91 (1999).
61. H. Dai, E. W. Wang, and C. M. Lieber, *Nature* 384, 147 (1996).
62. H. Nishijima, S. Kamo, S. Akita, Y. Nakayama, K. I. Hohmura, S. H. Yoshimura, and K. Takeyasu, *Appl. Phys. Lett.* 74, 4061 (1999).
63. K. Hanabusa, T. Miki, Y. Taguchi, T. Koyama, and H. Shirai, *J. Chem. Soc., Chem. Commun.* 1382 (1993).
64. C. Gregolire, S. Marco, J. Thimonier, L. Duplan, E. Jaurine, J.-P. Cheuvin, B. Michael, V. Peyort, and J.-M. Verdier, *EMBO. J.* 20, 3313 (2001).
65. J. H. Jung and S. Shinkai, *Journal of Inclusion Phenomena and Macrocyclic Chemistry* 41, 53 (2001).
66. S. Mabrey and J. M. Sturtevant, *Proc. Natl. Acad. Sci. U.S.A.* 73, 3862 (1976).
67. T. Kajiyama, A. Kumano, M. Takayanagi, Y. Okahata, and T. Kunitake, *Chem. Lett.* 645 (1979).
68. Y. Okahata and T. Kunitake, *Ber. Bunsenges. Phys. Chem.* 84, 550 (1980).
69. T. Kunitake, H. Ihara, and Y. Okahata, *J. Am. Chem. Soc.* 105, 6070 (1983).
70. H. Ihara, H. Hachisako, C. Hirayama, and K. Yamada, *J. Chem. Soc. Chem. Commun.* 1244 (1992).
71. T. Kimura and S. Shinkai, *Chem. Lett.* 1035 (1998).
72. T. Kato, G. Kondo, and K. Hanabusa, *Chem. Lett.* 193 (1998).
73. H. Ihara, M. Yoshitake, M. Takafuji, T. Yamada, T. Sagawa, C. Hirayama, and H. Hachisako, *Liq. Cryst.* 26, 1021 (1999).
74. T. Kunitake, N. Nakashima, M. Shimomura, Y. Okahata, K. Kano, and T. Ogawa, *J. Am. Chem. Soc.* 102, 6642 (1980).
75. M. Shimomura and T. Kunitake, *Chem. Lett.* 1001 (1981).
76. H. Ihara, H. Hachisako, C. Hirayama, and K. Yamada, *Liq. Cryst.* 2, 215 (1987).
77. J. E. Sohna and F. Fages, *Chem. Commun.* 327 (1997).
78. N. Amanokura, K. Yasumasa, and S. Shinkai, *J. Chem. Soc. Perkin Trans.* 2 1995 (1999).
79. K. Kano, A. Romero, B. Djermouni, H. J. Ache, and J. H. Fendler, *J. Am. Chem. Soc.* 101, 4030 (1979).
80. T. Sagawa, S. Fukugawa, T. Yamada, and H. Ihara, *Langmuir* 18, 7223 (2002).
81. T. Nagamura, S. Mihara, Y. Okahata, T. Kunitake, and T. Matsuo, *Ber. Bunsenges. Phys. Chem.* 82, 1093 (1978).
82. M. Shimomura, R. Ando, and T. Kunitake, *Ber. Bunsenges. Phys. Chem.* 87, 1134 (1983).
83. H. Hachisako, H. Ihara, T. Kamiya, C. Hirayama, and K. Yamada, *Chem. Commun.* 19 (1997).
84. J. M. Schnur, B. R. Ratna, J. V. Seliger, A. Singh, G. Jyothi, and K. R. K. Easwaran, *Science* 264, 945 (1994).
85. N. Nakashima, H. Fukushima, and T. Kunitake, *Chem. Lett.* 1207 (1981).
86. N. Nakashima and T. Kunitake, *J. Am. Chem. Soc.* 104, 4261 (1982).
87. T. Arimura, M. Shibata, H. Ihara, and C. Hirayama, *Anal. Sci.* 9, 401 (1993).
88. A. D. Bungham, M. M. Standish, and J. C. Watkins, *J. Mol. Biol.* 13, 238 (1965).
89. K. Yamada, H. Shosenji, H. Ihara, and O. Hotta, *Chem. Lett.* 43 (1983).
90. T. Kunitake, N. Kimizuka, N. Higashi, and N. Nakashima, *J. Am. Chem. Soc.* 106, 1978 (1984).
91. J.-H. Fuhrhop, P. Schnieder, J. Rosenberg, and R. Boekema, *J. Am. Chem. Soc.* 109, 3387 (1987).
92. J.-H. Fuhrhop, S. Svenson, C. Boettcher, E. Rossler, and H.-M. Vieth, *J. Am. Chem. Soc.* 112, 4307 (1990).
93. J. Koning, C. Boettcher, H. Winkler, E. Zeitler, Y. Talmon, and J.-H. Fuhrhop, *J. Am. Chem. Soc.* 115, 693 (1993).
94. T. Imae, Y. Takahashi, and H. Muramatsu, *J. Am. Chem. Soc.* 114, 3414 (1992).

95. R. Oda, I. Huc, M. Schmutz, S. J. Candau, and F. C. MacKintosh, *Nature* 399, 566 (1999).
96. R. Oda, I. Huc, and S. J. Candau, *Angew. Chem. Int. Ed.* 37, 2689 (1998).
97. W. Helfrich and W. Harbich, *Chem. Scr.* 25, 32 (1985).
98. W. Helfrich, *J. Chem. Phys.* 85, 1085 (1986).
99. W. Helfrich and J. Prost, *Phys. Rev. A* 38, 3065 (1988).
100. W. Helfrich, *Langmuir* 7, 567 (1991).
101. J. V. Selinger, M. S. Spector, and J. M. Schnur, *J. Phys. Chem. B* 105, 1757 (2001).
102. N. Nandi and B. Bagchi, *J. Am. Chem. Soc.* 118, 11208 (1996).
103. T. Kunitake, N. Nakashima, and M. Kunitake, *Macromolecules* 22, 3544 (1989).
104. S. Zhang and M. Altman, *Reactive & Functional Polymers* 41, 91 (1999).
105. M. R. Caplan, E. M. Schwartzfarb, S. Zhang, R. D. Kamm, and D. A. Lauenburger, *Biomaterials* 23, 219 (2002).
106. Y.-C. Yu, P. Berndt, M. Tirrell, and G. B. Fields, *J. Am. Chem. Soc.* 118, 12515 (1996).
107. M. Altman, P. Lee, A. Rich, and S. Zhang, *Protein Science* 9, 1095 (2000).
108. M. R. Caplan, P. N. Moore, S. Zhang, R. D. Kamm, and D. A. Lauffenburger, *Biomacromolecules* 1, 627 (2000).
109. L. A. Scrocchi, Y. Chen, S. Waschuk, F. Wang, S. Cheung, A. A. Darabie, J. McLaurin, and P. E. Fraser, *J. Mol. Biol.* 318, 697 (2002).
110. T. Kowalewski and D. M. Holtzman, *Proc. Natl. Acad. Sci. U.S.A.* 96, 3688 (1999).
111. R. Jayakumar, M. Murugesan, C. Asokan, and M. A. Scibioh, *Langmuir* 16, 1489 (2000).
112. H. A. Lashuel, S. R. LaBrenz, L. Woo, L. C. Serpell, and J. W. Kelly, *J. Am. Chem. Soc.* 122, 5262 (2000).
113. K. Tenidis, M. Waldner, J. Bernhagen, W. Fiscle, M. Bermann, M. Weber, M.-L. Merkle, W. Voelte, H. Brunner, and A. Kapurniotu, *J. Mol. Biol.* 295, 1055 (2000).
114. M. M. Martinez-Senac, J. Villalain, and J. C. Gomez-Fernandez, *Eur. J. Biochem.* 265, 744 (1999).
115. T. Sakurai, M. Koga, M. Takafuji, and H. Ihara, *Chem. Lett.* 152 (2003).
116. M. S. Specter, J. V. Selinger, A. Singh, J. M. Rodriguez, R. R. Price, and J. M. Schnur, *Langmuir* 14, 3493 (1998).
117. M. Hatano, M. Yoneyama, Y. Sato, and Y. Kawamura, *Biopolymers* 12, 2423 (1973).
118. H. Ihara, M. Shibata, and C. Hirayama, *Chem. Lett.* 1731 (1992).
119. M. Shibata, H. Ihara, and C. Hirayama, *Polymer* 34, 1106 (1993).
120. Y. Osada and A. R. Khokhlov, "Polymer Gels and Networks," Marcel Dekker, Inc., New York (2001).
121. J. P. Cohen Addad, "Physical Properties of Polymeric Gels," John Wiley & Sons, Inc., Hoboken, NJ (1996).
122. K. Almdal, J. Dyre, S. Hvidt, and O. Kramer, *Polym. Gels Networks* 1, 5 (1993).
123. P. Terech and R. Weiss, *Chem. Rev.* 97, 3133 (1997).
124. Y.-C. Lin and R. G. Weiss, *Macromolecules* 20, 414 (1987).
125. T. Brotin, R. Utermohlen, F. Fages, H. Bouas-Laurent, and J. P. Desvergne, *J. Chem. Soc., Chem Commun.* 416 (1991).
126. Y. Ishikawa, H. Kuwahara, and T. Kunitake, *J. Am. Chem. Soc.* 111, 8530 (1989).
127. N. Ide, T. Fukuda, and T. Miyamoto, *Bull. Chem. Soc. Jpn.* 68, 3423 (1995).
128. T. Tachibana and H. Kambara, *J. Am. Chem. Soc.* 87, 3016 (1965).
129. T. Tachibana, T. Mori, and K. Mori, *Bull. Chem. Soc. Jpn.* 53, 1741 (1980).
130. A. T. Polishuk, *J. Am. Soc. Lubn. Eng.* 22, 133 (1977).
131. P. Terech, A Coutin, and A. M. Giround-Godquin, *J. Phys. Chem. B* 101, 6810 (1994).
132. K. Hanabusa, K. Okui, K. Karaki, T. Koyama, and H. Shirai, *J. Chem. Soc., Chem. Commun.* 1371 (1992).
133. K. Hanabusa, K. Hiratsuka, M. Kimura, and H. Shirai, *Chem. Mater.* 11, 649 (1999).
134. K. Hanabusa, J. Tange, Y. Taguchi, T. Koyama, and H. Shirai, *J. Chem. Soc., Chem. Commun.* 390 (1993).
135. K. Hanabusa, M. Yamada, M. Kimura, and H. Shirai, *Angew. Chem. Int. Ed. Engl.* 35, 1949 (1996).
136. K. Hanabusa, K. Shimura, K. Hirose, M. Kimura, and H. Shirai, *Chem. Lett.* 885 (1996).
137. P. Thiyagarajan, F. Zeng, C. Y. Ku, and S. C. Zimmermanb, *J. Mater. Chem.* 7, 1221 (1997).
138. R. J. H. Hafkamp, B. P. A. Kokke, I. M. Danke, H. P. M. Geurts, A. E. Rowan, M. C. Feiters, and R. J. M. Nolte, *Chem. Commun.* 545 (1997).
139. A. Friggeri, O. Gronwald, K. J. C. Bommel, S. Shinkai, and D. N. Reinhoudt, *Chem. Commun.* 2434 (2001).
140. O. Gronwald and S. Shinkai, *Chem. Eur. J.* 7, 4328 (2001).
141. O. Gronwald, E. Snip, and S. Shinkai, *Current Opinion in Colloid & Interface Science* 7, 148 (2002).
142. I. Furman and R. G. Weiss, *Langmuir* 9, 2084 (1993).
143. R. Mukkamala and R. G. Weiss, *J. Chem. Soc., Chem. Commun.* 375 (1995).
144. R. Mukkamala and R. G. Weiss, *Langmuir* 12, 1474 (1996).
145. D. J. Abdallah and R. G. Weiss, *Adv. Mater.* 12, 1237 (2000).
146. L. Lu, D. L. Cocker, R. E. Bachman, and R. G. Weiss, *Langmuir* 16, 20 (2000).
147. U. Maitra, P. V. Kumar, N. Chandra, L. J. D'Souza, M. D. Prasanna, and A. R. Raju, *Chem. Commun.* 595 (1999).
148. F. Placin, M. Colomes, and J.-P. Desvergne, *Tetrahedron Lett.* 38, 2665 (1997).
149. J.-L. Pozzo, G. M. Clavier, and J.-P Desvergne, *J. Mater. Chem.* 8, 2575 (1998).
150. G. M. Clavier, J.-F. Brugger, H. Bouas-Laurent, and J.-L. Pozzo, *J. Chem. Soc., Perkin Trans.* 2 2527 (1998).
151. P. Terech, C. Chachaty, J. Gaillard, and A. M. Giroud-Godquin, *J. Phys. Fr.* 48, 663 (1987).
152. A. M. Godquin-Giroud, *J. Phys. Lett.* 45, L-387 (1984).
153. A. M. Godquin-Giroud, M. M. Gauthier, G. Sigaud, F. Hardouin, and F. Achard, *Mol. Cryst. Liq. Cryst.* 132, 35 (1986).
154. A. M. Godquin-Giroud and J. C. Marchon, *J. Phys. Lett.* 45, L-681 (1984).
155. G. S. Attard and P. R. Cullum, *Liq. Cryst.* 8, 299 (1990).
156. H. Abied, D. Guillon, A. Skoulios, P. Weber, A. M. Godquin-Giroud, and J. C. Marchon, *Liq. Cryst.* 2, 269 (1987).
157. H. Ihara, K. Shudo, M. Takafuji, C. Hirayama, H. Hachisako, and K. Yamada, *Jpn. J. Polym. Sci. Technol.* 52, 606 (1995).
158. M. de Loos, J. van Esch, I. Stokroos, R. M. Kellogg, and B. L. Feringa, *J. Am. Chem. Soc.* 119, 12675 (1997).
159. M. Masuda, T. Honda, K. Yase, and T. Shimizu, *Macromolecules* 31, 9403 (1998).
160. M. Masuda, T. Honda, Y. Okada, K. Yase, and T. Shimizu, *Macromolecules* 33, 9233 (2000).
161. K. Hanabusa, Y. Maesaka, M. Suzuki, M. Kimura, and H. Shirai, *Chem. Lett.* 1168 (2000).
162. W. Gu, L. Lu, G. B. Chapman, and R. G. Weiss, *Chem. Commun.* 543 (1997).
163. M. Takafuji, A. Ishiodori, and H. Ihara, "Proc. Int. Symp. Bioorg. Chem.," 2002, Toronto, Canada.
164. Y. Ono, K. Nakashima, M. Sano, Y. Kanekiyo, K. Inoue, J. Hojo, and S. Shinkai, *Chem. Commun.* 1477 (1998).
165. Y. Ono, K. Nakashima, M. Sano, J. Hojo, and S. Shinkai, *Chem. Lett.* 1119 (1999).
166. J. H. Jung, Y. Ono, and S. Shinkai, *Langmuir* 16, 1643 (2000).
167. J. H. Jung, Y. Ono, and S. Shinkai, *J. Chem. Soc., Perkin Trans.* 2 1289 (1999).
168. J. H. Jung, Y. Ono, K. Sakurai, M. Sano, and S. Shinkai, *J. Am. Chem. Soc.* 122, 8648 (2000).

169. M. Takafuji, H. Ihara, C. Hirayama, H. Hachisako, and K. Yamada, *Liq. Cryst.* 18, 97 (1995).
170. H. Ihara, K. Shudo, H. Hachisako, K. Yamada, and C. Hirayama, *Liq. Cryst.* 20, 807 (1996).
171. M. Takafuji, T. Sakurai, T. Hashimoto, N. Kido, T. Yamada, T. Sagawa, H. Hachisako, and H. Ihara, *Chem. Lett.* 7223 (2002).
172. H. Ihara, T. Sakurai, T. Yamada, T. Hashimoto, M. Takafuji, T. Sagawa, and H. Hachisako, *Langmuir* 18, 7120 (2002).
173. T. Ishi-i, J. H. Jung, and S. Shinkai, *J. Mater. Chem.* 10, 2238 (2000).
174. T. Ishi-i, R. Iguchi, E. Snip, M. Ikeda, and S. Shinkai, *Langmuir* 17, 1825 (2001).
175. K. Murata, M. Aoki, T. Nishi, A. Ikeda, and S. Shinkai, *J. Chem. Soc., Chem. Commun.* 1715 (1991).
176. K. Murata, M. Aoki, T. Suzuki, T. Harada, H. Kawabata, T. Komori, F. Ohseto, K. Ueda, and S. Shinkai, *J. Am. Chem. Soc.* 116, 6664 (1994).
177. S. Shinkai and K. Murata, *J. Mater. Chem.* 8, 485 (1998).
178. A. Halperin and S. Alexander, *Macromolecules* 22, 2403 (1989).
179. Z. Zhou, B. Chu, and D. G. Peiffer, *Macromolecules* 26, 1876 (1993).
180. Z. Tuzer and P. Kratochvil, *Surface Colloid Sci.* 15, 1 (1993).
181. R. Nagarajan and K. Ganesh, *J. Chem. Phys.* 90, 5843 (1989).
182. F. S. Bates and G. H. Fredrickson, *Ann. Rev. Phys. Chem.* 15, 584 (1990).
183. F. S. Bates, *Science* 251, 898 (1991).
184. E. L. Thomas, D. M. Anderson, C. S. Henkee, and D. Hoffman, *Nature* 334, 598 (1988).
185. S. Sakurai, *Trends Polym. Sci.* 3, 90 (1995).
186. L. Qi, H. Cölfen, and M. Antonietti, *Angew. Chem. Int. Ed.* 39, 604 (2000).
187. K. de Moel, G. O. R. Ekenstein, H. Nijland, E. Polushkin, and G. Brinke, *Chem. Mater.* 13, 4580 (2001).
188. H. Fong and D. H. Reneker, *J. Polym. Sci. B* 37, 3488 (1999).
189. G. Liu, J. Ding, L. Qiao, A. Guo, B. P. Dymov, J. T. Gleeson, T. Hashimoto, and K. Saijo, *Chem. Eur. J.* 5, 2740 (1999).
190. G. Liu, L. Qiao, and A. Guo, *Macromolecules* 29, 5508 (1996).
191. T. Fujiwara and Y. Kimura, *Macromol. Biosci.* 2, 11 (2002).
192. X. Yang, G. Liu, F. Liu, B. Z. Tang, H. Peng, A. B. Pakhomov, and C. Y. Wong, *Angew. Chem. Int. Ed.* 40, 3593 (2001).
193. K. L. Wooley, *J. Polym. Sci., Polym. Chem.* 38, 1397 (2000).
194. G. Che, B. B. Lakshmi, C. R. Martin, and E. R. Fisher, *Chem. Mater.* 10, 260 (1998).
195. M. Gao, S. Huang, L. Dai, G. Wallace, R. Gao, and Z. Wang, *Angew. Chem.* 112, 3810 (2000).
196. Z. Liu, Y. Sakamoto, T. Ohsuna, K. Hiraga, O. Terasaki, C. H. Ko, H. J. Shin, and R. Ryoo, *Angew. Chem.* 112, 3237 (2000).
197. A. G. MacDiarmid, W. E. Jones, I. D. Norris, J. Gao, A. T. Jhonson, N. J. Pinto, J. Hone, B. Han, F. K. Ko, H. Okazaki, and M. Llaguno, *Synthetic Metals* 119, 27 (2001).
198. B. Yu and H. Li, *Materials Science and Engineering* A325, 215 (2002).
199. L. Feng, S. Li, H. Li, J. Zhai, Y. Song, L. Jiang, and D. Zhu, *Angew. Chem. Int. Ed.* 41, 1221 (2002).
200. J. Bico, C. Marzolin, and D. Quéré, *Europhys. Lett.* 47, 220 (1999).
201. S. Shibuichi, T. Yamamoto, T. Onda, and K. Tsujii, *J. Colloid Interface Sci.* 208, 287 (1998).
202. A. F. Thüemann, *Langmuir* 16, 824 (2000).
203. T. Nishino, M. Meguro, K. Nakamae, and M. Matsushita, *Langmuir* 15, 4321 (1999).
204. P. Kovacic and M. B. Jones, *Chem. Rev.* 87, 357 (1987).
205. G. Grem, G. Leditzky, B. Ullrich, and G. Lesising, *Adv. Mater.* 4, 36 (1992).
206. P. Kovacic and A. Kyriakis, *J. Am. Chem. Soc.* 85, 454 (1963).
207. A. Formhals, U.S. Patent No. 1 975, 504 (1944).
208. J. Doshi and D. H. Reneker, *J. Electrost.* 35 (1999).
209. P. W. Gibson, H. L. Schreuder-Gibson, and D. Riven, *AIChE J.* 45 (1999).
210. D. H. Reneker, A. L. Yarin, H. Fong, and S. Koombhongse, *J. Appl. Phys.* 87, 4531 (2000).
211. I. D. Norris, M. M. Shaker, F. K. Ko, and A. G. MacDiarmid, *Synthetic Metals* 114, 109 (2000).
212. D. T. Bong, T. D. Clark, J. R. Granju, and M. R. Ghadiri, *Angew. Chem. Int. Ed.* 40, 988 (2001).
213. H. Hachisako, H. Nakayama, and H. Ihara, *Chem. Lett.* 1165 (1999).
214. J. H. Esch and B. L. Feringa, *Angew. Chem. Int. Ed.* 39, 2263 (2000).
215. K. Yoza, Y. Ono, K. Yoshihira, T. Akao, H. Shinmori, M. Takeuchi, S. Shinkai, and D. N. Reinhoudt, *J. Chem. Soc., Chem. Commun.* 907 (1998).
216. M. Smith and D. E. Katsoulis, *J. Mater. Chem.* 5, 1899 (1995).
217. Y. Ishikawa, H. Kuwabara, and T. Kunitake, *J. Am. Chem. Soc.* 116, 5579 (1994).
218. T. Kimura, T. Yamashita, K. Koumoto, and S. Shinkai, *Tetrahedron Lett.* 40, 6631 (1999).
219. N. Mizoshita, T. Kutsuna, K. Hanabusa, and T. Kato, *Chem. Commun.* 781 (1997).
220. X. Luo, C. Li, and Y. Liang, *Chem. Commun.* 2091 (2000).
221. R. Wang, C. Geiger, L. Chen, B. Swanson, and D. G. Whitten, *J. Am. Chem. Soc.* 122, 2399 (2000).
222. X. Luo, B. Liu, and Y. Liang, *Chem. Commun.* 1556 (2001).
223. S. Bhattacharya and Y. Krishnan-Ghosh, *Chem. Commun.* 185 (2001).
224. H. M. Willemen, T. Vermonden, A. T. M. Marcelis, and E. J. R. Sudholter, *Langmuir* 18, 7102 (2002).
225. S. A. Ahmed, X. Sallenave, F. Fages, G. Mieden-Gundert, W. M. Muller, F. Vogtle, and J.-L. Pozzo, *Langmuir* 18, 7096 (2002).
226. H. Ihara, K. Shudo, M. Takafuji, C. Hirayama, and K. Yamada, *J. Chem. Soc., Chem. Commun.* 1244 (1992).
227. M. de Loos, J. H. Esch, I. Stokroos, R. M. Kellogg, and B. L. Feringa, *J. Am. Chem. Soc.* 119, 12675 (1997).
228. H. Hachisako, T. Murata, and H. Ihara, *J. Chem. Soc., Perkin Trans.* 2, 2569 (1999).
229. H. Ihara, M. Yoshitake, M. Takafuji, T. Yamada, T. Sagawa, and C. Hirayama, *Liq. Cryst.* 26, 1021 (1999).
230. F. S. Schoonbeek, J. H. Esch, B. Wegewijs, D. B. A. Rep, M. P. de Haas, T. M. Klapwijk, R. M. Kellogg, and B. L. Feringa, *Angew. Chem. Int. Ed.* 38, 1393 (1999).
231. K. Ariga, J. Kikuchi, M. Naito, E. Koyama, and N. Yamada, *Langmuir* 16, 4929 (2000).
232. M. Kimura, T. Kitamura, T. Muto, K. Hanabusa, H. Shirai, and N. Kobayashi, *Chem. Lett.* 1088 (2000).
233. M. Suzuki and C. C. Waraksa, *Chem. Commun.* 2012 (2001).
234. K. Tomioka, T. Sumiyoshi, S. Narui, Y. Nagaoka, A. Iida, Y. Miwa, T. Taga, M. Nakano, and T. Handa, *J. Am. Chem. Soc.* 123, 11817 (2001).
235. M. de. Loos, J. H. Esch, R. M. Kellogg, and B. L. Feringa, *Angew. Chem. Int. Ed.* 40, 613 (2001).
236. G. Wang and A. D. Hamilton, *Chem. Eur. J.* 8, 1954 (2002).
237. S. Laan, B. L. Feringa, R. M. Kellogg, and J. H. Esch, *Langmuir* 18, 7136 (2002).
238. C. F. Nostrum, S. J. Picken, A.-J. Schouten, and R. J. M. Nolte, *J. Am. Chem. Soc.* 117, 9957 (1995).
239. U. Beginn, G. Zipp, and M. Moler, *Adv. Mater.* 12, 510 (2000).
240. J. Loisea, M. Lescanne, A. Colin, F. Fages, J.-B. Verlhac, and J.-M. Vincent, *Tetrahedron* 58, 4049 (2002).
241. Y. Lin and R. G. Weiss, *Macromolecules* 2, 414 (1987).
242. H. J. Tian, K. Inoue, K. Yoza, T. Ishi-I, and S. Shinkai, *Chem. Lett.* 871 (1998).

243. J. H. Jung, H. Kobayashi, K. J. C. Bommel, S. Shinkai, and T. Shimizu, *Chem. Mater.* 14, 1445 (2002).
244. J. Massey, K. N. Power, I. Manners, and M. A. Winnik, *J. Am. Chem. Soc.* 120, 9533 (1998).
245. A. Ajayaghosh and S. J. George, *J. Am. Chem. Soc.* 123, 5148 (2001).
246. Y. Tsabba, D. M. Rein, and Y. Cohen, *J. Polym. Sci. B.* 40, 1087 (2002).
247. N. Amanokura, Y. Kanekiyo, S. Shinkai, and D. N. Reinhoudt, *J. Chem. Soc., Perkin Trans.* 2 1995 (1999).
248. M. S. Vollmer, T. D. Clark, C. Steinem, and M. R. Ghadiri, *Angew. Chem. Int. Ed.* 38, 1598 (1999).
249. K. Hanabusa, M. Matsumoto, M. Kimura, A. Kakehi, and H. Shirai, *J. Colloid Interface. Sci.* 224 (2000).
250. A. Carre, P. L. Grel, and M. B. Floch, *Tetrahedron Lett.* 42, 1887 (2001).
251. A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov, and N. Boden, *Proc. Natl. Acad. Sci. U.S.A.* 98, 11857 (2001).
252. J. H. Esch, R. M. Kellogg, and B. L. Feringa, *Tetrahedron Lett.* 38, 281 (1997).
253. K. Inoue, T. Ono, Y. Kanekiyo, T. Ishi-i, K. Yoshihara, and S. Shinkai, *J. Org. Chem.* 64, 2933 (1999).
254. F. S. Schoonbeek, J. H. Esch, R. Hulst, R. M. Kellogg, and B. L. Feringa, *Chem. Eur. J.* 6, 2633 (2000).
255. M. de Loos, A. G. J. Ligtenbarg, J. H. Esch, H. Kooijman, A. L. Spec, R. Hage, R. M. Kellogg, and B. L. Feringa, *Eur. J. Org. Chem.* 3675 (2000).
256. H. M. Willemen, T. Vermonden, A. T. M. Marcelis, and E. J. R. Sudholter, *Eur. J. Org. Chem.* 2329 (2001).
257. M. Kolbel and F. M. Menger, *Chem. Commun.* 275 (2001).

www.aspbs.com/enn

Self-Assembled Nanostructures at Silicon Surfaces

D. Y. Petrovykh

University of Maryland, College Park, Maryland, USA and Naval Research Laboratory, Washington, DC, USA

F. J. Himpsel

University of Wisconsin, Madison, Wisconsin, USA

CONTENTS

1. Introduction
2. Macroscopic Templates
3. Self-Assembly and Heteroepitaxy on Silicon
4. Self-Assembled Nanostructures on Vicinal Surfaces
5. Molecular and Atomic Self-Assembly
6. Nanostructures and Biology

Glossary

References

1. INTRODUCTION

1.1. Self-Assembly—A Key Aspect of Nanotechnology

In recent years the field of nanotechnology has experienced a veritable explosion of ideas, which resulted in many potential applications for nanostructured materials. In fact, nanostructures are seen as the key that will enable practical devices in the future of information storage and processing, communications, and biotechnology [1–3]. Since these are part of the nano-, bio-, and information technology triad [4] that is expected to power a multi-prong industrial revolution of the 21st century [5–9], the importance of nanostructures would be difficult to overestimate.

The key feature that nanostructures offer for such a diverse range of potential applications is the ability to tailor the electronic [10–21], optical [10, 12, 17, 19, 21–23], and magnetic [14, 17, 24–26] structure and properties of materials. Coincidentally for a variety of properties, this ability emerges at essentially the same length scale in the single-digit nm range [18, 27]. If devices are required to operate at room temperature, the presence of thermal fluctuations on the order of $kT = 25$ meV sets the energy scale. Transport and confinement of electrons determine most of the properties of interest, so the appropriate scales of nanostructures can be estimated for a single electron with mass m and charge e.

Quantum confinement. The quantum-mechanical energy difference between the lowest two quantum well states for an infinite potential well of width l is

$$E_{QM} = \frac{3h^2}{8ml^2} \tag{1}$$

If we require $E_{QM} \geq kT$ to prevent thermal excitations, the width of the well must be $l \leq 7$ nm.

Single-electron "electrostatics." The Coulomb energy for a single electron on a sphere of radius r surrounded by medium with a dielectric constant ε is

$$E_C = \frac{e^2}{\varepsilon \cdot r} \tag{2}$$

Using ε of silicon and $E_C \geq kT$ requirement again, we obtain a limit for the diameter of that sphere (nanoparticle) $2r \leq 9$ nm.

The superparamagnetic limit. The minimum size of a particle used in magnetic data storage is determined from the requirement that its magnetization cannot be changed by thermal fluctuations. The estimate in that case involves a model more complicated than the above ones, but the result is about 3 nm [25, 27].

Another aspect common to most potential applications of nanostructures is that typically more than one such structure is required; ordered nanoscale assemblies [17, 23] of

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

size-controlled structures are usually preferred for optimal performance. Perhaps the most stringent requirements for relative and absolute positioning are inherent for devices based on just a few nanostructures, for example, single-electron transistors or collections of qubits for quantum computing [15, 28–30], because interactions between the nanostructures need to be precisely controlled [29] and registration with the next level of the device (e.g., contacts) has to be maintained. On the other hand, for some optical and electronics applications no coherence is required, and thus the high packing density and size uniformity of nanostructures become more significant than their ordering [19, 31]. For the ultimate data storage media [14, 16, 25] or for materials with novel properties [10, 17, 23, 32, 33], one needs both high density and precise positioning over macroscopic areas and/or volumes. Overall, whether for manipulating the interactions or for increasing the packing density, the positioning of the nanostructures needs to be controlled on the scale comparable to or smaller than the size of the structures themselves.

Both defining and positioning structures with sub-10 nm accuracy is beyond the limits of traditional photolithography, so accordingly a large number of alternative lithography techniques [1, 3] are being currently developed. There is also a considerable effort in increasing the throughput [34, 35] of electron beam lithography, a serial method used to prepare masks for a large number of alternative lithography approaches [1, 3]. Increased availability and quality of scanning probe microscopy (SPM) instruments in recent years propelled the development of a number of SPM-based lithography approaches [1, 3] that allow to produce complex surface architectures from constituents as small as individual atoms [36]. Notably, the dip-pen nanolithography (DPN) [37, 38] routinely allows to achieve sub-100 nm resolution [38, 39]. The inherent advantage of using an SPM-based approach is that the structures can be made, imaged, and corrected by the same instrument. Current SPM instruments, however, use a single probe tip to define the structures, so the writing speed limits the applications to making small arrays of nanostructures for research, or perhaps using SPM instead of a costlier electron beam writer for mask definition. But the progress in development of probe-tip arrays [40] suggests that some of the speed limitations can be overcome by writing with multiple tips in parallel [38].

Even with the impressive advances in sub-100 nm fabrication techniques [3], it is evident that all their inherent limitations become more difficult to negotiate when the size of the nanostructures falls below 10 nm—the range that holds the most potential for achieving the novel phenomena important for both basic research and practical applications. This is one of the primary reasons why a self-assembly approach becomes a more attractive alternative for creating and controlling nanostructures.

In self-assembly, natural interactions between atoms, nanostructures, and their environment produce kinetically or thermodynamically preferred structures and patterns, so the method is inherently parallel and thus should be able to produce macroscopic amounts of materials. An important distinction needs to be made between self-assembly and self-organization [41]. Hereafter, the spontaneous formation of structures with sizes on the order of 10 nm and a well-defined size distribution will be referred to as self-assembly (SA) of nanostructures. A natural tendency to form ordered assemblies will be referred to as self-organization (SO). In general, SA is required to produce large numbers of individual particles with unique and/or tailored properties, while SO provides a pathway for combining nanostructures into novel materials, or for establishing connections between nanostructures and the macroscopic world. Thus in general, systems that possess both self-assembly and self-organization properties are desirable to realize the full potential of nanotechnology.

1.2. Self-Assembly on Surfaces

There are several compelling reasons why surfaces became and will remain an important part of the nanotechnology "playground" [2]. Surfaces offer a quasi-two-dimensional (2D) environment ("quasi-2D" because usually more than one atomic layer is involved even in surface processes) and thus represent macroscopic objects with somewhat reduced complexity. Since the overwhelming majority of nanofabrication techniques [1, 3] are surface-based, there is a natural synergy between basic and applied research into nanoscale surface processes. Most lithography-based approaches [1, 3], as well as molecular beam epitaxy (MBE) and its variants [42], allow to create 3D structures by stacking multiple layers [10, 17], but even in those cases, the surface processes on the topmost layer play the key role during growth.

The practical importance of surface processes is partly responsible for the development within the last few decades of a large number of surface characterization techniques [43]. Numerous microelectronics applications and wide availability of high-quality single-crystal substrates made semiconductor surfaces, especially those of silicon, perhaps the most studied and the best understood surface systems. Scanning tunneling microscopy studies [44, 45] coupled with advanced theoretical modeling [46] have been particularly fruitful in elucidating details of silicon surface structures [47–50] with real-space atomic resolution, and recent instrumentation developments may soon allow surface electronic transport measurements on the same scale [51, 52].

Surface atomistic processes ultimately define the dynamics [53] and thermodynamics [54–56] of nanostructures on surfaces and thus their self-assembly. On a slightly larger scale, surface stress [53, 54] is another factor crucial in both self-assembly and self-organization of nanostructures on surfaces. Surface defects, in particular atomic steps [57], extend self-organization up to macroscopic scales [10, 27, 41, 58]. This availability of multiple scales of physical processes and interactions is extremely important for achieving directed self-assembly and self-organization, since different parameters during the growth and post-processing of nanostructures can be independently controlled to selectively enable particular mechanism(s) to be dominant at appropriate stages. On the other hand, the interplay between several mechanisms is often more beneficial than each one of them acting separately. In the following we provide a number of examples of self-assembly and self-organization on silicon surfaces, with a particular focus on the role of the multiscale phenomena as means of controlling the evolution of these systems.

1.3. Outline

This review proceeds by considering the processes and methods on a continuously decreasing scale, to reflect the multi-scale nature of the self-assembly and self-organization phenomena. In Section 2 the idea of using templates to guide SA and SO is introduced at a relatively coarse (macroscopic) scale. The templates themselves can be either artificially patterned (Section 2.2), for example, by photolithography, or a product of a SO process, for example, atomic step arrays (Section 4.1), or a combination of both approaches (Section 2.2.3). Thus, the formation and applications of such templates are important issues in their own right. Section 3 considers nanoscale processes associated with SA of nanostructures, and role of surface strain in forming nanostructures and interactions between them. Sections 4.2–4.4 demonstrate how nanoscale self-assembly can be combined with self-organized templates to produce ordered arrays of uniform nanowires and nanodots. In Section 5 molecular and atomic structures are considered, with surface reconstructions as SO atomic templates with well-defined chemical and structural properties. These can be used to guide SA of elemental and molecular adsorbates (Section 5.1), to produce low-dimensional structures with novel properties (Section 5.2), and to test concepts of devices on atomic scale (Section 5.3). In addition to their many potential device applications in electronics, macroscopic templates with nanoscale features also represent structures that are both relevant for interfaces with biomolecules and can benefit from bio-inspired assembly strategies. Since such biointerfaces are the cornerstone of the emerging synergy between the nano- and biotechnology, their discussion in Section 6 offers a fitting outlook into one of the most important future pathways opened by surface-based self-assembly.

2. MACROSCOPIC TEMPLATES

The templates considered in this section are (at least conceptually) macroscopic in size—a feature important for handling and integration with existing devices, and for ability to produce macroscopic amounts of SA and SO nanostructures. The lateral periodicity of these templates is in the $\approx$0.5–1.0 μm range, a size range that has several important implications. First, a wide variety of patterning methods can be used to define such templates. Second, there is the possibility of interfacing with microelectronics both in terms of matching scales and common processing methods. And third, one is able to produce nanostructures from "coarse" sub-micron scale templates through size-reduction, SA, and SO.

Surface templates are briefly introduced in Section 2.1, followed by an in-depth discussion of a large class of prepatterned surface templates in Section 2.2. Subsection 2.2.1 outlines the direct patterning and size-reduction approaches that can be used to form SA and SO nanostructures based on micron-scale templates. Subsections 2.2.2 and 2.2.3 describe the use of common processing techniques, such as etching and heating, in conjunction with prepatterned templates, and the types of SA and SO nanostructures that can be achieved by these approaches.

2.1. Templates and Interfaces

The template-based self-organization approaches are likely to be necessary in conjunction with almost any of the self-assembly methodologies under development today. From the practical perspective, the output of solid-state devices is based on macroscopic circuitry, and thus connections and alignment between the nanoscale elements and the appropriate "leads" must be provided; similar "bridging" of the gap in scales is needed in many other applications [3, 59]. To obtain arbitrary and complex patterns, the self-organization driven by properties of individual self-assembled particles may be aided by the top-to-bottom "instructions" provided by the template. Incidentally, the alternative view of such a process is that of a template as a basis for bottom-up assembly of simple structural elements, for example, by ensuring appropriate anchoring of molecular electronic components.

Conceptually, the templates used for surface patterning can be subdivided into two types: external and surface templates. The *external templates* are those prepared independently of the surface being patterned. The vast variety of noncontact masks for light and e-beam lithography [1, 3] fall into this category, and their discussion in any detail is beyond the scope of this article. External templates used as masks for directed atomic or molecular deposition are typically directly patterned, rather than SO or SA.

There are some examples of self-organized external masks, however. Atomic deposition through a standing light wave pattern [60–62] can be used to create arrays of uniformly spaced lines (e.g., for use as metrology standards [63]) and potentially more complex patterns [64]. Similar light wave patterns can also be created by mixing the incident laser beam with light scattered off the surface, for example, during pulsed-laser irradiation of silicon substrates [65]. These laser pulses create silicon nanoparticles, which under appropriate conditions adopt increasingly more ordered configuration and thus result in more ordered standing wave patterns. Therefore, even though a nominally external light wave template is used, because of the interactions between the template and the surface structures, SO arrays of SA nanoparticles can be created [65].

The rest of this section is focused on *surface templates* with macroscopic features. As the name implies, surface templates are typically formed by surface processes and thus are themselves subject to self-assembly, self-organization, as well as the more traditional direct patterning approaches, or any combination thereof.

2.2. Prepatterned Surface Templates

Perhaps the simplest way to prepare a surface template is to define the desired pattern using one's method of choice and then selectively add or subtract material(s) to form the structures. The deposited material may be in form of preassembled nanostructures (e.g., clusters), but, strictly speaking, deposition onto a prepatterned surface involves only rudimentary self-organization, that is, via direct selectivity of the deposition process to the predefined areas in the pattern. The caveat of this straightforward approach is that the chosen pattern definition method must be capable to produce structures at the required resolution [1, 3].

E-beam lithography allows direct patterning of submicron features, for example, to selectively grow InAs quantum dots within 100–200 nm wide lines [66], or to form laterally ordered arrays of $Si_{1-x}Ge_x$ nanostructures with similar spacing [67]. E-beam patterning of 2D arrays, however, is a time-consuming process for resolution below 100 nm, so alternative patterning techniques are being explored. Nanoimprinting [1, 3, 68, 69] has been used to define a 2D array of Si mesas with 250 nm pitch and 10–60 nm dimensions for ordered growth of Ge nanodots [70]. In this case, remnants of the mesas (after annealing to clean the sample surface) acted as nucleation centers for growing Ge nanodots. Since no mask, for example, oxide or resist, was used to ensure the selective deposition, the growth and nucleation of the Ge dots was in this case largely controlled by Ge surface diffusion. Thus to obtain an array with dots nucleated on each mesa and only on the mesas, a higher density array would be necessary [70]—a requirement difficult to fulfill using the direct patterning approaches.

"Size reduction" methods [3] can be used to extend the patterning resolution. Two of the most common reduction approaches are described in Section 2.2.1: shadow deposition and edge-selectivity. Although only relatively simple patterns and structures can be produced by direct patterning, their complexity (but generally not the surface number density) can be increased by postprocessing, for example, etching (Section 2.2.2) or heating (Section 2.2.3).

2.2.1. Size Reduction

Shadow deposition is an approach that achieves size reduction through the use of geometric factors, and thus allows deposition of periodic patterns of structures with sizes down to 10–100 nm size using micron-scale masks. *Nanosphere lithography* (NSL) is an example of a shadow deposition (or "natural lithography") technique [71, 72] that provides an excellent illustration of the main principles of the method (Fig. 1a). The small commercially available "nanospheres" that form the basis of NSL are typically about 500 nm in diameter, since their primary applications in biotechnology require optical observation and handling. These nanospheres can be suspended in solvents, and, while in suspension, spin or drop-coated on essentially arbitrary surfaces. As the solvent evaporates, the spheres are drawn together by capillary forces to form hexagonally close-packed arrays (Fig. 1a) with 10–100 μm^2 defect-free domains, which can be used as masks for deposition or etching of dots or triangles on 10–100 nm scale [72]. There are three modifications of NSL that push the size limit down into the 1–10 nm range: masks of smaller (<100 nm) nanospheres, double-layer (DL) masks, and angle-resolved (AR) NSL. Monodisperse smaller nanospheres are not as readily available and their assembly is harder to control, but both DL masks and AR NSL not only produce the smaller structures, but also allow variations on the simple hexagonal array motif [72]. In particular, AR NSL [72, 73], which uses deposition at off-normal angles (Fig. 1a), can produce nano-overlap and nanogap structures, as well as nanoparticle chains, because the size and position of the shadowed area continuously vary with the angle [72].

Another type of a template that can be used for size reduction via off-normal shadow deposition is a V-groove

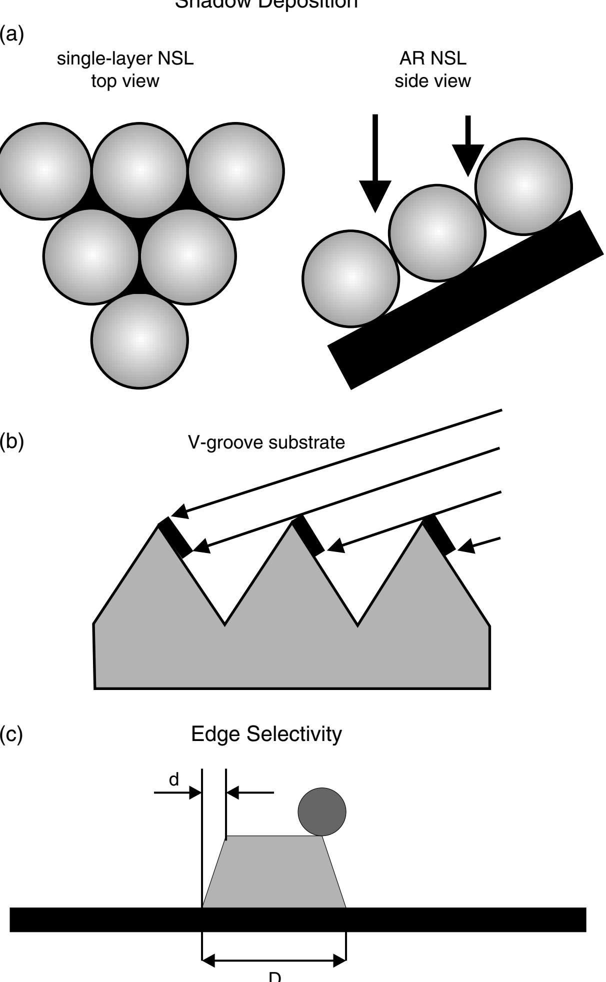


Figure 1. Common methods of size reduction. In shadow deposition, a specific geometry is used to produce features smaller than the original template, e.g., (a) an array of nanospheres [72], (b) a V-groove substrate [74]. Edge selectivity (c) of deposition or diffusion can also be used to confine nanostructures to edges of a pattern [75–77], which effectively achieves higher lateral resolution than the original patterning method.

pattern (Fig. 1b), for example, produced by combination of laser interference and anisotropic etching [74]. The steep ≈60° angle of the V-groove facets means that during a grazing angle evaporation neighboring ridges shadow all but the topmost area on each facet (Fig. 1b). Metal wires with 20 nm width have been thus achieved—a factor of 10 reduction from the original ≈200 nm V-groove periodicity [74].

A second commonly used method of size reduction is based on the fact that for any structure (Fig. 1c) the width of its edges d is always smaller than the size of the structure itself D; thus size reduction by a corresponding factor of D/d can be achieved via *edge-selectivity* of the deposition (or postprocessing). For example, Cu deposited on Si surface, patterned by conventional microlithography with lines of photoresist (2–5 μm width), selectively nucleated into sub-micron Cu clusters on the edges of the photoresist lines [75]. Because the Cu clusters were confined to edges of the pattern, the effective resolution (≈150 nm) in this case was better than the (≈1 μm) resolution of the patterning

technique [75]. In a similar manner, silicon mesa structures on Si(001) have been used as templates for growing linear arrays of self-assembled Ge islands. Because the islands aligned either along the edges of the top facet [76], or along the side facets of the mesas [77], the effective resolution of the patterning was beyond that of the conventional lithography used to define the mesa structures.

2.2.2. Etching of Patterned Surfaces

The resolution and speed limitations of direct patterning suggest that the patterning can be faster if only unstructured nucleation or anchoring points are patterned instead of the full structure. The nucleation points can then be transformed into more complex structures and patterns via postprocessing. Because some form of etching is always used in conjunction with patterning, a combination of an ordered 2D pattern and anisotropic etching is one of the most straightforward applications of this approach.

Nanopits, with facets determined by the etch anisotropy for different crystal orientations, can be formed by an anisotropic etch through windows opened by e-beam exposure of a protective oxide layer (Fig. 2a) [78]. Arrays of such "concave nanopyramids" have been demonstrated with densities up to 1.5 Terapits/in^2 (8 nm pit size, 20 nm pitch), but for the smallest pits uniformity was rather poor (determined by variation of the protective oxide thickness) [78]. One of the benefits of this approach is that the action of the e-beam was only required to open the windows in the oxide, and thus if an alternative means (e.g., DPN [37, 38] or nanoimprinting [68, 69]) can be employed, the bottleneck in speed associated with e-beam writing may be avoided. The anisotropic etch has also been used to produce nanopyramids (Fig. 2b) and grids [79, 80], but in this case the effect was due to a decrease in the etch rate with increasing ion-beam exposure, and thus the rate and the possibilities for scaling up were limited by the required minimum exposure per point. One can envision, however, using such an approach to define masters for soft lithography [1, 3], for example, nanoimprinting [68, 69].

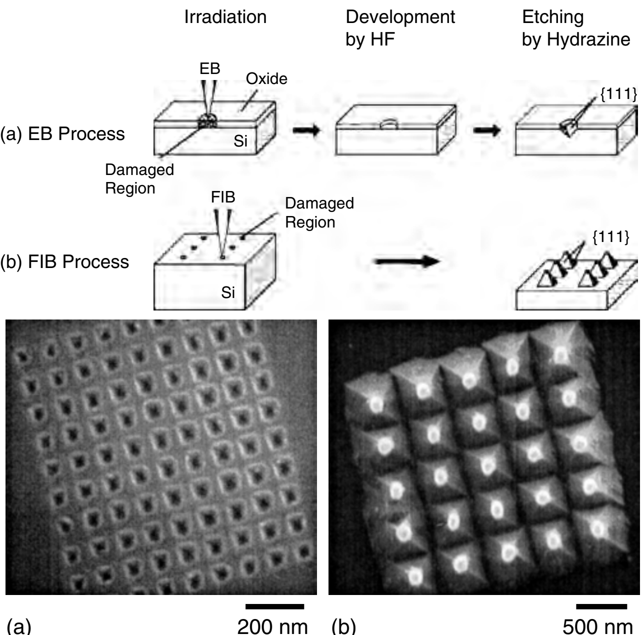


Figure 2. Nanopyramid arrays formed in Si(001) by (a) hydrazine etch through e-beam patterned oxide and (b) using focused-ion beam exposure to retard Si etch rate. Both the concave (a) and convex (b) nanopyramids are formed by ⟨111⟩ facets because anisotropic etch rate in hydrazine is lower for (111) than for (001) plane. Adapted with permission from [80], M. Koh et al., *Appl. Surf. Sci.* 162, 599 (2000). © 2000, Elsevier Science.

2.2.3. Heating of Patterned Surfaces

High-temperature annealing is often applied to semiconductor surfaces before processing, typically for surface cleaning. The requirement for an ultra-high vacuum (UHV) environment limits possible applications of any patterning method involving such a step; it is possible, however, to provide uniform heating across a whole Si wafer, so the limitation is essentially that of a batch mode processing. The transformation of Si surfaces during heating is one of the most in-depth-studied aspects of Si processing and the associated processes are, in general, well understood [57, 58, 81–84]. This understanding results in the ability to control the evolution of Si surface structures extremely well. For example, with properly adjusted surface temperature and Si deposition rate, arrays of atomically flat Si mesas tens of microns across can be created on lithographically patterned surfaces [85]. Similarly, a combination of MBE growth and heating has been used to produce nanostructures on patterned GaAs surfaces via preferential migration of material onto selected facets of micron-size holes or mesas [86].

When a silicon surface is heated in UHV to a high enough temperature, some silicon atoms can evaporate. Because of their weaker bonding, atoms at low coordination sites, such as edges of surface steps and islands, begin to evaporate at lower temperature than atoms incorporated in defect-free surface planes, for example, (111) or (001). When a surface is heated just enough to enable evaporation of the atoms with low coordination, this leads to the retreating motion of all atomic steps on that surface. On top of a mesa a few microns in size, this motion can eventually eliminate all the steps and achieve an atomically flat surface [85]. Conversely, steps at a bottom of a deep micron-size hole will retreat to form an atomically flat plane at its bottom [81]. In both cases, once the flat surface has formed, the evaporation rate from that area dramatically decreases [81, 85, 87] because all the remaining atoms have higher coordination than those at defect sites.

Step motion due to the selective evaporation of material from step edge offers many possibilities for self-organization of templates. For example, an array of holes patterned into a vicinal surface (Fig. 3a) stays put during step evaporation, but the multiple steps around the holes keep moving until they get pinned by the edges of the holes, and thus an ordered array of holes is transformed into an ordered array of bunched steps [81, 87, 88]. The benefit of combining the self-organization of the steps with a simple initial pattern is the possibility to produce intricate step bunch

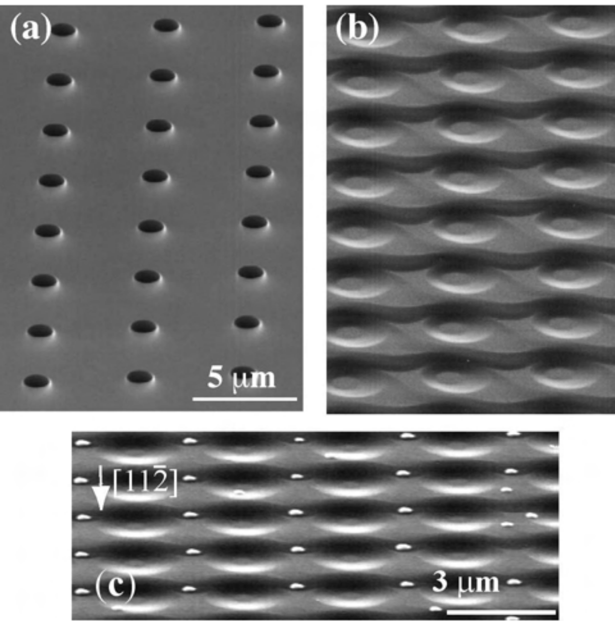


Figure 3. Step-band networks on Si(111) as templates for controlled Au island formation. The templates of ordered step bunches can be prepared from an initial hole pattern (a) by heating in UHV for 1–2 min at 1200–1300 °C (b). Examples of 0.8-μm holes patterned on 1.5° miscut Si(111) with the hole pattern rotated 7° azimuthally from the miscut direction are shown in (a)–(b). Deposition of a few ML of Au on the step-band templates, followed by annealing for 20–30 min at 400–600 °C, results in an ordered array of Au particles (c). The example shown in (c) started with a pattern of 0.6-μm holes. Adapted with permission from [90], Y. Homma et al., *J. Appl. Phys.* 86, 3083 (1999). © 1999, American Institute of Physics.

patterns (Fig. 3b) [88]. Such step bunch patterns can then be used as templates for selective deposition, for example, growth of GaAs on Si(111) vicinal surfaces, where site selectivity can be obtained via desorption or diffusion depending on the substrate temperature [88]. More features can be introduced into this type of a template if the surface is not heated long enough for the holes to completely disappear (fill-in). The resulting regular arrays of step bunches and holes can be used for selective deposition of metals (e.g., Au and Ga) and semiconductors (e.g., GaAs and Ge), which after appropriate annealing form ordered arrays of islands (Fig. 3c) [89–91]. A similar method has been also used to form device-quality arrays of quantum wires and dots on patterned high-index GaAs substrates [92–94].

The approach of combining lithography-defined patterns with self-organization on vicinal surfaces is thus particularly attractive, because it potentially offers a way to integrate self-assembled and self-organized nanostructures with micron-scale patterns and structures for device applications [92, 95]. This is also one of the few methods that produces ordered arrays of metal nanoclusters on Si surfaces (Fig. 3c), which can be used to define Si nanopillars through vapor-liquid-solid growth [90] (see also Section 3.2.1) or reactive-ion etching [96–98]. Conversely, because lithography is used for patterning, the inherent limitation of this process is the relatively low density (spacing on the order of 500 nm and above) of nanostructures in the resulting arrays (Fig. 3c).

3. SELF-ASSEMBLY AND HETEROEPITAXY ON SILICON

Heteroepitaxy is one of the main routes to growth of self-assembled nanostructures on silicon surfaces. Specifically, molecular beam epitaxy (MBE) [42] is the primary deposition technique used in research under UHV conditions. The two parameters that determine the thermodynamically favored growth modes are considered in Section 3.1: surface and interface energies, and surface strain. Section 3.2 provides examples of application of these general principles to heteroepitaxy of metals and semiconductors on silicon. The effects of the nonequilibrium kinetics and other deposition techniques are also discussed where applicable.

3.1. Growth Modes during Heteroepitaxy

The morphology of the surface formed during heteroepitaxy of lattice-matched materials is determined in the thermodynamic limit by the free surface energies γ of the two materials and the interface [17, 41, 99, 100]. Namely, high surface energy adsorbate will form random disjointed three-dimensional islands and leave the substrate exposed (also called *Volmer–Weber* (VW) growth, Fig. 4). More precisely, the corresponding condition is

$$\gamma_{\text{substrate}} < \gamma_{\text{adsorbate}} + \gamma_{\text{interface}} \tag{3}$$

A low energy adsorbate that satisfies the condition

$$\gamma_{\text{substrate}} > \gamma_{\text{adsorbate}} + \gamma_{\text{interface}} \tag{4}$$

will wet the substrate and thus form a continuous film (also called *Frank–van der Merwe* (FM) growth, Fig. 4). In ideal layer-by-layer growth, the nucleation of higher layers is strongly suppressed until the lower layer is completed. However, for subsequent layers the effective $\gamma_{\text{substrate}}$ is reduced

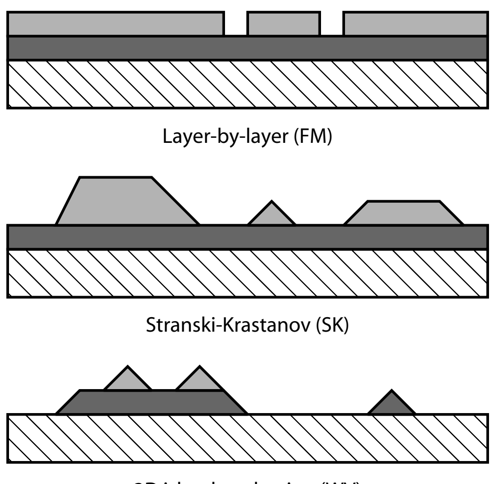


Figure 4. Equilibrium thin-film growth modes. In the layer-by-layer mode, a low surface energy adsorbate wets the substrate with a continuous film. In the SK mode, the first (or the first few) adsorbate layer reduces the substrate surface energy enough to stop the wetting behavior and proceed with nucleation of 3D islands. For an adsorbate with high surface energy, random 3D islands form directly on the substrate.

by the first layer, so instead of continuing in the layer-by-layer fashion the growth mode commonly changes to formation of islands on top of the first layer, resulting in *Stranski–Krastanov* (SK) mode (Fig. 4).

This relatively simple scheme, however, makes two important assumptions that often are not satisfied. First is the assumption of lattice-matched materials, which is in general not true, in particular in cases of technologically relevant material combinations (e.g., Ge/Si). Second is the assumption of a thermodynamically controlled process, which is not always satisfied even approximately (and never exactly, since all experimental processes are of finite duration).

In heteroepitaxy, if the lattice constant of the substrate is larger than that of the film, the resulting strain is tensile. In the opposite case, for example, for Ge/Si, the strain is compressive. In general, the difference in lattice constants results in SK-like growth mode, even if the surface energies favor the layer-by-layer growth, because roughening of the growth front allows to relax the elastic strain that otherwise would build up in the film [10, 17, 41, 99]. If this relaxation is not sufficient, plastic strain relief sets in, for example, creation of misfit dislocations and other defects in the film.

The dominance of kinetic effects versus thermodynamic stability is largely determined by the deposition flux and substrate temperature during growth. If the diffusion of the deposited atoms allows them to sample multiple bonding configurations, the energetically favored sites become populated with higher probability, and the resulting distribution tends towards the thermodynamically stable structures. If, however, either diffusion rate or diffusion length is limited, the deposited atoms are likely to be incorporated into the nearest site (typically an island edge) and thus kinetically preferred structures are formed.

While the variability of the growth modes that occurs in real heteroepitaxy was initially seen as an impediment to the smooth film growth required for device-quality materials, it soon became evident that these effects can be used to either enhance the film quality, or to produce self-assembled nanostructures [17, 41]. For example, if the flux and substrate temperature during the deposition are adjusted to allow the adsorbate atoms to diffuse to reach the nearest substrate step, the so-called step-flow growth mode is realized, which results in very smooth films; for example, homoepitaxy of Si under such conditions is routinely used to grow high-quality Si buffer layers. Additional examples of equilibrium and nonequilibrium driven SA in heteroepitaxy on Si are presented throughout Section 3.2.

3.2. Self-Assembly in Strained Heteroepitaxy

Formation of nanoscale islands via SK or VW heteroepitaxial growth is one of the most widely studied forms of self-assembly on a variety of substrates [17]. On silicon, the substrate of choice for microelectronics applications, the deposition of metals (Section 3.2.1) and semiconductors (Sections 3.2.2 and 3.2.3) has been examined for possible SA island formation. For metals, alloys, and semiconductors, both equilibrium and nonequilibrium processes can lead to SA.

3.2.1. Metal and Metal-Semiconductor Heteroepitaxy on Silicon

Metals tend not to form SA nanostructures when deposited directly on silicon, because of silicide formation and surface energy differences [48]. Thus formation of metallic nanostructures on silicon has to be guided by preferential deposition or diffusion on nanostructured and/or passivated surface templates [17, 27, 101–103], but not by heteroepitaxy-induced strain alone. One notable exception is the growth of several (primarily rare-earth) metal silicide nanowires on Si(001) (Fig. 5), which is driven by the asymmetric mismatch of the respective silicides with the Si lattice constant along orthogonal directions [102, 104–106]. The wires are formed in this case, because the growth proceeds uninterrupted along the direction where the constants match, but is limited in the perpendicular direction. An intriguing property of these wires is that they appear to be stable against atomic step motion during annealing; that is, moving steps get pinned around the wires, but the wires remain uninterrupted (Fig. 5) [102, 106]. While certainly very useful for creating individual long uninterrupted wires, this property, however, limits the possibilities of using stepped surface templates (Section 4.1) to control self-organization of many of such wires (Section 4.3).

Oxidation of metal films on silicon can result in SA metal-oxide nanostructures; in this case the volume change during the oxidation, rather than the lattice mismatch, is believed to be responsible for the SA process [107, 108]. One of the reasons metal oxides have attracted attention is that such SA structures can be magnetic [108] and thus offer a way to combine the benefits of silicon substrates and magnetic devices. SA 3D islands of ferromagnetic metal compounds that are used in semiconductor-based spintronics, for example, MnAs [109, 110], have also been demonstrated on Si(111) with SA achieved via the combination of the lattice mismatch and growth kinetics [110]. More complicated SA schemes are possible, for example, by metal deposition onto SA structures formed by the Ge/Si heteroepitaxy (Sections 3.2.2–3.2.3) which produces metal-semiconductor compound nanocrystals [111].

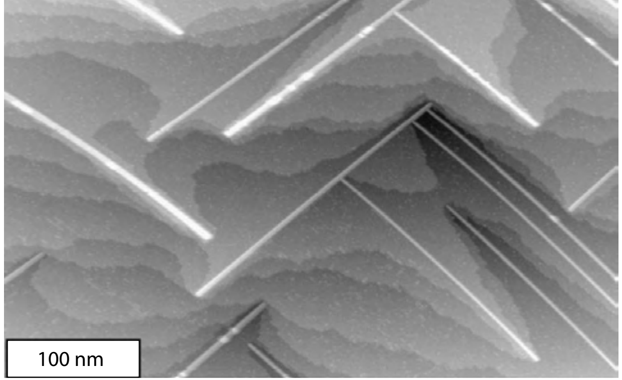
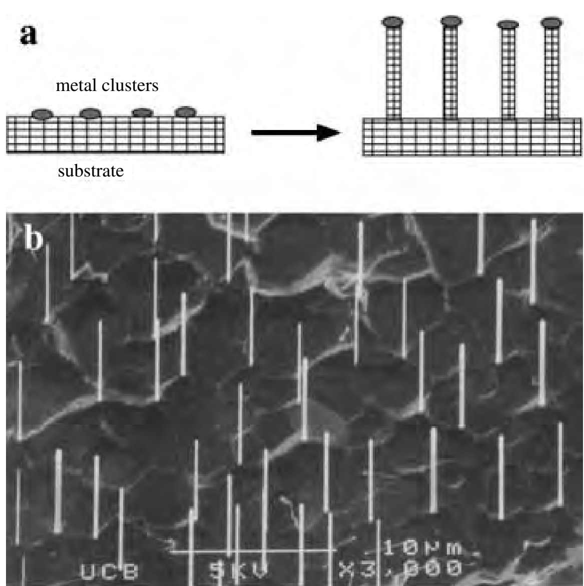


Figure 5. STM image of Ho silicide nanowires on stepped Si(001) surface. Note that multiple substrate steps stretch to accommodate uninterrupted nanowires. Reprinted with permission from [106], J. Nogami et al., *Phys. Rev. B* 63, 233305 (2001). © 2001, American Physical Society.

ZnO is a metal oxide that has attracted attention for optical device applications, because of the exciton binding energy of 60 meV, that is, larger than that of wide-bandgap semiconductors. ZnO SA quantum dots form in VW-like growth mode on silicon surfaces with intermediate layers, for example, silicon oxide [112] and GaSe bilayer [113]. For potential applications, for example, room-temperature short-wavelength nanolaser arrays, the main interest is in large island densities [114], which have been achieved by the vapor-liquid-solid (VLS) growth [115].

VLS [116] is a general method of growing semiconductor nanostructures [117] and nanowires [115, 118, 119] using metal nanoparticles as catalysts (Fig. 6a). VLS has been successfully used to grow bulk unsupported Si nanowires [119, 120] and oriented Si nanowires on Si surfaces (Fig. 6) [115, 121, 122]. Free-standing single-crystal nanowires in general do not form during semiconductor heteroepitaxy (Section 3.2.2), because such anisotropic crystals are thermodynamically unstable relative to bulk structures. The primary role of the metal particles is then to kinetically promote the growth of nanowires [115]. Their secondary role is to determine the diameter of the growing nanowires, which can thus be controlled if size-selected metal nanoparticles are used [115, 119]. The third possible function of the metal clusters is controlling the position of the resulting nanowires, which becomes meaningful if the clusters themselves can be positioned on a surface. During the VLS growth of Si, the metal (e.g., Au) particles partially melt and form a liquid alloy with the Si substrate. During the subsequent Si deposition, the liquid becomes a preferential adsorption site and the alloy quickly supersaturates with Si, which leads to Si precipitation and nanowire growth [115]. Silicon nanowires preferentially grow along the ⟨111⟩ direction, which produces *epitaxial* nanowires with unique orientation on Si(111) (Fig. 6b) and nanowires along the three equivalent ⟨111⟩ directions growing on Si(001) (Fig. 6c) [115].



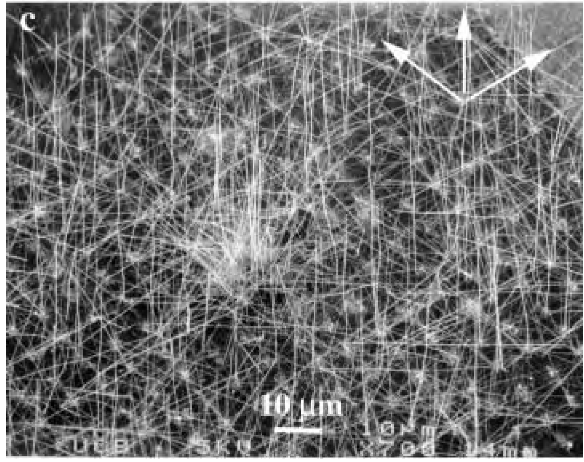


Figure 6. Silicon nanowire growth by vapor-liquid-solid epitaxy (VLSE). (a) VLSE schematic. (b) Vertical Si nanowires grown on Si(111) wafer. (c) Three sets of Si nanowires (preferred orientations indicated by arrows) grown on Si(001). Reprinted with permission from [115], Y. Y. Wu et al., *Chem. Eur. J.* 8, 1261 (2002). © 2002, Wiley-VCH.

VLS is thus a technique of choice for growing densely packed arrays of oriented single-crystal self-assembled Si nanowires. The seed metal particles, which terminate the wires, may be beneficial for some applications (e.g., for chemical selectivity), but also result in essentially unavoidable residual metal content—potential disadvantage for applications where high-purity Si is required (e.g., electronics). The self-assembled nanowires produced by VLS are not inherently self-organized (apart from the preferential orientation on epitaxial substrates), but because of its flexibility in terms of materials and conditions, VLS can potentially be combined with other approaches to produce self-organized structures.

3.2.2. Semiconductor Heteroepitaxy on Silicon: III-Vs and Ge

Semiconductor heteroepitaxy includes the three "canonical" material systems in which SA nanostructures with well-defined shapes and size distributions form under a variety of growth and postdeposition treatments: Ge/Si (≈4% mismatch), InAs/GaAs (≈7% mismatch), and InP/GaAs (≈3.8% mismatch) [17]. In addition to the widely studied elemental semiconductor heteroepitaxy of Ge/Si discussed below, SA of *compound semiconductors* on Si surfaces has been attempted as well, with the idea that the large lattice mismatches [123] that prevent smooth film growth in such combinations can result in SA of 3D islands. MBE of GaAs on Si(001) results in formation of SA islands with nucleation and evolution of shapes and sizes similar to those observed in the Ge/Si system, which suggests that a similar mechanism is responsible for SA [124]. High-quality SA GaAs quantum dots can be also produced by "droplet epitaxy" on GaSe-terminated Si(111) surface [125] or As-terminated Si(001) [126], whereby initially nanometer-size Ga droplets are formed under Ga flux and then they are transformed into GaAs by annealing under As flux [125, 126]. InAs is another III-V semiconductor for which the direct MBE growth on Si(001) results in SA quantum dots [127], but a more common method for growing quantum dots of InAs [128, 129] and other compound semiconductors [130] is to use H-terminated Si(001) substrates.

As already mentioned, Ge/Si is a *prototype for elemental semiconductor heteroepitaxy* that exhibits self-assembly and self-organization. The system is currently seen as one of the most promising for a variety of device applications; for example, see [21, 33, 131–133]. The extensive body of experimental and theoretical work on Ge/Si epitaxy and nanostructures is covered in several recent reviews [10, 21, 41, 54, 134]. In part because they are so widely studied, the mechanisms of SA nanostructures formation and evolution have been very controversial in this system and the debate offers an instructive example of development of the general understanding of the SA process.

Si and Ge have the same crystal structure, but Ge lattice constant is 4.2% larger, which results in compressive strain when Ge films grow pseudomorphically on Si. First, several uniformly strained full atomic layers of Ge form a wetting layer (Fig. 7a). The thickness of this wetting layer is largely determined by the ability of tetragonal distortion in vertical direction to relieve the strain. Since the lateral compressive strain still accumulates with increasing film thickness, the film becomes unstable against long-wavelength corrugation of the surface, a process particularly pronounced during growth on vicinal substrates (Fig. 7c), whereby ripples with more than 10 times the periodicity of substrate steps develop [41, 135]. Note that these structures, because of their large periodicity, can only contain very shallow facets. When such roughness is not sufficient for strain relief, steeper (higher-index) step bunches are introduced, leading to formation of 3D islands (Fig. 7b) with well-defined facets [10, 41].

3.2.3. Ge/Si Island Shapes and Evolution: Kinetics vs. Equilibrium

Similar to the surface roughening with increasing film thickness discussed previously, for classification purposes, the shapes of Ge islands on Si(001) (Fig. 8) can be described within the same framework of introducing progressively higher-index (steeper) facets with increasing island size [41, 54]. For the smallest islands, {105} facets with 11.3°

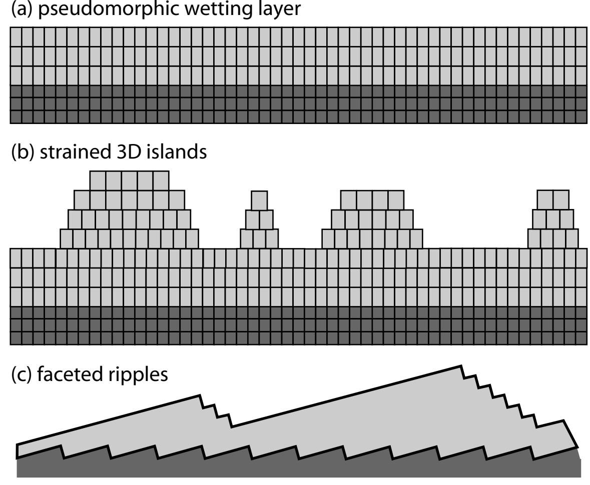


Figure 7. Strain relief in the Stranski–Krastanov growth mode. (a) Pseudomorphic wetting layer with tetragonal distortion. (b) Coherently strained 3D islands with steep facets. (c) On a vicinal substrate, faceting of the film through step bunching.

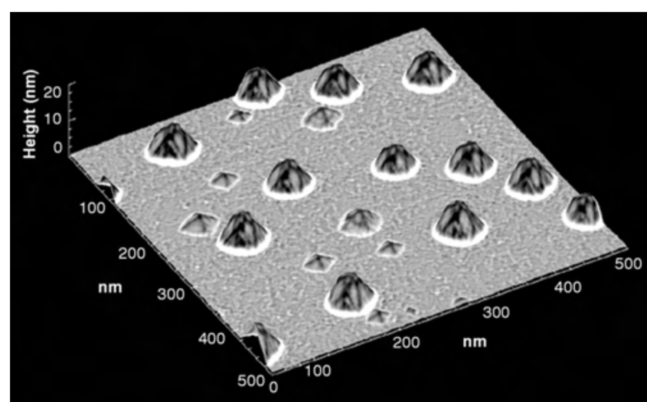


Figure 8. STM topograph of strained Ge nanocrystals on Si(001), showing both pyramids and domes. The gray scale is proportional to the local surface curvature as determined by the Laplacian $\nabla^2 h(x, y)$: positive curvature is white, flat areas are gray, and negative curvature is black. Reprinted with permission from [139], G. Medeiros-Ribeiro et al., *Science* 279, 353 (1998). © 1998, American Association for the Advancement of Science.

tilt relative to (001) plane appear. These islands come in two shapes: square-based *pyramids*, and elongated (in either ⟨100⟩ or ⟨010⟩ direction) "*huts*" with two trapezoidal and two triangular sides [41, 136]. For larger islands two more facets appear: {113} with 25.2° tilt relative to (001) plane and {518}. The latter was reassigned as {15 3 23} with 33.6° tilt [137] through comparison to a stable Ge surface [138]. Multifaceted islands with roughly symmetric bases are referred to as "*domes*" [54, 139] and consist mainly of {113} and {15 3 23} facets, but also contain small {105} and (001) facets. Finally, the largest islands are called "*superdomes*" [54, 140]; they are similar in shape to domes, but contain {111} and other steeper facets near edges.

Explaining the different island shapes (Fig. 8) and their relative stabilities, in an attempt to derive rational methods of growing SA nanostructures, became one of the most controversial problems in the field of SA systems. The huts were the first Ge/Si island type to be discovered [136], but now they are considered to be only metastable, because in annealing experiments huts transform into pyramids [141]. Pyramids and domes, on the other hand, co-exist (Fig. 8) over a wide range of coverage, temperature, growth, and annealing conditions, and thus the growth and shape transition mechanisms of these two island types have been extensively studied and debated for over a decade [10, 41, 54].

The main issue is whether the observed changes in island shapes are the result of a *kinetic roughening process* (e.g., Ostwald ripening [142]) or simply an *equilibrium distribution*, determined by the appropriate island energies. For practical purposes the equilibrium structures are preferred: they inherently evolve towards controlled uniform sizes, potentially can be grown by any deposition method, and are more likely to remain stable over the lifetime of devices (which may be orders of magnitude longer than the duration of the deposition process). If, on the other hand, the island types are determined by growth kinetics, then good understanding of the relative rates of the various processes is indispensable for controlling the appropriate growth parameters and stabilizing the structures (e.g., through surfactants or encapsulation).

Observable distinctions between kinetic and equilibrium processes are readily predicted by simple models. Classic Ostwald ripening, for example, leads to strictly unimodal, albeit not very narrow, particle size distribution and the average particle size growing monotonically with time [142]. For equilibrium models, in a system with several competing interactions, the existence of an optimal particle size (and shape) can be expected. Unfortunately, the actual experimental observations for the Ge/Si system are inconsistent with predictions from simple kinetic or equilibrium models. The problem then becomes that of choosing the appropriate model extensions and including additional surface processes, which in general tends to obscure the differences between model predictions. The development of the complex equilibrium models for this system has been extensively covered recently [54, 143, 144]; the somewhat more illustrative example of the kinetics models is briefly discussed in what follows.

The classic Ostwald ripening model and its extensions to 3D crystals on surfaces [53, 142, 145–148] predict that the chemical potential of an island decreases continuously with size because of the decreasing radius of curvature and surface/volume ratio. The atoms then have a higher probability to *detach* from smaller islands and to *attach* to larger ones; thus the larger islands grow at the expense of the smaller ones, which means that the average island size increases with time and the surface number density of islands decreases. The island size distribution on Ge/Si samples with co-existing pyramids and domes remains bimodal during growth and annealing [54, 141], so it clearly does not correspond to a simple Ostwald ripening. It has been suggested that a discontinuous change in the chemical potential of an island during its shape transformation could modify Ostwald ripening kinetics to produce a bimodal distribution [149, 150]. Quantitatively, the experimental size distribution [149] shows indications of size-limited behavior for both pyramids and domes, and thus narrower than expected width of these distributions. Therefore, an additional island size-dependent term in the adatom attachment rate is required in a kinetic model. Multiple such mechanisms have been suggested [54], for example, the increased strain for large islands at their edges and in surrounding substrates which favor detachment and flow away from these islands respectively [143].

Realistic equilibrium models [54] also have to include multiple energy terms: bulk strain, facet, interface, and edge energies for individual islands [151, 152], as well as interisland elastic interactions and ensemble thermodynamics [54]. As a result of the multiple terms in both kinetic and equilibrium models, their predictions depend on a number of unknown parameters and thus are difficult to compare to each other and experimental data. Other practical aspects of this system further increase the complexity; to name just two: above 650 °C all Ge islands are only metastable with respect to SiGe alloying [153], and transition shapes exist during growth but disappear upon cooling [137].

The general difficulty in assigning exclusively kinetic or equilibrium character to SA island formation has been noted in comparing growth behavior of Ge/Si, Co/Si, and Co/Ge/Si, where "in spite of the marked differences in surface thermodynamics and kinetic pathways... remarkably similar nanocrystal arrays" have been formed on the surface [111]. The main conclusion from the extensive research on the prototypical Ge/Si heteroepitaxy is that while the structures themselves appear to be thermodynamically stable, their growth is determined as much by kinetic pathways towards the equilibrium, as by the equilibrium configuration itself [54, 143, 154]. In other words, partially annealed configurations may exhibit strong nonequilibrium features, but with sufficient annealing distributions consistent with equilibrium theories emerge. One possible "general" scenario was suggested in [143]. Initial island nucleation and 2D–3D transitions are strongly kinetics-dependent. With increasing coverage, strain effects become relevant and island sizes and densities approach equilibrium values, but size-dependent kinetic terms are still significant. As the deposition flux is decreased and stopped, further annealing brings the system close to the equilibrium configuration, including saturated values of island sizes and density.

3.2.4. Ge/Si Multilayers

The ability to form *ordered multilayer* structures during the Ge/Si heteroepitaxy arises from interlayer interactions, which are discussed in detail in several excellent specialized reviews [10, 41, 134]. Here we focus on the importance of the multilayer ordering as a unique self-organization mechanism for self-assembled surface structures. Briefly, during the multilayer Ge/Si heteroepitaxy, nanostructures formed in the preceding layer affect the strain in the subsequent overlayer, which leads to self-alignment of nanostructures between the layers [10, 41, 135, 155]. In fact, these interactions can lead to improved size-uniformity and lateral ordering of the nanodots in the topmost layer even when the first layer was randomly nucleated [10, 41]. Almost perfect uniformity and lateral ordering is achieved when the first layer is prepatterned either artificially [133] or by self-organized template on the substrate [10, 41]. The latter cases offer some of the best practical realizations of truly self-organized arrays of self-assembled nanostructures (Section 4.4.2).

4. SELF-ASSEMBLED NANOSTRUCTURES ON VICINAL SURFACES

Vicinal surfaces and atomic steps are intimately related to self-assembly and self-organization on silicon surfaces. First, the vicinal surfaces themselves are subject to SO. Various aspects of these SO processes are described in Section 4.1: vicinal Si(111) and Si(001) used as substrates, formation of regular arrays of single atomic steps, step bunches, and domain boundaries. The traditional 3D heteroepitaxial growth modes are modified by the presence of multiple substrate steps, because in addition to the respective surface energies γ, substrate and adsorbate step energies σ have to be taken into account (Section 4.2). One of these growth modes, step decoration, based on the preferential adsorbate attachment at steps, provides the most intuitive way of using steps to guide SO (Section 4.3). But additional ordering, in particular in direction along the steps, can be obtained by combining vicinal templates and strained heteroepitaxy, as demonstrated in Section 4.4.

4.1. Vicinal Surfaces as Self-Organized Templates

Atomic steps are always present on macroscopic surfaces, so it is natural to develop self-organization schemes that benefit from their presence, rather than attempt to eliminate them. For example, step flow growth techniques take advantage of stepped surfaces for smooth film growth. Because atomistic processes in step dynamics can be controlled by macroscopic surface parameters, for example, wafer miscut angle, temperature, deposition and annealing rate, etc., there are many approaches that lead to self-organized step arrays. Such SO arrays of steps and step bunches represent surface templates macroscopic in one direction (along the steps) and nanoscale in two other directions (step-step separations and step heights).

Single-crystal silicon wafers with (001) and, to a lesser extent, (111) orientations are widely available for device fabrication and research purposes. Because of the increasingly stringent requirements of the semiconductor industry, production of these wafers is extremely well-controlled in terms of their composition, number of defects, and orientation. Since it is impossible to produce an atomically flat surface across a whole wafer, often it is beneficial to introduce a controlled miscut to produce a more ordered structure. The miscut orientation is usually defined in terms of its azimuthal ϕ (in-plane) and polar θ (out-of-plane) angle with respect to a low-index plane orientation, for example, [001] or [111]. Nominally (001) and (111) substrates are often randomly misoriented by about 0.5° in both ϕ and θ; accuracy in both angles of 0.1° and better is possible for customized substrates. Perhaps some of the best evidence for the ultimate wafer-scale control of the miscut and wafer orientation comes from recent advances in pattern formation by twisted wafer bonding (Fig. 9), with regular nanoscale patterns demonstrated over macroscopic areas [156, 157].

Steps are inherently linear objects, and thus the obvious goal for SO stepped templates is creation of uniform arrays of straight steps. While control of lateral ordering of SA nanostructures is possible even with straight steps (as discussed in Section 4.4.1), templates with 2D patterns are the most direct way to achieve lateral ordering. Similarly to the wafer bonding example (Fig. 9), by choosing a miscut rotated by ≥10° in both ϕ and θ angles from a low-index orientation, 2D patterns of atomic terraces or facets 10–100 nm in size are produced upon annealing. The large miscut in ϕ ensures that two equivalent step orientations act as boundaries of atomic terraces for each of the pattern elements, and determines whether the pattern is symmetric by the relative alignment of the miscut direction with respect to these two step orientations [41, 158]. The misorientation in ϕ in this case is an ordering parameter analogous to the misorientation between the hole pattern and the step direction in case of patterned templates [87, 95] discussed in Section 2.2.3 (Fig. 3). The large miscut in θ is mainly used to induce formation of steep (i.e., with tilt angles comparable to θ) facets [41, 158] which for miscut $\theta < 5°$ can be formed only by additional patterning on a larger scale (e.g., [88, 90]).

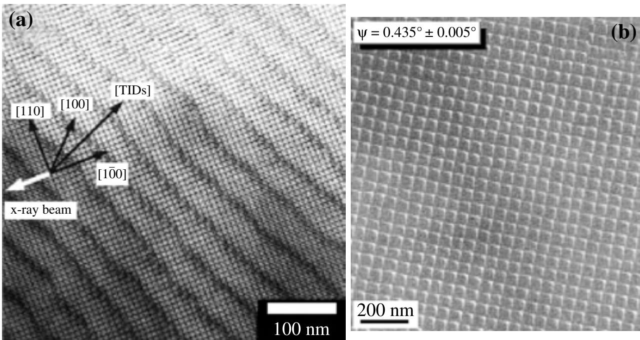


Figure 9. Pattern formation by twisted wafer bonding. TEM images of (a) a quasi-perfect square array of twist interfacial dislocations at the bonding interface, (b) a square array of pure twist dislocations fabricated by wafer bonding with high-precision (0.01°) control of the twist angle. Reprinted with permission from [156], D. Buttard et al., *IEEE J. Quantum Electron.* 38, 995 (2002). © 2002, IEEE.

For silicon surfaces that have been prepared via a high-temperature anneal in UHV, the structure of the atomic steps and facets is intimately related to surface reconstructions [159]. For Si(001) vicinal surfaces with a small polar miscut, this leads to formation of two types of steps: S_A and S_B, with the 2 × 1 reconstruction dimer rows in the upper terrace parallel and perpendicular to the steps, respectively [160–164]. This limits the applicability of Si(001) vicinal surfaces as stepped templates, because the S_B steps are always rough [164, 165]. Moreover, for polar miscut above 2° an increasing fraction of the steps is converted into double-height D_B steps with dimer rows perpendicular to step edges [160, 161]. For 4–5° miscut, full conversion to double-height steps is achieved and the D_B step edges become relatively straight because of step-step interactions, but they are never atomically straight over many lattice sites. Because of the structure of Si dimers, such surfaces have found applications for molecular self-assembly discussed in Section 5.1.2.

The 7 × 7 surface reconstruction, characteristic for Si(111) after a high-temperature anneal, is three-fold symmetric (Fig. 10a), so naïvely one may expect that straight steps would be even harder to form on this surface. However, atomic steps formed on slowly cooled vicinal Si(111) surfaces adopt a preferred displacement of the 7 × 7 reconstruction across the steps, and the terrace widths are quantized in units of the 7 × 7 unit cell (or rather 1/2 of it, as seen in Fig. 10a) [166]. The origin of this effect is in the large size and extreme energetic stability of a 7 × 7 unit cell, which strongly favor steps that preserve its structure [167, 168]. And while kinks on the Si(001) steps can be formed by thermal fluctuations [165], these fluctuations are insufficient to add or remove 1/2 of a 7 × 7 unit cell, leading to essentially atomically straight steps on well-annealed surfaces (Fig. 10a). As shown in Figure 10b, an improper miscut can force kink formation and rough steps on a Si(111) surface. Another potential source of defects on terraces and steps, also seen in Figure 10b, are the 7 × 7 domain boundaries [169, 170], but those can be avoided over micron-size areas by appropriate cooling sequence [171, 172] or otherwise controlled [87, 95]. In fact, because the relevant atomistic processes for Si(111) surfaces are well-understood [57, 58, 168, 173–176], several approaches for creating self-organized templates based on atomic steps on vicinal Si(111) surfaces have been developed, all of them achieved by a

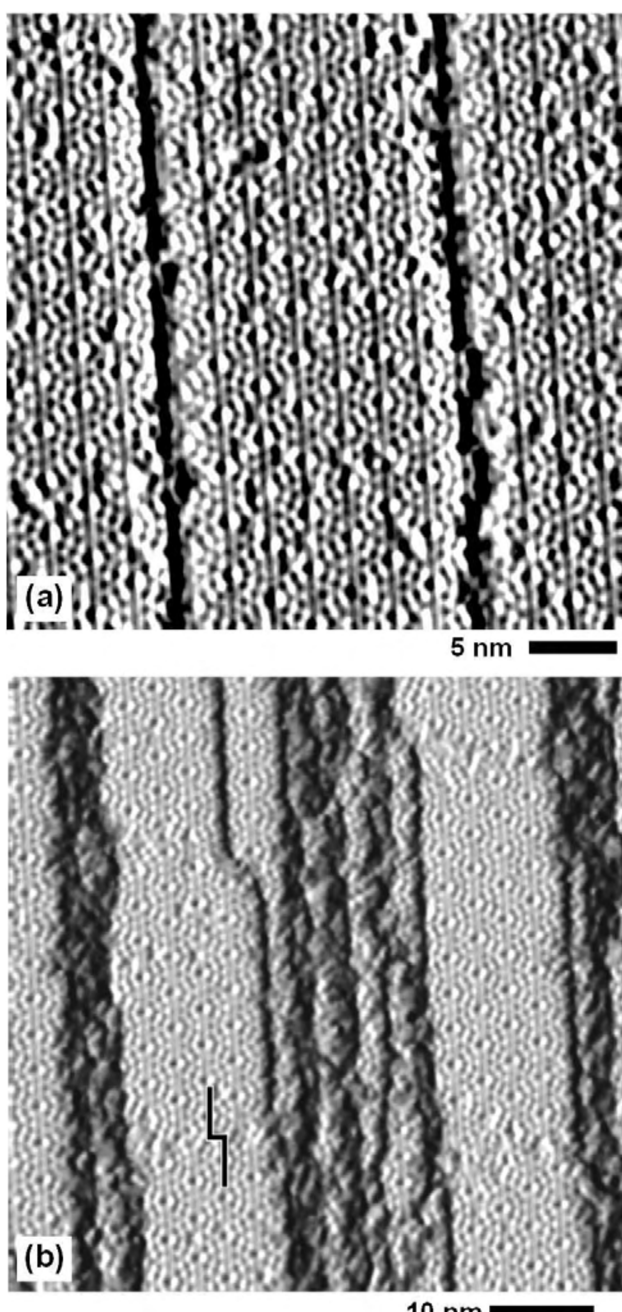


Figure 10. Atomic steps and 7×7 reconstruction on Si(111). (a) Atomically straight step edges (thick dark lines) incorporate rows of 7×7 corner holes. The faint lines parallel to the step edges emphasize the terrace width quantization in units of half a 7×7 unit cell (2.3 nm). (b) Three types of defects for a step array on Si(111): irregular step bunches, kinks along single steps, and 7×7 domain boundaries (jog in the black line). Downhill is to the right. The derivative of the STM topography is used to emphasize the reconstruction and steps in (a) and (b). Adapted in part with permission from [171], J. Viernow et al., *Appl. Phys. Lett.* 72, 948 (1998). © 1998, American Institute of Physics.

combination of controlled heating, annealing, or deposition: uniform arrays of straight atomic steps (Section 4.1.1), arrays of step bunches (Section 4.1.2), and arrays of steps and domain boundaries (Section 4.1.3).

4.1.1. Arrays of Single Steps and Step Interactions

The most stable step orientation on Si(111)7×7 surface is $[\bar{1}\ 1\ 0]$ direction, that is, the one that contains a row of corner holes on the upper terrace (Fig. 10a) [173, 176–183]. Such step edges are possible for two orientations of the surface normal: tilted from [111] towards $[1\ 1\ \bar{2}]$ or $[\bar{1}\ \bar{1}\ 2]$. The steps with the $[\bar{1}\ \bar{1}\ 2]$ tilt are taken as the most stable configuration since they are found to occur during Si/Si(111) homoepitaxy [178]. Atomic step templates with the kink densities as low as one 2.3-nm kink in 2×10^4 atom spacings ($\approx 8\ \mu$m) have been achieved [171] for vicinal surfaces cut 1.1° towards $[\bar{1}\ \bar{1}\ 2]$ (Fig. 11a). In addition to the exceptionally straight steps, another feature of these templates is the single-domain 7×7 reconstruction, which spans entire terraces over micron-sized areas, as evidenced by the lack of fractional kinks (i.e., those smaller than the 2.3 nm

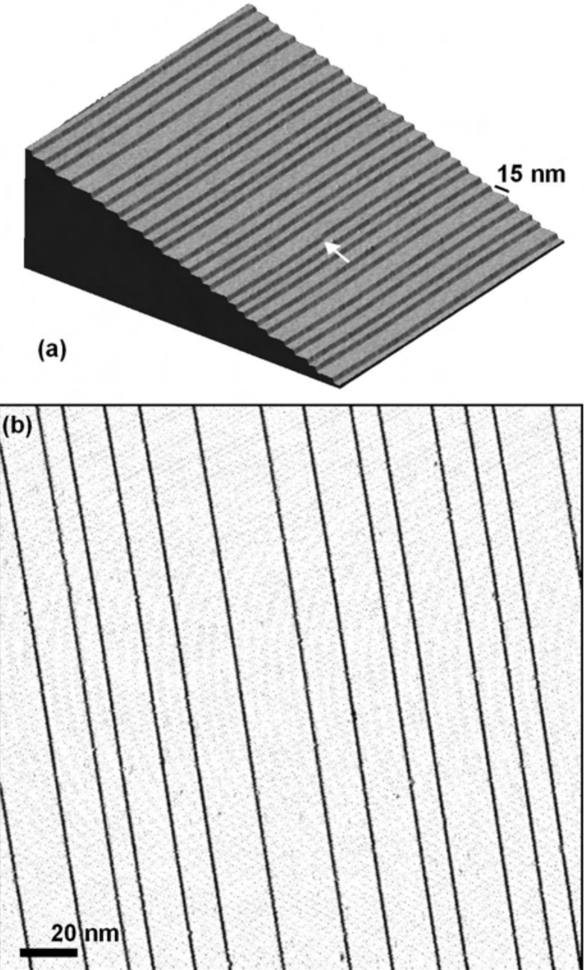


Figure 11. Arrays of single steps on Si(111). (a) Three-dimensional view of a high-quality array of single steps (0.3 nm height) prepared on Si(111) miscut by 1.1° towards the $[\bar{1}\ \bar{1}\ 2]$ direction. There is only one kink (arrow) in the 340×340 nm^2 area; the rest of the steps are atomically straight [171]. The vertical scale is enhanced by a factor of 20. (b) Faint traces of the 7×7 reconstruction pattern and the absence of kinks in this STM topography derivative image indicate that the 7×7 reconstruction is single-domain over the entire image.

1/2 of the 7 × 7 unit cell) [171, 184], direct STM observations of large areas without domain boundaries [171, 184] (Fig. 11b), and high-quality photoemission spectra achieved from such surfaces [172]. The presence of such large areas of single-domain reconstruction reflects the interplay between the 7 × 7 reconstruction helping to stabilize the straight step edges and step edges serving as nucleation sites for the single-domain reconstruction. By appropriately adjusting the preparation conditions, it is thus possible to either create single-domain templates for large-scale anisotropic surface reconstructions (Section 5.2), or SO templates with quasi-2D-periodicity (Section 5.1.3).

The uniformity of the terrace width distributions on the low-miscut surfaces does not match the perfection achieved along the step edges, because several of the possible 7 × 7 = 49 registrations occur between the 7 × 7 domains on adjacent terraces [166, 171, 182, 185]. Among the seven registrations parallel to the step edge, those that preserve the (1 $\bar{1}$ 0) mirror plane symmetry (i.e., those where the rows of corner holes perpendicular to the step edges continue from one terrace to the next) are prevalent (Fig. 10a), and the overall terrace width distribution exhibits a standard deviation of 3–4 nm (≈1–2 quanta). The average single-step separation is determined by the polar miscut angle and 0.31-nm step height; for example, the spacing is approximately 15 nm for 1.1° miscut (Fig. 11a) [171]. The equilibrium terrace width fluctuation is determined by a balance between the entropy cost of straight steps and the elastic energy cost of narrow terraces (i.e., small step separation) [186–188].

Elastic interactions between steps can be generalized according to their symmetry into monopole and dipole character. The monopole term originates from the presence of elastic-stress domains on reconstructed surfaces; on vicinal surfaces atomic steps act as domain walls and the corresponding domain-wall energy has a logarithmic dependence on their separation l [189]. The elastic-dipole interaction between steps arises to compensate for the nonzero moment of the surface capillary forces acting on upper and lower terraces (Fig. 12). The dipole interaction thus can have components perpendicular or parallel to the substrate (Fig. 12) and its energy has l^{-2} dependence [190, 191]. The dipole interaction is repulsive for steps of the same sign (i.e., the same downhill direction), so it helps stabilize step arrays on vicinal surfaces. The terrace width distribution on vicinal Si(111) surfaces is predominantly determined by the dipole interactions between steps [186]. Note that the dipole

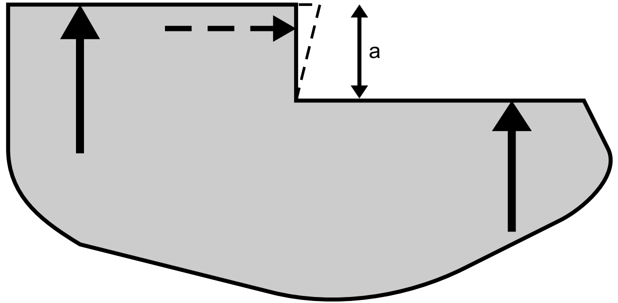


Figure 12. Surface capillary forces acting on the upper and lower terraces produce nonzero moment acting on the step, with components perpendicular and parallel to the surface.

interaction can be attractive in cases where materials with different elastic constants are involved; for example, adsorbate islands can be either attracted or repelled by the steps. When both the monopole and dipole interactions are included, one has to add to the total energy an attractive cross-term with a l^{-1} dependence [192, 193]. The combination of these elastic interactions is responsible, for example, for the minority terrace width distribution on vicinal Si(001) surfaces [194].

For polar miscut of 1–2°, long-range dipole repulsion stabilizes arrays of individual steps on both Si(001) and Si(111). The equilibrium terrace width fluctuations are determined for Si(001) and Si(111) primarily by the monopole and dipole terms, respectively, with fluctuations proportional correspondingly to $l^{1/2}$ and l [187, 188]. Smaller step separation l is thus expected to produce increasingly straight steps, but the resulting larger miscut angles often produce step bunching. On Si(001), as discussed earlier, step doubling produces the D_B steps with dimer rows perpendicular to step edges [160, 161] and in this configuration formation of individual kinks is difficult to suppress over multiple dimer sites. Si(111) steps become unstable against formation of triple steps [176] for polar miscut angles larger than about 2°. An isolated triple-height step is energetically unfavorable compared to three single steps, but with decreasing step spacing the elastic-dipole repulsion increases, and the total energy cost can be lowered by forming the triple steps (with correspondingly larger spacing) [186].

4.1.2. Step Bunches and Facets

Just as for single steps, vicinal surfaces with miscut towards [1 1 $\bar{2}$] or [$\bar{1}$ $\bar{1}$ 2] directions produce better ordered step bunches [184, 195] than other vicinal surfaces [58, 196–198]. In particular, on surfaces with polar miscut towards [1 1 $\bar{2}$], periodic arrays of (111) terraces and high-density facets are formed [184, 195]. Periodicity of these step arrays is independent of the miscut angle up to 6°, but the width of the stepped facets increases with the miscut [184, 195], which in principle should allow to create arbitrary ratios of the flat and stepped regions by adjusting the miscut angle (Fig. 13). Both the perfection and constant periodicity are attributed to the exceptionally stable structure of the facet that forms, which is characterized by an average facet angle of 12.7° [195] and is composed of narrow ledges with reconstructions similar to the 5 × 5 analog of the 7 × 7 reconstruction [184, 195]. It is important to note that because of the large width of the flat terraces and thus significantly reduced step interactions, formation of the regular arrays of bunched steps is much slower than that of the single steps [184].

The bunched step facets are energetically stable, but the step edge perfection of the single steps is lost during their formation. In the range of miscut angles between the (111) and (001), the evolution of stable structures is characterized by several stable planes: (5 5 12), (113), (114), (118), (331), and mixtures of various facets in between [158, 159, 196, 199–201]. Thus, in general, the local structural perfection is recovered at large miscut angles, but not the long-range periodicity of the steps.

There is, however, a number of vicinal Si(111) and Si(001) surfaces that lend themselves for use as self-organized templates (Fig. 14). Particularly with Au as an adsorbate [58],

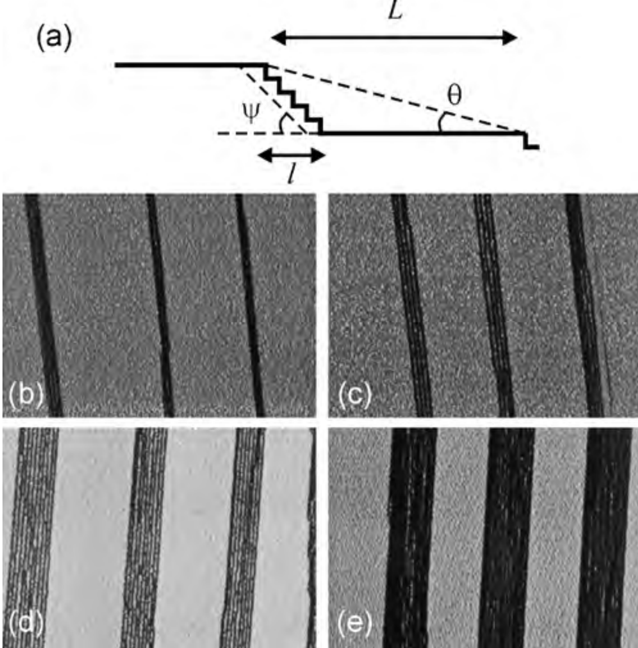


Figure 13. Arrays of step bunches on vicinal Si(111). (a) There are two parameters for uniform arrays of step bunches: the periodicity L defined by the polar miscut angle θ, and the terrace width (L–l) defined by the facet angle ψ of the bunches. For Si(111) miscut towards the $[\bar{2}\ 1\ 1]$ direction, the periodicity and the stable facet angle remain constant for $\theta < 6°$; therefore the terrace width systematically decreases in the miscut series of 1°, 2°, 4°, and 6° (b–e). The derivative of the STM topography is shown to emphasize the stepped regions. Downhill is to the right. Adapted in part with permission from [195], F. K. Men et al., *Phys. Rev. Lett.* 88, 096105 (2002). © 2002, American Physical Society.

a large fraction of the possible orientations has actually produced well-ordered structures, and many other possible crystallographic orientations have yet to be explored. Electronic properties of these surfaces are described in Section 5.2. In terms of its structure, perhaps the simplest is the case of a vicinal surface with 9.45° miscut towards $[\bar{1}\ \bar{1}\ 2]$ direction. It is composed of steps separated by (111) terraces

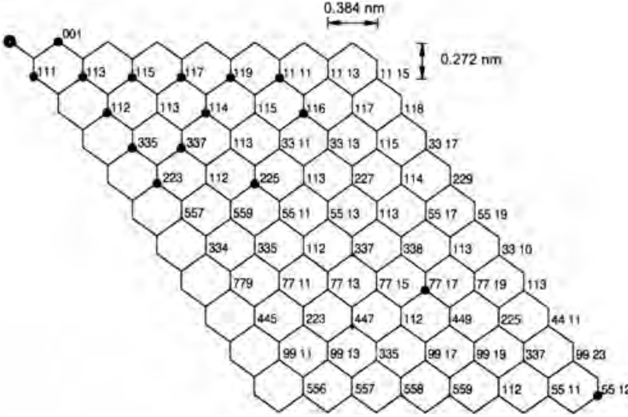


Figure 14. Silicon crystal lattice between (001) and (111) orientations. Single unit cells of bulk-terminated orientations are marked with solid circles. After [159], A. A. Baski et al., *Surf. Sci.* 392, 69 (1997). © 1997, Elsevier Science.

that accommodate a single row of the 7 × 7 reconstruction and one additional string of Si adatoms (Fig. 15) [202]. This is an example of a self-assembled atomically precise grating with the largest corrugation and unit cell achieved on Si surfaces [202]. The periodicity of this structure is expressed in terms of the Si lattice constant—one of the best-known lattice parameters, which suggests possible applications for nanoscale metrology [203, 204]. Atomically precise nature of this template also allows to create ordered atomic structures over at least hundreds of nanometers (Section 5.2.2).

4.1.3. Steps and Domain Boundaries

Stepped templates, discussed in the previous two subsections, took advantage of the near-equilibrium step structures, achieved through high-temperature annealing and stabilized by the 7 × 7 reconstruction or its analogs. Another well-studied process on silicon surfaces is the behavior and structure of steps and 7 × 7 domains during homoepitaxial growth [205–209], which is generally a nonequilibrium process, but can also result in ordered structures. Similarly to the annealing case, the ordering can be guided either by prepatterning the surface, or by controlling the deposition conditions.

On lithographically patterned elongated mesas oriented parallel to $[1\ \bar{1}\ \bar{2}]$ direction, step-flow growth results in terraces with stable steps along the two symmetric $\langle 211 \rangle$ directions, and under appropriate conditions dense and rather regular step arrays are formed (Fig. 16) [59, 210].

As discussed in Section 4.1.1, to produce atomically straight single steps, it was important to achieve equilibrated *single-domain* 7 × 7 reconstruction over entire terraces (Figs. 10, 11) [171, 184]. An alternative approach to step ordering is to allow random nucleation and growth of the 7 × 7 domains at multiple sites near the step edges. Because of the large unit cell, registration between random domains is unlikely, so when such domains collide, out-of-phase boundaries (OPBs) are formed [87]. Thus a 2D quasi-periodic network of atomic steps and OPBs is formed [87, 95], with the periodicities determined by step separation and 7 × 7 domain nucleation conditions. The uniformity of these networks can be dramatically improved by depositing additional Si. During Si homoepitaxy, atoms preferentially are

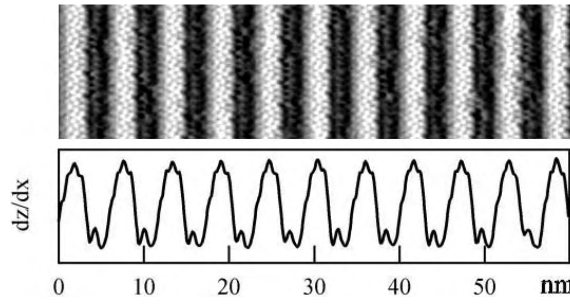


Figure 15. Atomically accurate Si grating. A series of steps with the perfect periodicity of 17 atomic rows (5.73 nm) for Si(557). STM image of the x derivative of the topography and cross-sectional profiles averaged over the y axis. Adapted with permission from [202], A. Kirakosian et al., *Appl. Phys. Lett.* 79, 1608 (2001). © 2001, American Institute of Physics.

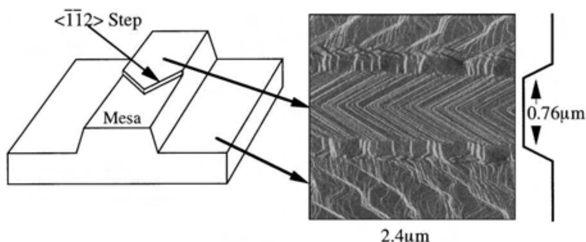


Figure 16. Uniform step array on a mesa top surface. Schematic view and an AFM image of an array produced by step-flow growth. Reprinted with permission from [59], T. Ogino et al., *Surf. Sci.* 514, 1 (2002). © 2002, Elsevier Science.

incorporated at cross-points of OPBs and steps, and combined with fluctuations of the growth rate this results in wavy growth fronts, with the periodicity of the waviness determined by the growth conditions. During such growth, the area of smaller domains tends to diminish, until they merge with larger neighbors, and thus a considerably more uniform network of steps and OPBs is formed [87, 95].

4.2. Heteroepitaxy in Two Dimensions

The step-flow growth is an example of a 2D equivalent of the layer-by-layer growth in three dimensions. The other 3D growth processes can also be generalized for cases of submonolayer deposition on stepped surfaces, where essentially 2D growth occurs [17, 41]. Instead of the surface free energies γ, however, one has to consider a set of respective step energies σ: substrate steps ($\sigma_{\text{substrate}}$), adsorbate island edges ($\sigma_{\text{adsorbate}}$), and the interface between a substrate step and an island attached to it ($\sigma_{\text{interface}}$). Replacing γ's with σ's in Eq. (3) for 3D growth, we get a set of conditions for the adsorbate to form disjointed islands along steps:

$$\sigma_{\text{substrate}} < \sigma_{\text{adsorbate}} + \sigma_{\text{interface}} \tag{5}$$

For continuous adsorbate stripes attached to substrate steps:

$$\sigma_{\text{substrate}} > \sigma_{\text{adsorbate}} + \sigma_{\text{interface}} \tag{6}$$

The familiar equilibrium growth modes that reappear in a 2D world can be classified as row-by-row, Stranski–Krastanov, and island growth in analogy with the 3D equivalents (Fig. 17). A new possibility arises if the following condition is satisfied:

$$\sigma_{\text{interface}} > \sigma_{\text{adsorbate}} + \sigma_{\text{substrate}} \tag{7}$$

in which case adsorbate islands will form, but avoid the step (Fig. 17). This is a uniquely 2D phenomenon, since in 3D growth this would correspond to adsorbate islands floating above the substrate.

In 2D growth just like in the 3D case, the general rule is to use low deposition flux and sufficiently high substrate temperature to approach the equilibrium distributions in experimentally observed growth modes. The next two sections discuss examples of the use of the different 3D and 2D growth phenomena in self-assembly and self-organization of nanostructures on Si surfaces.

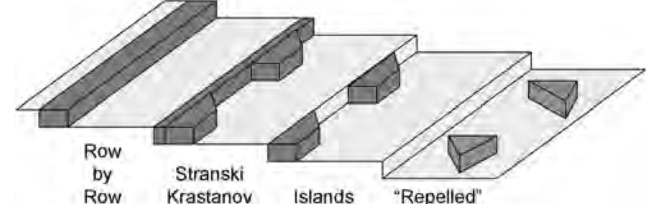


Figure 17. Equilibrium growth modes at steps. The first three regimes (left-to-right) are essentially 2D equivalents of the classic equilibrium growth modes—compare to Figure 4. The step energy rather than the surface free energy is the parameter responsible for the different growth modes in the 2D case. The last regime, with islands not attached to steps, is a uniquely 2D phenomenon (3D equivalent would be islands *floating above* the substrate surface).

4.3. Step Decoration and Self-Assembled Nanowires

Step decoration takes advantage of preferential attachment and/or growth of the deposited material at the atomic steps. The technique was pioneered almost half a century ago when it was used to visualize atomic-height steps on cleavage surfaces of sodium chloride by decorating them with gold [211]. Step selectivity of various surface processes can be utilized to produce step decoration; for example, for silver on graphite, step decoration can be achieved through: selective nanocluster growth [212], diffusion [212–214], and nanowire growth by electrochemistry [215, 216]. For epitaxial systems, step-flow growth results in nanowires of the adsorbate material attached to lower step edges of the substrate (Fig. 17). This growth mode has been first realized for the GaAs/GaAlAs system [217, 218], aided by the almost perfect chemical and structural match between these materials.

The step-flow growth of self-assembled nanowires can also be achieved for metal-on-metal combinations, if a substrate with high surface energy (e.g., W, Mo) and an adsorbate with low surface energy (e.g., Cu) that does not alloy with the substrate are used [219–225]. STM imaging with elemental contrast through spectroscopic states [222, 224, 226] allowed observation of the structure of these metal nanowires with atomic resolution and study of the corresponding growth modes [103]. Quantitative analysis of thus obtained nanowire width distributions produced two results important for rational size-control of SA nanowires: equilibrium structures exhibit the narrowest size distribution and the lower limit on such distribution is imposed by the roughness of substrate steps [103, 227].

Step arrays of vicinal Si(111) [171, 184] discussed in Section 4.1.1 thus represent an ideal substrate for self-assembly by step decoration because they provide atomically straight steps and essentially single-domain Si(111)7 × 7 terraces (Fig. 11)—both factors beneficial for producing self-assembled nanowires with narrow width distribution. Two material combinations have been primarily explored for step decoration on vicinal silicon surfaces: insulator-on-semiconductor CaF_2/Si [27, 103] and semiconductor-on-semiconductor Ge/Si [10, 41]. CaF_2/Si(111) is an excellent model system for studies of submonolayer growth and exploring structures and interactions in the <10-nm range, because of the combination of the high-quality vicinal

templates and atomically sharp interfaces. Ge/Si combination produces superior multilayer systems on a larger (>100 nm) lateral scale and, because inherently conducting SiGe structures are created in the process, it is better suited for characterization by optical and conductivity measurements, including looking for signatures of electron confinement at low temperatures.

4.3.1. CaF_2 Nanostripes on Vicinal Silicon Surfaces

CaF_2/Si(111) is a material combination that has been studied for nearly two decades both as a model heteroepitaxial system and as a possible device platform. For devices, it offers a possibility of epitaxial semiconductor-insulator-semiconductor and metal-insulator-semiconductor structures [228–230]. Model studies of growth modes in 2D heteroepitaxy [103, 231–236] also benefit from the nearly lattice-matched CaF_2/Si(111) combination that forms high-quality epitaxial films and atomically sharp interfaces [228]. CaF_2/Si(001) combination has not been as widely studied [237–240], but step decoration has been recently reported for this system as well [240].

Two types of CaF_2/Si(111) interfaces exist: a F-terminated interface at temperatures below 700 °C, and a Ca-terminated interface from 700 °C to the desorption of CaF_2 at 800 °C (for a monolayer (ML) coverage). A significant fraction of the interface remains F-terminated on a stepped surface, even at high annealing temperatures [241–243]. The F-terminated interface preserves the orientation of the Si(111) substrate (type A epitaxy) while the Ca-terminated interface rotates the structure azimuthally by 180° (type B epitaxy) [244–249]. The differences in bandgaps and band alignment between the clean Si substrate and the two types of interfaces [250, 251] allow to positively identify the different types of structures, even on surfaces with complex topography, via the chemical imaging STM mode [234, 252].

In general, one would expect a transition from island nucleation to step-flow growth with decreasing supersaturation, that is, increasing temperature, decreasing deposition rate, and decreasing step spacing. The results presented in Figure 18 are consistent with such transition: low temperature (550 °C) and low coverage result in randomly distributed small islands (Fig. 18a), higher temperature (600 °C) and coverage result in apparent step-flow growth (Fig. 18b). The distribution of the small CaF_2 islands across terraces in Figure 18a indicates that there is preferential attachment of CaF_2 islands at lower step edges, and a weak repulsion from upper steps edges [234], both factors beneficial for step-flow growth. The presence of residual islands on wider terraces in Figure 18b suggests that the higher deposition temperature corresponds to CaF_2 diffusion length of about 8–9 nm. The CaF_2 stripes in Figure 18b then are far from equilibrium even at 600 °C, which explains the substantial edge roughness. Interface conversion and Ca/Si(111) 3×1 formation [228, 253] prevent deposition at temperatures above 650 °C for step-flow-like growth, so smooth CaF_2 stripes cannot be achieved on Si(111) via a standard step-decoration approach. Moreover, as explained in the next section, the initial stages of growth in this system on vicinal surfaces with step separation <15 nm are strongly affected by the presence of the steps.

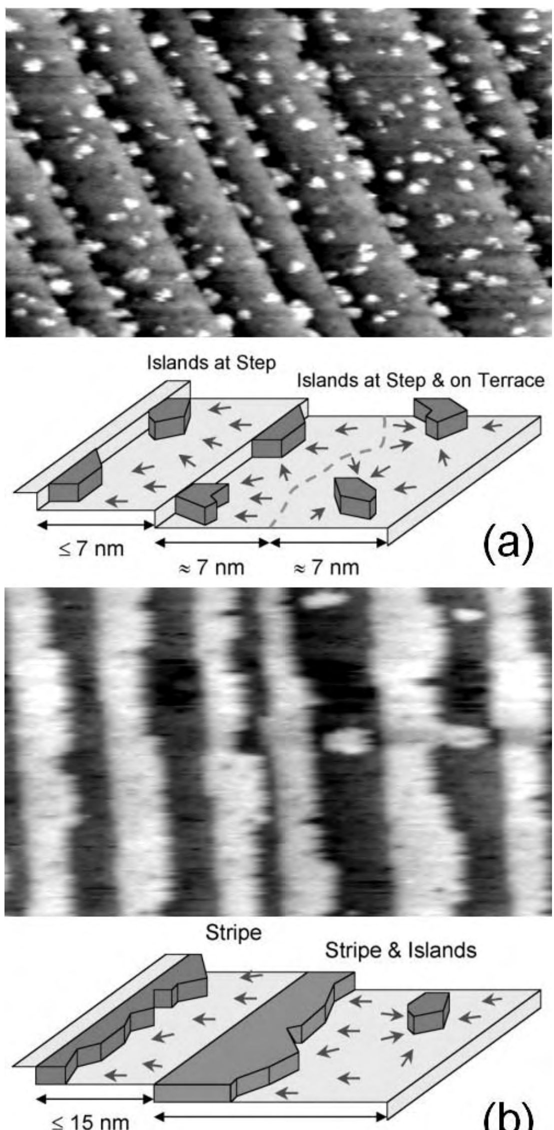


Figure 18. Nonequilibrium CaF_2 growth modes at Si(111) steps. (a) At 550 °C the diffusion length is ≈7 nm, that is, less than half the average terrace width, resulting in diffusion-limited growth. Small islands preferentially attach to steps, but also randomly nucleate on terraces. (b) At 600 °C the diffusion length of 8–9 nm allows step-flow growth on most terraces, except the widest ones where not all the CaF_2 molecules can reach the step. STM images 100 × 60 nm^2, downhill is to the right, CaF_2 appears bright, Si—dark. Adapted in part with permission from [234], D. Y. Petrovykh et al., *J. Vac. Sci. Technol. A* 17, 1415 (1999). © 1999, Elsevier Science.

Smooth CaF_2 stripes can, however, be grown via a process that takes advantage of the interface conversion. Deposition of 1.5 ML of CaF_2 followed by a post-anneal at 830 °C establishes a complete layer of Ca-terminated structure with CaF_2 stripes on top (Fig. 19) [27, 103, 234, 254]. Note that the stripes are located at upper rather than lower step edges; in fact, they avoid the lower edge to such an extent, that even for coverage very close to 2 ML an uninterrupted gap is maintained [234, 254]. This apparent repulsion is due to the 180° azimuthal rotation introduced into the CaF_2 film

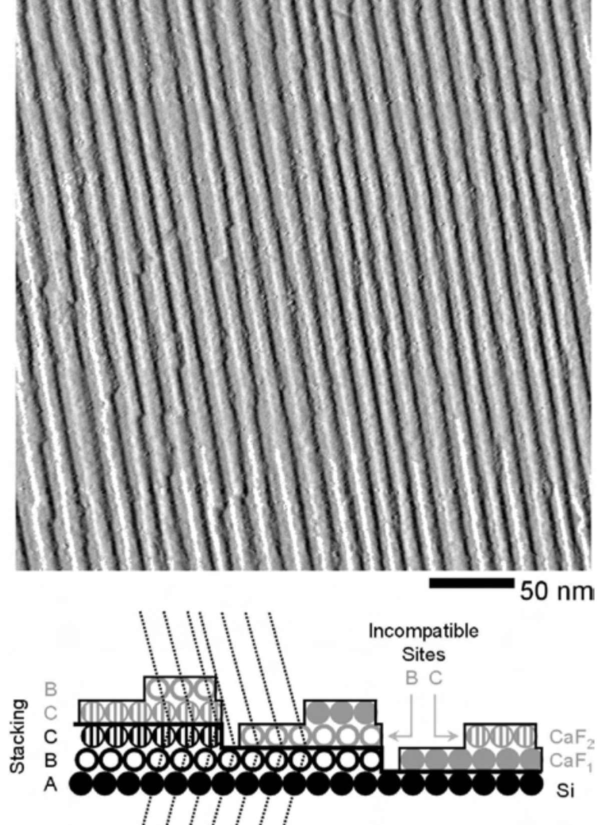


Figure 19. Reverse step-flow growth for CaF_2 on Si(111) at high temperature (830 °C post-anneal). Continuous CaF_2 stripes ≈7 nm wide start at the upper step edge and propagate in the uphill direction. This unusual growth mode is caused by a reversal in stacking at the interface, which causes a lateral incompatibility in the structure across a step. This prevents step flow and ensures that adjacent stripes strictly avoid each other. The derivative of the STM topography is shown to emphasize the steps and stripe edges. Downhill is to the right. Adapted with permission from [254], J. Viernow et al., *Appl. Phys. Lett.* 74, 2125 (1999). © 1999, American Institute of Physics.

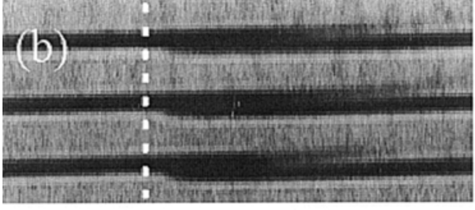
at the Ca-terminated interface [246] and the resulting mismatch in the orientation of Si steps and CaF_2 in the second layer (see diagram in Fig. 19) [234, 236, 254].

The stripe-and-trench CaF_2/Si(111) structure in Figure 19 is in fact very advantageous as a template for selective deposition, because the periodicity and the width of the trenches can be independently controlled by adjusting the miscut and the coverage, respectively. The stripes and trenches are too narrow for selective decoration by metals via direct deposition [255] or electron-stimulated desorption [256] that have been used for CaF_2 surfaces. Organic molecules [257] including chemical vapor deposition (CVD) precursors [258] can, however, be selectively deposited into the trenches due to their higher reactivity (Section 5), and thus Fe nanowires have been successfully grown on such templates via selective CVD [259]. The combination of the well-controlled self-assembled template and the widely applicable selective CVD deposition makes this approach promising for growing other types of self-assembled metal nanowires with width and periodicity <20 nm [14, 27, 103].

4.3.2. Step Decoration in Ge/Si Heteroepitaxy

In Ge/Si heteroepitaxy, most of the step-decoration research has been focused on growth using vicinal Si(001) substrates and primarily on multilayer structures [10, 41]. For submonolayer deposition of Ge, agglomeration at atomic steps has been demonstrated [260] and attributed to step-flow growth, lower step energy for Ge, and enhanced relaxation of Ge-Si lattice mismatch at steps [10]. Deposition of SiGe/Si multilayers on vicinal Si(001) results in *self-organized arrays of self-assembled SiGe nanowires* [261], the process driven by local strain relaxation [135, 262–264]. The nanowire-like structures within each SiGe layer are formed by thickness undulations within the ≈5-nm-thick layer with about 100 nm periodicity (Fig. 20) [261]. The facets of these wires are defined by Si and SiGe step bunches; thus the structural confinement of the individual wires can be improved by using templates with stronger bunching, such as Si(113) [261, 265]. The width and periodicity of these structures (intrinsic characteristics of the bunched Si(001) and Si(113) templates) are larger than those achievable for the CaF_2/Si(111) system discussed earlier. The larger size and the multilayer nature of the samples make conductivity measurements possible for this system, and the resistivity exhibits strong anisotropy (perpendicular vs. parallel to the wires) for these samples at temperature <100 K [10].

For Ge/Si(111) material combination, nanowires have been demonstrated on a Si(111) substrate miscut 0.5° towards [$\bar{1}$ 1 0], that is, direction perpendicular to that discussed in Section 4.1.1. The template thus consisted of atomic steps with ≈110 nm periodicity, and the nanowires with 4.5 nm average height and 43.4 nm average width formed on upper step edges [266, 267]. The material distribution in these wires appeared to be nonuniform, which together with the upper step position rules out step flow as the possible growth mode. The suggested mechanism included initial formation of SA Ge dots on top of a wetting



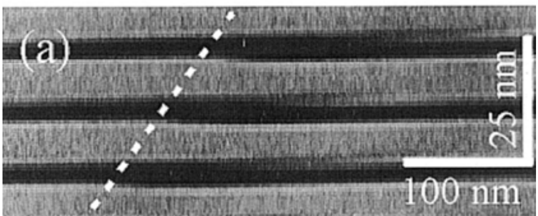


Figure 20. Self-organized arrays of self-assembled SiGe nanowires. A cross-sectional TEM image of SiGe faceted nanowires grown on vicinal Si(113) is shown. The wire stacking changes within the SiGe/Si multilayer between the central (a) and top (b) regions. Reprinted with permission from [261], K. Brunner et al., *Physica E* 7, 881 (2000). © 2000, Elsevier Science.

layer (i.e., SK growth mode), with the dots migrating towards upper step edges for strain relaxation. The wires were then completed by additional Ge migration into the gaps between the dots during the subsequent deposition [266, 267].

4.4. Strain and Steps—Two-Dimensional Control of Self-Assembly

To the extent that the two can be separated, the effects of surface strain and steps on heteroepitaxial growth modes have been separately considered in Sections 3.2 and 4.3, respectively. In general, strain provides a very useful mechanism for self-assembly of nanostructures with uniform size and shape distributions, while stepped templates are very effective in ordering structures over macroscopic areas. Here we demonstrate how the combination of the two approaches can result in systems of self-assembled nanostructures that also exhibit 2D self-organized ordering, using examples from the two systems introduced in the previous section: CaF_2/Si and Ge/Si. The 2D growth phenomena described for these prototype systems are quite general, as similar growth patterns are also observed in strained heteroepitaxy of III-V semiconductors [268–270].

4.4.1. CaF_2/Si(111) Quantum Platelets and Novel 2D Growth Mode

For the CaF_2/Si system, the possibility of unusual 2D growth modes is indicated by growth patterns such as shown in Figure 21. The equilibrium growth modes determined by the step energies alone (Section 4.2) would predict that the adsorbate must preferentially either wet the steps or form a broken interface, but not both simultaneously! Specifically for CaF_2/Si(111) combination, complete wetting (e.g., Fig. 21 left) is expected, because the surface energy of Si is almost three times higher than that of CaF_2 [228]. Clearly, some additional factor plays a significant role during heteroepitaxial growth on a vicinal surface.

As discussed in Section 3.2, strain often determines shapes and sizes of self-assembled nanostructures in heteroepitaxial systems; the elastic energy should then be also considered for an array of SA islands at steps [271]. Two model parameters describe strained islands on a vicinal surface and set the scales of interactions: terrace width W and optimal island size L_0. The terrace width is determined by the miscut of a sample and can be controlled in practice within a wide range of values (Section 4.1.1). The optimal size L_0 for an isolated strained island is determined by the balance of decreasing edge/area ratio and increasing strain energy with increasing island size; in practice L_0 is related to the lattice mismatch during heteroepitaxy [271, 272]. An accurate, albeit simple, estimate of L_0 is given by the size of a structure over which a mismatch of more than one lattice constant is accrued. For example, for a 2% mismatch typical under CaF_2/Si growth conditions [228], it is about 50 atoms $\approx$15 nm, in good agreement with the size of isolated islands observed in the CaF_2/Si(111) system. Other elastic interactions in the system, for example, edge/step repulsion, also would typically have a characteristic scale on the order of L_0, since the scales are related to the same set of substrate and adsorbate elastic parameters. The L_0/W ratio determines the growth mode for systems where strain energy dominates over step energies [271].

For the CaF_2/Si(111) system in Figures 21 and 22, the optimal island size L_0 is on the order of the terrace width W; that is, the interactions of island edges with substrate steps are strong even for low coverage. Islands attached to steps then initially grow to reach the width of $\approx 1/2W$, to minimize the repulsion between their free edges and both neighboring steps, and near-optimal length $\approx L_0$ along steps (Fig. 22a).

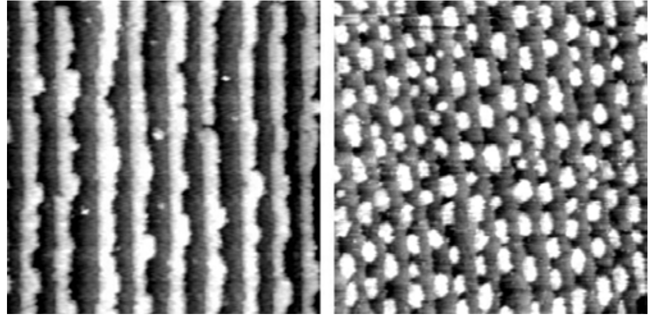


Figure 21. Step wetting and de-wetting behavior for Si(111) + CaF_2. Equilibrium growth predicts either wetting or de-wetting growth at steps, depending on the step and interface energies (Fig. 17). Elastic interactions in the CaF_2/Si(111) system, however, are strong enough to modify the equilibrium behavior, and realize both wetting and de-wetting growth depending on the coverage and step spacing [234, 254, 271].

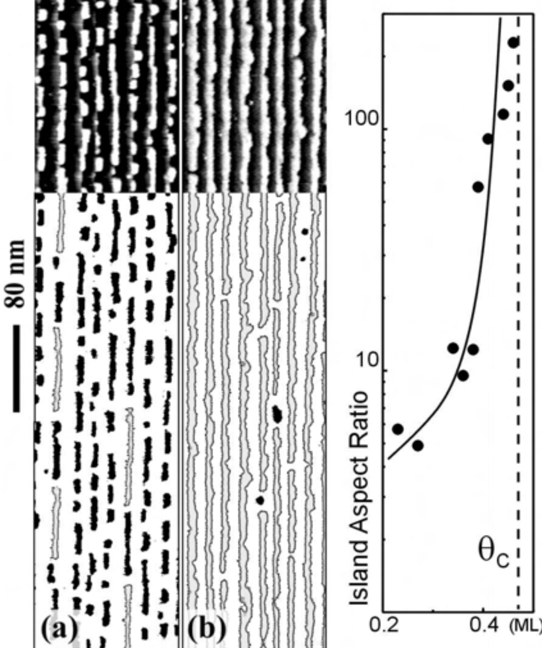


Figure 22. Island-to-stripe transition at a critical coverage. A drastic change in the CaF_2 growth mode is observed for 1.1° miscut Si(111) between 0.23 and 0.40 ML coverage. The growth front morphology changes from short islands (black) to long stripes (gray). The observed rough-to-smooth growth front transition is a feature unique to 2D growth. Such a critical coverage θ_C is predicted by a model that considers the effect of elastic interactions and step energies on the growth front morphology [271].

This apparent nonwetting behavior is the result of strain, rather than high step energy of the islands. With increasing coverage, islands grow along the steps but maintain near-constant width, because the edge-step repulsion dominates over intra- and interisland elastic energy gain. Only when a complete stripe is formed at a step, the free edge begins to advance towards the upper step edge, and the growth appears to be in step-flow mode with increasing coverage (Fig. 22b).

The rough-to-continuous interface transition is remarkably sharp in this system: there is less than 0.2 ML difference in coverage between the broken and continuous stripe configurations (Fig. 22). The "reverse SK mode" is also a strictly 2D growth phenomenon, since in three dimensions initially rough growth interfaces do not subsequently become smoother. The qualitative difference between the 2D and 3D growth modes that produces the novel behavior is the fact that in two dimensions the growth proceeds at multiple growth fronts *simultaneously*, whereas in three dimensions multilayers are grown sequentially. Both of the growth regimes can be potentially useful for forming continuous stripes (Figs. 21 left and 22b) and nanodots (Fig. 21 right), respectively. In the latter case, by properly adjusting the deposition parameters, arrays of CaF_2 nanodots can be produced with density on the order of 3×10^{11} cm^{-2} = 2 Teradots/in^2 and fairly uniform size distribution (Fig. 21 right) [234, 254].

4.4.2. Self-Organized Lateral Ordering of Ge/Si Nanostructures

Finding the right substrate was the major part of fulfilling the quest for self-organized 2D-ordered arrays of uniform Ge quantum dots. The search began with direct deposition of Ge on vicinal Si substrates with orientations previously used to grow SA Ge islands. On vicinal Si(113) this resulted in Ge dots aligned primarily at large (>12 ML) step bunches [261, 265]. The island formation was attributed to enhanced island nucleation at large Si step bunches, and higher strain produced Ge islands, rather than wires seen in partially relaxed SiGe growth (Section 4.3.2) [261]. Adjacent islands remained separated by narrow gaps in all cases, although there appeared to be a small amount of Ge accumulated in these gaps [265]. There was no correlation of island nucleation across terraces, most likely because the island size (80 nm) was much smaller than the terrace size (>300 nm) [261].

Similar results have been obtained for Ge growth on Si(001) miscut 4° towards [110], where Ge islands ≈100 nm in diameter formed at step bunches separating ≈360 nm (001) terraces [273]. In this case islands also showed no correlation across terraces and maintained separation along the steps, but whereas on the vicinal Si(113) islands appeared to be nucleating on the bunched steps [261], for vicinal Si(001) the islands were not as well aligned along the bunches, and upper side of (001) facets was determined to be the preferential nucleation site [273]. The same Si(001) 4° miscut surface prepared under conditions that avoid bunching consists of double-steps separated by about 4 nm [200]. Ge islands grown on this vicinal substrate were smaller (only ≈30 nm) but were not confined to individual terraces, and thus while the dense steps enhanced the nucleation, they did not affect ordering of the Ge islands [10].

The immediate conclusion from the above examples of Ge/Si heteroepitaxy is that the substrate steps and step bunches can control the positioning of Ge nanostructures, but for best results the substrates with the appropriate periodicity (typically <100 nm) would be required. SiGe alloy formation can be helpful for self-organization of nanostructures during Ge/Si heteroepitaxy, including their two-dimensional ordering, because co-deposition of Ge and Si on vicinal Si surfaces allows to control the strain and surface morphology as the growth proceeds, by adjusting the Ge/Si ratio, and by changing between continuous and interrupted growth.

Deposition of a single 2.5-nm $Si_{0.55}Ge_{0.45}$ layer on Si(001) substrate with $\theta = 4°$ polar miscut results in a surface covered with oriented triangular structures [41]. The structures are actually composed of three facets reminiscent of SA Ge pyramids on Si(001): a (001) square bound by $\langle 100 \rangle$ steps, and two {105} parallelograms (Fig. 23a) [41]. This pattern of SiGe surface faceting with three facet types is reported to be unique for a semiconductor surface [41]. The faceting mechanism has been attributed to a strain-driven transition from straight step bunches into energetically favored {105} facets. Specifically, for the polar miscut $\theta \approx 4°$ the step bunch tilt of ≈8° corresponds almost exactly to a $(\bar{1}\ \bar{1}\ 10)$ plane which is unstable with respect to breaking up into a zig-zag pattern of {105} facets and $\langle 100 \rangle$ steps [41]. Another important factor that favors the faceting for this particular type of step bunches is that, even though step length and surface area increase after faceting, no new steps have to be created. Substrates with $\theta \approx 2°$ miscut towards the same or 45° rotated azimuth ϕ do not show similar regular faceting,

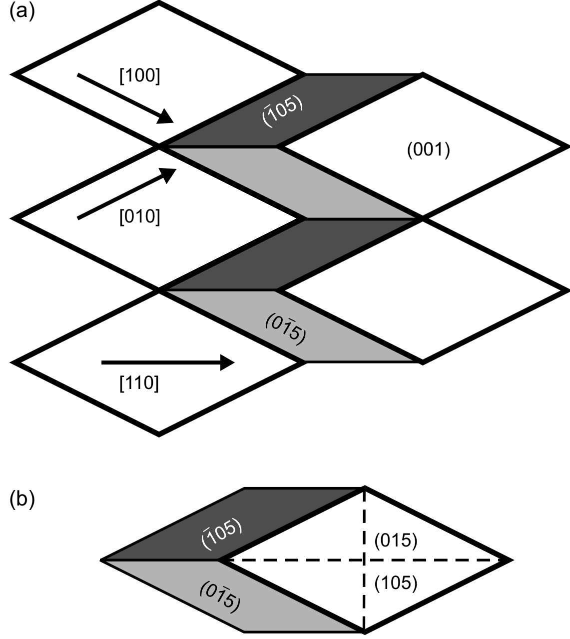


Figure 23. Faceting and island ordering for SiGe films on vicinal Si(001). (a) Ideal facet arrangement for $Si_{0.55}Ge_{0.45}$ film on Si(001) with 4° miscut. (b) $Si_{0.25}Ge_{0.75}$ islands form on the same substrate. Their shape can be obtained from the above pattern by adding a square {105} faceted pyramid on each (001) terrace. Adapted with permission from [41], C. Teichert, *Phys. Rep.* 365, 335 (2002). © 2002, Elsevier Science.

even though individual {105} facets do form [41]. {105} is not the only facet common to both SA Ge islands on Si(001) and vicinal Ge/Si. For higher miscut $\theta = 10°$, three types of steeper {113} facets form structures for $Si_{0.7}Ge_{0.3}$ grown and annealed on Si(118) [274].

On the same as above Si(001) substrate with $\theta = 4°$ polar miscut, deposition of a single 2.5-nm layer of a film with increased Ge concentration ($Si_{0.25}Ge_{0.75}$) results in formation of Ge-rich islands. The (001) facet of the $Si_{0.55}Ge_{0.45}$ morphology is replaced with two more triangular {105} facets, so the islands have hegaxonal base elongated in [110] direction, and four {105} facets: two trapezoidal and two triangular, and form a close-packed quasi-periodic array (Fig. 23b) [41]. Both the average size of the islands (about 50 nm) and their orientation are very similar to the above pattern of facets of the $Si_{0.55}Ge_{0.45}$ film, suggesting that one rational way to search for ordered SA Ge-rich islands is to look in Si-rich films for pattern formation on the same scale and with ⟨100⟩ preferred directions.

Another example of the correlation between the Si-rich film faceting and Ge-rich island ordering occurs for Si(001) substrate with $\theta = 2°$ and $\phi = 45°$ miscut. For 2.5-nm $Si_{0.55}Ge_{0.45}$ film, ripples along [010] direction with 70 nm separation are observed [275]. For Ge-rich $Si_{0.25}Ge_{0.75}$ film of the same nominal thickness, {105} faceted 3D islands arranged in chains parallel to the [010] form (Fig. 24) [275, 276]. In this case, island width of ≈35 nm has a narrow distribution and, compared to the 70-nm period of the above faceted $Si_{0.55}Ge_{0.45}$ surface, suggests island nucleation on both (001) terraces and step bunches [41]. The lengths of the islands, however, are more broadly distributed, reflecting little ordering of the rippled substrates along this direction.

As expected, in both of the above cases for islands grown on vicinal substrates, both the lateral ordering and the island size distribution were superior to those obtained on flat Si(001) substrates under the same conditions. However, even in case of the more uniform islands on the 4° miscut substrate, the width of the size distribution was still about 25%, and the lateral ordering was rather short-range. The structure uniformity was much better in terms of the orientation, which coupled with their statistical ordering allowed measurements of magnetic roughness and anisotropy for Cu/Co sandwiches [277] and arrays of nanomagnets [278] deposited on these nanostructured substrates.

It turns out that for the best ordering of Ge/Si structures, self-assembly of Ge islands on stepped substrates, discussed in the beginning of this section, had to be combined with uniformly faceted SiGe multilayers, discussed as templates for SiGe nanowire growth in the previous section. The multilayer $Si_{0.55}Ge_{0.45}$ wire arrays (Fig. 20) can provide a self-organized rippled substrate with 120 nm periodicity (1.5 nm modulation) [279] comparable to the size of Ge islands grown on vicinal Si. When Ge was deposited on top of a 10 nm buffer Si layer, the Ge islands exhibited remarkable 2D ordering (Fig. 25) [10, 279]. The islands were {105} faceted and truncated pyramids with approximately square bases; the absence of elongation similar to that in Figures 23 and 24 was attributed in part to the absence of [100] and [010] steps on these substrates [279]. Two substrate orientations corresponded to SiGe nanowire multilayers grown on Si(001) miscut 2° towards [100] and 1.5° towards [1 $\bar{1}$ 0], resulting in wires along [010] and [110], respectively [279]. On these two types of substrates, Ge islands formed close-packed arrays with simple-cubic and face-centered lattice, respectively, and size uniformity on the order of 10% (Fig. 25) [279]. Minimization of the strain energy of the whole system, taking into account the orientation of the

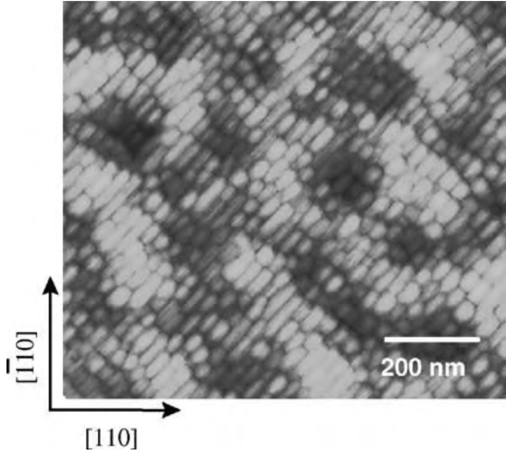


Figure 24. Lateral island ordering by matching substrate periodicity. The surface of a 2.5-nm $Si_{0.25}Ge_{0.75}$ film on Si(001) with 2° polar and 45° azimuthal miscut exhibits chains of mainly prism-like islands with the long axes exclusively aligned along [010]. Reprinted with permission from [41], C. Teichert, *Phys. Rep.* 365, 335 (2002). © 2002, Elsevier Science.

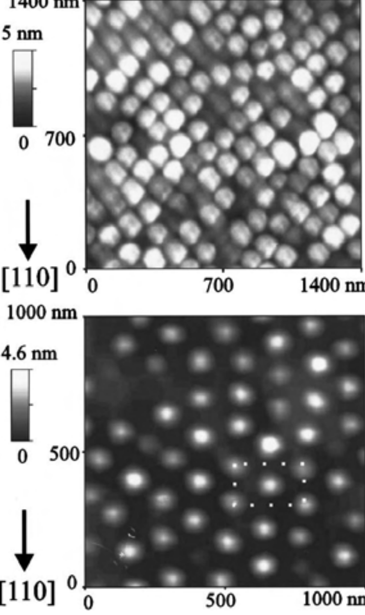


Figure 25. Self-organized uniform arrays of self-assembled Ge islands. When a self-organized periodic array of SiGe nanowires (Fig. 20) is used as a template for Ge deposition, uniform close-packed arrays of {105} faceted Ge islands form on the top layer. Depending on the substrate orientation, the wires in the template are oriented along [010] and [110] directions, which results in simple-cubic and face-centered lattice of the Ge island array, respectively. Adapted with permission from [279], J.-H. Zhu et al., *Appl. Phys. Lett.* 73, 620 (1998). © 1998, American Institute of Physics.

vicinal substrate and elastic dipole repulsion of neighboring islands, is believed to be responsible for the lateral ordering [10, 151, 279]. To date, this system represents perhaps the best example of an island array with uniform 2D ordering achieved through self-assembly on a self-organized substrate—a major milestone on the way towards practical SA and SO.

4.4.3. Ordering by Strain and Dislocation Networks

When elastic strain-relief mechanisms cannot accommodate lattice mismatch during heteroepitaxy, plastic dislocations set in [280]. The critical thickness for $Si_{1-x}Ge_x$ heteroepitaxy on Si(001) depends on the Ge concentration, and ranges between 1 and 100 nm, although theoretical and experimental values often disagree [41, 281, 282]. For $Si_{1-x}Ge_x$/Si(001) growth, the misfit dislocations propagate from the interface and manifest themselves on the surface as ridges and troughs along [110] and [$\bar{1}$ 1 0] directions [283, 284]. Variations in alloy compositions and growth temperature regimes produce networks of such intersecting dislocation lines (up to tens of microns in length) with different density [285–287], as described in detail in [41].

Deposition of Ge or SiGe on these dislocation networks results in nucleation of {105} faceted pyramid-like islands along the two orthogonal directions defined by the troughs and ridges. For low coverage, the islands are exclusively nucleated along the dislocations; with increasing coverage, close-packed chains of islands form along the dislocations, and the rest of the surface is filled with {105} faceted hut- and pyramid-like islands [41]. The range of elastic interactions within a dislocation network is comparable to the width of individual lines; thus, much like for islands on vicinal substrates from the previous section, to improve the lateral ordering the density of dislocations in the network has to be increased to the point where the individual lines touch each other. This has been demonstrated for $Si_{0.7}Ge_{0.3}$ growth combined with 1 keV Si^+ ion bombardment, where a checkerboard pattern of {105} faceted pyramids and pits with ≈190 nm periodicity formed on the surface [41]. Such patterned surfaces are potentially useful for applications, because the dislocation networks are inherently macroscopic objects, but currently it is still difficult to ensure the formation of close-packed arrays of dislocation lines over the scale comparable to the length of individual lines.

5. MOLECULAR AND ATOMIC SELF-ASSEMBLY

Self-assembly and self-organization on the scale of molecules and atoms is both the most ambitious and the most ambiguous goal of nanotechnology. The ability to manipulate elements and compounds on the level of their smallest structural units potentially opens up unlimited possibilities. On the other hand, for processes such as organic synthesis or surface reactions, for example, during growth of nanoclusters in solutions [19], the distinction between "self-assembly" and "traditional" chemistry is not easy to define. The following sections focus on SA and SO on surfaces with atomic-scale templates, in particular those with self-organized templates.

Several general approaches can push the controlled self-assembly to the limit of individual molecules and atoms. For example, self-assembled templates of CaF_2 stripes on vicinal Si(111) (Section 4.3.1) can be prepared with trenches between the stripes as narrow as 3–4 nm. Through chemical selectivity of such a template, individual molecules can then be adsorbed within the trenches, thus forming quasi-one-dimensional structures (Fig. 26) [257]. In another example, strain-relief patterns similar to those discussed in the previous section, but on nanometer rather than micron scale, can serve as templates for self-assembly of atomic clusters 1–10 nm in size for metals [288] or semiconductors [289]. Even though atomic-scale SO templates were used in both of the above examples, the resulting SA structures in general did not exhibit atomic or molecular precision in their sizes or ordering. Such precision is possible, however, if *atomically precise* self-organized templates are used.

5.1. Surface Reconstructions as Templates for Self-Assembly

Surface reconstructions are ideal candidates as templates for self-assembly. First, by definition they exhibit atomically precise periodic pattern over large areas. Second, the two major factors responsible for surface reconstructions [47, 49, 50, 55, 56, 99, 290], surface strain and broken bonds, are also the key factors in controlling self-assembly and positioning of nanostructures.

The key driving force towards atomic precision as self-assembled structures become smaller is the increasing interactions between building blocks. For example, step arrays on Si(111) have a step width that is quantized in units of 1/2 of the 7 × 7 unit cell (7 atomic rows or 2.3 nm, see Fig. 10a)

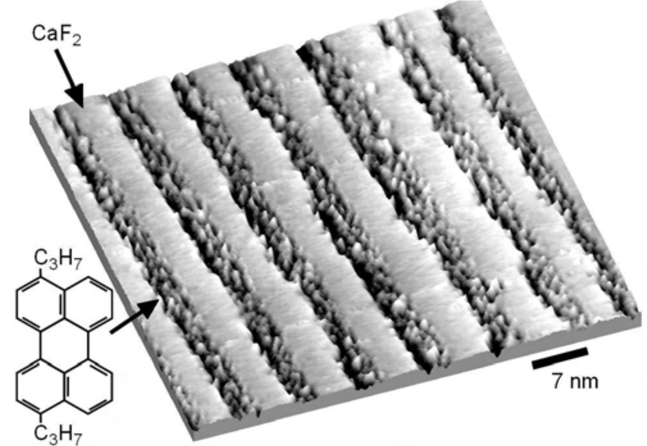


Figure 26. Quasi-one-dimensional confinement of organic molecules. Molecules of 3,10-di(propyl)perylene selectively adsorb between CaF_2 nanostripes on a self-organized CaF_2/CaF_1 template on vicinal Si(111) (Fig. 19). By adjusting the width of the CaF_2 nanostripes, the molecules can be confined in trenches as narrow as 3–4 nm (7–8 nm shown). Adapted with permission from [257], H. Rauscher et al., *Chem. Phys. Lett.* 303, 363 (1999). © 1999, Elsevier Science.

[166, 171]. At larger step spacings there is not enough interaction between adjacent steps to force a sharp width distribution (Section 4.1.1). For a $\approx 1°$ miscut angle with 15 nm average step spacing, one observes a coexistence of terraces with 5, 6, and 7 unit cells (Fig. 11) [171]. However, with a miscut of 10° the surface arranges itself into a regular pattern with a period of exactly 17 atom rows (5.73 nm) and a single 7×7 unit cell per terrace (Fig. 15) [202]. In fact, the existence of large unit cells, such as 7×7, points towards the possibility of creating both two- and one-dimensional structures with perfect periodicity. Such patterns can be used to form devices close to the atomic limit, such as a memory where bits is stored by the presence or absence of individual silicon atoms in 5×4 unit cells precisely lined along tracks 5 atom rows wide. Other examples of controlling self-assembly on the atomic scale include molecular adsorption and formation of "magic" clusters on reconstructed silicon surfaces.

5.1.1. Silicon Surface Chemistry and Functionalization

The combination of the versatility of silicon surfaces with essentially unlimited supply of custom-designed molecules and reactions provided by organic chemistry is an important pathway towards surface functionalization [291]. The sheer number of the different aspects of the silicon surface chemistry explored to date means that here they can only be enumerated; fortunately, a number of excellent recent reviews cover these topics in detail. Interaction of silicon surfaces with hydrogen and reactivity of hydrogen-terminated silicon surfaces under UHV conditions are reviewed in [292]. The technologically important silane chemistry is described in [293]. Reactions of small organic molecules with Si(001) are reviewed in [294, 295]. Si(111)7×7 and Si(001)2×1 surface reconstructions play a major role in determining chemistry of the silicon surfaces in model UHV studies. Reviews [291, 296] provide extensive descriptions of general principles of organic chemistry on silicon surfaces including: $[2+2]$ and $[4+2]$ (Diels Alder) cycloaddition in UHV, creation and reactivity of hydrogen- and halogen-terminated silicon surfaces, dry and wet hydrosilylation, and reactivity of various terminal functional groups.

Some of the major goals of surface functionalization are: surface passivation and protection, patterning surface chemistry or topography, surface-based sensing, molecular recognition, and molecular electronics. In general, the first three types of surface functionalization take advantage of the silicon surface reconstructions only in using the attachment chemistry that matches the reactivity of broken and double bonds available on reconstructed surfaces. The most advanced research goals in terms of molecular recognition and electronics intend to rely on the specific atomic arrangements on reconstructed surfaces much more closely, as briefly described in the next subsection.

Surface passivation typically involves covering the surface with a complete molecular layer and thus in general does not involve nanostructuring or self-assembly. However, similarly to self-assembled monolayers (SAMs) on gold [297, 298], layers of molecules covalently attached to silicon and presenting various terminal functionalities can be produced on silicon through gas-phase [293, 296] or wet chemistry [296, 299, 300]. In addition to providing chemical resistance or functionality, these layers are also being explored for applications as masks in nanolithography [1, 3] and for photopatterning [291, 296].

Using bare silicon surfaces for sensing is not practical, because they quickly oxidize or otherwise react in the ambient environment. Practical sensing applications typically rely on surfaces protected from the ambient environment by a functionalized molecular layer that has also been modified to selectively react with the sensing target; some of the relevant approaches are presented in Section 6.

5.1.2. Self-Assembly with Molecular Precision

A good example of a patterning approach that makes use of molecular self-assembly involves using block copolymers [301, 302], that is, polymers that consist of two or more chemically distinct fragments (blocks). If the constituents are immiscible, phase separation occurs [301, 302] on the scale determined by size of the fragments (typically 10–100 nm), which for block copolymer films on surfaces results in pattern formation [303]. With a proper choice of the components, these patterns can be used as self-assembled lithography or etch masks to produce nanoscale features on surfaces over macroscopic areas [304–307]. Alternatively, selective adsorption or reactivity of the copolymer components can be used as templates for self-assembly of nanostructures [308, 309]. Phase separation mechanism produces block copolymer patterns with inherent short-range (10–100 nm) ordering. Long-range ordering of the patterns can be achieved by using surface templates with periodicity comparable to the size of the copolymer components. The ordering can be induced by either topography [310, 311] or chemical heterogeneity [307, 312, 313] of the surface templates. Self-organized arrays of step bunches (Fig. 13, Section 4.1.2) have been particularly useful as templates for both of the above approaches [310–313].

The direct use of surface reconstructions as patterns for molecular adsorption is inherently restricted to model studies in UHV [294]. One potentially useful aspect of the reactivity of clean silicon surfaces is the selectivity in terms of chirality of small organic molecules [314, 315], which is intimately related to the atomic arrangement on reconstructed surfaces.

Si(001) surface offers compelling properties as a model substrate for molecular electronics: interfacing with conventional devices [316, 317], and a convenient target for attachment of organic molecules through double bonds in silicon dimers [294, 295]. In addition, on vicinal Si(001) surfaces with 4° miscut toward the [110] direction, arrays of (001) terraces with all the dimers oriented in the same direction can be prepared [162] (Section 4.1.1). The uniform orientation of the dimers on these templates has been used to produce organic layers with anisotropic electronic properties [295]. The atomic structures of Si(111)7×7 and Si(001)2×1 surfaces allow self-directed surface chain reactions to occur on these surfaces [318–320], in which reactions propagate along directions defined by the symmetry of respective surfaces creating lines of molecular adsorbates—suggesting a potential pathway towards surface-based molecular circuits.

5.1.3. Atomic Clusters Self-Assembly and Self-Organization in Two Dimensions

A variety of adsorbates form two-dimensional structures on silicon surfaces with unit cells up to a few nanometers in size, most of them based on the respective surface reconstructions [48]. Recently the Si(111)7×7 surface has been used as a template for adsorbing metal atoms on specific sites within the cell, which leads to atomically precise self-assembled monodispersed nanostructures, also referred to as surface magic clusters [321]. Both Si and Ge form such clusters on Si(111)7×7 [321–323]. For metal adsorbates, initial work used transition metals. Particularly regular patterns are observed with Group III metals, such as Ga [324], Tl [325, 326], Al [327], and In [328]. Figure 27 shows a spectacular example: the bright spots are small clusters of six Al atoms adsorbed at the centers of both halves of the 7×7 unit cell [327].

5.2. Electronic Properties of Low-Dimensional Surfaces

Because they are assembled with atomic precision, low-dimensional surface structures are some of the best model systems for examining the properties of low-dimensional electronic states. In addition to serving as atomic templates for these structures, silicon substrates also provide the convenience of a bandgap, which helps to decouple metallic surface states from bulk three-dimensional bands. The scanning probe tools that potentially would allow direct conductivity measurements on surface nanostructures are still only being developed [51], so macroscopic techniques such as surface conductivity measurements [52, 329], photoemission [330–332] and inverse-photoemission spectroscopies [333] are usually used to investigate properties of low-dimensional surface states.

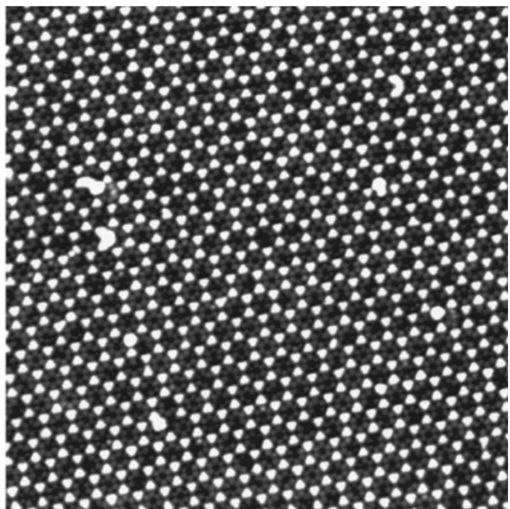


Figure 27. Atomically perfect Al nanocluster array on Si(111)7×7. The Al clusters are located at the centers of both faulted and unfaulted halves of the 7×7 unit cells and appear as bright dots on this STM image (50×50 nm^2). Each cluster contains six Al atoms. Both the structure and the positioning of the nanoclusters in this case are thus atomically precise. Reprinted with permission from [327], J. Jia et al., *Appl. Phys. Lett.* 80, 3186 (2002). © 2002, American Institute of Physics.

5.2.1. Two-Dimensional Patterns

A whole hierarchy of superlattices can be built up by noble metals on Si(111) [52, 329]. The starting point is the $\sqrt{3} \times \sqrt{3}$ structure of Ag and Au, which consists of honeycomb trimers [334, 335]. Adding additional Ag or Au atoms on top of a selected subgroup of trimers gives rise to structures, such as $\sqrt{21} \times \sqrt{21}$, and 6×6 [336, 337]. The $\sqrt{21} \times \sqrt{21}$ structure of Au on top of the $\sqrt{3} \times \sqrt{3}$-Ag structure is shown in Figure 28. These surfaces have the special property of being metallic, which is rather unusual for silicon surfaces [336, 338]. Although the stoichiometric $\sqrt{3} \times \sqrt{3}$ HCT structure with 1 ML of Ag is semiconducting [336], the addition of just a few percent of Ag or Au atoms provides enough electrons to fill a steep surface conduction band inside the bulk bandgap. Surface conductivity measurements, including four-point probes with independent STM tips, have been used to investigate the conductivity of the two-dimensional electron gas at a surface [52, 329]. The full complement of quantum numbers (energy E and momentum $\mathbf{p} = \hbar\mathbf{k}$) has been obtained by angle-resolved photoemission experiments [339]. Particularly important is the Fermi surface, which shows the location of the states at the Fermi level in $\mathbf{k}$-space. These states are responsible for conductivity and many other electronic properties, while electrons below the Fermi level cannot move due to Pauli's exclusion principle. Photoemission has revealed delicate patterns that are formed by the Fermi surfaces of these structures, as shown in Figure 29. A surface doping of a few percent produces small electron pockets, which give rise to small Fermi circles in two dimensions (Fig. 29a). The addition of Au in a $\sqrt{21} \times \sqrt{21}$ superlattice on top of the $\sqrt{3} \times \sqrt{3}$-Ag superlattice enlarges these Fermi circles by adding extra electrons to the band (Fig. 29b). In addition, the extra $\sqrt{21} \times \sqrt{21}$ reciprocal lattice vectors translate this Fermi circle to other parts of $\mathbf{k}$-space and form an intricate pattern of replicas. It can

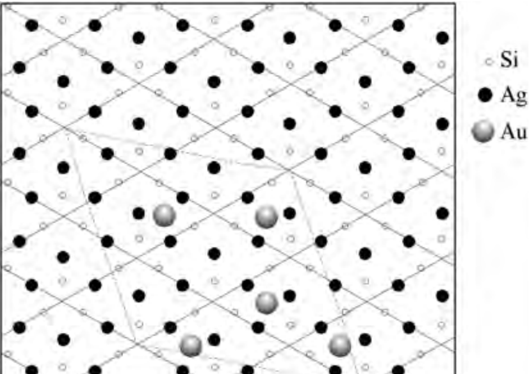


Figure 28. Superlattices formed by Ag and Au on Si(111). A monolayer of Ag forms a $\sqrt{3} \times \sqrt{3}$ superlattice (black dots), and an additional $5/21 = 0.24$ monolayer Au forms $\sqrt{21} \times \sqrt{21}$ superlattice on top of the $\sqrt{3} \times \sqrt{3}$ lattice (see the dotted unit cell containing 5 Au atoms). The corresponding Fermi surfaces are shown in Figure 29a and b. Adapted with permission from [339], J. N. Crain et al., *Phys. Rev. B* 66, 205302 (2002). © 2002, American Physical Society.

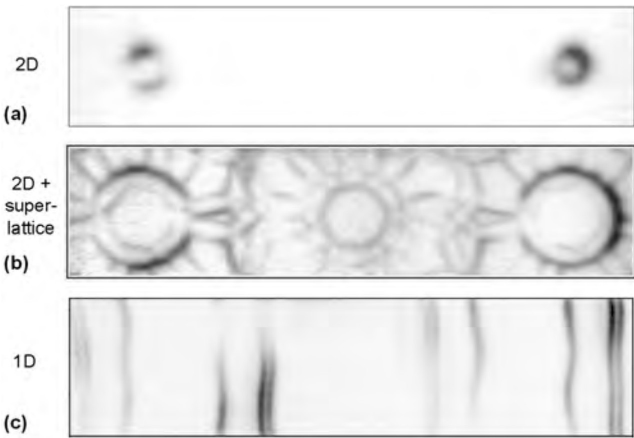


Figure 29. Fermi surfaces of two- and one-dimensional surface states on silicon, obtained by angle-resolved photoemission with energy and angle multidetection. The x- and y-axes are the **k**-vector components parallel to the surface, with $\mathbf{k} = 0$ at the center. A truly one-dimensional Fermi surface would consist of vertical lines since the energy does not depend on the wave vector perpendicular to the chains (vertical). (a) Si(111)$\sqrt{3} \times \sqrt{3}$-Ag (see Fig. 28) with a few percent excess Ag forming small electron pockets. (b) Si(111)$\sqrt{21} \times \sqrt{21}$-(Ag+Au) with a superlattice of Au atoms on top of the $\sqrt{3} \times \sqrt{3}$-Ag lattice forming large Fermi circles. The $\sqrt{21} \times \sqrt{21}$ superlattice causes faint replicas of the Fermi circles. (c) Si(553)-Au, a nearly one-dimensional structure containing a single Au chain per unit cell. Adapted with permission from [339], J. N. Crain et al., *Phys. Rev. B* 66, 205302 (2002) (a), (b); and from [362], J. N. Crain et al., *Phys. Rev. Lett.* 90, 176805 (2003) (c). © 2002 and 2003, American Physical Society.

be viewed as an electron diffraction pattern of the strong, primary Fermi circles that are located at the two $\sqrt{3} \times \sqrt{3}$ lattice points.

Metallic surface states on silicon offer the possibility of studying electrons in low dimensions, where exotic new phenomena have been observed and predicted [340–349]. Already in two dimension there are effects, such as the integer and fractional quantum Hall effect [345, 346]. In one dimension the predictions are even more exotic. The very existence of individual electrons is in question. Basically, electrons moving along a line in space cannot help penetrating each other and thereby forming many-body excitations. Furthermore, these excitations are predicted to separate in collective excitations of spins without charge (spinons) and charges without spin (holons) in a one-dimensional metal [344, 348]. A key for observing such effects is that the electrons are completely decoupled from the three-dimensional substrate. Since the Fermi level is located in the bandgap of silicon, metallic surface states cannot interact with states from the silicon substrate and thus their reduced dimensionality is preserved.

5.2.2. One-Dimensional Chains

One-dimensional structures can be formed by self-assembly on stepped surfaces, as reviewed in detail by Ortega and Speller [225]. Here we focus on chains of atoms adsorbed at silicon surface. They serve as the ideal quantum wires for testing the theoretical predictions of exotic phenomena in one dimension. One might think that the ultimate quantum wire would be a single string of atoms suspended freely in space. However, according to early arguments by Peierls, the atoms in such strings form pairs and lose their metallicity. It has become possible to produce metallic chains of metal atoms by self-assembly at stepped silicon surfaces [13, 27], where they line up parallel to the step edges. X-ray diffraction from the Si(557)-Au structure shows that gold atoms are incorporated rigidly into silicon lattice positions [350] without detectable Peierls distortion.

On Si(111), chain structures can be formed either by using stepped surfaces as templates or by spontaneous breaking of the three-fold symmetry of flat Si(111) into three domains of chain structures. A collection of four such structures is shown in Figure 30 [13]. Alkali metals [351], alkaline earths [253], noble metals [329], indium [22, 352, 353], and even magnetic rare earths, such as Gd [354], form single-domain chain structures on vicinal Si(111), as long as the coverage is well below 1 ML (typically 0.4 ML or 1–2 chains per unit cell). For a detailed overview of these chain structures see [13, 331]. The stability of these chain structures might be related to the stability of the honeycomb chain (HCC) [355], which is a structural element in many of them. This chain consists of a strip of graphitic, π-bonded Si. At a coverage approaching 1 ML the surfaces become two-dimensional and restore their three-fold symmetry, as shown in Section 5.2.1.

The Si(001) exhibits a native surface anisotropy that gives rise to dimer rows. The difficulty is selecting one of the two 90° rotated domains which are separated by a single surface step. A miscut angle of about 4° leads to the formation of double steps and has been used successfully for producing Si(001) with a dominant domain [162]. Other approaches include using rather flat surfaces and sophisticated growth methods [356, 357] or electromigration [358]. Substituting Si dimer rows by metals, such as In, creates dimer rows of adsorbates [359, 360]. Long chain structures of Si dimers

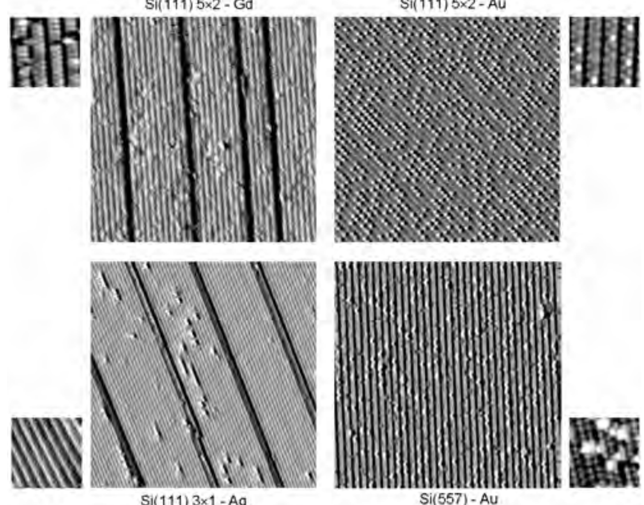


Figure 30. Chain structures formed by sub-monolayer deposition of metals on Si(111). The overview panels (60×60 nm^2) show the x-derivative of the STM topography, which produces dark lines at step edges (downhill to the right). The close-up panels (7×7 nm^2) show the topography itself. Adapted with permission from [13], F. J. Himpsel et al., *J. Phys.: Condens. Matter* 13, 11097 (2001). © 2001, Institute of Physics.

have been also observed on cubic SiC(001) [361]. So far, all row structures on Si(001) have been semiconducting and therefore less suitable for low-dimensional electron physics.

A number of vicinal Si(111) and Si(001) surfaces lend themselves as templates to the formation of one-dimensional chain structures, as shown in Figure 14 (Section 4.1.2) [159]. Particularly for Au as an adsorbate, a large fraction of the possible orientations has actually produced well-ordered structures with uninterrupted chains [362–369]. Only one of these structures (Si(557)-Au) has been analyzed structurally by X-ray diffraction [350] and total energy minimization [370, 371]. Surprisingly, the Au chain lies at the center of the Si(111) terrace, not at the step edge where one would expect higher coordination for adsorbates. Apparently, the simple idea of step decoration fails here. A more sophisticated strain-relief mechanism might be at work, which is not necessary at the step edge because atoms at the top of a step can relax laterally. The Au atoms substitute for a row of Si surface atoms and are thus locked in by three back-bonds, which explains their resilience against a Peierls distortion.

The coupling between the chains can be varied in a systematic way by changing the miscut angle. The transition between two- and one-dimensional chains happens at a chain spacing of about 5–6 atomic rows ($\approx$2 nm), judging from the vanishing of the two-dimensional band dispersion observed in photoemission [362]. For example, the Si(553)-Au surface with a chain spacing of 4-1/3 rows has a ratio of 10–40 for the intra- to interchain coupling, whereas the Si(557)-Au surface with 5-2/3 rows step spacing has a ratio $>$70 [362]. One-dimensional Fermi surfaces consist of straight lines perpendicular to the chain direction (Fig. 29c), as opposed to the two-dimensional circles in Figure 29a and b. The energy is independent of the **k**-component perpendicular to the chains. A slight waviness in Figure 29c indicates a residual two-dimensional character. A tight binding fit to the Fermi surface provides the ratio of the electron coupling parallel and perpendicular to the chains ($\approx$40:1 for the doublet and 12:1 for the singlet). Since wave functions decay exponentially, there is a wide range of couplings that can be covered by a narrow range of step spacings.

5.3. Atomic-Scale Memory

One of the first goals of nanotechnology, set by Richard Feynman in his pioneering 1959 talk at CalTech, is an atomic-scale memory. Using a cube 5 atoms wide to store one bit, Feynman estimates that "all of the information... accumulated in all the books in the world can be written... in a cube of material 1/200 inch wide." The tools are in place now to make an attempt at realizing this dream. Assembling atoms with the STM at low temperature has been possible for some time. Recently, we have been able to demonstrate an atomic memory operating at room temperature (Fig. 31 and [372]). Such a device is useful for finding the fundamental limits of data storage, irrespective of practicality. A chain structure induced by Au on flat Si(111) forms self-formatted "tracks" (dark horizontal stripes in Fig. 31). White dots on top of the tracks give the surface the appearance of a CD-ROM, except that the scale is in nanometers, not microns. The resulting storage density is a million times higher. It has been demonstrated that the white dots correspond to single Si atoms, by filling the vacant sites by Si deposition [372]. The extra Si atoms reside on a 5×4 lattice and can be used to store data if the presence of an atom is assigned to a one and its absence to a zero. Each bit thus occupies a territory of $5 \times 4 = 20$ atomic sites. A reduction of the area/bit is not possible for this particular system, as shown from the correlation function between adjacent bits and their analysis in terms of nearest-neighbor interactions [373]. The occupancy of the closest 5×2 site is highly suppressed, as shown in Figure 32.

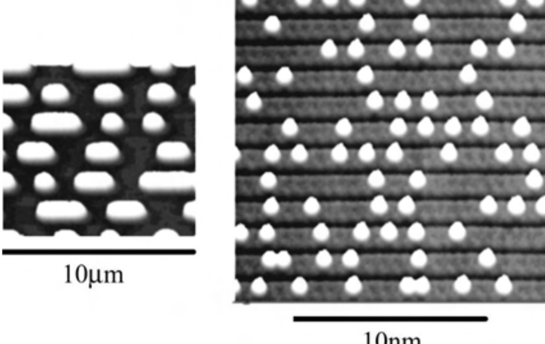


Figure 31. Atomic memory obtained by self-organization on a Si(111) surface with 0.4 monolayer of gold. Extra Si atoms (bright dots) occupy lattice sites on top of tracks that are five atoms wide (1.7 nm). Compared to a conventional CD-ROM (left), the scale is reduced from microns to nanometers, which leads to a 10^6 times higher density. Adapted with permission from [372], R. Bennewitz et al., *Nanotechnology* 13, 499 (2002). © 2002, Institute of Physics.

Such an atomic-scale device is useful for a deeper understanding of the limiting factors in data storage. In [372] the limits on storage density, retention, signal-to-noise, and readout speed were explored. The readout error rate is still uncharted, although the raw signal-to-noise figure is better than in hard disks. However, an optimum code for filtering and error correction needs to be worked out, analogous to

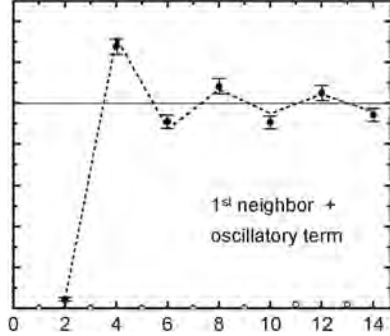


Figure 32. Correlations in a one-dimensional lattice fluid of Si atoms on top of the Si(111)5×2-Au chain structure (points). Theoretical modeling (dashed line) shows that a nearest-neighbor repulsion suppresses the occupancy at site 2 and causes a pile-up at site 4 (top). An extra oscillatory interaction is required to explain the subsequent oscillations beyond site 6. Such interactions between neighboring atoms limit the ultimate storage density achievable by an atomic-scale memory in Figure 31. Adapted with permission from [373], A. Kirakosian et al., *Phys. Rev. B* (submitted).

the partial response maximum likelihood (PRML) method that has been so successful for hard disks (see [372] and references therein). Such extremely selective filtering methods rely on a well-known signal shape, which allows all other waveforms to be filtered out. The signal from the STM readout of the model atomic memory is highly reproducible, but is different from that of a hard drive: it is unipolar in nature, whereas on a hard disk readout pulses alternate in sign and are less than a pulse width apart. These important distinctions indicate that a different noise reduction model will have to be used for the atomic memory, perhaps one similar to those used in long-distance communications through optical fibers.

Another useful information storage benchmark is the DNA molecule, which uses 32 atoms to store one bit. This compares well with the 20 atoms per bit in the Si-based atomic memory. Moreover, the current readout rate of the atomic memory by STM is roughly the same as that estimated from the DNA transcription rate—on the order of 100 bits/s [372]. The biological molecular machinery probably has not been optimized to simply maximize the transcription rate, and thus even the simplistic STM readout can potentially be several orders of magnitude faster [372]. However, the general trend of the drastic drop in readout rate with increasing storage density holds for both DNA and the atomic memory in comparison with current magnetic storage [372]. Such trends clearly have important implications for any future development of practical storage devices.

6. NANOSTRUCTURES AND BIOLOGY

The information storage density and readout rate comparison from the previous section is only one illustration of the fact that biology provides perhaps the best examples of systems that self-assemble and self-organize on scales from nanometers to meters. It is therefore not surprising that recently many biological processes are scrutinized in search for elements and methods that potentially can be duplicated in artificial systems [374–376]. At the same time, it is clear that nanotechnology will provide some of the most powerful tools and methods for studying and perhaps controlling the biology on a cellular and subcellular level. Accordingly, the studies of interfaces between various biological and bio-inspired structures and semiconductors have attracted much interest in both the life and physical science communities [376–378]. Higher-order biological systems, for example, proteins or cells, typically possess a three-dimensional structure with nanoscale features [375, 376, 379] and thus are affected by nanostructured surfaces [380–383]. Proper interfacing with biological structures then requires macroscopic templates with nanoscale features, that is, just the types of templates discussed in Sections 2 and 4.1, functionalized with appropriate surface chemistry (Sections 5.1.1–5.1.2).

6.1. Making Silicon Surfaces Bio-Compatible

Attaching nanometer-scale molecules to silicon surfaces opens up a whole new field, where a hierarchy of organic molecules is built up on top of a silicon template, as discussed in Sections 5.1.1–5.1.2. However, the chemistry of attaching biomolecules to silicon surfaces without denaturing them is still in its infancy. Short snippets of DNA have been successfully attached to Si(111) [384] and Si(001) [385, 386] surfaces via intermediary molecular layers, but if the development of the other types of DNA chips and sensors [387–390] is any indication, many additional opportunities are waiting to be explored.

While direct attachment of organic molecules to the silicon surface bonds is the most precise way to combine silicon with biotechnology, there are standard methods for creating bio-compatible surfaces via buffer layers. The most common method uses a Au buffer layer which makes a strong bond to thiol derivatives of organic molecules [388]. The method builds on the vast expertise in creating self-assembled monolayers (SAMs) on Au surfaces [297, 298]. At Si surfaces, Au forms large islands due to its low surface energy, and a Ti or Cr wetting layer is required between the Au and the Si to flatten it out [312, 391, 392]. It has been demonstrated that bunched step structures can be replicated with a fidelity of a few nanometers in this fashion [391]. There is a minimum Au coverage of 3 nm (about 12 monolayers) to reach the full coverage of alkane thiol SAMs, which is necessary for obtaining oriented SAMs [392]. Short snippets of DNA thiols have been attached in this fashion, and their orientation has been determined by near-edge X-ray absorption fine structure spectroscopy (NEXAFS) [392]. Another method for attaching organic molecules to silicon is the use of siloxane chemistry, which works on the native oxide surface [393].

Silicon surfaces can be used to induce self-organization in a number of biomimetic materials: block copolymers (as discussed in Section 5.1.2), biopolymers [394], and mushroom-shaped supramolecular nanostructures [395]. Recently it has also been demonstrated that the self-organized templates based on vicinal silicon surfaces (Section 4.1.2) can potentially be used in a biosensor. The biosensor platform in question is based on the preferential alignment of liquid crystals by surfaces that have submicron oriented structure [396, 397]. The approach has been originally developed based on statistically oriented obliquely deposited gold films [398–400]. The stepped silicon templates have an additional benefit of deterministic orientation (defined by the wafer miscut) and silicon surfaces prepared in UHV after proper passivation have been shown to induce alignment of liquid crystals [401]. Moreover, the nanostructured topography of silicon masters can be replicated in polymers through micromolding [402], which opens the possibilities for producing cheap disposable well-oriented substrates for research and applications.

6.2. Bio-Inspired Self-Assembly and Beyond

The area of nanostructured macroscopic templates is where the interplay between the biological world and that of artificial self-organized systems is possibly the most prominent. Biologically derived templates have been used to pattern surfaces [379, 403]. Biological molecules, in particular DNA, can be also used to add tailored and specific functionality to nanostructures and surfaces in order to guide self-organization [404–406]. Biological molecules

themselves can also be used as self-assembled scaffolding for inorganic nanostructures; particularly fruitful in that regard are quasi-one-dimensional molecules [407, 408] and organic nanotubes [409] which can be formed via a number of non-covalent interactions: metal coordination, hydrogen bonding, π-π stacking, hydrophobic effect, etc. While synthesis and self-assembly of these supramolecular structures has been very successful, their self-organization on a large scale (i.e., mats or arrays of 100–1000 fibers or nanotubes) is a more difficult task to achieve via only the short-range intermolecular interactions; therefore some of the properly functionalized silicon substrates can potentially be used to guide the self-organization on a larger scale.

More complicated pieces of the biological molecular machinery can be harnessed for self-assembly and self-organization as well. A very recent example is offered by the use of genetically engineered bacteriophages to create a composite film with self-assembled semiconductor nanodots self-organized within domains spanning from nanometer to centimeter scale [410]. The same type of a bacteriophage can be also engineered to selectively bind a number of semiconductors and oxides and even distinguish between crystal orientations [375, 411]. The recognition mechanism remains under investigation, but this development opens possibilities for creation of very specific customized biomimetic molecules that would serve as markers, as well as assembly and delivery vehicles for self-assembly and self-organization of nanostructures.

GLOSSARY

Epitaxy A growth process of a solid film on a crystalline substrate in which the atoms of the film replicate the arrangement of the atoms of the substrate. Homo- and heteroepitaxy respectively refer to the growth of a layer with chemical (and structural) parameters identical to and different from those of the substrate.

Miller indices Means of indexing crystal orientations and directions, for atomic planes and rows Miller indices are equal to reciprocals of the fractional intercepts with the crystallographic axes.

Self-assembly Spontaneous formation of structures with a well-defined size and shape distribution typically determined by thermodynamic stability of the structures and/or growth kinetics.

Self-organization Natural tendency of structures to form ordered arrays or assemblies.

Step decoration Nucleation of adsorbate structures at surface steps due to migration of adsorbed atoms across a terrace and preferential attachment at a step.

Surface energy The modification of the total energy of a solid due to the presence of its surface, that is, the difference between the energy of a truncated solid and the same number of atoms in the bulk. The term surface energy is often used for the surface free energy.

Surface reconstruction Rearrangement of the surface atoms driven by reduction of the surface energy. Reconstruction causes changes of the symmetry, periodicity, and ordering of the surface structure.

Surface states Modified electronic states at the surface, for example due to a rearrangement of broken bonds at semiconductor surfaces or due to the step in the inner potential at metal surfaces (see Section 5.2).

Vicinal surface A high Miller index surface, typically with low atomic density and high surface energy and thus unstable with respect to faceting. In practice, a vicinal surface is created by a controlled miscut with respect to a low-index surface.

REFERENCES

1. Y. Chen and A. Pepin, *Electrophoresis* 22, 187 (2001).
2. E. W. Plummer, Ismail, R. Matzdorf, A. V. Melechko, J. P. Pierce, and J. D. Zhang, *Surf. Sci.* 500, 1 (2002).
3. Y. N. Xia, J. A. Rogers, K. E. Paul, and G. M. Whitesides, *Chem. Rev.* 99, 1823 (1999).
4. For some of the expert predictions see essays by speakers of "The Next Twenty Years" conference series http://www.zdnet.com/special/filters/tnty/essays/.
5. Committee for the Review of the National Nanotechnology Initiative and National Research Council, "Small Wonders, Endless Frontiers: A Review of the National Nanotechnology Initiative." National Academy Press, Washington, DC, 2002.
6. M. C. Roco, *J. Nanopart. Res.* 3, 5 (2001).
7. M. C. Roco, *J. Nanopart. Res.* 3, 353 (2001).
8. M. C. Roco, *J. Nanopart. Res.* 4, 9 (2002).
9. M. C. Roco and W. S. Bainbridge, Eds., "Societal Implications of Nanoscience and Nanotechnology." Kluwer Academic, Dordrecht, 2001.
10. K. Brunner, *Rep. Prog. Phys.* 65, 27 (2002).
11. I. Eisele and W. Hansch, *Thin Solid Films* 369, 60 (2000).
12. W. Eberhardt, *Surf. Sci.* 500, 242 (2002).
13. F. J. Himpsel, K. N. Altmann, R. Bennewitz, J. N. Crain, A. Kirakosian, J. L. Lin, and J. L. McChesney, *J. Phys.: Condens. Matter* 13, 11097 (2001).
14. F. J. Himpsel, K. N. Altmann, G. J. Mankey, J. E. Ortega, and D. Y. Petrovykh, *J. Magn. Magn. Mater.* 200, 456 (1999).
15. M. A. Kastner, *Ann. Phys.* (*Leipzig*) 9, 885 (2000).
16. K. K. Likharev, *Nanotechnology* 10, 159 (1999).
17. P. Moriarty, *Rep. Prog. Phys.* 64, 297 (2001).
18. C. G. Smith, *Rep. Prog. Phys.* 59, 235 (1996).
19. A. N. Shipway, E. Katz, and I. Willner, *Chem. Phys. Chem.* 1, 18 (2000).
20. J. von Delft, *Ann. Phys.* (*Leipzig*) 10, 219 (2001).
21. J. Drucker, *IEEE J. Quantum Electron.* 38, 975 (2002).
22. J. F. McGilp, *Phys. Stat. Sol. A* 188, 1361 (2001).
23. F. Remacle and R. D. Levine, *Chem. Phys. Chem.* 2, 20 (2001).
24. S. D. Bader, *Surf. Sci.* 500, 172 (2002).
25. F. J. Himpsel, J. E. Ortega, G. J. Mankey, and R. F. Willis, *Adv. Phys.* 47, 511 (1998).
26. J. Shen and J. Kirschenr, *Surf. Sci.* 500, 300 (2002).
27. F. J. Himpsel, A. Kirakosian, J. N. Crain, J. L. Lin, and D. Y. Petrovykh, *Solid State Commun.* 117, 149 (2001).
28. G. Johansson, A. Kack, and G. Wendin, *Physica C* 368, 289 (2002).
29. B. E. Kane, *Fortschritte Phys.—Prog. Phys.* 48, 1023 (2000).
30. C. Single, R. Augke, E. E. Prins, D. A. Wharam, and D. P. Kern, *Superlattices and Microstructures* 28, 429 (2000).
31. O. P. Pchelyakov, Y. B. Bolkhovityanov, A. V. Dvurechenskii, L. V. Sokolov, A. I. Nikiforov, A. I. Yakimov, and B. Voigtlander, *Semiconductors* 34, 1229 (2000).
32. J. E. Han and V. H. Crespi, *Phys. Rev. Lett.* 86, 696 (2001).
33. K. Brunner and G. Abstreiter, *Jpn. J. Appl. Phys.* 40, 1860 (2001).
34. S. D. Berger and J. M. Gibson, *Appl. Phys. Lett.* 57, 153 (1990).
35. L. R. Harriott, *J. Vac. Sci. Technol. B* 15, 2130 (1997).

36. D. M. Eigler and E. K. Schweizer, *Nature* 344, 524 (1990).
37. R. D. Piner, J. Zhu, F. Xu, S. H. Hong, and C. A. Mirkin, *Science* 283, 661 (1999).
38. C. A. Mirkin, S. H. Hong, and L. Demers, *Chem. Phys. Chem.* 2, 37 (2001).
39. C. A. Mirkin, *MRS Bull.* 26, 535 (2001).
40. P. Vettiger, M. Despont, U. Drechsler, U. Durig, W. Haberle, M. I. Lutwyche, H. E. Rothuizen, R. Stutz, R. Widmer, and G. K. Binnig, *IBM J. Res. Dev.* 44, 323 (2000).
41. C. Teichert, *Phys. Rep.* 365, 335 (2002).
42. J. R. Arthur, *Surf. Sci.* 500, 189 (2002).
43. D. P. Woodruff, *Surf. Sci.* 500, 147 (2002).
44. J. A. Kubby and J. J. Boland, *Surf. Sci. Rep.* 26, 61 (1996).
45. H. Neddermeyer, *Rep. Prog. Phys.* 59, 701 (1996).
46. D. Drakova, *Rep. Prog. Phys.* 64, 205 (2001).
47. D. Haneman, *Rep. Prog. Phys.* 50, 1045 (1987).
48. V. G. Lifshits, A. A. Saranin, and A. V. Zotov, "Surface Phases on Silicon: Preparation, Structures, and Properties." Wiley, New York, 1994.
49. J. P. LaFemina, *Surf. Sci. Rep.* 16, 133 (1992).
50. G. P. Srivastava, *Rep. Prog. Phys.* 60, 561 (1997).
51. S. Hasegawa and F. Grey, *Surf. Sci.* 500, 84 (2002).
52. S. Hasegawa, X. Tong, S. Takeda, N. Sato, and T. Nagao, *Prog. Surf. Sci.* 60, 89 (1999).
53. M. Giesen, *Prog. Surf. Sci.* 68, 1 (2001).
54. R. S. Williams, G. Medeiros-Ribeiro, T. I. Kamins, and D. A. A. Ohlberg, *Annu. Rev. Phys. Chem.* 51, 527 (2000).
55. R. C. Cammarata, *Prog. Surf. Sci.* 46, 1 (1994).
56. R. C. Cammarata and K. Sieradzki, *Annu. Rev. Mater. Sci.* 24, 215 (1994).
57. H. C. Jeong and E. D. Williams, *Surf. Sci. Rep.* 34, 175 (1999).
58. K. Yagi, H. Minoda, and M. Degawa, *Surf. Sci. Rep.* 43, 49 (2001).
59. T. Ogino, Y. Homma, Y. Kobayashi, H. Hibino, K. Prabhakaran, K. Sumitomo, H. Omi, S. Suzuki, T. Yamashita, D. J. Bottomley, F. Ling, and A. Kaneko, *Surf. Sci.* 514, 1 (2002).
60. A. S. Bell, B. Brezger, U. Drodofsky, S. Nowak, T. Pfau, J. Stuhler, T. Schulze, and J. Mlynek, *Surf. Sci.* 435, 40 (1999).
61. K. S. Johnson, J. H. Thywissen, N. H. Dekker, K. K. Berggren, A. P. Chu, R. Younkin, and M. Prentiss, *Science* 280, 1583 (1998).
62. J. J. McClelland and R. J. Celotta, *Thin Solid Films* 367, 25 (2000).
63. J. H. Thywissen, K. S. Johnson, N. H. Dekker, A. P. Chu, and M. Prentiss, *J. Vac. Sci. Technol. B* 16, 3841 (1998).
64. M. Olshanii, N. Dekker, C. Herzog, and M. Prentiss, *Phys. Rev. A* 62, 033612 (2000).
65. J. D. Fowlkes, A. J. Pedraza, D. A. Blom, and H. M. Meyer, *Appl. Phys. Lett.* 80, 3799 (2002).
66. B. H. Choi, C. M. Park, S. H. Song, M. H. Son, S. W. Hwang, D. Ahn, and E. K. Kim, *Appl. Phys. Lett.* 78, 1403 (2001).
67. L. Vescan, *Mater. Sci. Eng. A* 302, 6 (2001).
68. S. Y. Chou, P. R. Krauss, and P. J. Renstrom, *J. Vac. Sci. Technol. B* 14, 4129 (1996).
69. S. Y. Chou, P. R. Krauss, and P. J. Renstrom, *Science* 272, 85 (1996).
70. T. I. Kamins, D. A. A. Ohlberg, R. S. Williams, W. Zhang, and S. Y. Chou, *Appl. Phys. Lett.* 74, 1773 (1999).
71. H. W. Deckman and J. H. Dunsmuir, *Appl. Phys. Lett.* 41, 377 (1982).
72. C. L. Haynes and R. P. Van Duyne, *J. Phys. Chem. B* 105, 5599 (2001).
73. C. L. Haynes, A. D. McFarland, M. T. Smith, J. C. Hulteen, and R. P. Van Duyne, *J. Phys. Chem. B* 106, 1898 (2002).
74. J. Jorritsma, M. A. M. Gijs, J. M. Kerkhof, and J. G. H. Stienen, *Nanotechnology* 7, 263 (1996).
75. J. Liu, J. C. Barnard, K. Seeger, and R. E. Palmer, *Appl. Phys. Lett.* 73, 2030 (1998).
76. T. I. Kamins, R. S. Williams, and D. P. Basile, *Nanotechnology* 10, 117 (1999).
77. L. Vescan, T. Stoica, and B. Hollander, *Mater. Sci. Eng. B* 89, 49 (2002).
78. M. Koh, S. Sawara, T. Goto, Y. Ando, T. Shinada, and I. Ohdomari, *Jpn. J. Appl. Phys.* 39, 5352 (2000).
79. M. Koh, S. Sawara, T. Goto, Y. Ando, T. Shinada, and I. Ohdomari, *Jpn. J. Appl. Phys.* 39, 2186 (2000).
80. M. Koh, S. Sawara, T. Shinada, T. Goto, Y. Ando, and I. Ohdomari, *Appl. Surf. Sci.* 162, 599 (2000).
81. T. Ogino, H. Hibino, and Y. Homma, *Appl. Surf. Sci.* 117, 642 (1997).
82. Y. Homma, H. Hibino, Y. Kunii, and T. Ogino, *Surf. Sci.* 445, 327 (2000).
83. T. Doi, M. Ichikawa, S. Hosoki, and H. Kakibayashi, *Appl. Phys. Lett.* 74, 3675 (1999).
84. A. V. Latyshev, A. L. Aseev, A. B. Krasilnikov, and S. I. Stenin, *Surf. Sci.* 213, 157 (1989).
85. D. Lee, J. M. Blakely, T. W. Schroeder, and J. R. Engstrom, *Appl. Phys. Lett.* 78, 1349 (2001).
86. Z. C. Niu, R. Notzel, H. P. Schonherr, J. Fricke, L. Daweritz, and K. H. Ploog, *J. Cryst. Growth* 187, 333 (1998).
87. T. Ogino, *Surf. Sci.* 386, 137 (1997).
88. P. Finnie and Y. Homma, *Appl. Phys. Lett.* 72, 827 (1998).
89. Y. Homma, P. Finnie, and T. Ogino, *Appl. Phys. Lett.* 74, 815 (1999).
90. Y. Homma, P. Finnie, T. Ogino, H. Noda, and T. Urisu, *J. Appl. Phys.* 86, 3083 (1999).
91. P. Finnie and Y. Homma, *J. Cryst. Growth* 202, 604 (1999).
92. K. H. Ploog and R. Notzel, *Physica E* 3, 92 (1998).
93. R. Notzel, U. Jahn, Z. C. Niu, A. Trampert, J. Fricke, H. P. Schonherr, T. Kurth, D. Heitmann, L. Daweritz, and K. H. Ploog, *Appl. Phys. Lett.* 72, 2002 (1998).
94. R. Notzel, Z. C. Niu, M. Ramsteiner, H. P. Schonherr, A. Tranpert, L. Daweritz, and K. H. Ploog, *Nature* 392, 56 (1998).
95. T. Ogino, H. Hibino, Y. Homma, Y. Kobayashi, K. Prabhakaran, K. Sumitomo, and H. Omi, *Acc. Chem. Res.* 32, 447 (1999).
96. P. A. Lewis, H. Ahmed, and T. Sato, *J. Vac. Sci. Technol. B* 16, 2938 (1998).
97. K. Seeger and R. E. Palmer, *J. Phys. D: Appl. Phys.* 32, L129 (1999).
98. T. Tada and T. Kanayama, *J. Vac. Sci. Technol. B* 16, 3934 (1998).
99. W. Haiss, *Rep. Prog. Phys.* 64, 591 (2001).
100. R. Koch, *J. Phys.: Condens. Matter* 6, 9519 (1994).
101. B. Rout, B. Sundaravel, A. K. Das, S. K. Ghose, K. Sekar, D. P. Mahapatra, and B. N. Dev, *J. Vac. Sci. Technol. B* 18, 1847 (2000).
102. J. Nogami, *Surf. Rev. Lett.* 7, 555 (2000).
103. F. J. Himpsel, T. Jung, A. Kirakosian, J. L. Lin, D. Y. Petrovykh, H. Rauscher, and J. Viernow, *MRS Bull.* 24, 20 (1999).
104. Y. Chen, D. A. A. Ohlberg, and R. S. Williams, *J. Appl. Phys.* 91, 3213 (2002).
105. Y. Chen, D. A. A. Ohlberg, G. Medeiros-Ribeiro, Y. A. Chang, and R. S. Williams, *Appl. Phys. Lett.* 76, 4004 (2000).
106. J. Nogami, B. Z. Liu, M. V. Katkov, C. Ohbuchi, and N. O. Birge, *Phys. Rev. B* 63, 233305 (2001).
107. S. Aggarwal, A. P. Monga, S. R. Perusse, R. Ramesh, V. Ballarotto, E. D. Williams, B. R. Chalamala, Y. Wei, and R. H. Reuss, *Science* 287, 2235 (2000).
108. S. Aggarwal, S. B. Ogale, C. S. Ganpule, S. R. Shinde, V. A. Novikov, A. P. Monga, M. R. Burr, R. Ramesh, V. Ballarotto, and E. D. Williams, *Appl. Phys. Lett.* 78, 1442 (2001).
109. A. G. Banshchikov, A. V. Kimel, B. B. Krichevtsov, A. A. Rzhevskii, N. S. Sokolov, and O. A. Yakubtsov, *Phys. Solid State* 41, 97 (1999).
110. A. M. Nazmul, A. G. Banshchikov, H. Shimizu, and M. Tanaka, *J. Cryst. Growth* 227, 874 (2001).
111. I. Goldfarb and G. A. D. Briggs, *Surf. Sci.* 454, 837 (2000).
112. S. W. Kim and S. Fujita, *Jpn. J. Appl. Phys.* 41, L543 (2002).
113. K. Ueno, H. Shirota, T. Kawamura, T. Shimada, K. Saiki, and A. Koma, *Appl. Surf. Sci.* 190, 485 (2002).

114. M. H. Huang, S. Mao, H. Feick, H. Q. Yan, Y. Y. Wu, H. Kind, E. Weber, R. Russo, and P. D. Yang, *Science* 292, 1897 (2001).
115. Y. Y. Wu, H. Q. Yan, M. Huang, B. Messer, J. H. Song, and P. D. Yang, *Chem. Eur. J.* 8, 1261 (2002).
116. R. S. Wagner and W. C. Ellis, *Appl. Phys. Lett.* 4, 89 (1964).
117. Q. Wan, T. H. Wang, T. Feng, X. H. Liu, and C. L. Lin, *Appl. Phys. Lett.* 81, 3281 (2002).
118. X. F. Duan and C. M. Lieber, *Adv. Mater.* 12, 298 (2000).
119. A. M. Morales and C. M. Lieber, *Science* 279, 208 (1998).
120. D. P. Yu, Z. G. Bai, Y. Ding, Q. L. Hang, H. Z. Zhang, J. J. Wang, Y. H. Zou, W. Qian, G. C. Xiong, H. T. Zhou, and S. Q. Feng, *Appl. Phys. Lett.* 72, 3458 (1998).
121. J. Westwater, D. P. Gosain, S. Tomiya, S. Usui, and H. Ruda, *J. Vac. Sci. Technol. B* 15, 554 (1997).
122. M. K. Sunkara, S. Sharma, R. Miranda, G. Lian, and E. C. Dickey, *Appl. Phys. Lett.* 79, 1546 (2001).
123. W. K. Liu, J. Winesett, W. L. Ma, X. M. Zhang, M. B. Santos, X. M. Fang, and P. J. McCann, *J. Appl. Phys.* 81, 1708 (1997).
124. P. Finnie and Y. Homma, *Phys. Rev. B* 59, 15240 (1999).
125. K. Ueno, K. Saiki, and A. Koma, *Jpn. J. Appl. Phys.* 40, 1888 (2001).
126. T. Chikyow and N. Koguchi, *J. Vac. Sci. Technol. B* 16, 2538 (1998).
127. R. Heitz, N. N. Ledentsov, D. Bimberg, A. Y. Egorov, M. V. Maximov, V. M. Ustinov, A. E. Zhukov, Z. I. Alferov, G. E. Cirlin, I. P. Soshnikov, N. D. Zakharov, P. Werner, and U. Gosele, *Appl. Phys. Lett.* 74, 1701 (1999).
128. L. Hansen, A. Ankudinov, F. Bensing, J. Wagner, G. Ade, P. Hinze, V. Wagner, J. Geurts, and A. Waag, *Phys. Stat. Sol. B* 224, 515 (2001).
129. T. Mano, H. Fujioka, K. Ono, Y. Watanabe, and M. Oshima, *Appl. Surf. Sci.* 132, 760 (1998).
130. M. Oshima, Y. Watanabe, S. Heun, M. Sugiyama, and T. Kiyokura, *J. Electr. Spectrosc. Relat. Phenom.* 80, 129 (1996).
131. T. I. Kamins, K. Nauka, and R. S. Williams, *Appl. Phys. A* 73, 1 (2001).
132. M. El kurdi, P. Boucaud, S. Sauvage, G. Fishman, O. Kermarrec, Y. Campidelli, D. Bensahel, G. Saint-Girons, I. Sagnes, and G. Patriarche, *J. Appl. Phys.* 92, 1858 (2002).
133. O. G. Schmidt, U. Denker, M. Dashiell, N. Y. Jin-Phillipp, K. Eberl, R. Schreiner, H. Grabeldinger, H. Schweizer, S. Christiansen, and F. Ernst, *Mater. Sci. Eng. B* 89, 101 (2002).
134. B. Voigtlander, *Surf. Sci. Rep.* 43, 127 (2001).
135. P. Venezuela, J. Tersoff, J. A. Floro, E. Chason, D. M. Follstaedt, F. Liu, and M. G. Lagally, *Nature* 397, 678 (1999).
136. Y. W. Mo, D. E. Savage, B. S. Swartzentruber, and M. G. Lagally, *Phys. Rev. Lett.* 65, 1020 (1990).
137. F. M. Ross, R. M. Tromp, and M. C. Reuter, *Science* 286, 1931 (1999).
138. Z. Gai, X. W. Li, R. G. Zhao, and W. S. Yang, *Phys. Rev. B* 57, R15060 (1998).
139. G. Medeiros-Ribeiro, A. M. Bratkovski, T. I. Kamins, D. A. A. Ohlberg, and R. S. Williams, *Science* 279, 353 (1998).
140. T. I. Kamins, G. Medeiros-Ribeiro, D. A. A. Ohlberg, and R. S. Williams, *J. Appl. Phys.* 85, 1159 (1999).
141. G. Medeiros-Ribeiro, T. I. Kamins, D. A. A. Ohlberg, and R. S. Williams, *Phys. Rev. B* 58, 3533 (1998).
142. M. Zinkeallmang, L. C. Feldman, and M. H. Grabow, *Surf. Sci. Rep.* 16, 377 (1992).
143. A. L. Barabasi, *Mater. Sci. Eng. B* 67, 23 (1999).
144. G. Medeiros-Ribeiro, *Phys. Stat. Sol. B* 230, 443 (2002).
145. I. M. Lifshitz and V. V. Slyozov, *J. Phys. Chem. Solids* 19, 35 (1961).
146. C. Wagner, *Z. Elektrochem.* 65, 581 (1961).
147. B. K. Chakraverty, *J. Phys. Chem. Solids* 28, 2413 (1967).
148. B. K. Chakraverty, *J. Phys. Chem. Solids* 28, 2401 (1967).
149. F. M. Ross, J. Tersoff, and R. M. Tromp, *Phys. Rev. Lett.* 80, 984 (1998).
150. I. Daruka, J. Tersoff, and A. L. Barabasi, *Phys. Rev. Lett.* 82, 2753 (1999).
151. V. A. Shchukin, N. N. Ledentsov, P. S. Kop'ev, and D. Bimberg, *Phys. Rev. Lett.* 75, 2968 (1995).
152. H. Uemura, M. Uwaha, and Y. Saito, *J. Phys. Soc. Jpn.* 71, 1296 (2002).
153. T. I. Kamins, G. Medeiros-Ribeiro, D. A. A. Ohlberg, and R. S. Williams, *Appl. Phys. A* 67, 727 (1998).
154. V. M. Kaganer and K. H. Ploog, *Phys. Rev. B* 6420, 205301 (2001).
155. P. Schittenhelm, G. Abstreiter, A. Darhuber, G. Bauer, P. Werner, and A. Kosogov, *Thin Solid Films* 294, 291 (1997).
156. D. Buttard, J. Eymery, F. Fournel, P. Gentile, F. Leroy, N. Magnea, H. Moriceau, G. Renaud, F. Rieutord, K. Rousseau, and J. L. Rouviere, *IEEE J. Quantum Electron.* 38, 995 (2002).
157. F. Fournel, H. Moriceau, N. Magnea, J. Eymery, D. Buttard, J. L. Rouviere, K. Rousseau, and B. Aspar, *Thin Solid Films* 380, 10 (2000).
158. B. Rottger, M. Hanbucken, and H. Neddermeyer, *Appl. Surf. Sci.* 162, 595 (2000).
159. A. A. Baski, S. C. Erwin, and L. J. Whitman, *Surf. Sci.* 392, 69 (1997).
160. B. S. Swartzentruber, N. Kitamura, M. G. Lagally, and M. B. Webb, *Phys. Rev. B* 47, 13432 (1993).
161. C. E. Aumann, J. J. Demiguel, R. Kariotis, and M. G. Lagally, *Surf. Sci.* 275, 1 (1992).
162. D. J. Chadi, *Phys. Rev. Lett.* 59, 1691 (1987).
163. D. E. Aspnes and J. Ihm, *Phys. Rev. Lett.* 57, 3054 (1986).
164. R. J. Hamers, R. M. Tromp, and J. E. Demuth, *Phys. Rev. B* 34, 5343 (1986).
165. B. S. Swartzentruber, Y. W. Mo, R. Kariotis, M. G. Lagally, and M. B. Webb, *Phys. Rev. Lett.* 65, 1913 (1990).
166. J. L. Goldberg, X. S. Wang, J. Wei, N. C. Bartelt, and E. D. Williams, *J. Vac. Sci. Technol. A* 9, 1868 (1991).
167. J. Wei, X. S. Wang, N. C. Bartelt, E. D. Williams, and R. T. Tung, *J. Chem. Phys.* 94, 8384 (1991).
168. E. D. Williams, R. J. Phaneuf, J. Wei, N. C. Bartelt, and T. L. Einstein, *Surf. Sci.* 294, 219 (1993).
169. Y. N. Yang and E. D. Williams, *Phys. Rev. B* 51, 13238 (1995).
170. R. J. Phaneuf, N. C. Bartelt, E. D. Williams, W. Swiech, and E. Bauer, *Surf. Sci.* 268, 227 (1992).
171. J. Viernow, J. L. Lin, D. Y. Petrovykh, F. M. Leibsle, F. K. Men, and F. J. Himpsel, *Appl. Phys. Lett.* 72, 948 (1998).
172. R. Losio, K. N. Altmann, and F. J. Himpsel, *Phys. Rev. B* 61, 10845 (2000).
173. E. D. Williams and N. C. Bartelt, *Science* 251, 393 (1991).
174. E. D. Williams, *Surf. Sci.* 300, 502 (1994).
175. N. C. Bartelt, E. D. Williams, R. J. Phaneuf, Y. Yang, and S. Dassarma, *J. Vac. Sci. Technol. A* 7, 1898 (1989).
176. R. J. Phaneuf and E. D. Williams, *Phys. Rev. B* 41, 2991 (1990).
177. R. S. Becker, J. A. Golovchenko, E. G. McRae, and B. S. Swartzentruber, *Phys. Rev. Lett.* 55, 2028 (1985).
178. R. J. Hamers, U. M. Kohler, and J. E. Demuth, *Ultramicroscopy* 31, 10 (1989).
179. R. Wiesendanger, G. Tarrach, D. Burgler, and H. J. Guntherodt, *Europhys. Lett.* 12, 57 (1990).
180. T. Kato, T. Takajyo, H. Tochihara, and W. Shimada, *Jpn. J. Appl. Phys.* 39, 4307 (2000).
181. W. Shimada, T. Kata, and H. Tochihara, *Surf. Sci.* 491, L663 (2001).
182. Y. F. Wang and T. T. Tsong, *Phys. Rev. B* 53, 6915 (1996).
183. B. Z. Olshanetsky and A. A. Shklyaev, *Surf. Sci.* 82, 445 (1979).
184. J. L. Lin, D. Y. Petrovykh, J. Viernow, F. K. Men, D. J. Seo, and F. J. Himpsel, *J. Appl. Phys.* 84, 255 (1998).
185. H. Tochihara, W. Shimada, M. Itoh, H. Tanaka, M. Udagawa, and I. Sumita, *Phys. Rev. B* 45, 11332 (1992).
186. X. S. Wang, J. L. Goldberg, N. C. Bartelt, T. L. Einstein, and E. D. Williams, *Phys. Rev. Lett.* 65, 2430 (1990).

187. N. C. Bartelt, T. L. Einstein, and E. D. Williams, *Surf. Sci.* 240, L591 (1990).
188. N. C. Bartelt, J. L. Goldberg, T. L. Einstein, and E. D. Williams, *Surf. Sci.* 273, 252 (1992).
189. O. L. Alerhand, D. Vanderbilt, R. D. Meade, and J. D. Joannopoulos, *Phys. Rev. Lett.* 61, 1973 (1988).
190. V. I. Marchenko and A. Y. Parshin, *JETP Lett.* 52, 129 (1980).
191. M. B. Gordon and J. Villain, *J. Phys. C: Solid State Phys.* 12, L151 (1979).
192. T. W. Poon, S. Yip, P. S. Ho, and F. F. Abraham, *Phys. Rev. Lett.* 65, 2161 (1990).
193. E. Pehlke and J. Tersoff, *Phys. Rev. Lett.* 67, 465 (1991).
194. M. B. Webb, *Surf. Sci.* 300, 454 (1994).
195. F. K. Men, F. Liu, P. J. Wang, C. H. Chen, D. L. Cheng, J. L. Lin, and F. J. Himpsel, *Phys. Rev. Lett.* 88, 096105 (2002).
196. B. Rottger, M. Hanbucken, I. Vianey, R. Kliese, and H. Neddermeyer, *Surf. Sci.* 309, 656 (1994).
197. B. Z. Olshanetsky and V. I. Mashanov, *Surf. Sci.* 111, 414 (1981).
198. M. Hanbucken, B. Rottger, and H. Neddermeyer, *Appl. Surf. Sci.* 164, 91 (2000).
199. S. G. J. Mochrie, S. Song, M. Yoon, D. L. Abernathy, and G. B. Stephenson, *Physica B* 221, 105 (1996).
200. J. Zhu, K. Brunner, and G. Abstreiter, *Appl. Surf. Sci.* 137, 191 (1999).
201. A. A. Baski and L. J. Whitman, *Phys. Rev. Lett.* 74, 956 (1995).
202. A. Kirakosian, R. Bennewitz, J. N. Crain, T. Fauster, J. L. Lin, D. Y. Petrovykh, and F. J. Himpsel, *Appl. Phys. Lett.* 79, 1608 (2001).
203. R. G. Dixson, N. G. Orji, J. Fu, V. W. Tsai, E. D. Williams, R. Kacker, T. V. Vorburger, H. Edwards, D. Cook, P. West, and R. Nyffenegger, *Proc. SPIE* 4344 (2001).
204. S. Gonda, H. Zhou, J. Fu, and R. M. Silver, *Proc. SPIE* 4608 (2002).
205. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, *Thin Solid Films* 306, 205 (1997).
206. A. V. Latyshev, A. B. Krasilnikov, and A. L. Aseev, *Phys. Stat. Sol. A* 146, 251 (1994).
207. W. Shimada and H. Tochihara, *Surf. Sci.* 311, 107 (1994).
208. A. V. Latyshev, A. B. Krasilnikov, A. L. Aseev, L. V. Sokolov, and S. I. Stenin, *Surf. Sci.* 254, 90 (1991).
209. A. V. Latyshev, A. L. Aseev, A. B. Krasilnikov, and S. I. Stenin, *Surf. Sci.* 227, 24 (1990).
210. H. Omi and T. Ogino, *Thin Solid Films* 369, 88 (2000).
211. G. A. Bassett, *Philos. Mag.* 3, 1042 (1958).
212. G. M. Francis, L. Kuipers, J. R. A. Cleaver, and R. E. Palmer, *J. Appl. Phys.* 79, 2942 (1996).
213. S. J. Carroll, K. Seeger, and R. E. Palmer, *Appl. Phys. Lett.* 72, 305 (1998).
214. S. J. Carroll, R. E. Palmer, P. A. Mulheran, S. Hobday, and R. Smith, *Appl. Phys. A* 67, 613 (1998).
215. E. C. Walter, B. J. Murray, F. Favier, G. Kaltenpoth, M. Grunze, and R. M. Penner, *J. Phys. Chem. B* 106, 11407 (2002).
216. M. P. Zach, K. Inazu, K. H. Ng, J. C. Hemminger, and R. M. Penner, *Chem. Mater.* 14, 3206 (2002).
217. P. M. Petroff, *Ultramicroscopy* 31, 67 (1989).
218. P. M. Petroff, M. S. Miller, Y. T. Lu, S. A. Chalmers, H. Metiu, H. Kroemer, and A. C. Gossard, *J. Cryst. Growth* 111, 360 (1991).
219. M. Mundschau, E. Bauer, and W. Swiech, *J. Appl. Phys.* 65, 581 (1989).
220. M. Paunov and E. Bauer, *Appl. Phys. A* 44, 201 (1987).
221. F. J. Himpsel, Y. W. Mo, T. Jung, J. E. Ortega, G. J. Mankey, and R. F. Willis, *Superlattices and Microstructures* 15, 237 (1994).
222. Y. W. Mo and F. J. Himpsel, *Phys. Rev. B* 50, 7868 (1994).
223. T. Jung, R. Schlittler, J. K. Gimzewski, and F. J. Himpsel, *Appl. Phys. A* 61, 467 (1995).
224. T. Jung, Y. W. Mo, and F. J. Himpsel, *Phys. Rev. Lett.* 74, 1641 (1995).
225. Ortega and Speller, in "Encyclopedia of Nanotechnology."
226. F. J. Himpsel, T. Jung, R. Schlittler, and J. K. Gimzewski, *Jpn. J. Appl. Phys.* 35, 3695 (1996).
227. D. Y. Petrovykh, F. J. Himpsel, and T. Jung, *Surf. Sci.* 407, 189 (1998).
228. M. A. Olmstead, in "Thin Films: Heteroepitaxial Systems" (W. K. Liu and M. B. Santos, Eds.), Vol. 15, p. 211. Series on Directions in Condensed Matter Physics, World Scientific, Singapore, 1999.
229. M. Watanabe, Y. Iketani, and M. Asada, *Jpn. J. Appl. Phys.* 39, L964 (2000).
230. D. H. Mosca, N. Mattoso, W. H. Schreiner, A. J. A. de Oliveira, W. A. Ortiz, W. H. Flores, and S. R. Teixeira, *J. Magn. Magn. Mater.* 231, 337 (2001).
231. U. Hessinger, M. Leskovar, and M. A. Olmstead, *Phys. Rev. Lett.* 75, 2380 (1995).
232. V. Mathet, F. Nguyenvandau, J. Olivier, and P. Galtier, *J. Cryst. Growth* 148, 133 (1995).
233. B. M. Kim, C. A. Ventrice, T. Mercer, R. Overney, and L. J. Schowalter, *Appl. Surf. Sci.* 104, 409 (1996).
234. D. Y. Petrovykh, J. Viernow, J. L. Lin, F. M. Leibsle, F. K. Men, A. Kirakosian, and F. J. Himpsel, *J. Vac. Sci. Technol. A* 17, 1415 (1999).
235. K. Kametani, K. Sudoh, and H. Iwasaki, *Jpn. J. Appl. Phys.* 41, 250 (2002).
236. J. Wollschlager, *Appl. Phys. A* 75, 155 (2002).
237. D. Loretto, F. M. Ross, and C. A. Lucas, *Appl. Phys. Lett.* 68, 2363 (1996).
238. T. Sumiya, T. Miura, H. Fujinuma, and S. Tanaka, *Surf. Sci.* 376, 192 (1997).
239. N. S. Sokolov and S. M. Suturin, *Appl. Surf. Sci.* 175, 619 (2001).
240. L. Pasquali, S. D'Addato, G. Selvaggi, S. Nannarone, N. S. Sokolov, S. M. Suturin, and H. Zogg, *Nanotechnology* 12, 403 (2001).
241. F. J. Himpsel, U. O. Karlsson, F. R. McFeely, J. F. Morar, D. Rieger, A. Taleb-Ibrahimi, and J. A. Yarmoff, *Mater. Sci. Eng. B* 1, 9 (1988).
242. D. Rieger, F. J. Himpsel, U. O. Karlsson, F. R. McFeely, J. F. Morar, and J. A. Yarmoff, *Phys. Rev. B* 34, 7295 (1986).
243. F. J. Himpsel, F. U. Hillebrecht, G. Hughes, J. L. Jordan, U. O. Karlsson, F. R. McFeely, J. F. Morar, and D. Rieger, *Appl. Phys. Lett.* 48, 596 (1986).
244. T. Asano and H. Ishiwara, *Appl. Phys. Lett.* 42, 517 (1983).
245. T. P. Smith III, J. M. Phillips, R. People, J. M. Gibson, L. Pfeiffer, and P. J. Stiles, *Mater. Res. Soc. Symp. Proc.* 54, 295 (1986).
246. R. M. Tromp and M. C. Reuter, *Phys. Rev. Lett.* 61, 1756 (1988).
247. C. C. Cho, H. Y. Liu, B. E. Gnade, T. S. Kim, and Y. Nishioka, *J. Vac. Sci. Technol. A* 10, 769 (1992).
248. C. C. Cho, T. S. Kim, B. E. Gnade, H. Y. Liu, and Y. Nishioka, *Appl. Phys. Lett.* 60, 338 (1992).
249. R. M. Tromp and M. C. Reuter, *Phys. Rev. Lett.* 73, 110 (1994).
250. F. J. Himpsel, T. F. Heinz, A. B. McLean, E. Palange, and E. Burstein, *J. Vac. Sci. Technol. B* 7, 879 (1989).
251. T. F. Heinz, F. J. Himpsel, E. Palange, and E. Burstein, *Phys. Rev. Lett.* 63, 644 (1989).
252. J. Viernow, D. Y. Petrovykh, A. Kirakosian, J. L. Lin, F. K. Men, M. Henzler, and F. J. Himpsel, *Phys. Rev. B* 59, 10356 (1999).
253. D. Y. Petrovykh, K. N. Altmann, J. L. Lin, F. J. Himpsel, and F. M. Leibsle, *Surf. Sci.* 512, 269 (2002).
254. J. Viernow, D. Y. Petrovykh, F. K. Men, A. Kirakosian, J. L. Lin, and F. J. Himpsel, *Appl. Phys. Lett.* 74, 2125 (1999).
255. M. Batzill, M. Sarstedt, and K. J. Snowdon, *Nanotechnology* 9, 20 (1998).
256. M. Batzill and K. J. Snowdon, *Appl. Phys. Lett.* 77, 1955 (2000).
257. H. Rauscher, T. A. Jung, J. L. Lin, A. Kirakosian, F. J. Himpsel, U. Rohr, and K. Mullen, *Chem. Phys. Lett.* 303, 363 (1999).
258. J. L. Lin, H. Rauscher, A. Kirakosian, F. J. Himpsel, and P. A. Dowben, *J. Appl. Phys.* 86, 5492 (1999).

259. J. L. Lin, D. Y. Petrovykh, A. Kirakosian, H. Rauscher, F. J. Himpsel, and P. A. Dowben, *Appl. Phys. Lett.* 78, 829 (2001).
260. H. Sunamura, N. Usami, Y. Shiraki, and S. Fukatsu, *Appl. Phys. Lett.* 68, 1847 (1996).
261. K. Brunner, J. Zhu, C. Miesner, G. Abstreiter, O. Kienzle, and F. Ernst, *Physica E* 7, 881 (2000).
262. Y. H. Phang, C. Teichert, M. G. Lagally, L. J. Peticolos, J. C. Bean, and E. Kasper, *Phys. Rev. B* 50, 14435 (1994).
263. J. Tersoff, Y. H. Phang, Z. Zhan, and M. G. Lagally, *Phys. Rev. Lett.* 75, 2730 (1995).
264. F. Liu, J. Tersoff, and M. G. Lagally, *Phys. Rev. Lett.* 80, 1268 (1998).
265. K. Brunner, J. Zhu, G. Abstreiter, O. Kienzle, and F. Ernst, *Thin Solid Films* 369, 39 (2000).
266. G. Jin, Y. S. Tang, J. L. Liu, and K. L. Wang, *J. Vac. Sci. Technol. A* 17, 1406 (1999).
267. G. Jin, Y. S. Tang, J. L. Liu, and K. L. Wang, *Appl. Phys. Lett.* 74, 2471 (1999).
268. L. Gonzalez, J. M. Garcia, R. Garcia, F. Briones, J. Martinez-Pastor, and C. Ballesteros, *Appl. Phys. Lett.* 76, 1104 (2000).
269. H. J. Kim, Y. J. Park, Y. M. Park, E. K. Kim, and T. W. Kim, *Appl. Phys. Lett.* 78, 3253 (2001).
270. Z. Jin, B. Z. Wang, Y. H. Peng, F. H. Zhao, W. Y. Chen, S. Y. Liu, and C. X. Gao, *Surf. Sci.* 423, L211 (1999).
271. A. Li, F. Liu, D. Y. Petrovykh, J. L. Lin, J. Viernow, F. J. Himpsel, and M. G. Lagally, *Phys. Rev. Lett.* 85, 5380 (2000).
272. K. O. Ng and D. Vanderbilt, *Phys. Rev. B* 52, 2177 (1995).
273. K. Sakamoto, H. Matsuhata, M. O. Tanner, D. W. Wang, and K. L. Wang, *Thin Solid Films* 321, 55 (1998).
274. I. Berbezier, B. Gallas, L. Lapena, J. Fernandez, J. Derrien, and B. Joyce, *J. Vac. Sci. Technol. B* 16, 1582 (1998).
275. C. Teichert, Y. H. Phang, L. J. Peticolas, J. C. Bean, and M. G. Lagally, in "Surface Diffusion: Atomistic and Collective Processes" (M. C. Tringides, Ed), Vol. 360, p. 297. Plenum, New York, 1997.
276. C. Teichert, J. C. Bean, and M. G. Lagally, *Appl. Phys. A* 67, 675 (1998).
277. J. F. MacKay, C. Teichert, D. E. Savage, and M. G. Lagally, *Phys. Rev. Lett.* 77, 3925 (1996).
278. C. Teichert, J. Barthel, H. P. Oepen, and J. Kirschner, *Appl. Phys. Lett.* 74, 588 (1999).
279. J. H. Zhu, K. Brunner, and G. Abstreiter, *Appl. Phys. Lett.* 73, 620 (1998).
280. P. Politi, G. Grenet, A. Marty, A. Ponchet, and J. Villain, *Phys. Rep.* 324, 271 (2000).
281. J. W. Matthews and A. E. Blakeslee, *J. Cryst. Growth* 27, 118 (1974).
282. J. C. Bean, L. C. Feldman, A. T. Fiory, S. Nakahara, and I. K. Robinson, *J. Vac. Sci. Technol. A* 2, 436 (1984).
283. M. Albrecht, S. Christiansen, J. Michler, W. Dorsch, H. P. Strunk, P. O. Hansson, and E. Bauser, *Appl. Phys. Lett.* 67, 1232 (1995).
284. M. A. Lutz, R. M. Feenstra, F. K. Legoues, P. M. Mooney, and J. O. Chu, *Appl. Phys. Lett.* 66, 724 (1995).
285. C. Teichert, C. Hofer, K. Lyutovich, M. Bauer, and E. Kasper, *Thin Solid Films* 380, 25 (2000).
286. S. Y. Shiryaev, F. Jensen, J. L. Hansen, J. W. Petersen, and A. N. Larsen, *Phys. Rev. Lett.* 78, 503 (1997).
287. E. V. Pedersen, S. Y. Shiryaev, F. Jensen, J. L. Hansen, and J. W. Petersen, *Surf. Sci.* 399, L351 (1998).
288. H. Brune, M. Giovannini, K. Bromann, and K. Kern, *Nature* 394, 451 (1998).
289. B. Voigtlander and N. Theuerkauf, *Surf. Sci.* 461, L575 (2000).
290. A. Patrykiejew, S. Sokolowski, and K. Binder, *Surf. Sci. Rep.* 37, 209 (2000).
291. S. F. Bent, *Surf. Sci.* 500, 879 (2002).
292. K. Oura, V. G. Lifshits, A. A. Saranin, A. V. Zotov, and M. Katayama, *Surf. Sci. Rep.* 35, 1 (1999).
293. H. Rauscher, *Surf. Sci. Rep.* 42, 207 (2001).
294. R. A. Wolkow, *Annu. Rev. Phys. Chem.* 50, 413 (1999).
295. R. J. Hamers, J. S. Hovis, S. K. Coulter, M. D. Ellison, and D. F. Padowitz, *Jpn. J. Appl. Phys.* 39, 4366 (2000).
296. J. M. Buriak, *Chem. Rev.* 102, 1271 (2002).
297. F. Schreiber, *Prog. Surf. Sci.* 65, 151 (2000).
298. D. K. Schwartz, *Annu. Rev. Phys. Chem.* 52, 107 (2001).
299. W. J. Royea, A. Juang, and N. S. Lewis, *Appl. Phys. Lett.* 77, 1988 (2000).
300. A. Bansal and N. S. Lewis, *J. Phys. Chem. B* 102, 1067 (1998).
301. F. S. Bates and G. H. Fredrickson, *Annu. Rev. Phys. Chem.* 41, 525 (1990).
302. F. S. Bates, *Science* 251, 898 (1991).
303. M. J. Fasolka and A. M. Mayes, *Ann. Rev. Mater. Res.* 31, 323 (2001).
304. M. Park, C. Harrison, P. M. Chaikin, R. A. Register, and D. H. Adamson, *Science* 276, 1401 (1997).
305. M. Park, P. M. Chaikin, R. A. Register, and D. H. Adamson, *Appl. Phys. Lett.* 79, 257 (2001).
306. K. W. Guarini, C. T. Black, K. R. Milkove, and R. L. Sandstrom, *J. Vac. Sci. Technol. B* 19, 2784 (2001).
307. R. D. Peters, X. M. Yang, and P. F. Nealey, *Macromolecules* 35, 1822 (2002).
308. R. E. Cohen, *Curr. Opin. Solid State Mat. Sci.* 4, 587 (1999).
309. W. A. Lopes, *Phys. Rev. E* 65, 031606 (2002).
310. M. J. Fasolka, D. J. Harris, A. M. Mayes, M. Yoon, and S. G. J. Mochrie, *Phys. Rev. Lett.* 79, 3018 (1997).
311. N. Rehse, C. Wang, M. Hund, M. Geoghegan, R. Magerle, and G. Krausch, *Eur. Phys. J. E* 4, 69 (2001).
312. L. Rockford, Y. Liu, P. Mansky, T. P. Russell, M. Yoon, and S. G. J. Mochrie, *Phys. Rev. Lett.* 82, 2602 (1999).
313. L. Rockford, S. G. J. Mochrie, and T. P. Russell, *Macromolecules* 34, 1487 (2001).
314. G. P. Lopinski, D. J. Moffatt, D. D. M. Wayner, M. Z. Zgierski, and R. A. Wolkow, *J. Am. Chem. Soc.* 121, 4532 (1999).
315. G. P. Lopinski, D. J. Moffatt, D. D. Wayner, and R. A. Wolkow, *Nature* 392, 909 (1998).
316. R. A. Wolkow, *Jpn. J. Appl. Phys.* 40, 4378 (2001).
317. V. Palermo, M. Buchanan, A. Bezinger, and R. A. Wolkow, *Appl. Phys. Lett.* 81, 3636 (2002).
318. G. P. Lopinski, D. D. M. Wayner, and R. A. Wolkow, *Nature* 406, 48 (2000).
319. P. Kruse, E. R. Johnson, G. A. DiLabio, and R. A. Wolkow, *Nano Lett.* 2, 807 (2002).
320. R. L. Cicero, C. E. D. Chidsey, G. P. Lopinski, D. D. M. Wayner, and R. A. Wolkow, *Langmuir* 18, 305 (2002).
321. Y. L. Wang and M. Y. Lai, *J. Phys.: Condens. Matter* 13, R589 (2001).
322. J. Schulze, T. Stimpel, H. Baumgartner, and I. Eisele, *Appl. Surf. Sci.* 162, 332 (2000).
323. Y. P. Zhang, L. Yan, S. S. Xie, S. J. Pang, and H. J. Gao, *Surf. Sci.* 497, L60 (2002).
324. M. Y. Lai and Y. L. Wang, *Phys. Rev. B* 6424, 241404 (2001).
325. L. Vitali, M. G. Ramsey, and F. P. Netzer, *Phys. Rev. Lett.* 83, 316 (1999).
326. L. Vitali, M. G. Ramsey, and F. P. Netzer, *Appl. Surf. Sci.* 175, 146 (2001).
327. J. F. Jia, J. Z. Wang, X. Liu, Q. K. Xue, Z. Q. Li, Y. Kawazoe, and S. B. Zhang, *Appl. Phys. Lett.* 80, 3186 (2002).
328. J. L. Li, J. F. Jia, X. J. Liang, X. Liu, J. Z. Wang, Q. K. Xue, Z. Q. Li, J. S. Tse, Z. Y. Zhang, and S. B. Zhang, *Phys. Rev. Lett.* 88, 066101 (2002).
329. S. Hasegawa, *J. Phys.: Condens. Matter* 12, R463 (2000).
330. F. J. Himpsel, *Adv. Phys.* 32, 1 (1983).
331. F. J. Himpsel, K. N. Altmann, J. N. Crain, A. Kirakosian, J. L. Lin, and A. Liebsch, *J. Electr. Spectrosc. Relat. Phenom.* 126, 89 (2002).
332. S. D. Kevan, *J. Electr. Spectrosc. Relat. Phenom.* 75, 175 (1995).
333. F. J. Himpsel, *Surf. Sci. Rep.* 12, 1 (1990).

334. M. Chester and T. Gustafsson, *Surf. Sci.* 256, 135 (1991).
335. Y. G. Ding, C. T. Chan, and K. M. Ho, *Surf. Sci.* 275, L691 (1992).
336. H. M. Zhang, K. Sakamoto, and R. I. G. Uhrberg, *Phys. Rev. B* 6424, 245421 (2001).
337. H. M. Zhang, T. Balasubramanian, and R. I. G. Uhrberg, *Phys. Rev. B* 6503, 035314 (2002).
338. H. M. Zhang, T. Balasubramanian, and R. I. G. Uhrberg, *Phys. Rev. B* 66, 165402 (2002).
339. J. N. Crain, K. N. Altmann, C. Bromberger, and F. J. Himpsel, *Phys. Rev. B* 66, 205302 (2002).
340. J. M. Luttinger, *J. Math. Phys.* 4, 1154 (1963).
341. V. Meden and K. Schonhammer, *Phys. Rev. B* 46, 15753 (1992).
342. J. Voit, *Rep. Prog. Phys.* 58, 977 (1995).
343. N. Shannon and R. Joynt, *J. Phys.: Condens. Matter* 8, 10493 (1996).
344. M. G. Zacher, E. Arrigoni, W. Hanke, and J. R. Schrieffer, *Phys. Rev. B* 57, 6370 (1998).
345. R. B. Laughlin, *Rev. Mod. Phys.* 71, 863 (1999).
346. H. L. Stormer, *Rev. Mod. Phys.* 71, 875 (1999).
347. M. Grioni, I. Vobornik, F. Zwick, and G. Margaritondo, *J. Electr. Spectrosc. Relat. Phenom.* 100, 313 (1999).
348. J. K. Jain, *Phys. Today* 53, 39 (2000).
349. J. Voit, *J. Electr. Spectrosc. Relat. Phenom.* 117, 469 (2001).
350. I. K. Robinson, P. A. Bennett, and F. J. Himpsel, *Phys. Rev. Lett.* 88, 096104 (2002).
351. A. A. Saranin, A. V. Zotov, V. G. Lifshits, M. Katayama, and K. Oura, *Surf. Sci.* 426, 298 (1999).
352. H. W. Yeom, S. Takeda, E. Rotenberg, I. Matsuda, K. Horikoshi, J. Schaefer, C. M. Lee, S. D. Kevan, T. Ohta, T. Nagao, and S. Hasegawa, *Phys. Rev. Lett.* 82, 4898 (1999).
353. I. G. Hill and A. B. McLean, *Phys. Rev. Lett.* 82, 2155 (1999).
354. A. Kirakosian, J. L. McChesney, R. Bennewitz, J. N. Crain, J. L. Lin, and F. J. Himpsel, *Surf. Sci.* 498, L109 (2002).
355. S. C. Erwin and H. H. Weitering, *Phys. Rev. Lett.* 81, 2296 (1998).
356. T. Sakamoto and G. Hashiguchi, *Jpn. J. Appl. Phys.* 25, L78 (1986).
357. T. Abukawa, T. Okane, and S. Kono, *Surf. Sci.* 256, 370 (1991).
358. R. Shioda and J. van der Weide, *Phys. Rev. B* 57, R6823 (1998).
359. A. A. Baski, J. Nogami, and C. F. Quate, *Phys. Rev. B* 43, 9316 (1991).
360. M. M. R. Evans and J. Nogami, *Phys. Rev. B* 59, 7644 (1999).
361. P. Soukiassian, F. Semond, A. Mayne, and G. Dujardin, *Phys. Rev. Lett.* 79, 2498 (1997).
362. J. N. Crain, K. N. Altmann, C. Bromberger, S. C. Erwin, A. Kirakosian, J. L. McChesney, J.-L. Lin, and F. J. Himpsel, *Phys. Rev. Lett.* 90, 176805 (2003).
363. A. A. Baski, K. M. Saoud, and K. M. Jones, *Appl. Surf. Sci.* 182, 216 (2001).
364. A. A. Baski and K. M. Saoud, *J. Clust. Sci.* 12, 527 (2001).
365. A. A. Baski, K. M. Jones, and K. M. Saoud, *Ultramicroscopy* 86, 23 (2001).
366. H. Minoda, *J. Cryst. Growth* 237, 21 (2002).
367. P. Mazurek and M. Jalochowski, *Opt. Appl.* 32, 247 (2002).
368. R. Zdyb, M. Strozak, and M. Jalochowski, *Vacuum* 63, 107 (2001).
369. M. Jalochowski, M. Strozak, and R. Zdyb, *Surf. Sci.* 375, 203 (1997).
370. D. Sanchez-Portal, J. D. Gale, A. Garcia, and R. M. Martin, *Phys. Rev. B* 65, 081401 (2002).
371. S. C. Erwin, personal communication.
372. R. Bennewitz, J. N. Crain, A. Kirakosian, J. L. Lin, J. L. McChesney, D. Y. Petrovykh, and F. J. Himpsel, *Nanotechnology* 13, 499 (2002).
373. A. Kirakosian, R. Bennewitz, F. J. Himpsel, and L. W. Bruch, *Phys. Rev. B* 67, 205412 (2003).
374. L. M. Greig and D. Philp, *Chem. Soc. Rev.* 30, 287 (2001).
375. N. C. Seeman and A. M. Belcher, *Proc. Natl. Acad. Sci. USA* 99, 6451 (2002).
376. C. M. Niemeyer, *Angew. Chem. Int. Ed.* 40, 4128 (2001).
377. D. G. Castner and B. D. Ratner, *Surf. Sci.* 500, 28 (2002).
378. B. Kasemo, *Surf. Sci.* 500, 656 (2002).
379. U. B. Sleytr, P. Messner, D. Pum, and M. Sara, *Angew. Chem. Int. Ed.* 38, 1035 (1999).
380. A. S. G. Curtis, B. Casey, J. O. Gallagher, D. Pasqui, M. A. Wood, and C. D. W. Wilkinson, *Biophys. Chem.* 94, 275 (2001).
381. P. F. Nealey, A. I. Teixeira, G. A. Abrams, and C. J. Murphy, *Abstr. Pap. Am. Chem. Soc.* 221, 118 (2001).
382. A. Prokop, in "Bioartificial Organs III: Tissue Sourcing, Immunoisolation, and Clinical Trials," Vol. 944, p. 472. Annals of the New York Academy of Sciences, 2001.
383. B. Muller, *Surf. Rev. Lett.* 8, 169 (2001).
384. T. Strother, W. Cai, X. S. Zhao, R. J. Hamers, and L. M. Smith, *J. Am. Chem. Soc.* 122, 1205 (2000).
385. T. Strother, R. J. Hamers, and L. M. Smith, *Nucleic Acids Res.* 28, 3535 (2000).
386. Z. Lin, T. Strother, W. Cai, X. P. Cao, L. M. Smith, and R. J. Hamers, *Langmuir* 18, 788 (2002).
387. M. C. Pirrung, *Angew. Chem. Int. Ed.* 41, 1277 (2002).
388. M. J. Tarlov and A. B. Steel, in "Biomolecular Films: Design, Function, and Applications" (J. F. Rusling, Ed.), Vol. 111. Marcel Dekker, New York, 2003.
389. J. Wang, *Nucleic Acids Res.* 28, 3011 (2000).
390. W. C. I. Homs, *Anal. Lett.* 35, 1875 (2002).
391. A. Kirakosian, J. L. Lin, D. Y. Petrovykh, J. N. Crain, and F. J. Himpsel, *J. Appl. Phys.* 90, 3286 (2001).
392. J. N. Crain, A. Kirakosian, J. L. Lin, Y. D. Gu, R. R. Shah, N. L. Abbott, and F. J. Himpsel, *J. Appl. Phys.* 90, 3291 (2001).
393. R. D. Peters, P. F. Nealey, J. N. Crain, and F. J. Himpsel, *Langmuir* 18, 1250 (2002).
394. A. K. Chakraborty and A. J. Golumbfskie, *Annu. Rev. Phys. Chem.* 52, 537 (2001).
395. J. Gunther and S. I. Stupp, *Langmuir* 17, 6530 (2001).
396. R. R. Shah and N. L. Abbott, *Science* 293, 1296 (2001).
397. V. K. Gupta, J. J. Skaife, T. B. Dubrovsky, and N. L. Abbott, *Science* 279, 2077 (1998).
398. D. L. Everitt, W. J. W. Miller, N. L. Abbott, and X. D. Zhu, *Phys. Rev. B* 62, R4833 (2000).
399. J. J. Skaife and N. L. Abbott, *Langmuir* 16, 3529 (2000).
400. J. J. Skaife and N. L. Abbott, *Chem. Mater.* 11, 612 (1999).
401. E. H. Lay, A. Kirakosian, J. L. Lin, D. Y. Petrovykh, J. N. Crain, F. J. Himpsel, R. R. Shah, and N. L. Abbott, *Langmuir* 16, 6731 (2000).
402. S. R. Kim, A. I. Teixeira, P. F. Nealey, A. E. Wendt, and N. L. Abbott, *Adv. Mater.* 14, 1468 (2002).
403. T. A. Winningham, H. P. Gillis, D. A. Choutov, K. P. Martin, J. T. Moore, and K. Douglas, *Surf. Sci.* 406, 221 (1998).
404. R. Bashir, *Superlattices and Microstructures* 29, 1 (2001).
405. J. J. Storhoff and C. A. Mirkin, *Chem. Rev.* 99, 1849 (1999).
406. N. C. Seeman, *Angew. Chem. Int. Ed.* 37, 3220 (1998).
407. J. D. Hartgerink, E. Beniash, and S. I. Stupp, *Proc. Natl. Acad. Sci. USA* 99, 5133 (2002).
408. J. D. Hartgerink, E. R. Zubarev, and S. I. Stupp, *Curr. Opin. Solid State Mat. Sci.* 5, 355 (2001).
409. D. T. Bong, T. D. Clark, J. R. Granja, and M. R. Ghadiri, *Angew. Chem. Int. Ed.* 40, 988 (2001).
410. S. W. Lee, C. B. Mao, C. E. Flynn, and A. M. Belcher, *Science* 296, 892 (2002).
411. S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, and A. M. Belcher, *Nature* 405, 665 (2000).

www.aspbs.com/enn

Self-Assembled Organic/Inorganic Nanocomposites

Byron McCaughey, J. Eric Hampsey, Donghai Wang, Yunfeng Lu

Tulane University, New Orleans, Louisiana, USA

CONTENTS

1. Introduction
2. Background of Self-Assembly
3. Synthesis of Organic/Inorganic Nanocomposites via Liquid Crystalline Templates
4. Other Assembly Techniques
5. Inorganic Particle/Polymer Nanocomposites
Glossary
References

1. INTRODUCTION

Composite materials are a very common part of everyday life. Examples of macroscale composites include plywood (composed of layered wood and glue), fiberglass (composed of glass fibers surrounded by an epoxy glue), reinforced concrete (composed of steel wires surrounded by concrete), socks (composed of interwoven elastic rubber and cotton fibers), and paint (inorganic pigment particles suspended in an organic mixture). Simply put, a nanocomposite is a material composed of two or more discrete, nanometer-sized (1 to 100 nm) components. Self-assembled organic/inorganic nanocomposites are composed of discrete nanoscale organic and inorganic components that have been spontaneously organized based on noncovalent interactions. These organic/inorganic nanocomposites often demonstrate interesting properties because of the nanoscale size effects in constituent phases, high interfacial area, and synergic properties of these components.

1.1. Size Effects

Material properties depend on the nature and sizes of the constituent components. For example, metal [1, 2], inorganic ceramic [3], semiconductor [3, 4], and other electrically and/or optically conductive material [5, 6] properties change from bulk toward quantum-derived as critical size dimensions are decreased toward the nanoscale. Metal nanoparticles, have overall properties between those of an individual molecule and bulk [7]. A decrease in particle size also decreases the number of neighboring metal atoms that facilitate metal–metal bonding due to an increase in the number of surface atoms [8, 9]. Because individual atoms do not hold core elections as strongly, metal properties such as the work function are altered. Inorganic nanoparticles with smaller particle sizes may show increased bandgap, and semiconductor nanoparticles may demonstrate quantum well effects that are useful for the control and storage of electrons [3]. For solids composed of arrays of nanoparticles, properties such as conduction of light and electricity depend heavily on the size and size distribution of the original particles [10]. For example, photonic crystals, based on ordered nanoparticle arrays, control the propagation of photons due to periodic variation in dielectric constant in an analogous fashion to semiconductor control of electrons [5, 6]. Natural opal is iridescent because the ordered, periodic array of nanoparticles with similar size as visible light wavelengths (400 to 700 nm) diffracts visible light [5]. Metallic arrays experience changes in electrical properties such as capacitance and conductivity, while magnetic materials experience magnetostatic interactions that change bulk magnetic properties [6].

As the components decrease in size, the interfacial or surface area increases dramatically. This can have a profound effect on the final system properties. For example, the melting point of a collection of nanoparticles is lower than bulk material because of the high surface area [11]. Higher pressure is also required to induce solid-state phase transitions in many nanoparticles due to a kinetic barrier for breaking surface bonds as opposed to those in the bulk [11]. In general, placing two materials with different properties in close, ordered proximity produces new controllable structures and properties because of beneficial interactions between the phases [12, 13].

1.2. Property Synergies

Organic/inorganic nanocomposites are attractive because properties of inorganics (e.g., rigidity, high surface area, thermal and mechanical stability) are combined with those

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

of organics (e.g., optical, electrical, and magnetic functionality, porosity, and hydrophobicity). Optical materials, magnetic materials [14], electrical materials such as low k thin films, solid electrolytes, structural materials, improved plastics, coatings, chemical sensors [14], catalysts, separation membranes [12, 13], and other advanced applications are possible as a result of these unique and often superior properties.

In many cases, optical nanocomposites have improved optical efficiency, mechanical strength, and sensing functionality [15] while maintaining optical transparency [16]. Materials with optical anisotropy have aligned organic molecules that are stabilized by an inorganic structural framework [17]. Addition of organic components such as monomers and photosensitive or fluorescent dyes to a transparent ceramic strengthens the material and introduces new functionality [17]. Finally, robust mechanical sensors composed of color-changing, sensing polymers encapsulated within a rigid framework have been constructed [18].

Normally, electrical components are composed of inorganic materials because of controllable conduction, magnetism, and dielectric transitions inherent to these materials [17]. However, a reduction of dielectric constant and brittleness may be possible through a combination of organic and inorganic components. As a result, nanocomposites would allow a decrease in microchip component size. High computing speeds may be possible due to the quantum control and storage of electrons in organic/inorganic structures [3]. Also, nanocomposite solid electrolytes with better ion transport properties due to changes in crystalline structure can be synthesized [19, 20].

Mechanical properties of structural components are improved by utilizing nanocomposites. These composites benefit from the combination of inorganic phase thermal stability and mechanical hardness with organic phase plasticity and synthetic flexibility. Nanocomposite materials may experience an increase in strength and a reduction of brittleness due to decreased crack propagation. For example, composite coatings that mimic the alternating biopolymer/$CaCO_3$ layers in nacre seashell have shown improved indention hardness [21]. Polymers containing nanoparticles typically have improved mechanical and thermal properties due to altered polymer chain [22] and nanoparticle mobility [23].

Other applications for nanocomposites include sensors, catalysts, and separation membranes. Sensors utilize high surface area mesoporous ceramics functionalized with organic sensing layers or constituents [24, 25]. For catalysts, nanocomposites composed of organometallic active centers on high surface area inorganic supports are more thermally stable than those on polymeric supports [26]. These catalytic materials often have greater selectivity, reactivity, and durability as a result [4]. Organic/inorganic separation membranes have better selectivity and performance because of the functionality, site specificity, and high surface areas [27–31].

1.3. Synthesis Methods

Effective nanocomposite synthesis depends on the ability to manipulate the dimension, composition, structure, and cooperative interactions of individual components [12]. In general, nanostructured materials are synthesized by breaking apart bulk materials through ball milling, lithography, or other techniques or by assembling nanoscale building blocks or scaffolds together. The former technique is used extensively in the semiconductor and electronics industries through electron beam lithography, reactive ion etching, and other techniques. Ball milling, in which mechanical abrasion physically breaks a bulk material apart, is often used to create nanoscale building blocks that can be subsequently assembled together [32]. Although nanocomposites have been synthesized for decades by both routes, the recent development of self-assembly has broadly expanded the interest and synthetic flexibility in the latter "bottom up" approach. In fact, self-assembly has emerged as one of the most promising techniques for the efficient fabrication of nanostructured materials.

The method of self-assembly spontaneously assembles and organizes various building blocks into periodic, hierarchical structures via noncovalent intramolecular or intraparticulate interactions [33]. The basic synthesis strategy has been outlined by Ozin et. al. [34]:

1. Building blocks such as molecules or particles are selected, localized, and concentrated into specific areas of the system.
2. Structure directing agents assemble, create defined spaces and architectures, and partition the building blocks into specified regions.
3. Reactions or assembly within the defined spatial regions under the influence of interfacial forces results in formation of self-assembled composites.

The advantage of self-assembly is that it starts from components in almost every conceivable ratio and uses mild, flexible synthetic steps that are versatile enough to form nanocomposites with a wide variety of material properties and property enhancements. Because self-assembly is a spontaneous process involving a balance of multiple noncovalent interactions between components, the final materials are often robust, stable, and possibly self-healing [35].

This chapter addresses organic/inorganic nanocomposites synthesized via direct self-assembling synthetic procedures or via postsynthetic procedures involving modification and incorporation of organic components into preformed, self-assembled structures. In the scope of this work, self-assembly does not refer to synthesis procedures that employ electric or magnetic fields to induce system organization. Langmuir–Blodgett, self-assembled monolayers, and other two-dimensional (2D) array materials are not included because of their inability to form consolidated engineering materials. The field of supramolecular chemistry also forms a wide variety of structures containing both organic and inorganic components. However, these materials will only be addressed briefly.

2. BACKGROUND OF SELF-ASSEMBLY

Self-assembly occurs with molecules (nanometer sized or less), colloidal particles (1–2000 nm) [36], or macroscale objects (mm) [37, 38]. The building blocks used in self-assembly are important because they carry information such as shape, topology, and surface properties that will define

material characteristics [5]. Controlling these building blocks will allow for materials with improved overall properties.

Biological processes that involve molecular recognition, self-assembly, and replication such as protein synthesis or DNA helix structures inspired much of the early work toward self-assembly and molecular templating of inorganic crystals [34]. In the formation of bone and shell, biological systems utilize self-assembly via calcium carbonate deposition and biomineralization around preformed, biologically controlled vesicle arrays [39]. Examples of other naturally occurring materials formed in this way are packed silica spheres in opal, diatom silica skeletons, and layered inorganic/biopolymer in nacre (abalone shell), as shown in Figure 1. The advantage of these biological materials is that they are organized on multiple length scales in a hierarchical fashion [40]. Today, a field called biomimetics [40] uses ideas that are inspired by observing natural solutions to materials science problems and extracting the essential operational features for use in novel synthesis procedures. For example, cystein and lysine amino acids, which are frequently used in sponges to direct inorganic growth, have been used to synthesize spherical and columnar silica [41].

Self-assembly has synthesized many useful materials over the past 40 years. For example, in the 1960s, zeolites, with 0.4 to 1.5 nm micropores, were synthesized via self-assembly of crystalline inorganic structures around organic structure directing agents such as alkylammonium ions [36, 43]. Zeolites have been used extensively in industry for catalysis and separations and several reviews are available [43–46]. However, many catalytic and separation applications require materials with larger pore channels. In 1992, MCM-41 became the first mesoporous material with large pore sizes to be synthesized by supramolecular self-assembly (also called supramolecular templating) [39, 47]. This mesoporous silicate contains a hexagonal array of pores templated by self-assembled surfactant liquid crystalline structures [48, 49]. The discovery of this approach greatly motivated research in this field. A series of ceramics such as silicates, aluminosilicates [50], titanium oxide [51], niobium oxide [51], tantalum oxide [51], tin oxide [51], tungsten oxide [51], alumina [52], or various mixed oxides such as aluminosilicates can be synthesized by this method. The pore sizes of these materials can be up to 50 nm and the pore channels can be arranged in hexagonal, cubic, lamellar, or other mesophases. However, unlike zeolites that contain crystalline walls, these mesoporous materials often have amorphous pore walls.

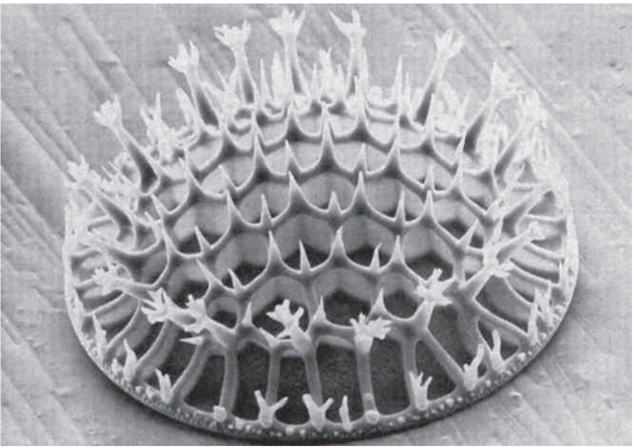


Figure 1. Single valve of an unidentified diatom found in a deep-sea core of an Eocene marine deposit. Reprinted with permission from [42], C. W. Li and B. E. Volcani, *Philos. Trans. Roy. Soc. London Ser. B* 304, 518 (1984). © 1984, The Royal Society of Chemistry.

2.1. Forces in Self-Assembly

Self-assembly is based on weak interactions such as hydrogen bonding (3–5 kcal/mol), electrostatic interactions, van der Waals forces (1 kcal/mol), π–π interactions, solvophobic (such as hydrophobic) interactions, and capillary forces. Hydrogen bonding, the most frequently used interaction in self-assembly [36], involves the strong attraction that occurs between an electronegative atom like nitrogen (N) or oxygen (O) and a hydrogen (H) that is bonded to another electronegative atom. Electrostatic forces attract oppositely charged species. Van der Waals attractive forces result from instantaneous, nonuniform distributions of electrons that are manifested as temporary and induced dipoles in adjacent molecules. Extended delocalized *p* systems like benzene rings or conjugated polymers experience attractive forces that encourage molecular stacking based on the favorable overlap of *p* orbitals. Because of unfavorable entropic ordering of polar molecules in a hydration sheath around hydrocarbons, solvophobic interactions arise that encourage hydrocarbon agglomeration and possible phase separation. Capillary forces are based on unequal pressure across curved liquid–vapor interfaces between nanoparticles and within pores. Because intermolecular forces such as van der Waals, hydrogen bonding, and π–π interactions usually are inversely proportional to higher powers of molecular distance, they are only relevant at very short distances. With the exception of hydrogen bonding and electrostatic interactions, all of these bonds have a similar energy to thermal oscillation at room temperature [35]. The self-assembly process usually requires several of these weak interactions to work in concert.

2.2. Self-Assembly of Surfactants

Amphiphiles self-organize into a wide variety of structures (also called association colloids) [53] such as micelles, emulsions, or liquid crystal aggregates [1]. This unique capability enables self-assembled amphiphiles to act as structure directing agents in the synthesis of organic/inorganic nanocomposites and other self-assembled materials.

2.2.1. Surfactants

Surfactants, a specific class of the amphiphiles that are often used as detergents and soaps, can be categorized based on their anionic, cationic, or nonionic hydrophilic headgroups. Anionic groups include sulfate, sulfonate, phosphate, and carboxylate while cationic surfactants often have pyridinium or quaternary ammonium headgroups. Nonionic surfactants usually contain hydrophilic poly(ethylene oxide) headgroups. Table 1 shows examples from each of these surfactant classes.

Table 1. Common surfactants used in self-assembly.

Surfactant	Name/class	Hydrophobic–hydrophillic
O—SO_3Na	sodium dodecyl sulfate	dodecane-sulfate
$N^+(CH_3)_2H_3C$ Br^-	cetyltrimethyl-ammonium bromide (CTAB)	alkyl chain-ammonium bromide
$HO-[CH_2CH_2O]_x-[CH_2CH(CH_3)O]_y-[OCH_2CH_2]_z-OH$	pluoronic	poly(oxypropylene)–poly(oxyethylene)
$[\ ]_x[O\ ]_y$OH	brij	polyethylene–poly(oxyethylene)
$H_3C-[CH_2]_x-C{\equiv}C-C{\equiv}C-[CH_2]_y-C(=O)OH$	diacetylene	alkyl chain-carboxylic acid

2.2.2. Micelles

Surfactant molecules have a tendency to aggregate at surfaces (hence the name surfactant from surface active). They can also self-organize into micelles and other structures when a critical concentration (termed the critical micelle concentration or CMC) is reached. The theory describing the formation of micelles was first proposed by McBrain and Salmon in 1920 [54]. As Figure 2 shows, micelles in an aqueous system are roughly spherical pseudophases that consist of 50 to 100 surfactant molecules. Surfactant molecules are arranged so that their hydrophilic heads are exposed at the micelle aqueous media interface. Surfactant hydrophobic tails are clustered and intertwined [53, 55] to form an internal environment that very closely resembles a hydrocarbon liquid such as decane but with higher viscosity [56, 57]. The hydrocarbon tails in micelles or liquid crystals are not arranged like radial spokes in a wheel; they are in fact jumbled and often bent so that interaction between tails and the aqueous phase is minimized [58]. These structures are deformable and they readily change size, shape, and morphology due to new environmental conditions such as temperature and salt or surfactant concentration [58]. As Figure 3 demonstrates, micelles are in a state of dynamic equilibrium in which surfactant molecules are added and removed from the micelle ($t_1 \approx$ microseconds) and entire micelles can be destroyed and recreated ($t_2 \approx$ milliseconds to seconds) [58]. Note that if a hydrophobic solvent is used instead of water, reverse micelles that project their tail outward can be formed.

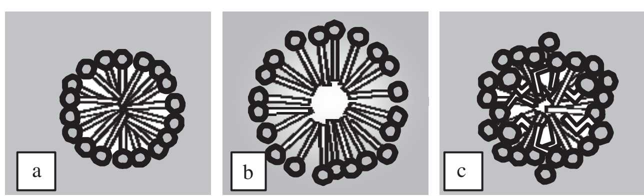


Figure 2. A schematic representation of micelle structure. Three possible tail arrangements are illustrated: (a) tails overlap at the center; (b) water penetrates the core; and (c) chain protrusion and bending correct deficiencies of (a) and (b). Panel (c) most closely resembles the actual physical structure of a micelle [53].

An equation [59] has been used to determine the number of surfactant molecules per micelle or aggregation number, Z, for many block copolymers and ionic or nonionic surfactants,

$$Z = Z_0 N_A^2 N_B^{-0.8} \tag{1}$$

where N_A and N_B are the chain lengths of the insoluble and soluble portions of the molecule respectively, and Z_0 is related to molecular interactions and volume (refer to [59] for list of values for common surfactants) [4]. The number of surfactant molecules per micelle (Z) increases as the surfactant hydrophobic chain length (N_A) increases [53].

Micellization is a spontaneous process with a negative change in Gibbs free energy (ΔG) given by $\Delta G = \Delta H - T\Delta S$, where ΔH refers to the change in enthalpy and ΔS refers to change in entropy. Example values for these thermodynamic quantities at 25 °C are: $\Delta H = -10$ to 20 kJ/mol, $\Delta S = 50$ to 150 kJ/mol K, and $\Delta G = -20$ to -30 kJ/mol [60]. As a result of a balance between hydrophobic attraction and headgroup repulsions (see Fig. 5), ΔH can be either positive or negative. The favorable replacement of surfactant–water interactions by surfactant tail–surfactant tail and water–water interactions during the micellization process contributes to a negative ΔH [53, 61, 62]. Experimental evidence has shown that only the first 2 methylene groups (out of 12 or more) of hydrophobic tails are exposed

$\bullet_1$

$\bullet_2$

Figure 3. Relaxation times $\bullet_1$ and $\bullet_2$ for surfactant removal from micelle and for micelle destruction respectively [58].

to an aqueous environment while the rest are effectively shielded [63]. However, micellization also creates unfavorable (positive ΔH), repulsive forces between neighboring headgroups such as the electrostatic repulsion between ionic headgroups with the same charge and the hydration sheath overlap around each head group [64]. The driving force for micellization is a large positive entropy change ($\Delta S > 0$). The micellization process decreases the entropy of the surfactant molecules by ordering them into precise arrangements. However, important changes also occur to the water molecules during this process. Strong hydrogen bonding between water molecules must be distorted to accommodate a hydrophobic solute molecule (or hydrophobic tail) into solution [61]. Water will preferentially orient and align with a certain degree of order around this solute in a hydration layer so that interwater hydrogen bonds are maintained [53]. The formation of micelles allows more of the water to return to its less ordered state, thereby contributing a large increase in entropy. Each micelle possesses approximately two layers of water in its hydration sheath, which is far less than the number of water molecules in the combined hydration sheaths of 50 surfactant tails [63].

2.2.3. Surfactant Liquid Crystalline Phases

A small increase in concentration above the CMC causes the number of micelles to increase but it does not change the surfactant aggregation number [65]. As the total number of micelles increases and the distance between micelles decreases, interactions like Coulombic repulsion between micelle counterion clouds [58] and the unfavorable interaction between layers of hydration surrounding each micelle create a driving force for spherical micelles to transform into liquid crystalline phases.

A liquid crystal is composed of aligned molecular arrays that can flow and deform like a liquid. Typically, molecules that form liquid crystals have length to diameter ratios or aspect ratios of 10–20 [62]. Liquid crystals can be composed of molecules that are aligned in layers (cholesteric and smectic) or in a random arrangement (nematic). A lyotropic liquid crystal is one that contains multiple components such as an amphiphile in solvent (soap and water) [66]. Since the molecules that form a liquid crystal have long-range order based on weak intermolecular interactions, they can undergo liquid deformation and at the same time exhibit crystalline optical properties such as the diffraction of light and optical anisotropy (direction dependant properties) [66]. The following discussion will deal exclusively with surfactant lyotropic liquid crystals because they can spatially organize a system into hydrophobic and hydrophilic regions [58] that are useful for the synthesis of hierarchically ordered organic or inorganic materials [67].

Lyotropic surfactant liquid crystals are based on the close packing of surfactant molecules. While micelles are spherical and monodispersed near or slightly above the CMC, an increase in surfactant concentration to 10 times the CMC [68] causes the micelle cores to grow and organize into cylindrical rods or lamellar disks with a wide range of surfactant aggregation numbers [53, 58]. These organized structures pack into periodic arrays in such a way as to expose hydrophilic head groups toward the hydrophilic solvent. Lamellar, hexagonal, or cubic structures with increasing 1- to 3D periodicity are based on the packing of layers, cylinders, or spheres respectively (Fig. 4) [66]. These structures maintain a disordered liquid-like arrangement of surfactant tails within a confined environment [69]. The lamellar structure consists of parallel alternating surfactant double-layer sheets and hydrophilic solvent in which the tails dissolve each other thus exposing the heads to the solvent. Hexagonal phases are built up from a number of long, parallel cylinders of surfactant that are packed together with the heads exposed outward. Uniform surfactant spheres pack together in a way that also exposes head groups outward to form an isotropic cubic arrangement. The cubic mesophase is bicontinuous in that the surfactant tail phase forms a continuous grid in three dimensions surrounded by a continuous grid of water. Hexagonal and cubic structures can be inverted to expose their hydrophobic tails outward if the solvent is hydrophobic instead of hydrophilic. Each mesophase has an interface curvature that depends on a balance of repulsive forces between headgroups and van der Waals interactions between tails (Fig. 5) [70].

Effect of Surfactant Molecule Structure on Liquid Crystalline Phases The shape of a surfactant molecule in the self-assembling system can be geometrically depicted using its cross-sectional head group area (a_0), tail volume (v), and length of hydrophobic tail (l_c) as shown in Figure 5 [53]. The head group area and tail volume are determined by dividing surfactant surface area and internal hydrophobic volume by the micelle aggregation number (number of surfactant molecules per micelle), respectively. The head group area depends on the hydrophilic chain length for nonionic surfactants. The length of the hydrophobic tail is determined empirically by the surfactant chain length in a hydrophobic media. A packing parameter, P, that describes the shape of micellular and liquid crystalline aggregates, is given by [53]

$$P = v/(a_0 l_c) \tag{2}$$

Figure 6 illustrates the relation between the packing parameter P and packing shapes of different surfactants and lipids [65]. Changes in pH, salt concentration, and temperature may affect the headgroup area and therefore the packing

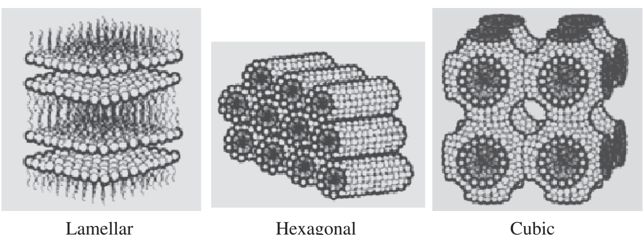


Figure 4. Schematic of liquid crystalline mesophases showing surfactant orientation in lamellar, hexagonal, and cubic morphologies. Reprinted with permission from [33], C. J. Brinker et al., *Adv. Mater.* 11, 580 (1999). © 1999, Wiley–VCH.

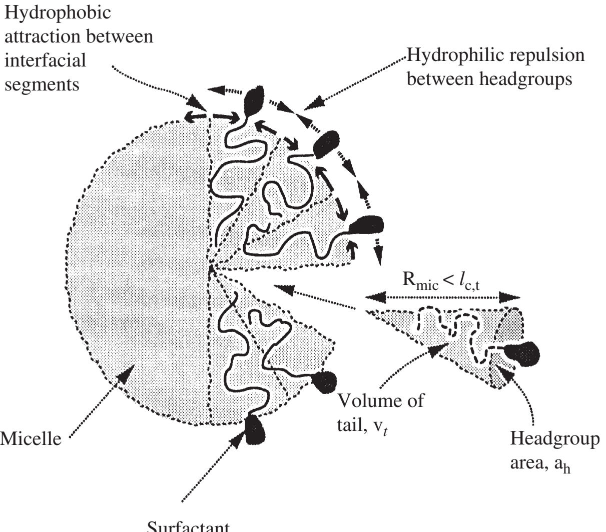


Figure 5. Surfactant geometric parameters and intermolecular interactions. Reprinted with permission from [65], J. N. Israelachvili, "Intermolecular and Surface Forces," 2nd ed. Academic Press, San Diego, 1991. © 1991, Academic Press.

parameter [53]. For example, a counterion salt in solution will act as a buffer between oppositely charged surfactant headgroups and reduce the effective headgroup area thereby changing the surfactant packing parameter [64]. An increase in temperature often reduces the repulsion between headgroups due to a smaller hydration sheath and this can result in a shift from a hexagonal to a lamellar structure [71]. Each liquid crystalline phase has a different mode of hydrocarbon tail packing. The volume fraction occupied by the surfactant in a lyotropic liquid crystalline system is also important in determining the liquid crystalline phase [3].

The use of a packing parameter provides an empirical route to understanding the self-assembling system; however, the packing parameter alone does not fully describe the complexity of surfactant packing. Surfactant concentration [62], system temperature, solvents, external fields [67], and specific surfactant interactions or morphologies may dramatically affect the final assembled structures. For example, an increase in temperature or surfactant concentration can cause a system to shift through the following sequence: spherical micelles, elongated micelles, lamellar liquid crystalline phase, and finally bicontinuous liquid crystalline phase [58]. The structure of the surfactant molecule plays an important role in determining the final structure formed [58]. Surfactants with special interactions and morphologies may form supramolecular assembly with rod or disk morphologies [67] or other complicated structures like gyroid, modulated lamellar and packed spheres [3].

Lipid	Critical packing parameter	Critical packing shape	Structures formed
Single-chained lipids (surfactants) with large head group areas: SDS in low salt	<1/3	Cone	Spherical micelles
Single-chained lipids with small head group areas: SDS and CTAB in high salt, nonionic lipids	1/3–1/2	Truncated cone	Cylindrical micelles
Double-chained lipids with large head group areas, fluid chains: phosphatidyl choline (lecithin), phosphatidyl serine, phosphatidyl glycerol, phosphatidyl inositol, phosphatidic acid, sphingomyelin, DGDG,[a] dihexadecyl phosphate, dialkyl dimethyl ammonium salts	1/2–1	Truncated cone	Flexible bilayers, vesicles
Double-chained lipids with small head group areas, anionic lipids in high salt, saturated frozen chains: phosphatidyl ethanolamine phosphatidyl serine + Ca^{2+}	~1	Cylinder	Planar bilayers
Double-chained lipids with small head group areas, nonionic lipids, poly (cis) unsaturated chains, high T: unsaturated phosphatidyl ethanolamine, cardiolipin + Ca^{2+} phosphatidic acid + Ca^{2+} cholesterol, MGDG[b]	>1	Inverted truncated cone or wedge	Inverted micelles

Figure 6. Packing parameter and its relation to shapes of aggregates. Reprinted with permission from [65], J. Israelachvili, Intermolecular and Surface Forces." 2nd ed. Academic Press, San Diego, 1991. © 1991, Academic Press.

3. SYNTHESIS OF ORGANIC/INORGANIC NANOCOMPOSITES VIA LIQUID CRYSTALLINE TEMPLATES

A large variety of nanocomposites can be synthesized by a self-assembly process that spontaneously organizes and arranges different building blocks into ordered self-assembled structures. Often, materials can be synthesized that have order, functionality, and periodicity on multiple scales using surfactants (refer to Fig. 7). On the molecular scale, chemical design or modification of structural parts determines overall molecular architecture. On the nanometer scale, synthesis conditions determine liquid crystalline conformation, size, and periodicity. On the micrometer

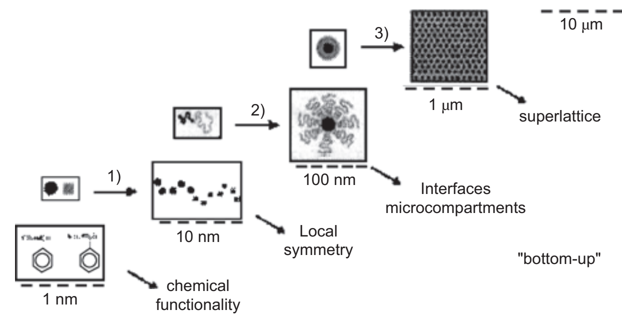


Figure 7. (1) Chemical functional groups and molecules on the atomic scale for (2) self-assembly into larger nanometer sized aggregates based on shape and interactions. These aggregates (3) template micrometer-sized periodic arrays for the formation functional devices. Modified with permission from [3], S. Forster and T. plantenberg, *Angew. Chem. Int. Ed.* 41, 688 (2002). © 2002, Wiley–VCH.

scale, external electrical, magnetic, and mechanical (flow) fields influence the formation of extended superlattices [3].

A crucial step in the formation of hierarchical materials with order on multiple scales involves the liquid–crystalline template synthesis procedure. A commonly used organic/inorganic nanocomposite synthesis process involves the simultaneous co-assembly of surfactants and other building blocks like inorganic silicate species. This co-assembly technique often involves the sol–gel formation of inorganic clusters [18, 21].

3.1. Sol–Gel Process

The sol–gel process utilizes low-cost, commercially available precursors to synthesize inorganic materials and organic/inorganic hybrid materials at low temperature and mild reaction conditions [73]. In general, a sol is a solution containing metal oxide colloidal particles formed by hydrolysis and condensation of the appropriate metal alkoxide or metal halide precursors [74]. A gel contains clusters that span the entire container or clusters that are intertwined so that the viscosity is greatly increased. Commonly used metal alkoxides, $M(OR)_n$, contain metal atoms (M) such as Si, Ti, Al, Zr, and La connected to alkoxide groups (OR) (e.g., CH_3O and CH_3CH_2O). Metal halides MX_n employ the same metal atoms connected to halides (X) such as Cl or Br. The sol–gel process provides a versatile way to synthesize various materials such as optical coatings, low-density aerogels, inorganic membranes, membrane reactors, and colloidal powders with controlled sizes and morphologies.

3.1.1. Sol–Gel Chemistry

Hydrolysis and condensation reactions convert discrete precursors in solution into extended clusters capable of forming engineering materials. Although the sol–gel reaction mechanism depends on which metal precursor is utilized, silica based sol–gel reactions are described as an example because silica materials are most common.

The silica-based sol–gel process involves the following reactions (1) to (3). Reaction (1) involves the hydrolysis of a tetraalkoxysilane with water to generate silanol groups. Alcohol condensation reactions (2) between a tetraalkoxysilane and a silanol group release an alcohol molecule while water condensation reactions (3) between two silanol groups release a water molecule. Reaction (4) shows a hydrolysis reaction of an organosilane that contains a nonhydrolyzable ligand R′. These hydrolyzed organosilane molecules can undergo alcohol or water condensation reactions similar to those shown in reactions (2) and (3). Also, copolymerization of organosilanes $[R'_xSi(OR)_{4-x}]$ with tetraalkoxysilanes $[Si(OR)_4]$ via cross-condensation reactions may lead to a hybrid silica framework containing molecular level mixing of the organic ligands. The reaction kinetics are controlled by the choice of solvent, choice of alkoxide groups, system pH, and the ratio of water to inorganic precursor [33, 73, 75]. Larger alkoxide groups such as butoxide usually result in a slower reaction rate because of steric hindrance.

Reaction Scheme 1. Silica hydrolysis and condensation reactions.

$$\equiv Si(OR) + H_2O \longleftrightarrow \equiv Si(OH) + ROH \text{ Hydrolysis} \quad (1)$$

$$\equiv Si(OR) + (HO)Si\equiv \longleftrightarrow \equiv Si{-}O{-}Si\equiv + ROH \text{ Condensation} \quad (2)$$

$$\equiv Si(OH) + (HO)Si\equiv \longleftrightarrow \equiv Si{-}O{-}Si\equiv + H_2O \text{ Condensation} \quad (3)$$

$$= Si\mathbf{R}'(OR) + H_2O \longleftrightarrow = Si\mathbf{R}'(OH) + ROH \text{ Hydrolysis} \quad (4)$$

Symbols "≡" and "=" represent three and two siloxane bonds that connect to the Si atoms.

3.1.2. Sol–Gel Dip Coating

These sol–gel reactions usually occur in solution and result in silicate colloidal solutions that can prepare various inorganic and hybrid thin films, membranes, powders, particles and fibers through conventional processing techniques such as spin-coating, dip-coating, spraying, aerosol processing, and fiber extrusion.

Sol–gel dip-coating techniques offer a convenient method to form thin films on either planar or nonplanar substrates. The process can be divided into five steps: immersion, start-up, deposition, drainage, and evaporation, as shown in Figure 8. Formation of silica thin films by the dip-coating process has been well developed [76]. Beginning with a silicate sol containing weakly branched silicate species, solvent (e.g., ethanol and water) evaporation during the coating process enriches silicate species within the sol trapped on the substrate surface. This enrichment promotes silicate aggregation, gelation, and formation of thin films on the substrate.

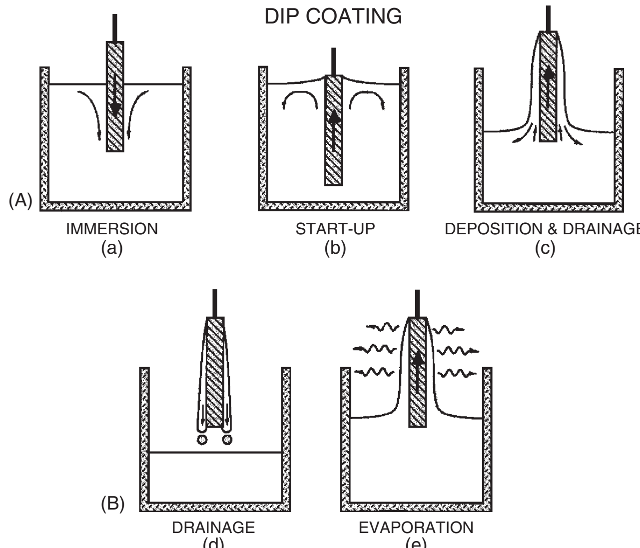


Figure 8. Schematic illustration of a sol–gel dip-coating process. Reprinted with permission from [76], C. J. Brinker and G. W. Scherer, "Sol–Gel Science: The Physics and Chemistry of Sol–Gel Processing," p. 789. Academic Press, San Diego, 1990. © 1990, Academic Press.

A balance between capillary pressure and silica network modulus determines pore sizes of the thin films. Higher capillary pressure and lower network modulus usually result in denser thin films. The thickness of the deposited films can be determined using Eq. (3), where U is the substrate speed, ρ is the density of the sol, η is the viscosity, and γ_{LV} is the liquid vapor interfacial energy [76]:

$$h = c_1(\eta U/\rho g)^{1/2} \tag{5}$$

When the liquid viscosity and coating speed are high enough to lower the curvature of the meniscus, Eq. (4) is used to express the thickness. Usually, a higher coating speed and sol viscosity lead to a thicker film [76]:

$$h = 0.94(\eta U)^{2/3}/\gamma_{LV}^{1/6}(\rho g)^{1/2} \tag{6}$$

3.1.3. Co-Assembly of Surfactant and Silica

As discussed, surfactants self-assemble into micelles and liquid crystalline mesophases in aqueous media at high concentrations. In 1992, Mobil scientists demonstrated that surfactant and silicate can co-assemble and form nanocomposites containing ordered surfactant lyotropic liquid crystalline phases via van der Waals and electrostatic interactions. Surfactant removal results in mesoporous silica (MCM-41) replicas of the liquid crystalline surfactant assembly (see Fig. 9). This co-assembling synthesis approach has been extended to create a large variety of mesoporous and nanocomposites through van der Waals, electrostatic, hydrogen bonding, or other noncovalent interactions. Examples of these mesoporous materials include mesoporous SiO_2 [77–80], SiO_2 containing interstitial Ti [81] or Al [82], aluminosilicates [83], $AlPO_4$ [84–88], Al_2O_3 [89], TiO_2 [90], Nb_2O_5 [91], SnO_2 [92, 93], ZrO_2 [94], mixed oxides [95], and metal Pt [96].

The synthesis procedures usually start with a solution containing surfactant, inorganic precursors, and solvent. There has been debate about whether micelles or liquid crystalline phase formation and inorganic precursor association with the surfactant molecules are sequential or simultaneous processes [48]. One possible self-assembly mechanism involves the initial formation of a liquid crystalline phase. This mechanism assumes that the concentration of the surfactant is sufficiently high to not only form micelles but, in some cases, ordered mesophases. Subsequent association and reaction of inorganic clusters solidifies the mesostructure. This route involves mineralization around already formed liquid crystalline arrays at higher surfactant concentrations and diffusion of inorganic precursors into preformed liquid crystals [97].

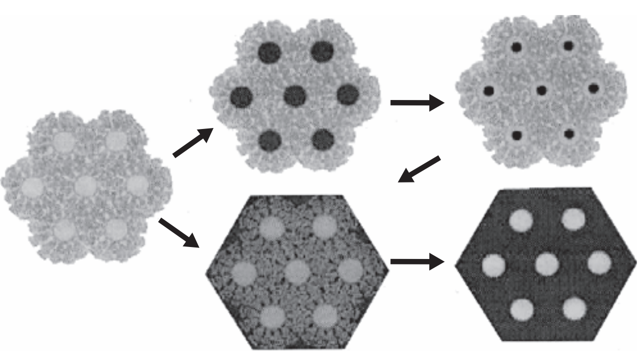


Figure 9. Overview of liquid crystalline templating routes to nanoparticle and mesoporous material formation: (1) Swelling liquid crystal interior with hydrophobic reactants for (2) nanoparticle formation and (3) reaction of hydrophilic reactants followed by (4) surfactant removal to form a mesoporous material. Reprinted with permission from [3], S. Forster and T. plantenberg, *Angew. Chem. Int. Ed.* 41, 688 (2002). © 2002, Wiley–VCH.

Although this synthesis route may hold true in some instances, a growing amount of evidence implies the simultaneous formation mechanism is more likely. During the formation of inorganic clusters in solution, alcohol byproducts increase surfactant CMC and reduce the long-range order of the system [98]. Also, nuclear magnetic resonance (NMR) evidence does not support the sequential mechanism for many commonly used silica systems [49]. Since some synthesis techniques such as evaporation induced self-assembly start to assemble well below the CMC, it is believed that surfactant–inorganic interactions play an important role in the simultaneous formation of a periodic structure [6].

A widely accepted mechanism for many surfactant/sol–gel syntheses of inorganic materials is shown in Figure 10. This theory states that the liquid crystalline phase is created based on cooperative self-assembly of the clusters and surfactant before substantial solid formation has occurred. As the inorganic clusters form, they begin interacting via electrostatic or hydrogen bonding, or other noncovalent interactions with surfactant head groups present in solution. These charged clusters grow and encourage the packing of liquid crystalline structures into arrays such as rods with short-range order and with weak framework walls [52]. Interactions between phases transmit the liquid crystalline structural information to the inorganic clusters [49]. The inorganic/organic arrays then combine to form extended composite structures. In thin films, the cooperative assembly of inorganic associated micellular complexes occurs first at a microscopic gas–liquid or liquid–solid interface [97].

The self-assembly of organic and inorganic components occurs because of favorable interactions [70] and entropic gain that result from inorganic components replacing the water hydration sheath surrounding the surfactant [49]. Synthesis conditions must be optimized to allow for favorable interaction between the liquid crystal surface and inorganic clusters and to allow a lowering of the interfacial energy [99, 100]. This occurs by matching the density of electrostatic charge or hydrogen bonding sites present at the interface between organic and inorganic phases [101, 102]. Metal oxide surfaces typically consist of metal hydroxy groups MOH. Charge matching between this growing surface and the surfactant interface depends on the pH of the system, the valence and content of the metal oxides, the pKa of the surfactant, and the ratio between silica and surfactant. Initially [39, 99], cationic surfactants were used so that the negatively charged inorganic clusters would be charge matched by positively charged surfactant headgroups [22, 48]. It is also possible to use anionic surfactants with positively charged clusters, surfactants, and clusters with

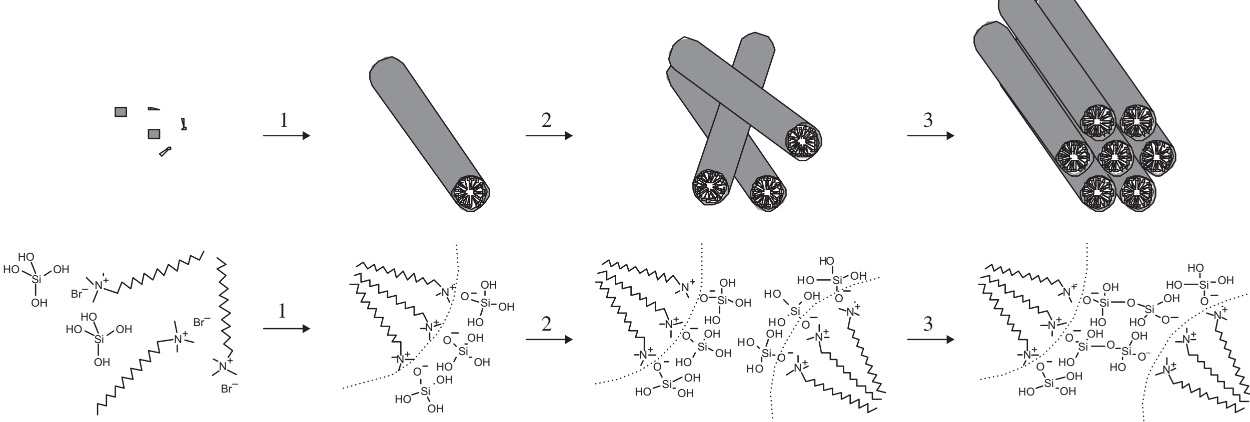


Figure 10. Templated mesostructured inorganic synthesis via (1) co-assembly of surfactant and inorganic precursors into cylindrical micelles that are (2) forced into close proximity based on solvent removal or precipitation allowing interactions and condensation reactions between growing inorganic clusters to finally form a (4) hexagonally arrayed surfactant/inorganic network containing cross-linked inorganic [49].

the same charge brought together by counterions, or nonionic surfactants with neutral inorganic precursors brought together by hydrogen bonds [52]. Changing the pH above or below the isoelectric point of the inorganic clusters can vary the inorganic cluster charge. Nonionic surfactants utilize hydrogen bonding from metal oxide surface hydroxyl groups to the surfactant headgroups as opposed to electrostatic interactions. The forces between phases transmit self-assembled surfactant information to the growing inorganic matrix and control the morphology of the composites [6, 103].

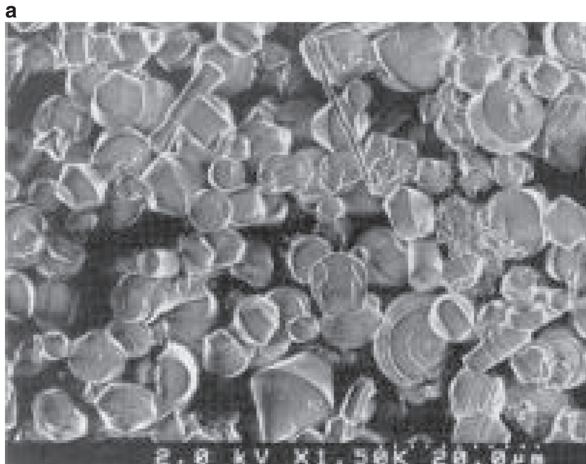
3.1.4. Synthesis Methods

Two major synthesis methods have been developed to synthesize mesoporous materials and nanocomposites: solution precipitation and evaporation induced self-assembly (EISA).

Solution Precipitation The solution precipitation method involves precipitation of co-assembled products in the form of particles, powders, or thin films from solutions that were prepared using inorganic precursor and surfactant [90, 97, 102, 104–107]. The precipitation mechanism [1, 47], which mimics natural biomineralization, involves interactions between the inorganic clusters and surfactant molecules and deposition of the inorganic clusters based on the liquid crystalline scaffolds [33, 47, 108]. Important aspects of this synthesis include concentration of surfactant and inorganic precursors, temperature, and solution pH, which may determine the rate of inorganic cluster formation and the electrostatic charge of the cluster surface [1]. The liquid crystalline array constrains crystal growth by a sterically imposed template mechanism, controls mass transport of inorganic precursors [109, 110], and serves as sites for crystal nucleation based on locally high precursor concentrations at the surfactant interface. Once nucleation has occurred, the growing crystal, which sometimes orients itself based on interfacial forces, increases in size by the addition of adjacent ions [64]. As shown in Figure 11, this method typically results in the formation of small, micrometer sized particles that are difficult to use for membrane applications [33].

Evaporation Induced Self-Assembly EISA combines the simplicity, readiness, flexibility, and affordability of sol–gel processing technology with the efficiency and spontaneity

a



b

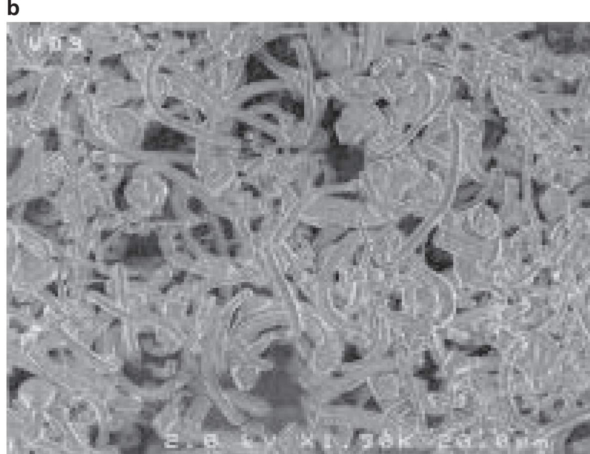


Figure 11. Scanning electron microscope (SEM) images of mesoporous silica obtained by solution precipitation. Reprinted with permission from [111], H. Yang et al., *Nature* 386, 692 (1997). © 1997, Macmillan Magazines Ltd.

of silica/surfactant cooperative assembly. EISA simultaneously and spatially organizes organic and inorganic components into two- or three-dimensional nanostructures during conventional processes like dip-coating, spin-coating, aerosol processing, fiber extruding, and microwriting. Mesostructured thin films, particles, and patterns (or arrays) containing different mesophases can be achieved readily and rapidly by controlling the various parameters such as initial surfactant concentration. Figure 12 shows representative transmission electron microscopy (TEM) images of a thin film prepared by a dip-coating process, lamellar particles by an aerosol process, and micropatterns by microwriting and inkjet printing techniques, respectively.

Dip-Coating The formation of nanostructured thin films through EISA and a sol–gel dip-coating process begins with a homogenous solution of soluble silica and surfactant in ethanol/water solvent. Solvent evaporation during the coating process enriches silica and surfactant and induces their co-assembly into liquid crystalline mesophases. During the drying process, polymerization of silica freezes the mesophase and results in nanostructured thin films exhibiting hexagonal, cubic, or lamellar mesostructures. A similar EISA process involves spin-coating solutions onto substrates to form mesostructured thin films. The advantages and unique features of these processes include:

(1) Mesostructured thin films can be readily fabricated within a time scale of seconds. Research conducted using *in-situ* small angle X-ray scattering and ellipsometry techniques suggests that this coating process involves the formation of many intermediate structures [76]. The rapid formation process kinetically freezes these intermediate structures, resulting in thin films containing metastable structures. Control of sol–gel reaction kinetics, inorganic cluster sizes, pH, and solvent evaporation rate is necessary in order to achieve a highly ordered mesostructured thin film [76]. Mesostructure is determined based on the choice of surfactant and surfactant concentration.

(2) This is a continuous process that allows synthesis of defect-free thin films for membranes, sensors, low-dielectric constant (low-K) materials, and other applications. In fact, low-K mesoporous thin films prepared by EISA have been used in the current semiconductor industry for next generation microelectronic applications.

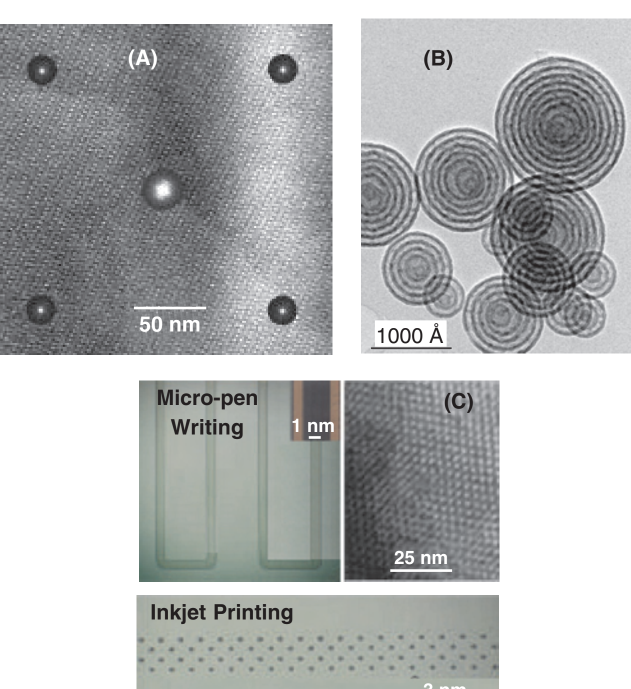


Figure 12. TEM images of nanostructured thin films, particles and patterns showing precise periodic ordering of pores. (a) Plain view of a cubic film coated on a silicon substrate. Reprinted with permission from [33], C. J. Brinker et al., *Adv. Mater.* 11, 582 (1999). ©1999, American Chemical Society. (b) Nanoparticles exhibiting vesicular mesophases. (c) Hierarchical nanostructured patterns and arrays by microwriting or inkjet printing processing. Reprinted with permission from [113], H. Fan et al., *Nature* 405, 57 (2000) © 2000, Macmillan Magazines Ltd.

Aerosol Process Aerosol techniques have been widely used to generate different sizes of droplets for drug delivery, humidification, particle syntheses, and other applications. Figure 13 schematically depicts the formation of nanostructured particles by an aerosol EISA process. Using the surfactant containing sol as precursor solution, an aerosol apparatus generates droplets that pass through the drying and heating zones. The solidified particles are then collected using a filter. During this process, solvent evaporation from the droplet air–liquid interface enriches surfactant and silicate, resulting in their co-assembly and the formation of mesostructures growing from the interface toward the interior. Further drying and solidification processes result in the formation of various mesostructured particles.

Micro-Writing and Printing Microwriting and printing techniques use the surfactant containing sols as ink to write or inkjet print various EISA mesostructured patterns and arrays. The micropen or printing cartridge injects surfactant containing sol droplets on the substrates (e.g., paper, transparency, and glass slides). Co-assembly of silicate and surfactant induced by solvent evaporation and coalescing of these enriched droplets create various hierarchical patterns and arrays on the substrates.

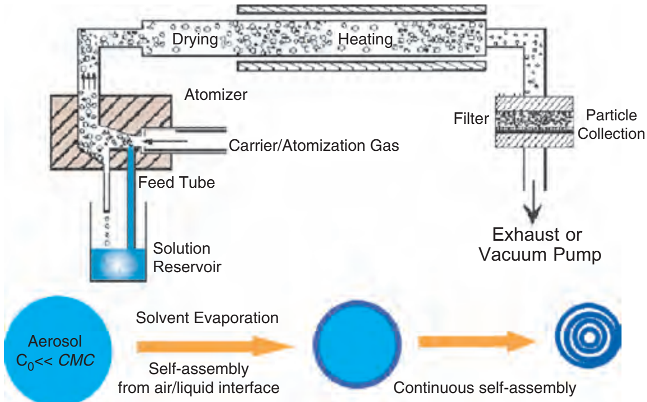


Figure 13. Schematic formation of nanostructured particles by EISA and an aerosol process. Reprinted with permission from [112], Y. Lu et al., *Nature* 398, 224 (1999). © 1999, Macmillan Magazines Ltd.

3.2. Design of Nanocomposites

The EISA approach has been extended to synthesize various nanocomposites. An example phase diagram for a surfactant/hydrophobic component/hydrophilic component system at a given temperature and pressure is shown in Figure 14. This phase diagram demonstrates the utility of the self-assembly methodology for incorporating a variety of components such as monomers or inorganic precursors (silica) into different mesophases. Using the surfactant phase diagram as a conceptual guide, one can consider oil as a wide variety of hydrophobic additives (e.g., monomers, hydrophobic colloidal particles, or functional polymers.) that are incorporated within the hydrophobic micellar cores. In a process similar to that of dishwashing, micelle formations spatially separate and organize inorganic silicates and the additives. Further self-organization of silicate and the micelles into hexagonal, cubic, or lamellar mesophases simultaneously positions the silica, surfactant and additives into precise three-dimensional arrangements. A wide variety of additives (e.g., metal ions, organic functional ligands and molecules, pigments, monomers, oligmers, polymers, metal colloidal particles, catalyst particles, biomolecules, and cells) can be readily introduced to the surfactant/silicate solutions.

As an alternative approach, also shown in Figure 14, direct co-assembly of silicate and functional surfactant micelles can also lead to the formation of nanocomposites with spatially defined functionality. For example, polymerizable groups on a surfactant can react so that the liquid crystal structure is retained after self-assembly has taken place [114, 115]. When used in tandem with the sol–gel process, polymerizable surfactants can easily form nanocomposites with a variety of architectures. Organization of these molecules into arrays is beneficial because of the alignment and condensation of reactive groups on the surfactant tails. Studies have shown that this association increases the rate of polymerization because of the more effective intermolecular collisions [64]. This process can form polymers with controlled mesostructure such as spheres. Sometimes it is difficult to maintain the mesostructure of lyotropic liquid crystals during the polymerization process [64, 114].

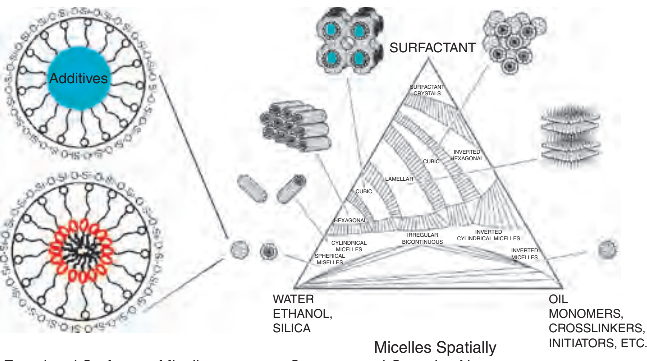


Figure 14. Generic detergent (surfactant–oil–water) phase diagram showing common surfactant mesophases and the ability of surfactants or functional surfactants to spatially separate organic additives and functional groups. Reprinted with permission from [33], C. J. Brinker et al., *Adv. Mater.* 11, 582 (1999). © 1999, Wiley–VCH.

The lamellar phase is the easiest to maintain during polymerization. Other research has been able to synthesize polymers that retain the original hexagonal, cubic [114], and inverted hexagonal mesophase [115].

3.3. Mesoporous Inorganic Nanocomposites

3.3.1. Functional Mesoporous Inorganic

Since the discovery of templated mesoporous silica synthesis, great efforts have been made to functionalize the surface of these materials. Organic functionalities often provide the mesoporous materials with better hydrophilicity and stability to compression [116], better compatibility and more favorable interactions with other components like polymers, improved hydrothermal stability [117], decreased water adsorption [118, 119], less agglomeration between particles [22], and the capability to graft new reactive complexes such as organometallic catalysts to the surface [119–122]. Organic groups such as methyl [116, 123–125] alkyl chains [126], vinyl [14, 127], methacrylate [128], amino [129], thiol [130–132], sulfonic [129], phenyl [133, 134], and metal/ligand complexes [120, 135] are just some examples of moieties that have been functionalized onto the surface of mesoporous silica. In general, either surface grafting or direct synthesis techniques functionalize mesoporous silica.

Surface Grafting Techniques Surface grafting techniques attach organic molecules to the pore surface of preformed mesoporous silica by silylation of surface SiOH groups. Most commonly, the surface modification reaction involves an organosilane molecule R-SiX_3, where X represents an alkoxide or halide group and R is a nonhydrolyzable ligand such as an alkane [130]. Also, molecules can be adsorbed onto the internal surface of the mesoporous material to change the surface properties [136]. This modification coats the internal surface in a similar fashion to self-assembled monolayer formation on flat surfaces [130]. Bourgeat-Lami [22] has compiled an extensive list of references covering the surface modification procedure including: coupling agent chemistry [137] inorganic surface reactivity [138–140], and modified surface properties [141–143] such as stability toward various solvents [144]. Table 2 lists several MCM-41 type materials that have been functionalized by surface grafting.

Direct Synthesis Techniques The direct synthesis approach involves co-condensation of organosilanes that contain nonhydrolyzable functional ligands with a silica precursor like tetraethoxysilane (TEOS) in the presence of the templating surfactant. Hydrolyzed organosilanes containing hydrophobic ligands often serve as co-surfactants by inserting their ligands into the hydrophobic micellar cores through hydrophobic interactions. In this way, organosilane molecules preferentially concentrate at the inorganic/surfactant interface. Surfactant removal leads to the formation of a mesoporous material with organic ligands covalently bonded on the surface. Table 3 lists more examples of organically modified mesoporous silica and their applications.

Another class of mesoporous organic/inorganic materials contains organic moieties integrated within the pore wall.

Table 2. Surface grafted organic functionalization in MCM-41 type materials.

Grafted organic group	Technique	Characteristics	Ref.
Methyl ($—CH_3$)	react silica with trimethylchlorosilane, hexamethyldisilazane, and tetramethyldisilazane	pore surfaces become hydrophobic and more stable; ambient dried silica aerogels can be synthesized	[116, 123–125]
Vinyl ($—CH{=}CH_2$)	react silica with vinyltrichlorosilane or hydrolysis and condensation of vinyltriethoxysilane	deduce intra-channel functionalization by studying kinetics of alkene bromination	[14, 127]
2,4,6-triphenylpyrylium cation	cation formed inside MCM-41 by reaction of chalcone and acetophenone or by ion exchange of 2,4,6-triphenylpyrylium tetrafluoroborate	photoinduced electron transfer reagent for isomerization of *cis*-stilbene to *trans*-stilbene	[145]
Substituted triarylmethylium	zeolite and MCM-41 exposed to benzaldehyde and *N*,*N*-dimethylaniline or anisole to create large interstitial methylium cations	heterogeneous photosensitizers for 1,3-cyclohexadiene dimerization	[146]
Thiol ($—(CH_2)_3SH$)	tris(methoxy)mercaptopropylsilane modified the silica pore surface; Thiol end groups self-assembled into monoloayers	thiol binds heavy metals or other functional molecules; extremely efficient in removing heavy metals such as mercury from both aqueous and organic waste streams	[130, 131]
Manganese (Mn) and 1,4,7-triazacyclononane (TACN) complexes	(3-glycidyloxypropyl)trimethoxysilane TACN ligands, and then Mn dopants used to sequentially modify silica	catalytically active, nonaromatic Mn complex on silica pore surface	[135]
Ferrocenyl ($—(C_5H_4)_2Fe$)	treated MCM-41 with a pentane (1,1′-ferrocenediyl) dimethylsilane	NMR and EXAFS spectroscopy indicate structure is tethered to the silicate wall via Si–O linkages	[120]
Ethylenediamine, diethylenetriamine, and ethylenediaminetriacetic acid salt	grafted to surface using ($RSi[OMe]_3$) type synthetic reagent	readily binds metals such as cobalt	[147]
Amine and cyanide	3-aminopropyl, 2-cyanoethyl organosilanes grafted to silica surface	demonstration of organic modification by adsorption, IR, and TGA	[148, 149]
Amine ($—NH_2$ or $—NR_2$)	surface modification using either 3-aminopropyltriethoxysilane or 3-chloropropyltriethoxysilane; surface chloro groups further modified to produce surface piperidine groups	highly active and regioselective catalyst for the epoxide ring-opening of glycidol by fatty acid addition	[150]
Globular enzymes: cytochrome *c* (bovine heart), papain (papaya latex) and trypsin (bovine pancreas)	immobilization of enzyme onto solid support in aqueous solution	improve chemical sensing and biocatalysts	[151]
Anthracene, naphthalene, and thianthrene	silica modified using polynuclear aromatic compound in iso-octane and tert-butyl alcohol	catalyst	[152]
Thianthrene	chemical vapor deposition of thianthrene under argon; radical cations formed	methodology developed for control of chemical reactivity in constrained environment	[153]
Alkyl halides, alkanes, and water	liquid-phase adsorption at room temperature	study of the variation in intensity of the observed X-ray diffraction peaks as a function of sorbate scattering power	[136]

Source: Modified with permission from [154], G. A. Ozin et al., *Current Opinion Colloid Interface Sci.* 3, 181 (1998). © 1998, Elsevier Science.

Table 3. Direct organic functionalization in MCM-41 type materials.

Organic guest	Technique	Characteristics	Ref.
Fluorinated alkane-, mecapto-, and amino-functionalized organosilanes	organosilane added to surfactant rich silica sols made from TEOS; these sols were then used as inks in three rapid EISA printing processes	patterened, highly ordered, functionalized three-dimensional structures were formed within second	[113]
Thiol($-(CH_2)_3-SH$)	mixture of (3-mercaptopropyl) triethoxysilane (MPTS) and tetramethoxysilane (TMOS) were hydrolyzed and condensed in the presence of CTAB and organic waste streams	thiol binds heavy metals or other functional molecules; extremely efficient in removing heavy metals such as mercury from both aqueous	[132]
Sulfonic surface groups ($-SO_3H$)	MPTS and TEOS hydrolyzed and condensed in the presence of surfactant, SH then converted to sulfonic acid by mild oxidation with H_2O_2 followed by acidification	polyol esterification catalyst	[129]
Straight chain alkanes ($-CH_n$)	organic–inorganic thin films were prepared by co-condensation of alkyltriethoxysilanes (alkyl chain = 8–18) and TEOS using an EISA dip-coat process	highly ordered, multilayered structure depended on the length of the alkyl chain and the solution temperature	[126]
Phenyl and *n*-octyl	co-condensation using organosilane	catalyst	[133, 134]
Methacrylate ($CH_2{=}C(CH_3)CO_2-$)	co-condensation of TMOS and 3-(trimethoxysilyl)propyl methacrylate in the presence of CTAB; additional monomer incorporation and reaction forms silica/poly (methyl methacrylate) nanocomposite	grafted methacrylate groups accessible to bromination and polymerization	[128]

Source: Modified with permission from [154], G. A. Ozin et al., *Current Opinion Colloid Interface Sci.* 3, 181 (1998). © 1998, Elsevier Science.

Bridged silsesquioxanes (Fig. 15) and other silica precursors hydrolyze, co-condense, and co-assemble with surfactant to form functionalized mesostructured materials. Each of these silsesquioxane precursors contains a nonhydrolysizable organic ligand that covalently bridges two alkoxysilane groups. Co-assembly organizes surfactants and these organic moieties into spatially defined arrays. Surfactant removal leads to the formation of a mesoporous material with organic moieties integrated within the pore walls. Bridging organic groups such as aromatic rings [13], alkyne [155], alkene [156], and alkane chains [157] have been incorporated within a silicate matrix. After removal of the surfactant, silicates with highly ordered pore structure, crystal-like wall structures, and/or high internal surface area have been synthesized by this method [158]. For example, mesoporous silica containing ethane or ethylene incorporated within the silica network demonstrated a high >1200 m^2/g surface area [156]. Not only could these materials absorb organic compounds unlike unmodified MCM-41, but incorporated ethylene groups could also be brominated with little or no damage to the pore structures. Ethane molecules incorporated uniformly within a silica matrix have also been used to produce a variety of highly ordered mesophases that can be used as catalysts, adsorbents, and hosts for other nanocomposites [157]. mesoporous silsesquioxanes have many applications including the production of low dielectric constant materials, catalysts, sensors, adsorbents, and as host for other nanocomposites. Table 4 describes many of the procedures used to create ordered structures from these silsesquioxanes.

These modification techniques can be combined with solvent EISA techniques (see Section 3.1.4) to produce organic functional thin films, particles, and patterned arrays using conventional dip-coating or spin-coating, aerosol processing, and microwriting or inkjet printing techniques, respectively [13].

Applications of Organic Modified Mesoporous Silica in the Microelectronics Industry Organic modified mesoporous silica has a broad range of applications ranging from catalysts, adsorbent, separation, sensors, to molecular host materials. However, because of the emerging importance of these materials to the microelectronics industry, this section will focus on low dielectric constant materials based on organic modified mesoporous silica.

Continuous miniaturization of devices and circuits has driven rapid growth in the semiconductor industry, resulting in better device performance and lower cost per bit. The decrease of feature sizes and the increase of metal level result in an increase of interconnect resistance (R) and the associated total capacitance (C). As the device dimensions shrink toward 0.1 μm, the interconnect delay, defined as the RC delay, becomes a dominant factor in determining device performance [160, 161]. Reducing the RC delay by lowering the ratio of total resistance or capacitance to device delay has become a challenge for both materials development and interconnect architecture design. Although optimization

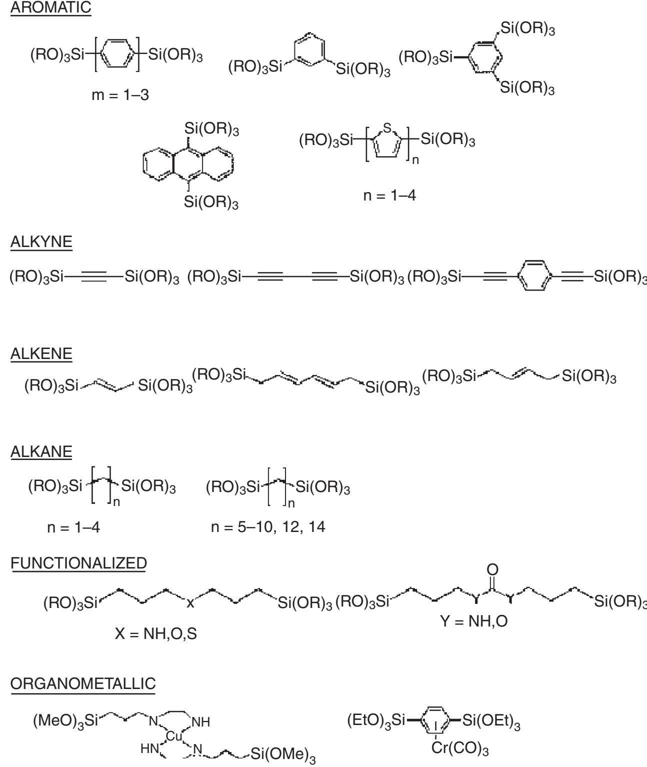


Figure 15. Representative bridged polysilsesquioxanes used in direct internal pore wall functionalization. Reprinted with permission from [159], K. J. Shea and D. A. Loy, *Chem. Mater*. 13, 3306 (2001). © 2001, American Chemical Society.

of the interconnect design can lower a large fraction of this delay, viable solutions have not yet been found. Thus, efforts have been concentrated on replacing the current interlayer dielectrics (ILDs) ($K \sim 4$) with lower-K ILDs, and on changing interconnects composed of Al/W to Cu-based interconnects [162].

The dielectric constant can be described using a basic capacitor made of two conducting parallel plates separated by an insulator, also known as the dielectric (refer to Fig. 16) [163].

The electrical charge (Q) stored on either plate can be found from the equation

$$Q = C * V \tag{7}$$

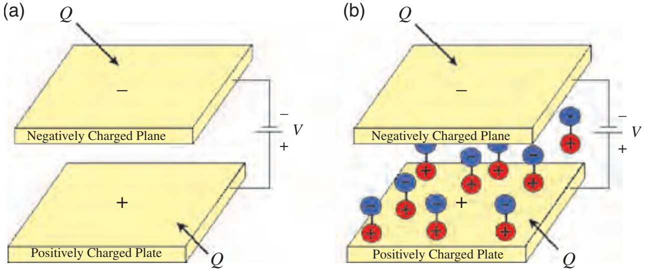


Figure 16. Schematic of a basic capacitor: (a) the dielectric between the two parallel plates is air or a vacuum, and (b) the dielectric is a material that is capable of being polarized.

where V is the applied voltage between the two plates and C is a proportionality constant defined as capacitance. The capacitance can be determined from the expression

$$C = \varepsilon(A/d) \tag{8}$$

where ε is the permittivity of the dielectric material, A is the area of one plate, and d is the distance between the two plates. The permittivity of a material describes the ability of a material to become polarized by an electric field. In Figure 16, there are two capacitors, one with air as the dielectric and the other with a material capable of being polarized as the dielectric. The polarized material has a permittivity that can be found from the relationship

$$\varepsilon = \varepsilon_0 + (P/\xi) \tag{9}$$

where ε_0 is the permittivity of a vacuum (8.85×10^{-12} F/m), P is the polarization (vector sum of all dipoles per unit volume), and ξ is the electric field strength. The relative permittivity ($\varepsilon/\varepsilon_0$) is defined as the dielectric constant (κ). Interlayer materials with lower dielectric constants allow smaller device dimensions due to the less interconnect delay and faster on/off component switching.

Low-K materials are typically designed by the addition of fluorine to the material or by the introduction of pores to lower the density. To date, various new low-k dielectrics have been extensively studied, including fluorine-doped SiO_2 ($K \sim 3.4$), hydrogen silsesquiosxane (e.g., XLK^{TM}) ($K \sim 3.0$–3.2), spin-on organic polymer (e.g., Silk^{TM}), spin-on inorganic dielectrics (e.g., Nanoglass^{TM}) ($K \sim 2.5$–3.0), chemical vapor deposition carbon-doped SiO_2 (e.g., Black Diamond^{TM}) (K$\sim$ 2.5–3.0), fluropolymer, and surfactant templated SiO_2 ($K \sim 1.8$–2.3). Among these materials, spin-on SiO_2-based dielectrics are among the most promising candidates because they are easily manufactured and integrated into circuits. However, traditional SiO_2-based spin-on low-K films (e.g., Nanoglass^{TM}) usually contain a highly disordered, often fractal, gel network with large pores (over 10 nm) and a broad pore size distribution. These network characteristics dramatically deteriorate the dielectric and mechanical performance. In contrast, surfactant templated SiO_2 dielectrics are much better candidates for low-K applications because of unimodal 2–10 nm pore size distribution, controllable porosity, a complete connectivity of the silica framework, absence of drying stress, and controllable pore surface chemistry.

Mesoporous methylsilsesquioxane (MSSQ) (Fig. 17) and methyl modified mesoporous silica are currently being examined by the semiconductor industry as low-K interlayers. These materials are prepared using TEOS and the appropriate methyl precursors through the surfactant templating approach described earlier. The incorporation of methyl groups provides thin film hydrophobicity, prevents water adsorption, and thus decreases the dielectric constant (water has a dielectric constant of 80). Controlled porosity and pore surface chemistry of MSSQ systems provide low dielectric constants between 2.1 and 2.8 [164, 165]. Table 4 describes many of the procedures used to create ordered structures from these silsesquioxanes.

Figure 17. Ideal structure of methlysilsesquioxane [165].

3.3.2. Incorporation of Organics during Co-assembly Process

Surfactants are such useful tools for the synthesis of nanocomposites because they partition hydrophobic molecules inside micelles or liquid crystal assemblies, resulting in spatially ordered arrangements of these components [167]. Organic components can be added to the initial solution during the sol–gel co-assembly process so that they are incorporated within the newly formed inorganic structure. This incorporation technique can form mesoporous materials with larger pore sizes or nanocomposites based on the addition of swelling agents or monomers.

Swelling Agents The original work used liquid crystal templates swollen by 1,3,5-trimethylbenzene to increase the mesoporous silica pore size from 2 to 10 nm [39, 47, 168]. Organic compounds such as alkanes [169] or alkylamines [170, 171] partition into the liquid crystal template during sol–gel co-assembly to enlarge the final mesoporous material pore size. In fact, it is possible to synthesize an inorganic matrix with pores on the order of 10 to 25 nm instead of the typical 3 nm pore size [172, 173]. Even preformed silica can have its pore size expanded with swelling agents such as amines [170]. The pore structures in these expanded materials generally become less ordered as the pore size is increased [170]. These systems are useful for the production of high surface area catalysts with large pore sizes as well as drug delivery devices for the controlled release of incorporated organic components. Current research is creating materials with larger pores and ordered structures so that catalysts can be synthesized to react with larger organic chemicals [169, 172, 174].

Functional Component Incorporation Because functional components like dyes have different properties based on the specific surfactant microenvironment, these materials can be used for optical applications or to probe and gain fundamental understanding of the self-assembly process, solute partitioning [175, 176], solute reactivity [177], surfactant packing [178], and internal inorganic structure [179, 180]. Solute molecules can either associate near the inorganic interface, within the mostly hydrophilic chain sections, or be solubilized inside the hydrophobic cores [175]. Small amounts of chlorophyll demonstrate a similar ultraviolet (UV) adsorption pattern to that of chlorophyll in solution, indicating solubilization within the cores for this particular solute [176]. Reactive molecules incorporated within mesoporous materials demonstrate that the reactivity

Table 4. Mesoporous materials synthesized with bridged silisesquioxanes.

Bridging group	Technique	Characteristics	Ref.
Methyl	co-assemble MSSQ with star-shaped hydroxy-terminated poly(ϵ-caprolactone), star PMMA, and triblock copolymer poly(ethylene oxide-b-propylene oxide-b-ethylene oxide) respectively; surfactant removed by heating to 400 °C	dielectric constants of 2.1, 2.2, and 1.5 respectively	[164–166]
Benzene, ferrocene, acetylene, thiophene, and bithiophene	bridged silisesquioxane precursors synthesized and hydrolyzed in the presence of surfactant to produce periodic mesoporous materials	benzene and thiophene materials had highest degree order	[155]
Alkyl, vinyl, and benzene	evaporation-induced self-assembly produced thin films and particles	highly ordered materials had increased mechanical properties and decreased dielectric constant	[13]
Ethane and ethylene	bridged silisesquioxane precursors synthesized and hydrolyzed in the presence of CTAB to produce periodic mesoporous materials with uniform organic bridge incorporation	produced highly ordered mesophases with a wormlike morphology and surface areas as high as 1200 m^2/g; no damage to pore structure during bromination or organic compound adsorption	[156, 157]
Benzene	precipitation in basic aqueous solutions containing alkyltrimethylammonium surfactant	material had hexagonal mesopores with crystal-like pore walls that were composed of alternating benzene and silica layers	[158]

and transport of organic guest into the surrounding media depends on the surfactant and salt concentrations [177]. As shown in Figure 18, amphiphilic fluorescent metal complexes inside an inorganic framework provide information on the surfactant microenvironment [178]. Specific binding of the dye to inorganic structural sites with different coordination numbers allows inorganic material information to be obtained [179].

Monomer and Polymer Incorporation Simultaneous polymer/inorganic nanocomposite formation utilizes the same principles mentioned for mesoporous materials with expanded pore sizes. Hydrophobic components like monomers and certain catalytic complexes can be incorporated into the hydrophobic cores of micelles or liquid crystalline structures, resulting in spatially ordered arrangements of hydrophobic and hydrophilic sol–gel components. Surfactant and inorganic components co-assemble and form organic/inorganic nanocomposites containing ordered surfactant lyotropic liquid crystalline phases (see Fig. 14). Subsequent polymerization of the incorporated monomer such as methacrylate [21, 181] with incorporated catalyst or by thermal or photoinitiation results in an interpenetrating polymer/inorganic nanocomposite that still contains the structure directing surfactant. For example, Sellinger et al. spatially positioned hydrophobic monomers within micellar cores during EISA and subsequently polymerized these monomers to synthesize lamellar silica/polymer nanocomposite thin films that mimic the structure of nacre (see Fig. 19) [21]. These synthetic nacre films exhibit increased indentation hardness and resistance to wear [21, 182]. Thus, the monomer incorporation method utilizes surfactant-assisted co-assembly of inorganic clusters, surfactants, and organic monomers to spatially separate and organize these components into hexagonal, cubic, lamellar, or other precise arrangements.

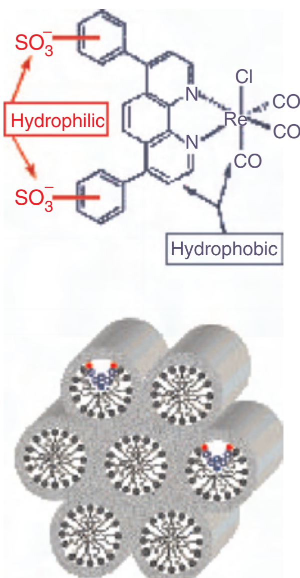


Figure 18. Structure and location of a fluorescent molecule used to probe internal environment of a surfactant/inorganic nanocomposite. Reprinted with permission from [178], J. Wu et al., *Nano. Lett.* 1, 27 (2001). © 2001, American Chemical Society.

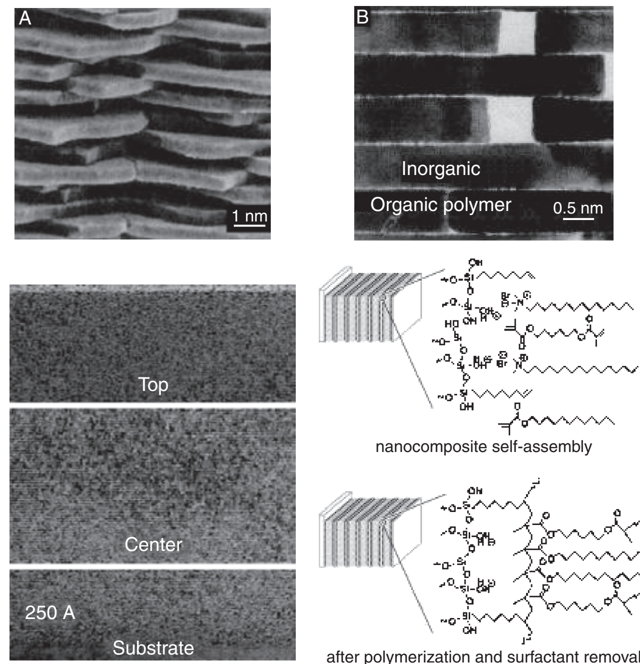


Figure 19. SEM of the Nacre of Abolnia (TOP) and TEM images of the lamellar polymer/silica nanocomposites prepared by EISA and sol–gel dip-coating. Top two images reprinted with permission from [183], I. A. Aksay et al., *Science* 273, 892 (1996). © 1996, Science Permissions Department. Bottom images reprinted with permission from [33]. C. J. Brinker et al., *Adv. Mater.* 11, 583 (1999). © 1999, Wiley–VCH.

Co-assembly using liquid crystal arrays swollen with preformed polymers such as styrene [184] or polyaniline [185] allows improved material properties. For example, conductive polyaniline incorporated into mesoporous metal oxides has increased conductivity at higher temperatures because of increased thermal stability and favorable interactions with the structural matrix [185].

3.3.3. Surfactant/Inorganic Nanocomposites

After sol–gel co-assembly synthesis of mesostructured materials such as MCM-41, the surfactant is often removed to allow for incorporation of more robust materials. However, these inorganic/surfactant composites have several beneficial properties and uses in their own right such as the study of the self-assembly process and of the internal environment after material formation, improvement in material properties such as stability [186], novel catalysts with good selectivity, isolation of magnetic components, improved separation membranes, and improved optical characteristics for optical applications. The synthesis of inorganic/surfactant nanocomposites involves the same synthetic methods used for mesoporous inorganic materials except that there is no surfactant removal step [187–189].

Normal Surfactants Although surfactants are typically not mechanically robust, their incorporation within a mesoporous material allows synergistic improvement of properties. In terms of structure, a silica framework creates a protected environment for the surfactant molecules allowing them to become fully rigid at high pressures [178].

The surfactant can actually improve silica properties allowing structural elasticity, reduction of cracks [190], reversible deformations, and bulk moduli similar to solid silica [178]. Also, further improvement in material properties should be possible when stronger interactions between surfactant and the inorganic surface are employed. However, these materials are still not stable enough for structural applications.

Other interesting inorganic/surfactant materials with interesting magnetic, catalytic, separation, and optical properties have been synthesized. Magnetic materials such as iron oxide separated by layers of surfactant demonstrate novel properties such as superantiferromagnetic domain structures [191]. Silica/surfactant nanocomposites also have high catalytic activity due to active surface sites such as $=SiO^-$ on the silica [192] or electrostatically bonded surfactant [176]. Surfactants can coat specially designed inorganic solids via adsorption allowing effective interactions with components from different solvents. In this way, a membrane composed of a hydrophobic conductive backbone surrounded by surfactant can filter water-soluble, electroactive compounds [193]. Also, the surfactant can improve optical properties based on controlled nanophase separations [175, 194]. In fact, optically transparent films are possible by leaving the surfactant inside the composite [194].

Functional Surfactant Functional surfactants are molecules that combine the amphiphilic properties of surfactants with specially designed functional groups such as polymerizable groups or optically active centers. Since these surfactants are synthesized prior to the self-assembly process, there is a high degree of synthetic freedom concerning both the choice of functional center and its location on the amphiphile. Polymerizable surfactants lock in the mesostructure after the co-assembly process by cross-linking between amphiphiles. These surfactants contain reactive groups such as diacetylene [18], methacrylate [195], or ethylene [196]. Advantages of polymerizable surfactants in the synthesis of nanocomposites are that:

1. By changing the amphiphile structure, the extent of polymerization reaction can be varied and the mesostructure can be altered between cubic, hexagonal, and lamellar [18].
2. A simple one-pot synthesis allows the inorganic and organic phases to cross-link and polymerize simultaneously or sequentially.
3. Mesostructure is often maintained during amphiphile polymerization [115, 197].
4. The synthesis can proceed without additional monomers that may change the packing efficiency and structure of the liquid crystal.
5. There is no need for postsynthesis surfactant removal that may damage or alter the structure and interactions between phases.
6. The combination of inorganic structure and cross-linking between amphiphiles strengthens and improves the overall material properties such as material strength and optical transparency [18, 21].
7. Functional and conjugate polymers such as poly(diacetylene) are aligned and protected within the inorganic framework allowing for the possibility of controlled electrical conduction and chromatic reversibility [18].

Optically active centers can be placed in different locations on a synthesized co-polymer surfactant to control the fluorescent properties based on its specific chemical environment [198]. As Figure 20 shows, an amphiphile containing a photoelectronic endgroup produces an interesting inorganic/surfactant composite in an economically more efficient way than pore wall modification or surfactant adhesion [176]. Applications for this type of material include nonlinear optical switching materials and photoconversion materials [176]. A limitation of this technique is that the active center must have specific interactions with itself and with the solvent system in order to self-assemble.

Another interesting example is the formation of mesostructured chromatic organic/inorganic nanocomposites using functional nonionic surfactants that contain diacetylene moieties within the hydrophobic portion of the molecules and hydrophilic oligmeric ethylene oxide headgroups [18]. As shown in Figure 21, co-assembly of silica and the surfactants leads to the formation of nanocomposites with lamellar, cubic, or hexagonal mesostructures. Topochemical polymerization of the diacetylene groups by UV irradiation results in conjugated nanocomposites that show responsive color changes to the external stimuli (e.g., temperature, chemical, or mechanical stimuli). Figure 21 shows the chromatic transformation of the conjugated polydiacetylene/silica nanocomposite thin films due to the temperature stimuli. While as coated nanocomposite thin films are transparent and colorless, UV irradiation through a mask polymerizes the diacetylene moieties within the irradiated area and creates a blue thin film containing a colorless pattern. The

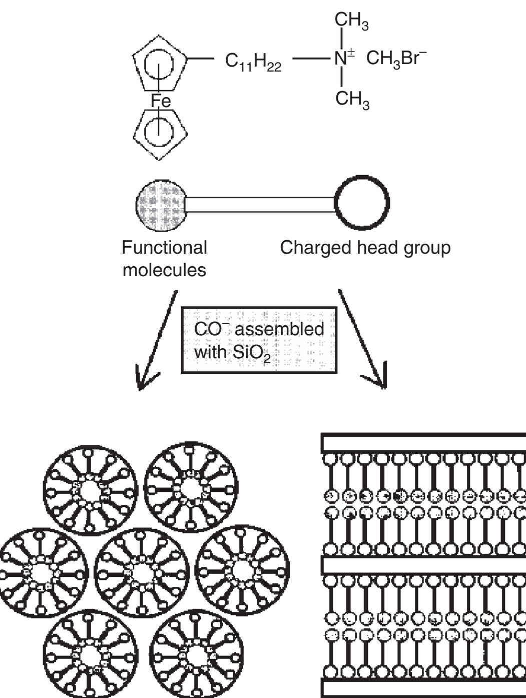


Figure 20. Lamellar and hexagonal mesostructured materials formed by self-assembly of a functional surfactant containing electrostatic and functional molecular portions into. Reprinted with permission from [176], I. Honma and H. S. Zhou, *Adv. Mater.* 10, 1532 (1998). © 1998, Wiley-VCH.

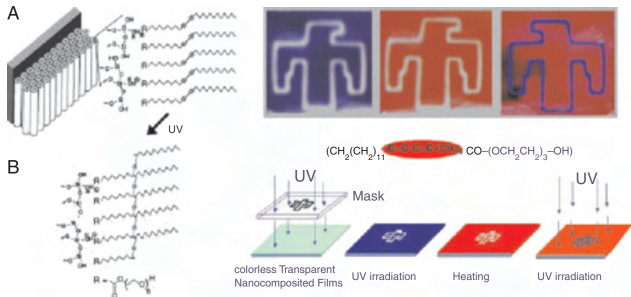


Figure 21. Silica/diacetylene nanocomposite is formed by (A) surfactant sol–gel co-assembly. Polymerization of surfactant (B)(C) through a UV mask created (D) a patterned blue film. (E) Heating the film changed blue portions to red allowing (F) a novel blue/red patterned film after further irradiation. Reprinted with permission from [18]. Y. Lu et al., *Nature*, 410, 913 (2001). © 2001, Nature.

blue color is due to adsorption of the red portion of visible light by the conjugated diacetylene. When subject to a high temperature (e.g., 60 °C) for seconds, the blue film transforms its color to red due to the configuration change of the polymer and other factors. Such chromatic transformation may be reversible. Further polymerization leads to the red film with a blue color pattern. These chromatic properties are useful for sensing applications [18].

4. OTHER ASSEMBLY TECHNIQUES

Other nanocomposite synthesis methods include infiltration [199, 200], surface modification, layer-by-layer assembly [201], intercalation, and nanoparticle array approaches. The infiltration approach incorporates and polymerizes monomers within a preformed inorganic mesoporous template that has been prepared by a surfactant templating technique or by self-assembly of colloidal silica particles [200, 202, 203]. The layer-by-layer method utilizes electrostatic interactions of positive and negative components to fabricate layer structures one layer at a time. Intercalation involves the filling of voids between swollen layers of material. Finally, nanoparticles can be assembled into periodic arrays with a variety of methods such as precipitation or particle confinement.

4.1. Infiltration Methods

Infiltration involves introduction and subsequent polymerization of organic monomers within the pore channels of a preformed inorganic matrix such as mesoporous silica or microporous zeolites. Typically, monomers are introduced via liquid infiltration, vapor adsorption and condensation, or sublimation methods. After reaction, polymers are confined within the rigid inorganic environment. An alternate method involves the infiltration of preformed polymers into the mesopores [204].

Polymerization mechanisms such as free radical polymerization, addition reaction, metal catalytic reaction, and ring-opening polymerization react incorporated monomers and synthesize a large variety of nanocomposites. Reaction within the mesoporous material dramatically affects kinetics and the final polymer structures (e.g., molecular weight and molecular weight distribution). However, free radical polymerization may be initiated with UV irradiation or conventional initiators [205–210]. Because of constrained molecular motions and isolation of the active species in the confined pore channel environment, the propagation reaction may be accelerated at elevated temperatures while the termination reaction may slow down, resulting in higher molecular weight polymers. For example, Ng and co-workers utilized radical polymerization of methyl methacrylate within mesoporous silica using benzoyl peroxide as initiator to obtain poly(methyl methacrylate) (PMMA) with a much higher molecular weight than formed under identical solution conditions [210]. However, polyaniline and polyacrylonitrile formed within the nanocomposites through an oxidation and free-radical polymerization mechanism respectively demonstrate considerably lower molecular weights than those of bulk material. The limited growth may be due to the constrained diffusion of monomers into the pore channels [205–208].

4.1.1. Infiltration of Mesoporous Silica

Silica is the most commonly used host material because of its ease of synthesis and the wide variety of possible pore morphologies, sizes, and surface functionalization. Early work, which infiltrated aniline into mesoporous silica via vapor phase condensation followed by polymerization, formed silica/polyaniline nanocomposites [205]. Other silica based nanocomposites contain polymers such as poly(2-methoxy-5-(2'-ethyl-hexyloxy)-1,4-phenylene vinylene) (MEH-PPV), acrylonitrile, polystyrene, PMMA, poly(phenol-formaldehyde), polyethylene, and ferrocenlysilanes. Poly(phenol-formaldehyde)/silica nanocomposites have been prepared by addition reaction of phenol and formaldehyde within the mesostructured pore channels (see Fig. 22) [211]. Silica/polyethylene nanocomposites were prepared by a metal-catalyzed atom transfer radical polymerization (see Fig. 23) [212]. Ring-opening polymerization of silicaferrocenophane inside MCM-41 was reported to form poly(ferrocenylsilane)/silica nanocomposites (see Fig. 24) [120, 213]. Table 5 summarizes polymer/mesoporous silica nanocomposites prepared by the infiltration methods. The unique features and application of these nanocomposites are also included in Table 5.

4.1.2. Infiltration of Zeolites

Zeolites have also been used as hosts for inorganic/polymer nanocomposite synthesis. Historically, the most extensively studied system involves radical polymerization of polar vinyl monomers within zeolites [214]. Other conducting polymer systems include poly(methylacetylene) [215], polypyrrole [216] and polyaniline [217]. A problem with zeolite infiltration is that the host pore sizes are often too small for effective and continuous polymerization.

4.2. Layer-by-Layer Sequential Assembly

Layer-by-layer sequential assembly utilizes electrostatic interactions between anionic and cationic components to fabricate heterostructure ultrathin films. This is conducted

Figure 22. TEM image of poly(phenol-formaldehyde) mesofibers synthesized by polyaddition/condensation of phenol and formaldehyde within silica mesochannels. Reprinted with permission from [211]. S. A. Johnson et al., *J. Mater. Chem.* 8, 13 (1998). © 1998, Royal Society of Chemistry.

by sequentially dipping a substrate into a polyanion and polycation solution (see Fig. 25) [218]. The initial step involves cleaning and charging the underlying substrate in preparation for the initial layer. This substrate is then

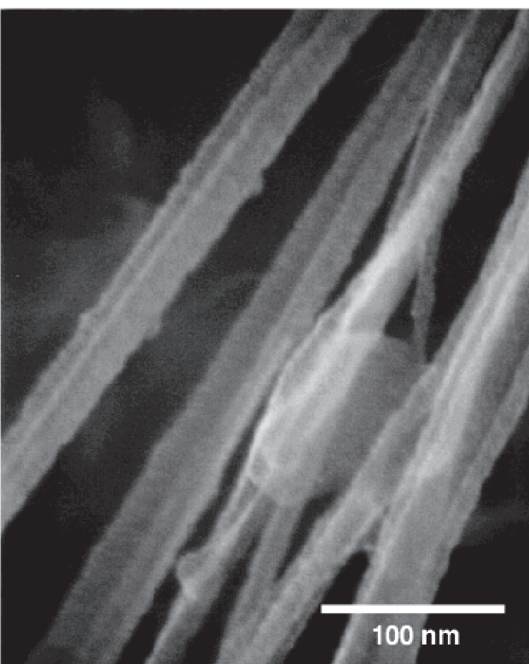


Figure 23. SEM image of crystalline polyethylene fibers formed by extrusion polymerization from supported titanocene mesoporous silica. Reprinted with permission from [212]. K. Kageyama et al., *Science* 285, 2113 (1999). © 1999, American Association for the Advancement of Science.

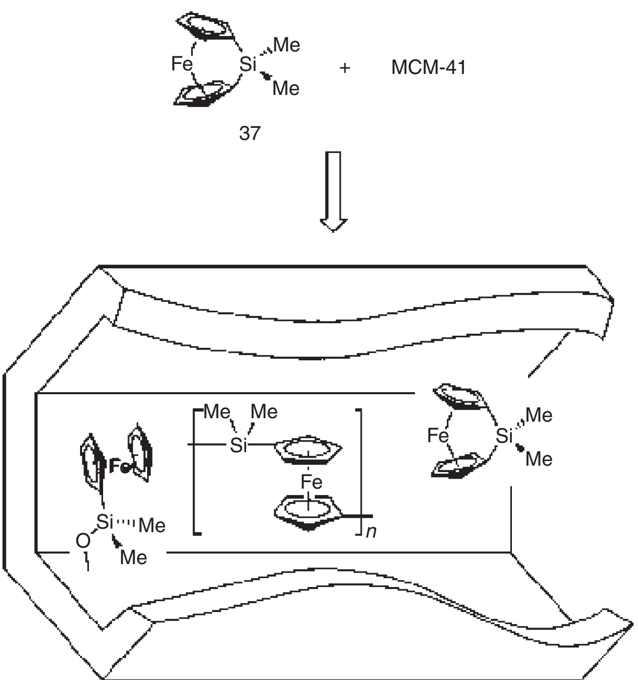


Figure 24. Illustration of ring-opening polymerization of silaferrocenophane within MCM-41 channels. Reprinted with permission form [213], M. J. MacLachlan et al., *Adv. Mater.* 10, 144 (1998). © 1998, Wiley–VCH.

dipped into a solution containing soluble but oppositely charged components such as polyelectrolytes or charged quantum dots [22] that attach themselves to the substrate based on Coulombic attraction [219]. The newly formed layer is then exposed to an oppositely charged electrolyte that adsorbs based on favorable interactions. The relative rates of component adsorption and desorption must be controlled to obtain a homogeneous film with the correct thickness [220]. Film thickness depends mostly on the number of deposited layers (each layer is approximately 1.5 nm thick) [221]. The synthetic freedom inherent in this method allows a wide variety of surfaces to be used as the original substrate, and an extensive list of possible materials for incorporation includes metal or ceramic nanoparticles, polyelectrolytes, conjugated polymers, supramolecules, dyes, fullerenes, biological molecules such as proteins, and electrostatically active components such as chromophores [201, 218, 222–226].

Precise control of material properties is possible through variation of the sequence and composition of the deposited layers. For instance, anionic polymer and ceramic nanoparticles can be layered with polycations in a variety of patterns such as ABAB, ABAC, and other patterns. The possibility of incorporating different layers allows precise tuning of the index of refraction at different depths. Also, electrostatic interactions from the underlying molecules may align dipoles or polymers and result in polarized components and material anisotropy without the use of electric fields [218]. Such unique multilayer films are used for microelectronics, X-ray optics, chemical and biological sensors, and light emitting devices [201, 221–223].

Table 5. Polymer/silicate nanocomposites formed by infiltration.

Guest	Host	Technique	Characteristics	Ref.
Polyaniline	aluminosilicate MCM-41	aniline vapor adsorption in pores followed by oxidative peroxydisulfate initiation	confined, conductive polymer filaments confined in MCM-41	[205–207]
Acrylonitrile, graphite	MCM-41	acrylonitrile vapor adsorption in pores; polymerization initiated by aqueous addition of $K_2S_2O_8$ and $NaHSO_3$ radical initiators; polymer pyrolysis yields graphite filaments	confined graphite has microwave conductivity 10 times higher than bulk pyrolyzed polyacrylonitrile	[208]
Polyethylene	titanocene-mounted mesoporous silica	extrusion polymerization of ethylene on titanocene/methyl-alumoxane active centers	crystalline linear polyethylene nanofibers obtained	[212]
Polymethyl-methacrylate, polystyrene, or polyvinylacetate	MCM-41	monomer vapor adsorption in pores followed by benzoyl peroxide and UV light initiation	Pore size depends on the glass transition temperature of polyvinylacetate in MCM-41	[209, 210]
MEH-PPV	surface modified, hexagonal mesoporous silica	hydrophobic silica pores infiltrated with MEH-PPV polymer by thermal cycling	control of energy transfer in semiconducting polymers	[204]
Poly(phenol-formaldehyde)	MCM-41	phenol and formaldehyde vapor adsorption in pores followed by acid-catalyzed polymerization; removal of mesoporous host yielded polymer fibers	20–30 Å polymer fibers possible	[211]
Poly(ferrocenlysilane), Fe clusters	MCM-41	sublimation of ferrocenophane $Fe(C_5H_4)_2SiMe_2$ into pores followed by pyrolysis	forms magnetic composite	[120, 213]

Source: Reprinted with permission from [154], G. A. Ozin, *Current Opinion Colloid Interface Sci.* 3, 181 (1998). © 1998, Elsevier Science.

The greatest disadvantage in the electrostatic self-assembly method is that a great number of time-consuming steps are required to form a material with any considerable thickness. However, this method has an advantage over Langmuir–Blodgett film formation in that the layer-by-layer technique forms films that are held together with electrostatic attraction as opposed to van der Waals forces [221].

4.3. Intercalation

A slight variation of electostatic self-assembly involves intercalation. The galleries (voids) between lamella (sheets) of material such as clay, graphite, transition metal chalcogenide, or mica can easily be intercalated (filled) with a wide variety of organic guest molecules based on interactions between sheet and filler, as shown in Figure 26 [35, 109]. Because each sheet of host material is weakly bound to neighboring sheets, the sheets are swollen by the addition of reactive chemicals, change in pH, or introduction of polar chemicals. Ionic clays are often used because of the ease and flexibility in forming ion pairs [227]. Before modification or swelling, most layered host materials are negatively charged and they are balanced with removable cations like Na^+ or Ca^{2+} [228]. For instance, potassium, which balances the overall negative charge in mica, can be replaced with other cations or polycations in an ion exchange reaction [35]. This involves exposing a cleaned suspension of clay or graphite to a suitable chemical that swells and replaces the ions [229]. Often, large cations such as tetraalkylammonium ions form pillars between the sheets, thereby creating high surface area materials for catalytic applications [230]. Also, monomers often adsorb onto the surface and react to

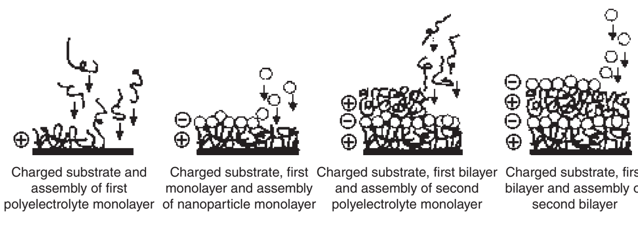


Figure 25. Layer by layer self-assembly of cationic poylelectrolyte and anionic nanoparticles. Reprinted with permission from [218], K. C. Cooper et al., *Proc SPIE*, 4512, 93 (2001). © 2001, SPIE—The International Society for Optical Engineering.

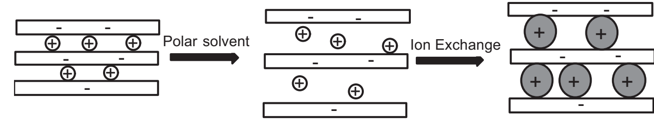


Figure 26. Intercalation of negatively charged lamellar structure by polar swelling and ion exchange of interstitial cations with new cationic species such as polymers.

form a nanocomposite [22]. For example, the incorporation of conducting polymers produces materials with interesting anisotropic conductivity [231, 232].

The advantage of intercalation is that atoms, ions, or particles are reversibly included and sometimes aligned within a structurally preserved microreactor at room temperature [228, 233]. Once again, interactions between phases such as electrostatic and hydrogen bonding self-assemble the material together [227]. Large surface area allows a high degree of interaction between phases thereby strengthening the material to avoid disintegration [228].

4.4. Nanoparticle Self-Assembly

Nanoparticles or colloids self-assemble into arrays with different properties based on interparticle interactions and particle shape, size, and size distribution [10, 109]. A colloid refers any particle with a size from 10 Å to 1 μm and a colloidal solution refers to a stable suspension of these particles [22, 53]. When packed together to form an array, spherical nanoparticles can fill a 3D space in a myriad of different conformations. However, if there is a driving force (such as pressure) for the spheres to occupy as much of the space as possible and become close-packed [234], there are only two acceptable configurations: cubic-close-packed (ccp) and hexagonally close-packed (hcp) (Fig. 27) [235]. Often ccp is preferentially formed instead of hcp because of the inherent entropic difference between the two structures [236]. In this way, 74% of the space is filled by spheres that each touch 12 nearest neighbors and leave void spaces that form a sphere-void interpenetrating network (meaning each phase continuously extends across the system in all directions). For other shapes, such as surfactant-coated nanoparticles that contain facets or flat edges, specific binding of like-sized faces preferentially occurs because of the favorable formation of a

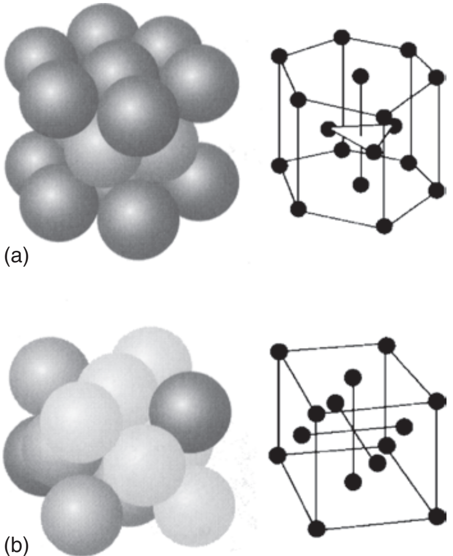


Figure 27. Schematic showing (a) hexagonally closed packed and (b) cubic close packed arrays. Reprinted with permission from [239], D. F. Shriver et al., "Inorganic Chemistry." Freeman, New York, 1999. © 1999, Freeman.

surfactant bilayer between particles [237]. It is also possible to use cubic particles to form cubic superlattices [238].

Colloidal particles can be used to form ordered, 3D arrays of nanoparticles by gravity- or centrifuge-induced particle sedimentation, filtration, solvent evaporation, confinement within membranes or gels, use of high pressures, and application of electrical or magnetic fields [5, 6, 22, 240–248]. The best routes for assembling colloids into these colloidal crystals involve particle sedimentation, intraparticle electrostatic interactions, or physical confinement to control stacking and periodicity [5].

Sedimentation usually involves ordering spherical colloids into 3D arrays by controlling particle density, size, and rate of sedimentation [5]. The sedimentation velocity, which normally depends on particle size and density, controls the order of the colloidal crystal based on kinetics and the ability of particles to find energy minima. Larger and denser particles such as amorphous silica (density = 2.19 g/cm^3) naturally have a greater tendency to precipitate but do not always form ordered matrices because the sedimentation velocity is too high. Particles must be larger than approximately 0.7 μm so that gravity forces them into minimum energy sites. Otherwise these objects would take extended periods of time to settle because the particles remain in solution [245]. Using sedimentation, silica has been produced in ccp lattices that mimic natural, iridescent opal [249]. Although this process can form lattices as large as 1 cm^3, sedimentation is problematic because the process can take months or years to complete for submicrometer particles and there is a lack of control over crystallinity, morphology, and the number of layers produced [5]. Techniques such as simultaneous solvent evaporation [250] and applied electric field [251] have attempted to compensate for the shortcomings inherent in sedimentation. Solvent evaporation forms ordered arrays of nanoparticles as long as the evaporation rate is slow enough [6, 22]. Slow evaporation allows particle diffusion into the proper crystalline position while fast evaporation produces disordered arrangements. Electrophoretic assisted particle settling involves the application of an electric field to control the settling velocity of the spheres [245, 251]. Variation of the electric field on charged particles alters the settling velocity and forms ordered arrays in reasonable time frames. Often, colloids are ordered into periodic arrays due to physical confinement within mircofluidic flow, overlapping sound waves (sonication), or a structural scaffold [5, 235].

After the ordered array has been formed, the interstitial regions of the nanoparticle array can be infiltrated with inorganic or organic precursors to form a nanocomposite [234, 248] (also see the infiltration section). Sometimes, the simple introduction of a precursor solution to the dried solid will allow sufficient infiltration. However, it is often necessary use several freezing, evacuation, and then heating (or freeze–pump–thaw) cycles to remove gasses in the solid and those dissolved in the monomer [234]. Nanoparticle thin films and membranes are used for filtration, solar cells, batteries, photoconductive material, photo- or electroluminescent materials, liquid and gaseous chemical sensors, electrodes, and absorbents [11].

5. INORGANIC PARTICLE/POLYMER NANOCOMPOSITES

Inorganic particle/polymer nanocomposites often utilize ceramic or metallic colloids with diameters in the 1–100 nm range that are coated or associated with a polymer layer or continuous polymer phase. A layer of polymeric material surrounding inorganic nanoparticles creates interesting aggregates, allows new packing morphologies, and introduces novel confinement properties to the system. This incorporation can result from covalent attachment of the polymers to the particle, encapsulation of the particles by polymerization or melting, or by simultaneous reaction of the organic and inorganic precursors to form the nanocomposite [224]. Nanoparticles within a polymer matrix allow the nanocomposites to have many advantageous properties such as increased toughness due to the mobility of the nanoparticles [23]. Another attractive characteristic of these nanocomposites is transparency of the composite material [252]. Synthesizing nanoparticles with defined particle size and morphology is an important first synthetic step in the creation of these nanocomposites.

5.1. Synthesis of Nanoparticles

5.1.1. Background

Many methods have been developed to synthesize a large variety of monodispersed nanoparticles [4, 253] by high-energy ball milling [32], chemical, or other approaches. The chemical approach usually involves reacting organic or inorganic precursors inside confined environments created by dendrimers [254], membrane pores [5, 199, 255], close-packed lattices of spherical colloids [241, 256], carbon arrays [257], gels [1], zeolites [258], lipid vesicles [259], Langmuir–Blodgett films, colloids [260], microemulsions [261–265], liquid crystals [3], and polymers [22, 266]. Other synthesis approaches like combustion flame [267, 268], plasma [269, 270], laser abilation, chemical vapor deposition [271], spray pyrolysis [272], electrospray [273], plasma spray [274, 275], aerosol [112, 276–278], and sonolysis [11, 279] have also been used to produce various nanoparticles.

Nanoparticles are commonly composed of calcium carbonate [280]; Ag [281], Fe/Cu alloys [282], and other metals [266, 283]; SiO_2 [78, 142, 284], TiO_2 [285, 286], ZnO [264, 278], WO_3, Nb_2O_5, and other metal oxides [287]; II–VI chalcogenides such as CdS, CdSe [11], CdZnS [288], and CdMnS [289]; and AgCl [290], Ag_2Se [291], and other metal salts. A current trend in nanoparticle synthesis is toward polymers [5, 292], such as monodispersed latex nanoparticles [293], and nanocomposites [261, 262, 294, 295]. These nanocomposites are often layered particles containing a core of one material surrounded by a shell of another. For example, it is possible to polymerize monomers around metal or inorganic particle exteriors to form a core–shell morphology [293, 294, 296, 297]. Besides spherical particles, other shapes such as cubes [238], rectangular prisms [298], nanotubes [32], nanofibers [199], cylinders [299], disks, discoids, and toroids [111] have been synthesized [299].

5.1.2. Nanoparticle Synthesized Using Self-Assembled Surfactants

Microemulsions, formed by thermodynamically stable oil in water micelles or reverse micelles of water droplets stabilized in oil with surfactant, have been used as nanoreactors with controlled size and chemical environment for the formation of nanoparticles (Fig. 28) [58, 299]. Because microemulsions may consist of monodispersed spherical, cylindrical, or sponge-like structures with features on the order of 10 nm [58], they can template the synthesis of a wide variety of particle shapes [1, 4, 10, 293, 299]. However, it is often crucial to synthesize nanoparticles with a narrow size distribution as well as controlled interfacial properties, morphology, shape, and composition [32]. These factors greatly affect the final material properties because they exert a great influence over the ability of particles to form periodic arrays. Particle size is generally several times larger than the original micelle size, and it is controlled by the concentrations of surfactant and aqueous reactants and the ratio between them [299]. Morphology of the nanoparticle can be tailored using specially designed surfactants [1].

The basic particle formation steps include particle nucleation either inside the micelle [220] or in the bulk solution; particle growth which involves continuous surfactant stabilization, micellular growth, and surfactant and reactant diffusion though the reaction media; and termination of particle growth. For either inorganic or organic nanoparticles, a short burst of particle nucleation is achieved by the sudden mixing of inorganic precursors [299] or by the introduction of an initiator [5]. The high interfacial surfactant surface areas (100 m^2/ml) form favorable sites for particle nucleation [220] and exert control over the growth [58, 300], internal structure, and surface properties of the nanoparticles [10]. After particle nucleation has occurred, particles grow via precursor addition to the particle surface or by agglomeration of adjacent particles. Reaction kinetics and particle size depend on the type and amount of surfactant used, temperature, particle environment, precursor concentration and transport limitations [301]. As this reaction progresses, surfactant diffuses to and stabilizes [301] particle-containing micelles at the expense of surfactant in solution and empty micelles. Because agglomeration yields particle polydispersity and unstability in the colloid system toward premature precipitation [32, 218], this growth mechanism is usually suppressed by control of overall particle concentration in solution and by the proper choice of surfactant stabilization agents. Particle growth is terminated by depletion of reactant, surface attachment, steric influences, or precipitation. Naturally, a limited amount of reactant

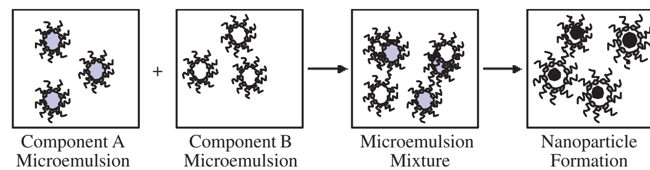


Figure 28. Formation of nanoparticles using microemulsions (water-in-oil) as nanoreactors. The water droplets continually collide, coalesce, and break up upon mixing of two microemulsions containing reactants [58].

Table 6. Silica particle/polymer nanocomposites.

Polymer	Technique	Characteristics	Ref.
Polyaniline (PANI)	aqueous phase polymerization in the presence of 35 nm diameter colloidal silica particles	electrical conductivity of 37.8 wt% PANI /silica nanocomposite (6.1×10^{-2} S cm^{-1}) higher than pure PANI (10^{-1} S cm^{-1})	[306]
Polypyrrole	polymerization in the presence of ultrafine silica sols	precipitating polymer coats and attaches particles together in a 100 to 300 nm raspberry morphology	[295]
4-vinylpyridine (4VP), acrylates, and styrene	interaction between basic 4VP and acidic silica or methacrylate-capped poly(ethylene glycol) produced favorable interactions between phases; composite formed by co-polymerization of 4VP with methyl methacrylate, butyl acrylate, or styrene	functional surfactant and co-polymers stabilize colloidal particles and allows a higher silica content; highly transparent, free-standing nanocomposite films were synthesized; nanocomposite particles and films have a "currant-bun" morphology	[307–310]
Poly(methyl methacrylate)	graft or adsorb initiator onto the silica surface followed by polymerization of the MMA monomer	core-shell silica/PMMA latex nanocomposites have raspberry-like morphology	[261, 262, 311, 312]
Polystyrene	polymer nanoparticles and modified silica colloids aggregate due to specific chemical (amine-aldehyde) or biological (avidin-biotin) interactions	core-shell nanocomposite materials consist of a silica core and a polystyrene shell	[296]
	silica particles are functionalized with DPE initiator and activated using *n*-BuLi; styrene anionic polymerization initiated by particle surface in benzene solution	living polymerization of styrene	[316]
Poly(methyl acrylate)	graft 3-(trimethoxysilyl)propyl methacrylate groups onto the surface of the silica colloids; system self-assembles into hexagonally ordered when more monomer added; monomer and surface groups polymerize	aggregated arrangements have significantly improved mechanical properties	[313–315]

present will stop particle growth [293]. Attachment of polymers on the nanoparticle surface caps the growing particle and prevents further growth by steric and electrostatic repulsions [220]. Sometimes specific binding agents like thiols are added to the solution to terminate cluster growth by attaching to the growing particle surface [220]. Steric termination occurs when there is no more surfactant in solution available to diffuse to the micelle and increase its size further [299]. It is important to note material density for colloidal synthesis techniques because gravity is a driving force for particle precipitation. Typical inorganic densities range from 2 to 4 g/cm^3 for materials such as silica (2 g/cm^3), cadmium sulfide (2.5 g/cm^3), calcite (2.6 g/cm^3), zinc oxide (4 g/cm^3), zinc sulfide (4 g/cm^3), and titanium oxide (4 g/cm^3) [22].

5.2. Nanoparticle Composite Synthesis Techniques

For nanoparticles, the sequential composite synthesis approach creates a second phase in a system of either preformed nanoparticles or polymer. Preformed nanoparticles form inorganic/polymer nanocomposites by either (1) directly polymerizing monomers in the presence of an inorganic sol or (2) modifying the surface of the silica particles with functional groups followed by polymerization of the monomers. Sometimes, polymerization preferentially forms a precipitated polymer shell on the individual particles based on specific interactions between the silica surface

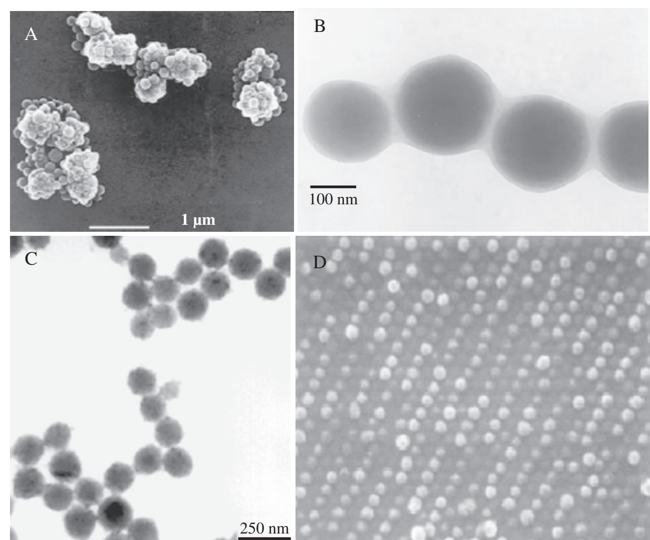


Figure 29. Different morphologies of silica colloid/polymer nanocomposites. (A) Raspberry morphology. Reprinted with permission from [306], J. Stejskal et al., *Macromolecules* 29, 6814 (1996). © 1996, American Chemical Society. (B) Core-shell morphology. Reprinted with permission from [261], J.-L. Luna-Xavier et al., *J. Colloid Interface Sci.* 250, 90 (2002). © 2002, Elsevier Science. (C) Currant-bun morphology. Reprinted with permission from [307], C. Barthet et al., *Adv. Mater.* 11, 408 (1999). © 1999, Wiley-VCH. (D) Ordered hexagonal array. Reprinted with permission from [315], J. M. Jethmalani et al., *Langmuir* 13, 3342 (1997). American Chemical Society.

Table 7. Metal particle/polymer nanocomposites

Metal/polymer	Technique	Characteristics	Ref.
Platinum/polystyrene	polymerize styrene in a system containing dispersed, 1–2 nm platinum nanoparticles, dialkyldithiophosphates, and octadecanethiol	platinum colloids self-assemble within the polystyrene matrix into structures such as hollow shell; morphology depends mainly0n on the chain lengths of the dialkyldithiophosphates	[317]
Silver/block copolymer PS-PVP	Silver nanoparticles initially adsorb onto PVP domains exposed on a preformed block copolymer film surface; after annealing of the films, nanoparticles migrate to the PS block of the polymer due to the lower surface energy of PS over PVP	Nanoparticles incorporate into the polymer film	[319]
Silver/HDPE	add HDPE to films of dodecanethiol modified silver nanoparticles in ρ-xylene; heated, homogeneous mixture drawn on a hot shoe at 120 °C	pearl-necklace-type arrays of spherical particles aligned in the drawing direction	[318]
Gold/ultrahigh molecular-weight polyethylene	mix ultrahigh-molecular-weight polyethylene with hot solution of dodecanethiol-modified gold nanoparticles; solid state drawing forms nanocomposite.	uniaxially oriented arrays of gold nanoparticles	[320]
Gold or copper/ poly(amidoamine) dendrimers	*in-situ* reduction of $HAuCl_4$ or copper precursor in the presence of poly(amidoamine) dendrimers	synthesize 2–3 nm gold or copper colloids encapsulated and stabilized by dendrimers	[321, 323]
Gold/poly(4-vinylpyridine) and poly(acrylic acid)	modify gold nanoparticles with pyridine, carboxyl, and flavin functional groups; poly(4-vinylpyridine) and poly(acrylic acid) then added in a layer by layer approach based on hydrogen bond interactions	core/shell gold nanoparticles containing organic multilayer shells	[322]

and the new shell [10]. In other cases, functional or reactive surface groups are required to increase the adhesion between phases or form a covalent link between polymer and particles. Hydrogen bonding, electrostatic interactions, and covalent bonds are required for close association of the phases.

Preformed polymer acts as a structure-directing agent during the formation of organic/inorganic nanocomposites. Often, reduction of a metal salt or metal alkoxide precursor in the presence of a self-assembled polymer system forms particles inside the swollen polymer matrices [22, 266, 302]. Initially, an inorganic precursor solution diffuses and swells a cross-linked polymer. Subsequent sol–gel [95, 303] or ion exchange [220] reactions form metal oxide or metal particles within the polymer matrix. The properties of the polymer and particle interfaces are important because they will determine the amount of adhesion between phases. Polymer/particle interactions can be increased by covalently linking the two phases or by using hybrid polymers with metal containing units [304]. This technique is very useful for mechanical reinforcement of elastic polymer because the nanoparticles act as cross-linking points in the polymer [266]. Also, if metal nanoparticles are combined with conducting polymer, an increase in conductivity is experienced [305].

Other methods for the formation of an organic/inorganic nanoparticle system include infiltration and adsorption of the metal nanoparticles into the polymer system, layer-by-layer self-assembly of nanoparticles and polyelectrolytes, melting of the polymer around the metal nanoparticles, and simultaneous polymer and inorganic particle formation. In the last case, inorganic precursors, monomer, and coupling agents can be combined and reacted simultaneously to give an interpenetrating network [22]. Different morphologies are obtained depending on relative speeds of either formative reaction.

5.2.1. Silica Nanoparticle/Polymer

Silica is one of the most commonly used colloids for the syntheses of inorganic particle/organic nanocomposites. Polymer systems such as polyaniline [306], polypyrrole [295], 4-vinylpyridine [307–310], PMMA [261, 262, 311, 312], poly(methacrylate) [313–315], and polystyrene [296, 316] have all been combined with silica nanoparticles to create a wide variety of nanocomposites. Table 6 summarizes several different silica particle/polymer nanocomposites, the experimental methods that synthesized them, and novel characteristics.

These techniques create composites with different morphologies such as core–shell, raspberry, and currant-bun. As shown in Figure 29, hexagonally ordered arrays of silica

Table 8. Other particle/polymer nanocomposites

Particle/polymer	Technique	Characteristics	Ref.
Titanium dioxide/PSS	layer-by layer self-assembly of cationic titania nanoparticles and anionic PSS polymers on *N*-2(2-aminoethyl)-3-aminopropane modified substrate	nanocomposite film of multiple alternating titania and PSS layers	[244]
Iron oxyhydroxide/κ-carrageenan and cellulose sulfate	κ-carrageenan and cellulose sulfate polysaccharides stabilize surface of 10 nm iron oxyhydroxide nanoparticles	core/shell iron oxyhydroxide nanoparticles with tight crosslinked polysaccharide layer	[324]
Iron oxide/PMMA	thin layer of silica coats magnetic iron oxide rods as small as 70 nm; addition of nanoparticles to PMMA solution and subsequent solvent evaporation results in solid film	nanoparticle agglomeration avoided by magnetizing iron oxide; final nanocomposite thin film composed of iron oxide/PMMA	[325]
Cadmium sulfide (CdS)/PMMA	CdS nanoparticles coated with silica and modified with initiator, polymerize methyl methacrylate around nanoparticles or disperse them in a PMMA matrix; cast nanocomposites into films	nanocomposite films retain the photoluminescence of the CdS quantum dots and show homogeneous nanoparticle dispersal	[297]
CdS/MMA and DDAMA	CdS nanoparticles synthesized inside didecyldimethylammonium methacrylate (DDAMA) reverse micelles in MMA organic phase; after nanoparticle formation, MMA and DDAMA were polymerized	transparent solid polymer matrix containing nonaggregated 20–80 nm CdS nanoparticles was formed	[326]
CdS or ferrite/AOT and isooctane organogel	CdS or ferrite nanoparticles stabilized in sodium bis(2-ethylhexyl) sulfosuccinate (AOT)/isooctane reverse micelle system; addition of ρ-chlorophenol formed organogel	phenols self-assemble into stacked fibers encapsulate CdS or superparamagnetic ferrite nanopaticles	[327]
Clay/styrene	Silica gel and clay nanoparticles functionalized with 1,1-diphenylethylene (DPE) initiator; modified surfaces initiate styrene polymerization; other monomers such as isoprene added later	living anion polymerization demonstrated by formation of styrene-isoprene block copolymer/nanoparticle composite material; bound polymer has low molecular weight and high molecular weight distribution	[294]

particles in polymer have also been demonstrated. Possible applications of these silica particle/polymer nanocomposites include biological assays, abrasive-resistant coatings, and catalyst supports.

5.2.2. Metal Nanoparticle/Polymer

Metal nanoparticles have many unique properties due to quantum size effects and a large number of unsaturated surface atoms. Dispersing these metal nanoparticles into a polymeric matrix provides processability, solubility, and stability [266]. Platinum [317], silver [318, 319], gold [320–322], and copper(323) nanoparticles have been incorporated into polystyrene, poly(2-vinylpyridine), poly(styrene-*b*-2-vinyl pyridine) (PS-PVP), high density polyethylene (HDPE), ultra-high-molecular-weight polyethylene, polypyrrole, poly(amidoamine) dendrimer, and poly(acrylic acid) polymer systems. Table 7 summarizes experimental methods and characteristics of several different metal particle/polymer nanocomposites.

5.2.3. Other Inorganic Nanoparticles/Polymer

Besides silica and metal colloids, nanoparticles such as titanium dioxide [244], iron oxide [324, 325], cadmium sulfide [297, 326, 327], and clay [294] have all been incorporated into polymer systems such as poly(sodium 4-styrenesolfonate), polysaccharides such as κ-carrageenan and cellulose sulfate, PMMA, polystyrene, and sodium bis(2-ethylhexyl) sulfosuccinate (AOT)/isooctane organogels.

The applications of these nanocomposites range from photovoltaics, catalysts, sensors, and reverse osmosis membranes [286]. One of the major applications of semiconductor nanoparticle/polymer composites is in the field of photovoltaics. Akimov et al. first demonstrated synthesis of conductive polymer nanocomposites containing 2 to 50 nm CdS nanocrystals, which were prepared in poly(vinyl alcohol), poly(vinylpyridine), and photographic gelatin. The composites exhibited good photosensitivity and photoconductivity. Also, nonlinear optical properties of quantum confined semiconductors are enhanced in polymer systems. Table 8 summarizes experimental methods and characteristics of these particle/polymer nanocomposites.

GLOSSARY

Amphiphile Molecule containing hydrophilic and hydrophobic regions.

Association colloid Micelle, emulsion, or liquid crystalline aggregation of amphiphiles based on weak interactions such as solvophobic interactions.

Biomimetic Observing natural solutions, extracting the essential operational features, and applying them to new engineered systems.

Bridged silsesquioxane Organometallic compound composed of two alkoxysilance groups covalently bound together by an organic ligand.

Capillary pressure Pressure differential inside a partially filled pore with a narrow diameter due to a curved interface. This pressure difference causes structural collapse in some mesoporous materials during solvent removal.

Critical micelle concentration (CMC) Minimum surfactant concentration for the formation of micelles.

Cubic mesostructure Bicontinuous two phase system that can be formed from the close packing of spheres into a periodic array.

Dip coating Immersion and controlled removal of a solid from a solution to cover the surface with a uniform layer.

Evaporation induced self-assembly (EISA) Evaporation of a solution containing inorganic precursors and surfactant allows co-assembly of inorganic clusters with liquid crystalline phases in the synthesis of a mesostructured inorganic material.

Gel A solution containing a solute that has formed a large, entangled network so that the system viscosity is greatly increased.

Hexagonal mesostructure A two phase system that consists of long parallel cylinders packed together in an array so that each cylinder is surrounded by six neighboring cylinders with the other phase in the voids.

Intercalation The infiltration and filling of a lamellar material with a guest species.

Lamellar mesostructure Consists of parallel alternating layers of two phases.

Layer-by-layer sequential assembly Alternating layers of anionic and cationic species are deposited on a substrate by a cycle of dip-coats in the corresponding solvents.

Liquid crystal Aligned molecular array that can flow and deform like a liquid.

Lyotropic liquid crystal An aligned molecular array that contains multiple components such as aligned surfactant in water.

Mesopore Pore with a 1 to 50 nm pore size.

Mesoscale Size scale of 1 to 50 nm.

Mesostructure Arrangement of material parts on the 1 to 50 nm size scale.

Micelle Roughly spherical pseudophase in an aqueous solution consisting of an aggregation of amphiphilic molecules that align to expose hydrophobic tails inward and hydrophilic heads outward.

Microemulsion Two phase system composed of small, 100 nm or less, droplets stabilized in a solvent by a surfactant.

Nanocomposite Material composed of two or more components that have at least one dimension between 1 to 100 nm.

Reverse micelle Roughly spherical pseudophase in an organic solution consisting of an aggregation of amphiphilic molecules that align to expose hydrophilic heads inward and hydrophobic tails outward.

Self-assembly Spontaneous organization of components into periodic or hierarchical structures based on noncovalent interactions.

Sol A solution containing dissolved metal oxide clusters such as colloidal particles.

Sol–gel process Useful synthesis route for metal oxide solids reacting soluble metal alkoxide or metal halide precursors to form metal oxide sol clusters in solution. Cluster size increases until a gel is created.

Solution precipitation For sol–gel systems, particles are formed from a solution of inorganic precursors and surfactant that interact and self-assemble together.

Supramolecule Atoms or molecules that aggregate together based on a series of noncovalent interactions to create a specific structure.

Surfactant A specific class of amphiphile often used as detergents and soaps. Usually a surfactant contains a hydrophobic head and a dydrophilic tail.

REFERENCES

1. W. Meier, *Curr. Opin. Colloid Interface Sci.* 4, 6 (1999).
2. E. S. Steigerwalt, 2001.
3. S. Forster and T. Plantenberg, *Angew. Chem. Int. Ed.* 41, 688 (2002).
4. S. Foerster and M. Antonietti, *Adv. Mater.* 10, 195 (1998).
5. B. D. Gates, University of Washington, 2001.
6. J. Yin, Georgia Institute of Technology, 2000.
7. C. Hwang, Y. S. Fu, Y. L. Lu, S. W. Jang, P. T. Chou, C. R. Wang, and S. J. Yu, *J. Catal.* 195, 336 (2000).
8. T. Terenishi and M. Miyake, *Chem. Mater.* 10, 594 (1998).
9. P. Chen, X. Wu, J. Lin, and K. L. Tan, *J. Phys. Chem. B* 103 (1999).
10. E. Bourgeat-Lami, *J. Nanosci. Nanotech.* 2, 1 (2002).
11. D. Meisel, *Curr. Opin. Colloid Interface Sci.* 2, 188 (1997).
12. H. Eckert and M. Ward, *Chem. Mater.* 13, 3059 (2001).
13. Y. Lu, H. Fan, N. Doke, D. A. Loy, R. A. Assink, D. A. LaVan, and C. J. Brinker, *J. Am. Chem. Soc.* 122, 5258 (2000).
14. M. H. Lim and A. Stein, *Chem. Mater.* 11, 3285 (1999).
15. C. Sanchez, F. Ribot, L. Rozes, and B. Alonso, *Molec. Cryst. Liq. Cryst. Sci. Tech. A* 354, 143 (2000).
16. J. Wen and G. L. Wilkes, *Chem. Mater.* 8, 1667 (1996).
17. D. B. Mitzi, K. Chondroudis, and C. R. Kagan, *IBM J. Res. Dev.* 45, 29 (2001).
18. Y. Lu, Y. Yang, A. Sellinger, M. Lu, J. Huang, H. Fan, R. Haddad, G. Lopez, A. R. Burns, D. Y. Sasaki, J. Shelnutt, and C. J. Brinker, *Nature* 410, 913 (2001).
19. N. F. Uvarov and E. F. Hairretdinov, *Diffusion Defect Data Solid State Data B* 39–40, 27 (1994).
20. K. Adachi and N. Imanaka, *Jpn. Kokai Tokkyo Koho Japan JP* 3 (1999).
21. A. Sellinger, P. M. Weiss, N. Anh, Y. Lu, R. A. Assink, W. Gong, and C. J. Brinker, *Nature* 394, 256 (1998).

22. E. Bourgeat-Lami, *Microspheres Microcapsules Liposomes* 5, 149 (2002).
23. D. Gersappe, *Phys. Rev. Lett.* 89, 058301/1-058301/4 (2002).
24. T. Takagi, T. Ishihara, H. Nishiguchi, and Y. Takita, *Proc. Electrochem. Soc.* 99-23, 396 (1999).
25. K. Suri, S. Annapoorni, and R. P. Tandon, *J. Sci. Indust. Res.* 60, 724 (2001).
26. J. H. Clark and D. J. Macquarrie, *Chem. Commun.* 853–860 (1998).
27. C. G. Guizard and A. C. Julbe, *Membrane Sci. Tech. Ser.* 6, 435 (2000).
28. Z.-K. Xu, L. Xiao, J. Wang, and Y.-Y. Xu, *Polym. Mater. Sci. Eng.* 85, 92 (2001).
29. T. Uragami, H. Matsugi, and T. Miyata, *Nettowaku Porima* 22, 157 (2001).
30. S.-Y. Lu, H.-Y. Huang, and K.-H. Wu, *J. Mater. Res.* 16, 3053 (2001).
31. Y. Kong, H. Du, J. Yang, D. Shi, Y. Wang, Y. Zhang, and W. Xin, *Desalination* 146, 49 (2002).
32. R. W. Siegel, E. Hu, and M. C. Roco, "Nanostructure Science and Technology." Kluwer Academic, Dordrecht, 1999.
33. C. J. Brinker, Y. Lu, A. Sellinger, and H. Fan, *Adv. Mater.* 11, 579 (1999).
34. G. A. Ozin, D. Khushalani, and S. Oliver, *NATO ASI Ser. C* 499, 323 (1997).
35. E. Giannelis, *Adv. Chem. Ser.* 245, 259 (1995).
36. C. N. R. Rao, *Curr. Sci.* 81, 1030 (2001).
37. N. Bowden, A. Terfort, J. Carbeck, and G. M. Whitesides, *Science* 276, 233 (1997).
38. A. Terfort, N. Bowden, and G. M. Whitesides, *Nature* 386, 162 (1997).
39. J. S. Beck, J. C. Vartuli, W. J. Roth, M. E. Leonowicz, C. T. Kresge, K. D. Schmitt, C. T. W. Chu, D. H. Olson, E. W. Sheppard, et al., *J. Am. Chem. Soc.* 114, 10834 (1992).
40. M. Sarikaya, H. Fong, D. W. Frech, and R. Humbert, *Mater. Sci. Forum* 293, 83 (1999).
41. J. N. Cha, G. D. Stucky, D. E. Morse, and T. J. Deming, *Nature* 403, 289 (2000).
42. C.-W. Li and B. E. Volcani, *Philos. Trans. Roy. Soc. London Ser. B* 304, 518 (1984).
43. C. N. Satterfield, "Heterogeneous Catalysis in Industrial Practice." 2nd ed. Krieger, Malabar, 1996.
44. N. Y. Chen and W. E. Garwood, *Catal. Rev. - Sci. Eng.* 28, 185 (1986).
45. N. Y. Chen, W. E. Garwood, and F. G. Dwyer, *Chem. Ind.* 36 (1989).
46. H. W. Haynes, Jr., *Catal. Rev. - Sci. Eng.* 17, 273 (1978).
47. C. Kresge, M. Leonowicz, W. Roth, C. Vartuli, and J. Beck, *Nature* 359, 710 (1992).
48. D. M. Antonelli and J. Y. Ying, *Curr. Opin. Colloid Interface Sci.* 1, 523 (1996).
49. M. E. Davis, C.-Y. Chen, S. L. Burkett, and R. F. Lobo, *Mater. Res. Soc. Symp. Proc.* 346, 831 (1994).
50. G. Fu, C. A. Fyfe, W. Schwieger, and G. T. Kokotailo, *Angew. Chem. Int. Ed.* 34, 1499 (1995).
51. P. Yang, D. Zhao, D. I. Margolese, B. F. Chmelka, and G. D. Stucky, *Nature* 396, 152 (1998).
52. P. T. Tanev and T. J. Pinnavaia, *Science* 267, 865 (1995).
53. P. C. Hiemenz and R. Rajagopalan, "Principles of Colloid and Surface Chemistry," 3rd ed. Dekker, New York, 1997.
54. J. W. McBain and C. S. Salmon, *J. Am. Chem. Soc.* 42, 426 (1920).
55. F. Grieser and C. J. Drummond, *J. Phys. Chem.* 92, 5580 (1988).
56. R. Zana, *Surf. Sci. Ser.* 22, 479 (1987).
57. F. M. Menger and J. M. Jerkunica, *J. Am. Chem. Soc.* 100, 688 (1978).
58. D. O. Shah and C. A. Stokes, "Micelles, Microemulsions, and Monolayers: Science and Technology." Dekker, New York, 1998.
59. S. Foerster, M. Zisenis, E. Wenz, and M. Antonietti, *J. Chem. Phys.* 104, 9956 (1996).
60. J. H. Fendler and E. J. Fendler, "Catalysis in Micellar and Macromolecular Systems." Academic Press, New York, 1975.
61. J. M. Prausnitz, R. N. Lichtenthaler, and E. G. D. Azevedo, "Molecular Thermodynamics of Fluid-Phase Equilibria," 3rd ed. Prentice Hall, Upper Saddle River, 1999.
62. S. Kumar, J. Brock, D. Finotello, M. Fisch, C. Garland, J. Ho, M. Neubert, B. Padulka, P. Photinos, S. Sinha, and P. Ulkeja, "Liquid Crystals: Experimental Study of Physical Properties and Phase Transitions." Cambridge Univ. Press, Cambridge, UK, 2001.
63. B. Halle and G. Carlstroem, *J. Phys. Chem.* 85, 2142 (1981).
64. S. Eftekharzadeh, 1998.
65. J. Israelachvili, "Intermolecular and Surface Forces," 2nd ed. Academic Press, San Diego, 1992.
66. G. H. Brown and J. J. Wolken, "Liquid Crystals and Biological Structures." Academic Press, New York, 1979.
67. T. Kato, *Science* 295, 2414 (2002).
68. M. Aamodt, M. Landgren, and B. Joensson, *J. Phys. Chem.* 96, 945 (1992).
69. E. Hackett, E. Manias, and E. P. Giannelis, *J. Chem. Phys.* 108, 7410 (1998).
70. Q. Huo, D. I. Margolese, and G. D. Stucky, *Chem. Mater.* 8, 1147 (1996).
71. J. Sjoeblom, P. Stenius, and I. Danielsson, *Surf. Sci. Ser.* 23, 369 (1987).
72. S. Forster, M. Rehahn, and J. P. Rabe, *Macromolec. Chem. Phys.* 202, 2802 (2001).
73. Z. Ahmad and J. E. Mark, *Mater. Sci. Eng. C* 6, 183 (1998).
74. J. D. F. Ramsay, *Curr. Opin. Colloid Interface Sci.* 1, 208 (1996).
75. L. L. Hench, in "Encyclopedia of Chemical Technology" (J. I. Kroschwitz and M. Howe-Grant, Eds.), Vol. 22, pp. 497–528. Wiley Interscience, New York, 1998.
76. C. J. Brinker and G. W. Scherer, "sol–gel Science: The Physics and Chemistry of sol–gel Processing." Academic Press, San Diego, 1990.
77. Z. Luan, H. He, W. Zhou, and J. Klinowski, *J. Chem. Soc., Faraday Trans.* 94, 979 (1998).
78. H. Yang, G. Vovk, N. Coombs, I. Sokolov, and G. A. Ozin, *J. Mater. Chem.* 8, 743 (1998).
79. X. Chen, L. Huang, and Q. Li, *J. Phys. Chem. B* 101, 8460 (1997).
80. M. T. Anderson, P. S. Sawyer, and T. Rieker, *Microporous Mesoporous Mater.* 20, 53 (1998).
81. A. Corma, Q. Kan, and F. Rey, *Chem. Commun.* 579-580 (1998).
82. A. A. Romero, M. D. Alba, and J. Klinowski, *J. Phys. Chem. B*, 102, 123 (1998).
83. M. Templin, A. Franck, A. Du Chesne, H. Leist, Y. Zhang, R. Ulrich, V. Schadler, and U. Wiesner, *Science* 278, 1795 (1997).
84. T. Kimura, Y. Sugahara, and K. Kuroda, *Chem. Commun.* 559 (1998).
85. Z. Luan, D. Zhao, H. He, J. Klinowski, and L. Kevan, *J. Phys. Chem. B*, 102, 1250 (1998).
86. A. Firouzi, F. Atef, A. G. Oertli, G. D. Stucky, and B. F. Chmelka, *J. Am. Chem. Soc.* 119, 3596 (1997).
87. S. Oliver, A. Kuperman, and G. A. Ozin, *Angew. Chem. Int. Ed.* 37, 46 (1998).
88. S. R. J. Oliver and G. A. Ozin, *J. Mater. Chem.* 8, 1081 (1998).
89. M. Yada, H. Kitamura, M. Machida, and T. Kijima, Langmuir 13, 5252 (1997).
90. S. Cabrera, J. El Haskouri, A. Beltran-Porter, D. Beltran-Porter, M. D. Marcos, and P. Amoros, *Solid State Sci.* 2, 513 (2000).
91. V. F. Stone, Jr. and R. J. Davis, *Chem. Mater.* 10, 1468 (1998).
92. G. J. Li and S. Kawi, *Talanta* 45, 759 (1998).
93. L. Qi, J. Ma, H. Cheng, and Z. Zhao, *Langmuir* 14, 2579 (1998).
94. G. Pacheco, E. Zhao, A. Garcia, A. Sklyarov, and J. J. Fripiat, *J. Mater. Chem.* 8, 219 (1998).

95. K. C. Kwiatkowski and C. M. Lukehart, in "Handbook of Nanostructured Materials and Nanotech.: Synthesis and Processing" (H. S. Nalwa, Ed.), Vol. 1, pp. 387–421. Academic Press, San Diego, 2000.
96. G. S. Attard, P. N. Bartlett, N. R. B. Coleman, J. M. Elliott, J. R. Owen, and J. H. Wang, *Science* 278, 838 (1997).
97. H. Yang, N. Coombs, and G. A. Ozin, *J. Mater. Chem.* 8, 1205 (1998).
98. M. T. Anderson, J. E. Martin, J. G. Odinek, and P. P. Newcomer, *Chem. Mater.* 10, 1490 (1998).
99. M. Antonietti, *Curr. Opin. Colloid Interface Sci.* 6, 244 (2001).
100. R. Ulrich, A. C. Finnefrock, A. DuChesne, S. Mahajan, P. F. W. Simon, S. M. Gruner, and U. Wiesner, *Polym. Mater. Sci. Eng.* 84, 788 (2001).
101. A. Corma, *Chem. Rev.* 97, 2373 (1997).
102. M. Roca, J. E. Haskouri, A. Beltran-Porter, J. Alamo, and P. Amoros, *Chem. Commun* 1883 (1998).
103. I. Honma, H. S. Zhou, D. Kundu, and A. Endo, *Adv. Mater.* 12, 1529 (2000).
104. D. Zhao, J. Feng, Q. Huo, N. Melosh, G. H. Frederickson, B. F. Chmelka, and G. D. Stucky, *Science* 279, 548 (1998).
105. D. Zhao, J. Sun, Q. Li, and G. D. Stucky, *Chem. Mater.* 12, 275 (2000).
106. D. Zhao, Q. Huo, J. Feng, B. F. Chmelka, and G. D. Stucky, *J. Am. Chem. Soc.* 120, 6024 (1998).
107. S. Valange, J. L. Guth, F. Kolenda, S. Lacombe, and Z. Gabelica, *Microporous Mesoporous Mater.* 35–36, 597 (2000).
108. P. V. Braun, P. Osenar, and S. I. Stupp, *Nature* 380, 325 (1996).
109. T. E. Mallouk, S. W. Keller, C. M. Bell, H.-G. Hong, H.-N. Kim, Y. Kim, II, D. M. Kaschak, D. L. Feldheim, P. J. Ollivier, and S. A. Johnson, in "Proc. Robert A. Welch Found. Conf. Chem. Res.," pp. 123–137, 1995.
110. P. Feng, X. Bu, G. D. Stucky, and D. J. Pine, *J. Am. Chem. Soc.*, 122, 994 (2000).
111. H. Yang, N. Coombs, and G. A. Ozin, *Nature* 386, 692 (1997).
112. Y. Lu, H. Fan, A. Stump, T. L. Ward, T. Rieker, and C. J. Brinker, *Nature* 398, 223 (1999).
113. H. Fan, Y. Lu, A. Stump, S. T. Reed, T. Baer, R. Schunk, V. Perez-Luna, G. P. Lopez, and C. J. Brinker, *Nature* 405, 56 (2000).
114. R. Laversanne, *Macromolecules* 25, 489 (1992).
115. W. Srisiri, Y.-S. Lee, T. M. Sisson, B. Bondurant, and D. F. O'Brien, *Tetrahedron* 53, 15397 (1997).
116. K. A. Koyano, T. Tatsumi, Y. Tanaka, and S. Nakata, *J. Phys. Chem. B* 101, 9436 (1997).
117. P. Van Der Voort, M. Baltes, and E. F. Vansant, *J. Phys. Chem. B* 103, 10102 (1999).
118. V. Antochshuk and M. Jaroniec, *Chem. Commun.* 2373 (1999).
119. C. P. Jaroniec, M. Kruk, M. Jaroniec, and A. Sayari, *J. Phys. Chem. B* 102, 5503 (1998).
120. S. O'Brien, J. Tudor, S. Barlow, M. J. Drewitt, S. J. Heyes, and D. O'Hare, *Chem. Commun.* 641 (1997).
121. J. Tudor and D. O'Hare, *Chem. Commun.* 603 (1997).
122. J. F. Dý'az, K. J. Balkus Jr., F. Bedioui, V. Kurshev, and L. Kevan, *Chem. Mater.* 9, 61 (1997).
123. R. Anwander, I. Nagl, M. Widenmeyer, O. G. Engelhardt, C. Palm, and T. Rolser, *J. Phys. Chem. B* 104, 3532 (2000).
124. S. S. Prakash, C. J. Brinker, A. J. Hurd, and S. M. Rao, *Nature* 375, 431 (1995).
125. X. S. Zhao and G. Q. Lu, *J. Phys. Chem. B* 102, 1556 (1998).
126. A. Shimojima and K. Kuroda, *Langmuir* 18, 1144 (2002).
127. M. H. Lim, C. F. Blanford, and A. Stein, *J. Am. Chem. Soc.* 119, 4090 (1997).
128. K. Moller, T. Bein, and R. X. Fischer, *Chem. Mater.* 11, 665 (1999).
129. W. M. Van Rhijn, D. E. De Vos, B. F. Sels, W. D. Bossaert, and P. A. Jacobs, *Chem. Commun.* 317 (1998).
130. J. Liu, X. Feng, G. E. Fryxell, L.-Q. Wang, A. Y. Kim, and M. Gong, *Adv. Mater.* 10, 161 (1998).
131. X. Feng, G. E. Fryxell, L. Q. Wang, A. Y. Kim, J. Liu, and K. M. Kemner, *Science* 276, 923 (1997).
132. M. H. Lim, C. F. Blanford, and A. Stein, *Chem. Mater.* 10, 467 (1998).
133. S. L. Burkett, S. D. Sims, and S. Mann, *Chem. Commun.* 1367 (1996).
134. S. D. Sims, S. L. Burkett, and S. Mann, *Mater. Res. Soc. Symp. Proc.* 431, 77 (1996).
135. Y. V. S. Rao, D. E. De Vos, T. Bein, and P. A. Jacobs, *Chem. Commun.* 355 (1997).
136. B. Marler, U. Oberhagemann, S. Vortmann, and H. Gies, *Microporous Mater.* 6, 375 (1996).
137. E. P. Plueddemann, "Silane Coupling Agents," 1982.
138. S. Naviroj, S. R. Culler, J. L. Koenig, and H. Ishida, *J. Colloid Interface Sci.* 97, 308 (1984).
139. C. P. Tripp and M. L. Hair, *Langmuir* 11, 149 (1995).
140. N. Nishiyama, T. Asakura, and K. Horie, *J. Colloid Interface Sci.* 124, 14 (1988).
141. R. Laible and K. Hamann, *Adv. Colloid Interface Sci.* 13, 65 (1980).
142. A. P. Philipse and A. Vrij, *J. Colloid Interface Sci.* 128, 121 (1989).
143. S. Brandriss and S. Margel, *Langmuir* 9, 1232 (1993).
144. J. W. Goodwin, R. S. Harbron, and P. A. Reynolds, *Colloid Polym. Sci.* 268, 766 (1990).
145. A. Corma, V. Fornes, H. Garcia, M. A. Miranda, and M. J. Sabater, *J. Am. Chem. Soc.* 116, 9767 (1994).
146. M. L. Cano, A. Corma, V. Fornes, H. Garcia, M. A. Miranda, C. Baerlocher, and C. Lengauer, *J. Am. Chem. Soc.* 118, 11006 (1996).
147. J. F. Diaz, K. J. Balkus, Jr., F. Bedioui, V. Kurshev, and L. Kevan, *Chem. Mater.* 9, 61 (1997).
148. D. J. Macquarrie, *Chem. Commun.* 1961 (1996).
149. D. J. Macquarrie and D. B. Jackson, *Chem. Commun.* 1781 (1997).
150. A. Cauvel, G. Renard, and D. Brunel, *J. Org. Chem.* 62, 749 (1997).
151. J. F. Diaz and K. J. Balkus, Jr., *J. Molec. Catal. B* 2, 115 (1996).
152. E. Armengol, A. Corma, H. Garcia, and J. Primo, *Appl. Catal. A* 149, 411 (1997).
153. A. Corma, V. Fornes, H. Garcia, V. Marti, and M. A. Miranda, *Chem. Mater.* 7, 2136 (1995).
154. G. A. Ozin, E. Chomski, D. Khushalani, and M. J. Maclachlan, *Curr. Opin. Colloid Interface Sci.* 3, 181 (1998).
155. C. Yoshina-Ishii, T. Asefa, N. Coombs, M. J. MacLachlan, and G. A. Ozin, *Chem. Commun.* 2539 (1999).
156. B. J. Melde, B. T. Holland, C. F. Blanford, and A. Stein, *Chem. Mater.* 11, 3302 (1999).
157. S. Inagaki, S. Guan, Y. Fukushima, T. Ohsuna, and O. Terasaki, *J. Am. Chem. Soc.* 121, 9611 (1999).
158. S. Inagaki, S. Guan, T. Ohsuna, and O. Terasaki, *Nature* 416, 304 (2002).
159. K. J. Shea and D. A. Loy, *Chem. Mater.* 13, 3306 (2001).
160. R. C. Kenneth, in "ACS Division of Polymer Chemistry," 1999.
161. S. I. Association, "International Technology Roadmap for Semiconductors," 1994.
162. S. I. Association, "International Technology Roadmap for Semiconductors," 1999.
163. J. P. Schaffer, A. Saxena, S. D. Antolovich, J. Thomas H. Sanders, and S. B. Warner, "The Science and Design of Engineering Materials," 1995.
164. S. Yang, P. A. Mirau, C.-S. Pai, O. Nalamasu, E. Reichmanis, E. K. Lin, H.-J. Lee, D. W. Gidley, and J. Sun, *Chem. Mater.* 13, 2762 (2001).
165. C. V. Nguyen, K. R. Carter, C. J. Hawker, J. L. Hedrick, R. L. Jaffe, R. D. Miller, J. F. Remenar, H.-W. Rhee, P. M. Rice, M. F. Toney, M. Trollss, and D. Y. Yoon, *Chem. Mater.* 11, 3080 (1999).
166. A. Heise, C. Nguyen, R. Malek, J. L. Hedrick, C. W. Frank, and R. D. Miller, *Macromolecules* 33, 2346 (2000).
167. P. Plucinski and W. Nitsch, *J. Phys. Chem.* 97, 8983 (1993).

168. C. T. Kresge, J. C. Vartuli, W. J. Roth, M. E. Leonowicz, J. S. Beck, K. D. Schmitt, C. T. W. Chu, D. H. Olson, E. W. Sheppard et al., *Stud. Surf. Sci. Catal.* 92, 11 (1995).
169. J. L. Blin, C. Otjacques, G. Herrier, and B.-L. Su, *Langmuir* 16, 4229 (2000).
170. A. Sayari, Y. Yang, M. Kruk, and M. Jaroniec, *J. Phys. Chem. B* 103, 3651 (1999).
171. M. Kruk, M. Jaroniec, and A. Sayari, *J. Phys. Chem. B* 103, 4590 (1999).
172. A. Sayari, *Agnew. Chem. Int. Ed.* 39, 2920 (2000).
173. A. Sayari and S. Hamoudi, *Chem. Mater.* 13, 3151 (2001).
174. D. Zhao, P. Yang, N. Melosh, J. Feng, B. F. Chmelka, and G. D. Stucky, *Adv. Mater.* 10, 1380 (1998).
175. B. J. Scott, G. Wirnsberger, and G. D. Stucky, *Chem. Mater.* 13, 3140 (2001).
176. I. Honma and H. S. Zhou, *Adv. Mater.* 10, 1532 (1998).
177. J. D. Badjic and N. M. Kostic, *J. Phys. Chem. B* 105, 7482 (2001).
178. J. Wu, M. M. Abu-Omar, and S. H. Tolbert, *Nano. Lett.* 1, 27 (2001).
179. R. Hoppe, A. Ortlam, J. Rathousky, G. Schulz-Ekloff, and A. Zukal, *Microporous Mater.* 8, 267 (1997).
180. M. H. Huang, B. S. Dunn, and J. I. Zink, *J. Am. Chem. Soc.* 122, 3739 (2000).
181. Y. Wei, Q. Feng, J. Xu, H. Dong, K.-Y. Qiu, S. A. Jansen, R. Yin, and K. K. Ong, *Adv. Mater.* 12, 1448 (2000).
182. S.-M. Zhang, Y.-J. Gao, Z.-J. Zhang, H.-X. Dang, W.-M. Liu, and Q.-J. Xue, Huaxue Xuebao 60, 1497 (2002).
183. I. A. Aksay, M. Trau, S. Manne, I. Honma, N. Yao, L. Zhou, P. Fenter, P. M. Eisenberger, and S. M. Gruner, *Science* 273, 892 (1996).
184. Q. W. Feng, J. G. Xu, H. Dong, S. Li, and Y. Wei, *J. Mater. Chem.* 10, 2490 (2000).
185. T. Hori, N. Kuramoto, H. Tagaya, M. Karasu, J.-I. Kadokawa, and K. Chiba, *J. Mater. Res.* 14, 5 (1999).
186. A. F. Gross, E. J. Ruiz, and S. H. Tolbert, *J. Phys. Chem. B* 104, 5448 (2000).
187. M. Ogawa and H. Sasabe, A silica-surfactant nanocomposite and its manufacture, in Eur. Pat. Appl., Institute of Physical and Chemical Research, Japan., 1996, p. 12.
188. M. Ogawa, *Chem. Commun.* 1149 (1996).
189. M. Ogawa, *J. Am. Chem. Soc.* 116, 7941 (1994).
190. C. P. Kao, H. P. Lin, and C. Y. Mou, *J. Phys. Chem. Solid* 62, 1555 (2001).
191. S. H. Tolbert, P. Sieger, G. D. Stucky, S. M. J. Aubin, C.-C. Wu, and D. N. Hendrickson, *J. Am. Chem. Soc.* 119, 8652 (1997).
192. Y. Kubota, Y. Nishizaki, and Y. Sugi, *Chem. Lett.* 998 (2000).
193. L. F. da Silva, L. P. R. Profeti, N. R. Stradiotto, and H. P. Oliveira, *J. Non-Cryst. Solid* 298, 213 (2002).
194. M. Ogawa, T. Igarashi, and K. Kuroda, *Bull. Chem. Soc. Jpn.* 70, 2833 (1997).
195. J. H. Norton and C. A. Guymon, *Polym. Preprint ACS Div. Polym. Chem.* 42, 753 (2001).
196. A. Shimojima and K. Kuroda, *Chem. Lett.* 1310 (2000).
197. D. F. O'Brien, B. Armitage, A. Benedicto, D. E. Bennett, H. G. Lamparski, Y.-S. Lee, W. Srisiri, and T. M. Sisson, *Acc. Chem. Res.* 31, 861 (1998).
198. G. Wirnsberger, B. J. Scott, H. C. Huang, N. A. Melosh, P. Yang, B. Chmelka and G. D. Stucky, Inorganic/block copolymer-dye composites and dye doped mesoporous materials for optical and sensing applications, in PCT Int. Appl., The Regents of the University of California, USA, 2002, p. 60.
199. C. R. Martin, *Science* 266, 1961 (1994).
200. T.-Q. Nguyen, J. Wu, S. H. Tolbert, and B. J. Schwartz, *Adv. Mater.* 13, 609 (2001).
201. E. R. Kleinfeld and G. S. Ferguson, *Science* 265, 370 (1994).
202. R. Gangopadhyay and A. De, *Chem. Mater.* 12, 608 (2000).
203. J. Wu, A. F. Gross, and S. H. Tolbert, *J. Phys. Chem. B* 103, 2374 (1999).
204. T.-Q. Nguyen, J. Wu, V. Doan, B. J. Schwartz, and S. H. Tolbert, *Science* 288, 652 (2000).
205. C. G. Wu and T. Bein, *Science* 264, 1757 (1994).
206. C. G. Wu and T. Bein, *Stud. Surf. Sci. Catal.* 84, 2269 (1994).
207. C.-G. Wu and T. Bein, *Chem. Mater.* 6, 1109 (1994).
208. C.-G. Wu and T. Bein, *Science* 266, 1013 (1994).
209. P. L. Llewellyn, U. Ciesla, H. Decher, R. Stadler, F. Schueth, and K. K. Unger, *Stud. Surf. Sci. Catal.* 84, 2013 (1994).
210. S. M. Ng, S. Ogino, T. Aida, K. A. Koyano, and T. Tatsumi, *Macromolec. Rapid Commun.* 18, 991 (1997).
211. S. A. Johnson, D. Khushalani, N. Coombs, T. E. Mallouk, and G. A. Ozin, *J. Mater. Chem.* 8, 13 (1998).
212. K. Kageyama, J.-I. Tamazawa, and T. Aida, *Science* 285, 2113 (1999).
213. M. J. MacLachlan, P. Aroca, N. Coombs, I. Manners, and A. O. Geoffrey, *Adv. Mater.* 10, 144 (1998).
214. J. P. Quaegebeur, T. Seguchi, H. Le Bail, and C. Chachaty, *J. Polym. Sci., Polym. Chem. Ed.* 14, 2703 (1976).
215. S. D. Cox and G. D. Stucky, *J. Phys. Chem.* 95, 710 (1991).
216. T. Bein and P. Enzel, *Angew. Chem. Int. Ed.* 101, 1737 (1989).
217. P. Enzel and T. Bein, *J. Phys. Chem.* 93, 6270 (1989).
218. K. L. Cooper, R. O. Claus, J. B. Mecham, K. Huie, and R. Swavey, *Proc. SPIE* 4512, 93 (2001).
219. R. O. Claus, Y. Liu, and K. L. Cooper, *Proc. SPIE* 4134, 1 (2000).
220. K. E. Gonsalves, S. P. Rangarajan, and J. Wang, in "Handbook of Nanostructured Materials and Nanotech.: Synthesis and Processing" (H. S. Nalwa, Ed.), Vol. 1, pp. 1–56. Academic Press, San Diego, 2000.
221. D. Li, D. Yang, Y. Jiang, X. Deng, Z. Wu, and X. Chen, *Proc. SPIE* 3324, 301 (1998).
222. A. C. Fou, O. Onitsuka, M. Ferreira, M. F. Rubner, and B. R. Hsieh, *J. Appl. Phys.* 79, 7501 (1996).
223. D. Ingersoll, P. J. Kulesza, and L. R. Faulkner, *J. Electrochem. Soc.* 141, 140 (1994).
224. J. H. Fendler, *Chem. Mater.* 8, 1616 (1996).
225. M. Sano, Y. Lvov, and T. Kunitake, *Annu. Rev. Mater. Sci.* 26, 153 (1996).
226. R. G. Freeman, K. C. Grabar, K. J. Allison, R. M. Bright, J. A. Davis, A. P. Guthrie, M. B. Hommer, M. A. Jackson, P. C. Smith et al., *Science* 267, 1629 (1995).
227. M. Z. Hussein and H. K. Seng, *J. Mater. Sci. Lett.* 20, 1787 (2001).
228. A. Lerf, in "Handbook of Nanostructured Materials and Nanotech.: Organics, Polymers, and Biological Materials" (H. S. Nalwa, Ed.), Vol. 5, pp. 1–166. Academic Press, San Diego, 2000.
229. J.-R. Butruille and T. J. Pinnavaia, in "Characterization of Catalytic Materials" (I. E. Wachs and L. E. Fitzpatric, Ed.), pp. 149–163. Butterworth-Heinemann, Boston, 1992.
230. R. M. Barrer and D. M. MacLeod, *Trans. Faraday Soc.* 51, 1290 (1955).
231. Y. J. Liu, D. C. DeGroot, J. L. Schindler, C. R. Kannewurf, and M. G. Kanatzidis, *J. Chem. Soc., Chem. Commun.* 593–6 (1993).
232. R. Bissessur, D. C. DeGroot, J. L. Schindler, C. R. Kannewurf, and M. G. Kanatzidis, *J. Chem. Soc., Chem. Commun.* 687–9 (1993).
233. T. Cassagneau, G. B. Hix, D. J. Jones, P. Maireles-Torres, M. Rhomari, and J. Roziere, *J. Mater. Chem.* 4, 189 (1994).
234. S. A. Johnson, P. J. Ollivier, and T. E. Mallouk, *Science* 283, 963 (1999).
235. S. Neser, C. Bechinger, P. Leiderer, and T. Palberg, *Phys. Rev. Lett.* 79, 2348 (1997).
236. A. van Blaaderen, R. Rue, and P. Wiltzius, *Nature* 385, 321 (1997).
237. S. Mann, S. A. Davis, S. R. Hall, M. Li, K. H. Rhodes, W. Shenton, S. Vaucher, and B. Zhang, *Dalton* 3753–3763 (2000).
238. S. Vaucher, M. Li, and S. Mann, *Angew. Chem. Int. Ed.* 39, 1793 (2000).

239. D. F. Shriver and P. W. Atkins, "Inorganic Chemistry," W. H. Freeman, and Company, New York, 1999.
240. B. Gates and Y. Xia, in "Materials Research Society Symposium Proceedings," 1999.
241. A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S. W. Leonardo, C. Lopez, F. Meseguer, H. Miguez, J. P. Mondia, G. A. Ozin, O. Toader, and H. M. Van Driel, *Nature* 405, 437 (2000).
242. I. I. Tarhan and G. H. Watson, *Phys. Rev. Lett.* 76, 315 (1996).
243. J. W. Goodwin, R. H. Ottewill, and A. Parentich, *J. Phys. Chem.* 84, 1580 (1980).
244. Y. Liu and R. O. Claus, *Proc. SPIE* 3242, 118 (1997).
245. H. Miguez, F. Meseguer, C. Lopez, M. Holgado, G. Andreasen, A. Mifsud, and V. Fornes, *Langmuir* 16, 4405 (2000).
246. S. H. Park and Y. Xia, *Chem. Mater.* 10, 1745 (1998).
247. K. Busch and S. John, *Phys. Rev. Lett.* 83, 967 (1999).
248. O. D. Velev, T. A. Jede, R. F. Lobo, and A. M. Lenhoff, *Chem. Mater.* 10, 3597 (1998).
249. H. Miguez, C. Lopez, F. Meseguer, A. Blanco, L. Vazquez, R. Mayoral, M. Ocana, V. Fornes, and A. Mifsud, *Appl. Phys. Lett.* 71, 1148 (1997).
250. J. F. Bertone, P. Jiang, K. S. Hwang, D. M. Mittleman, and V. L. Colvin, *Phys. Rev. Lett.* 83, 300 (1999).
251. M. Holgado, F. Garcia-Santamaria, A. Blanco, M. Ibisate, A. Cintas, H. Miguez, C. J. Serna, C. Molpeceres, J. Requena, A. Mifsud, F. Meseguer, and C. Lopez, *Langmuir* 15, 4701 (1999).
252. G. Carotenuto, X. Xuejun, and L. Nicolais, *Polym. News* 25, 6 (2000).
253. E. L. Hu and D. T. Shaw, in "Nanostruct. Science and Technology" (R. W. Siegel, Ed.), Vol., pp. 15–33. Kluwer Academic, Dordrecht, 1998.
254. M. Zhao and R. M. Crooks, *Chem. Mater.* 11, 3379 (1999).
255. C. R. Martin, *Chem. Mater.* 8, 1739 (1996).
256. J. E. G. J. Wijnhoven and W. L. Vos, *Science* 281, 802 (1998).
257. S. H. Joo, S. J. Choi, I. Oh, J. Kwak, Z. Liu, O. Terasaki, and R. Ryoo, *Nature* 412, 169 (2001).
258. Y. Wang and N. Herron, *J. Phys. Chem.* 91, 257 (1987).
259. F. Z. Cui, Y. Zhang, and Q. Cai, in "Mater. Res. Soc. Symp. Proc.," 2000.
260. M. P. Pileni, Ber. Bunsenges. *Phys. Chem. Chem. Phys.* 101, 1578 (1997).
261. J.-L. Luna-Xavier, A. Guyot, and E. Bourgeat-Lami, *J. Colloid Interface Sci.* 250, 82 (2002).
262. J. L. Luna-Xavier, E. Bourgeat-Lami, and A. Guyot, *Colloid Polym. Sci.* 279, 947 (2001).
263. V. Pillai, P. Kumar, M. J. Hou, P. Ayyub, and D. O. Shah, *Adv. Colloid Interface Sci.* 55, 241 (1995).
264. S. Hingorani, V. Pillai, P. Kumar, M. S. Multani, and D. O. Shah, *Mater. Res. Bull.* 28, 1303 (1993).
265. C. Petit, A. Taleb, and M. P. Pileni, *Adv. Mater.* 10, 259 (1998).
266. A. B. R. Mayer, *Polym. Adv. Tech.* 12, 96 (2001).
267. R. L. Axelbaum, D. P. Defaux, C. A. Frey, and S. M. L. Sastry, *Metall. Mater. Trans. B* 28, 1199 (1997).
268. S. E. Pratsinis, C. Artelt, S. Sanders, and P. T. Spicer, *J. Aerosol Sci.* 28, S753 (1997).
269. N. P. Rao, H. J. Lee, M. Kelkar, D. J. Hansen, J. V. R. Heberlein, P. H. McMurry, and S. L. Girshick, *Nanostruct. Mater.* 9, 129 (1997).
270. J. Costa, in "Handbook of Nanostructured Materials and Nanotech.: Synthesis and Processing" (H. S. Nalwa, Ed.), Vol. 1, pp. 57–158. Academic Press, San Diego, 2000.
271. B. H. Kear and G. Skandan, *Nanostruct. Mater.* 8, 765 (1997).
272. A. Piciacchio, S.-H. Lee, and G. L. Messing, *J. Am. Ceram. Soc.* 77, 2157 (1994).
273. J. F. de la Mora and L. de Juan, in "210th ACS National Meeting," Chicago, 1995.
274. H. Jiang, M. Lau, V. L. Telkamp, and E. J. Lavernia, in "Handbook of Nanostructured Materials and Nanotech.: Synthesis and Processing" (H. S. Nalwa, Ed.), Vol. 1, pp. 159–209. Academic Press, San Diego, 2000.
275. J. Karthikeyan, C. C. Berndt, J. Tikkanen, S. Reddy, and H. Herman, *Mater. Sci. Eng. A* 238, 275 (1997).
276. R. Uyeda, *Progr. Mater. Sci.* 35, 1 (1991).
277. S. K. Friedlander, in "WTEC Workshop Report on R&D Status and Trends in Nanoparticles, Nanostructured Materials, and Nanodevices in the United States," Arlington, 1998.
278. M. Singhal, V. Chhabra, P. Kang, and D. O. Shah, *Mater. Res. Bull.* 32, 239 (1997).
279. K. S. Suslick, T. Hyeon, M. Fang, and A. A. Cichowlas, *Adv. Catal. Nanostructured Mater.* 197 (1996).
280. H. Coelfen and M. Antonietti, *Langmuir* 14, 582 (1998).
281. M. P. Pileni, A. Taleb, and C. Petit, *J. Dispersion Sci. Tech.* 19, 185 (1998).
282. N. Duxin, N. Brun, C. Colliex, and M. P. Pileni, *Langmuir* 14, 1984 (1998).
283. L. M. Bronstein, O. A. Platonova, A. N. Yakunin, I. M. Yanovskaya, P. M. Valetsky, A. T. Dembo, E. E. Makhaeva, A. V. Mironov, and A. R. Khokhlov, *Langmuir* 14, 252 (1998).
284. W. Stoeber, A. Fink, and E. Bohn, *J. Colloid Interface Sci.* 26, 62 (1968).
285. V. Chhabra, V. Pillai, B. K. Mishra, A. Morrone, and D. O. Shah, *Langmuir* 11, 3307 (1995).
286. S.-Y. Kwak, S. H. Kim, and S. S. Kim, *Env. Sci. Tech.* 35, 2388 (2001).
287. H. Mattoussi, G. P. Anderson, J. M. Mauro, M. G. Bawendi, and V. C. Sundar, Inorganic particle conjugates, in "PCT Int. Appl.," Massachusetts Institute of Technology, Naval Research Laboratory, 2001. p. 48.
288. J. Cizeron and M. P. Pileni, *J. Phys. Chem. B* 101, 8887 (1997).
289. L. Levy, N. Feltin, D. Ingert, and M. P. Pileni, *J. Phys. Chem. B* 101, 9153 (1997).
290. R. P. Bagwe and K. C. Khilar, *Langmuir* 13, 6432 (1997).
291. V. Buschmann, G. Van Tendeloo, P. Monnoyer, and J. B. Nagy, *Langmuir* 14, 1528 (1998).
292. J. Eastoe and B. Warne, *Curr. Opin. Colloid Interface Sci.* 1, 800 (1996).
293. L. A. Estroff and A. D. Hamilton, *Chem. Mater.* 13, 3227 (2001).
294. Q. Zhou, S. Wang, X. Fan, J. Mays, R. Advincula, G. Sakellariou, S. Pispas, and N. Hadjichristides, *Polym. Preprint ACS Div. Polym. Chem.* 42, 59 (2001).
295. S. F. Lascelles, G. P. McCarthy, M. D. Butterworth, and S. P. Armes, *Colloid Polym. Sci.* 276, 893 (1998).
296. Q. Zhou, S. Wang, X. Fan, R. Advincula, and J. Mays, *Langmuir* 18, 3324 (2002).
297. S. C. Farmer and T. E. Patten, *Chem. Mater.* 13, 3920 (2001).
298. M. Li, H. Schnablegger, and S. Mann, *Nature* 402, 393 (1999).
299. W. F. C. Sager, *Curr. Opin. Colloid Interface Sci.* 3, 276 (1998).
300. S. G. Dixit, A. R. Mahadeshwar, and S. K. Haram, *Colloid Surface A* 133, 69 (1998).
301. R. H. Kore, J. S. Kulkarni, and S. K. Haram, *Chem. Mater.* 13, 1789 (2001).
302. J. Pyun and K. Matyjaszewski, *Chem. Mater.* 13, 3436 (2001).
303. S. Aggarwal, S. B. Ogale, C. S. Ganpule, S. R. Shinde, V. A. Novikov, A. P. Monga, M. R. Burr, R. Ramesh, V. Ballarotto, and E. D. Williams, *Appl. Phys. Lett.* 78, 1442 (2001).
304. C. Sanchez and F. Ribot, *New J. Chem.* 18, 1007 (1994).
305. A. Bhattacharya, K. M. Ganguly, A. De, and S. Sarkar, *Mater. Res. Bull.* 31, 527 (1996).

306. J. Stejskal, P. Kratochvil, S. P. Armes, S. F. Lascelles, A. Riede, M. Helmstedt, J. Prokes, and I. Krivka, *Macromolecules* 29, 6814 (1996).
307. C. Barthet, A. J. Hickey, D. B. Cairns, and S. P. Armes, *Adv. Mater.* 11, 408 (1999).
308. J. I. Amalvy, M. J. Percy, S. P. Armes, and H. Wiese, *Langmuir* 17, 4770 (2001).
309. M. J. Percy, C. Barthet, J. C. Lobb, M. A. Khan, S. F. Lascelles, M. Vamvakaki, and S. P. Armes, *Langmuir* 16, 6913 (2000).
310. F. Tiarks, K. Landfester, and M. Antonietti, *Langmuir* 17, 5775 (2001).
311. G. Boven, M. L. C. M. Oosterling, G. Challa, and A. J. Schouten, *Polymer* 31, 2377 (1990).
312. M. J. Percy and S. P. Armes, *Langmuir* 18, 4562 (2002).
313. Z. Pu, J. E. Mark, J. M. Jethmalani, and W. T. Ford, *Chem. Mater.* 9, 2442 (1997).
314. J. M. Jethmalani and W. T. Ford, *Chem. Mater.* 8, 2138 (1996).
315. J. M. Jethmalani, W. T. Ford, and G. Beaucage, *Langmuir* 13, 3338 (1997).
316. M. S. Fleming, T. K. Mandal, and D. R. Walt, *Chem. Mater.* 13, 2210 (2001).
317. M. Gianini, W. R. Caseri, and U. W. Suter, *J. Phys. Chem. B* 105, 7399 (2001).
318. Y. Dirix, C. Bastiaansen, W. Caseri, and P. Smith, *Adv. Mater.* 11, 223 (1999).
319. K. S. Iyer, J. Moreland, I. Luzinov, S. Malynych, and G. Chumanov, in "224th ACS National Meeting," Boston, 2002.
320. W. Heffels, C. Bastiaansen, W. Caseri, and P. Smith, *Molec. Cryst. Liq. Cryst. Sci. Tech. A* 353, 191 (2000).
321. M. E. Garcia, L. A. Baker, and R. M. Crooks, *Anal. Chem.* 71, 256 (1999).
322. E. Hao and T. Lian, *Mater. Res. Soc. Symp. Proc.* 648, 6.17/1 (2001).
323. M. Zhao, L. Sun, and R. M. Crooks, *J. Am. Chem. Soc.* 120, 4877 (1998).
324. F. Jones, H. Colfen, and M. Antonietti, *Colloid Polym. Sci.* 278, 491 (2000).
325. K. E. Gonsalves, H. Li, and P. Santiago, *J. Mater. Sci.* 36, 2461 (2001).
326. F. M. Pavel and R. A. Mackay, *Langmuir* 16, 8568 (2000).
327. B. Simmons, S. Li, V. T. John, G. L. McPherson, C. Taylor, D. K. Schwartz, and K. Maskos, *Nano. Lett.*, in press.

Self-Assembled Porphyrin Arrays

Kazuya Ogawa, Yoshiaki Kobuke

CREST of Japan Science and Technology Corporation, Ikoma, Nara, Japan

CONTENTS

1. Introduction
2. Self-Organization via Coordination to Outer Metal Ions
3. Self-Assembled Porphyrins by Axial Coordination
4. Self-Assembled Porphyrins by Hydrogen Bonds
5. Spectroscopic Characterization of Self-Assembled Porphyrin Arrays
6. Properties of Self-Assembled Porphyrin Arrays
7. Summary
Glossary
References

1. INTRODUCTION

Porphyrin [1–3] is a macrocyclic molecule consisting of four pyrrolic rings bridged by four methine groups. The pyrrolic and methine CH groups are usually called as β- and *meso*-positions, respectively, as shown in Figure 1A, and various substituents can be introduced into these positions in place of hydrogen atoms. Porphyrin is a highly conjugated 18π-electron system, which causes decrease of the HOMO-LUMO bandgap energy, and therefore it is interesting in view of visible-light absorbing and emitting materials as desirable for electronic and photonic applications as well as artificial photosynthetic building units. Another characteristic feature is that porphyrin can incorporate various metal ions in the central core surrounded by four pyrrolic nitrogen atoms to provide metalloporphyrins (Fig. 1A). In many cases, the central metal ion such as zinc, magnesium, iron, cobalt, and so on, can accept fifth or sixth ligation (denoted as L_n in Fig. 1B) at axial position(s) to allow construction of desirable supramolecular porphyrin architectures. The porphyrin skeleton is found in biological systems [4–15] such as chlorophylls or hemes, which play important roles such as light-harvesting, energy and electron transfer to convert the light energy to chemical energy in photosynthesis [4–12], transport and activation of molecular oxygen, and catalysis of oxidation-reduction reactions [13–15].

Supramolecular chemistry [16–20] has developed rapidly since the 1980s. Even complicated structures aiming novel functions have been successfully assembled from simple molecular units by the use of noncovalent interactions such as hydrogen bonding, ligand-to-metal coordination, hydrophobic, van der Waals interactions, and π-π interaction between aromatic molecules as well. These organization processes involving thermodynamic equilibria between simple units are called "self-assembly" or "self-organization." The most important feature of the supramolecular approach is that small and simple subunits, which can be easily synthesized and do not show any significant functions by themselves, are automatically assembled to produce macromolecular systems exhibiting desirable functions. The self-assembly plays important roles especially in biological systems and finds so many relevant examples. Most typically, complementary hydrogen bond formation is the source of gene replication operating in DNA/RNA molecular recognitions [13–15]. Three-dimensional structure formation of polypeptides is based on their hydrogen-bonded α-helices and β-sheets [13–15]. Hydrophobic interactions are the fundamental driving force of substrate binding to the active site of many enzymes. Nature also uses coordination bonds as the strongest among intermolecular forces. In the photosynthesis system, chlorophylls are fixed to membrane proteins by coordination of histidine-imidazole to their central metal to construct the light-harvesting antenna complex [4–6] and the special pair of chlorophylls [7–12].

Self-assembly allows the construction of complicated large molecular system even in the porphyrin chemistry, in which sometimes it may be difficult to synthesize such a complex system by using covalent linking. In the 1990s, many ideas of self-assembled porphyrins [20] have been proposed to construct mostly the light-harvesting antenna complex and the special pair models of photosynthesis and to reveal the relationship between their structures and biological roles. Although supramolecular approaches to synthesize such porphyrin systems contain noncovalent interactions as described, most systems have employed the idea of

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (561–591)

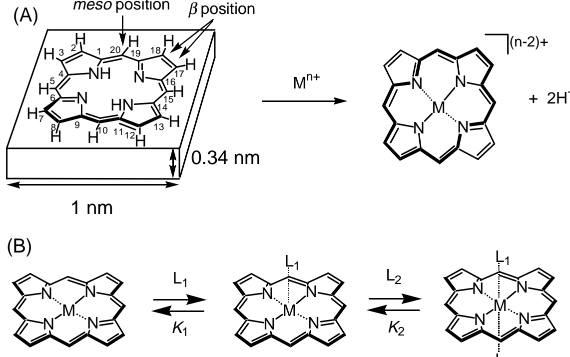


Figure 1. (A) Schematic representation of porphyrin and metalloporphyrin. (B) Stepwise complexation of metalloporphyrin with axial ligands (L_n). K_n is defined as stepwise equilibrium constants.

ligand coordination to the outer metal or central metal of porphyrin. Hydrogen-bonded porphyrin assemblies have also been reported. Since porphyrins are usually dissolved in nonpolar solvent such as chloroform, benzene, toluene, and so on, linking of porphyrins by hydrogen bond is more effective and stable in such nonpolar media. Other interactions such as hydrophobic, van der Waals, and π-π interactions cannot be main driving forces, but may be used secondarily or cooperatively to stabilize the porphyrin assembly.

Recent attentions are also directed to nanometer-scale molecular fabrication to push forward the idea of molecular electronic and photonic devices [21–25]. Since miniaturization of LSI built by metal-oxide-semiconductor (MOS) is close to its absolute limit [24], a molecular-scale device, which may provide high-performance computers [25], is now definitely required. One of the important characteristics required for the material to build up such a device may be highly conductive organic molecules such as carbon nanotube [26, 27] and multi-porphyrin array [28]. Since the size of porphyrin molecule is around 1 nm square with a 0.34-nm height, nanometer-scaled architecture of technological requirements can be built up by assembling less than 100 porphyrin building blocks.

At the same time, third-order nonlinear optical (NLO) materials have also attracted attention because of potential utilities for photonic applications such as ultrafast optical switching and modulations [29–31]. Various types of organic compounds have so far been designed and synthesized to obtain materials with large third-order susceptibilities $\chi^{(3)}$ [32–34]. Porphyrins are one of the most promising candidates for such NLO materials in the view not only of the largest class of the values attained hitherto, but of allowance of versatile structural modifications and variations of central metal ions to design ideal properties [28, 35–43].

As mentioned earlier, porphyrin has great potential to demonstrate excellent properties as optical and electronic materials toward the goal of single-molecular electronics. Efficient conversion of light energy to chemical or electric energy is certainly the most fascinating scientific challenge to be met in this century. Porphyrin as a functional mimic of chlorophyll should play central roles in elucidating the mechanism of photosynthesis as well. In order to achieve these targets, large numbers of chemists have been trying to obtain multi-porphyrin systems based on covalent and supramolecular approaches. In this review, we will introduce methodologies and recent advances of self-assembled multi-porphyrin systems by the use of noncovalent interactions, especially focused on the self-assembly by introducing unique substituents to the central porphyrin matrix in the past few years.

2. SELF-ORGANIZATION VIA COORDINATION TO OUTER METAL IONS

Coordination of *meso*-substituted ligand, such as pyridyl, 2,2′-bipyridyl, phenanthroline, terpyridyl, crown ethers, and 8-hydroxyquinoline (oxine) as shown in Figure 2, to outer metal ion(s), other than the central metal ion of porphyrin, has been employed in many literatures to construct supramolecular architectures. Geometries of porphyrins in the resulting supramolecule are determined by the coordination structure around the outer metal ion and can be controlled by adjusting the coordination direction of ligand to the metal ion. In general, the ligand used in constructing a supramolecular porphyrin array contains electron-donating atom such as oxygen, nitrogen, and sometimes sulfur, whose affinity to the metal ion is determined by the rule of hard soft acid base relation. In this section, self-assembly by complexation of the ligand of porphyrin substituent with outer metal ions will be introduced.

2.1. Coordination of Pyridyl Ligand to Outer Metal Ion

Supramolecular formation by pyridyl coordination to Pd(II) or Pt(II) has been extensively developed by Fujita [44–46] to varieties of macrocyclic compounds. Porphyrin-containing pyridines behaved similarly and multi-porphyrin macrocycles were reported along the line. One of the examples using *meso*-substituted 4-pyridyl porphyrin is shown in Figure 3 [47]. When porphyrin **1** having two 4-pyridyl substituents at 5,10-positions reacted with one equivalent of *cis*-$Pd(NCPh)_2Cl_2$, the pyridyl ligand coordinated to Pd ion by exchanging PhNC to form dimer **2**. On the other hand, reaction with *trans*-$Pd(NCPh)_2Cl_2$ provided tetrameric square **3** with four porphyrins at the corners of the square. Since Pd(II) or Pt(II) prefers square planar arrangement with four ligands and PhNC is more labile than chloride

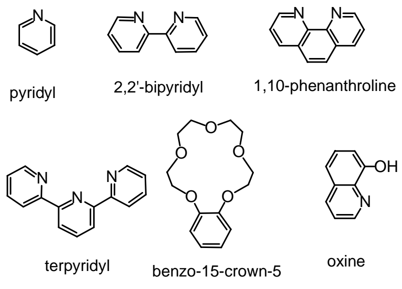


Figure 2. Structures of ligand for coordination to outer metal ion.

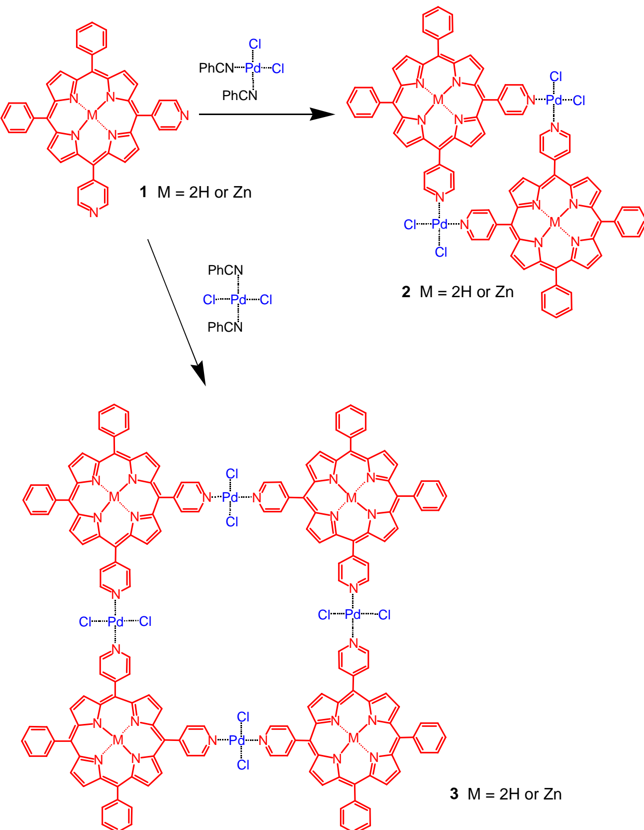


Figure 3. Formation of porphyrin dimer **2** and tetramer **3** from 5,10-bis(4-pyridyl)porphyrin **1** with *cis*-$Pd(NCPh)_2Cl_2$ and *trans*-$Pd(NCPh)_2Cl_2$, respectively [47].

or pyridyl, *cis*- and *trans*-$Pd(NCPh)_2Cl_2$ afford two respective supramolecules, **2** and **3**. Contrary to **3**, 5,15-bis(4-pyridyl)porphyrin **4**, in which two pyridyls are placed at two facing *meso*-positions, gave different square organization **5**, as shown in Figure 4 [47]. In this case, porphyrins and *cis*-$PdCl_2$ parts occupy the edges and the corners, respectively.

The square made by 5,15-bis(4-pyridyl)porphyrinatozinc and $Re(CO)_3Cl$ incorporated a 5,10,15,20-tetrakis(4-pyridyl)porphyrin (H_2TPyP) by axial coordinations of four pyridyl groups to central zinc ions (Fig. 5) [48]. Formation of pentamer **6** was confirmed by fluorescence measurement with an association constant of 4×10^7 M^{-1} in CH_2Cl_2. The fluorescence quenching of zincporphyrin moiety (~90%) and the growth of H_2TPyP fluorescence with the corresponding time scale were observed, indicating the occurrence of energy transfer from zinc moiety to H_2TPyP within the complex **6**.

The square porphyrin nonamer **7** (Fig. 6) was synthesized from a mixture of *trans*-$[PdCl_2(NCPh)_2]$ and three different porphyrins: one central X-shaped unit (TPyP), four T-shaped units on the edges (tris(4-pyridyl)porphyrin), and four L-shaped units at the corners (5,10-bis(4-pyridyl)porphyrin) [49]. The nonameric array was formed in 90% yield at room temperature in a dilute chloroform solution (total concentration below 10 mM). The nonameric structure was supported by ^1H NMR titration, light-scattering measurement with a radius of 5 ± 7 nm,

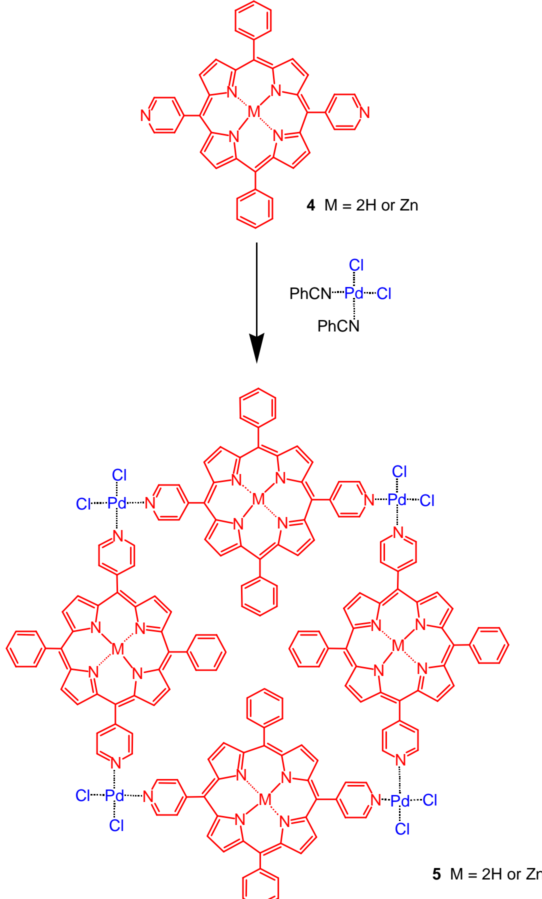


Figure 4. Formation of porphyrin tetramer **5** from 5,15-bis(4-pyridyl)-porphyrin **4** with *cis*-$Pd(NCPh)_2Cl_2$ [47].

side view of porphyrin plane

Figure 5. Porphyrin pentamer **6** [48].

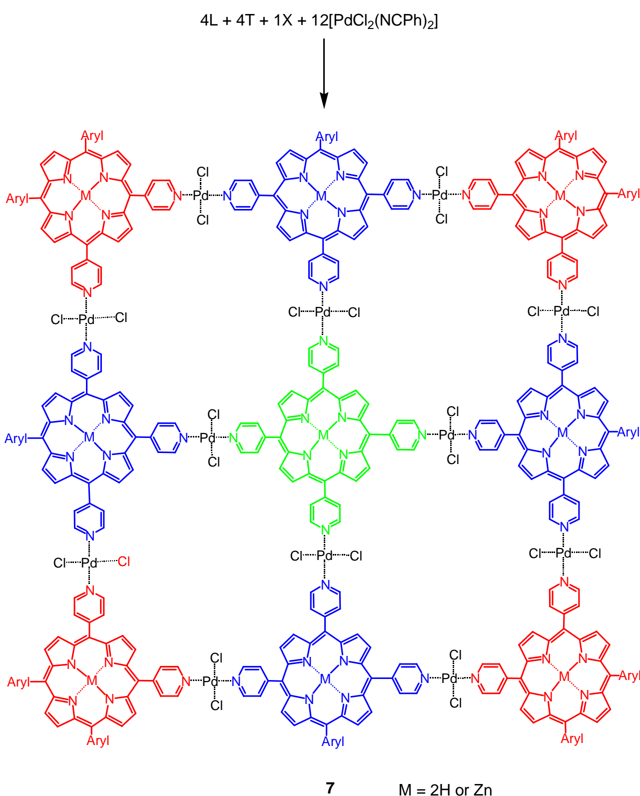


Figure 6. Formation of porphyrin nonamer **7** [49].

size-exclusion chromatography, vapor phase osmometry (VPO), and electrospray mass (ESI-MS) spectra with tricationic and tetracationic molecular ions. The fluorescence of a stoichiometric mixture of porphyrins without Pd(II) was quenched over 90% by titration of Pd(II) salt to form the nonamer **7**. The fluorescence lifetime measurements also showed that the average lifetime of a mixture of porphyrins without Pd(II) in toluene was 12 ns and that of the array was less than 1 ns due to the heavy-atom effect. Porphyrin tapes **8** and **9** (Fig. 7) were also synthesized in a similar manner to that of **7**. Addition of two equivalents of O-shaped mono(4-pyridyl)porphyrin to one equivalent of I-shaped 5,15-bis(4-pyridyl)porphyrin two equivalents of *trans*-[$PdCl_2(NCPh)_2$] gave Pa(II)-assisted tape **8** as a mixture of dimer, trimer, and tetramer. Tape **9** was made by mixing X-, L-shaped porphyrins, and *trans*-[$PtCl_2(NCPh)_2$] (2:2:6) for 4–5 days at 40 °C as a mixture of trimer, tetramer, and pentamer. These formations were confirmed by ESI-MS measurement. There have been some reports on supramolecular porphyrins with similar dimeric or tetrameric structure by using 4-pyridyl ligand and different metal sources, such as Pd(dppp), Pt(dppp), Ru(II), and so on, where dppp denotes diphenylphosphine [50–58].

Recently, Fujita et al. reported porphyrin trimer **11** constructed by complexation of 5,10,15,20-tetrakis(3-pyridyl)porphyrin **10** with Pd(II) ion (Fig. 8) [59]. The reaction of porphyrin **10** with two equivalents of [$Pd(en)(ONO_2)_2$] in H_2O/MeCN at 80 °C for one day allowed a purple solution of **11**, which was isolated in a 96% yield by adding acetone to the solution. The trimeric structure maintained in a MeCN/H_2O (1:1) solution as

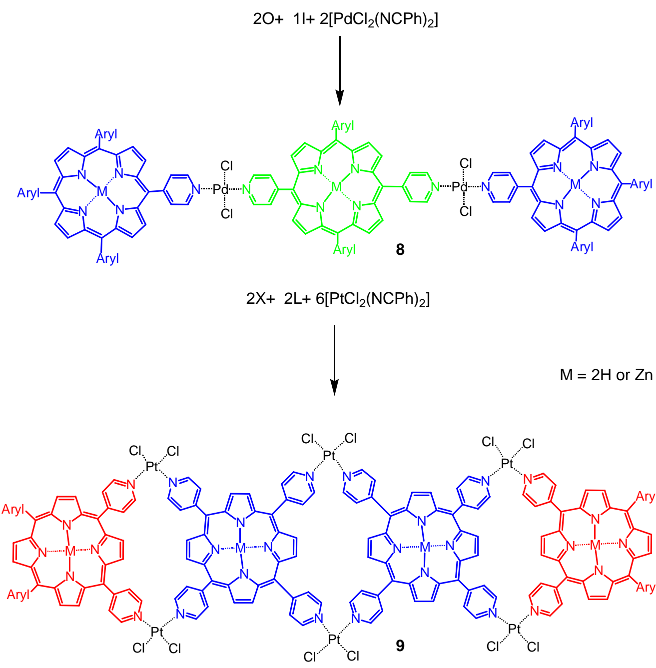


Figure 7. Self-assembled porphyrin tapes **8** and **9** [49].

supported by the evidences of ^{1}H NMR and cold spray-ionization mass (CSI-MS) spectroscopic analyses. One NO_3^- ion could not be exchanged by PF_6^- ion even with a large excess of NH_4PF_6 and was trapped by the trimer supported by mass spectroscopic data. X-ray crystallography revealed the prism structure of **11**, in which six Pd(II) ions were located at apical positions of the prism. The axial-to-equatorial flipping of two Py-Pd-Py sites at a diagonal position allowed the formation of a cavity, which would be expected to incorporate a small organic guest such as pyrene in an aqueous solution.

2.2. Chelate Formation to Give Stable Assembly

Chelating complexations of 2,2′-bipyridyl, 1,10-phenanthrolinyl, and terpyridyl with specific metal ions provide a method to fix the porphyrin geometry with large stability constants. The orientation of the porphyrins directly linked by the chelating substituent should also depend on the geometry around the chelated metal ion. Porphyrin dimer **12**, in which two porphyrin units were connected directly to symmetrical 4,4′-positions of the bipyridyl group, was designed and synthesized (Fig. 9) [60].

Figure 8. Porphyrin trimer **11** given from 5,10,15,20-tris(3-pyridyl)porphyrin **10** and *cis*-$Pd(en)(ONO_2)_2$ (en = ethylenediamine) [59].

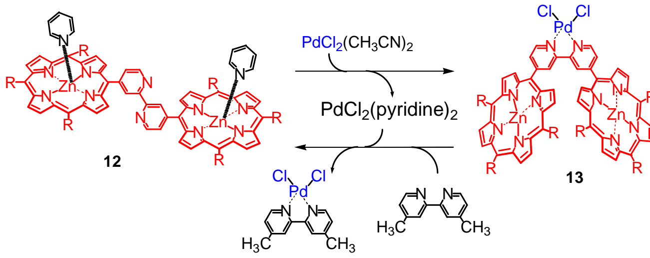


Figure 9. Reversible conversion between **12** and **13** by Pd(II) ion and 2,2′-bipyridyl [60].

Addition of 2.2 eq. of $PdCl_2$ to **12** converted the freely rotating bipyridyl conformation to the cofacial one giving rise to the formation of bisporphyrin **13** with a dihedral angle of 60°. The subsequent addition of 4,4′-dimethyl-2,2′-bipyridine regenerated back the initial bisporphyrin **12**. Spatial geometries of two porphyrins could be regulated by reversible complexation of the bipyridyl part with $PdCl_2$.

Porphyrin dimer bridged by 1,10-phenanthroline moiety has been synthesized and reacted with Cu(I) ion to afford self-assembled porphyrin tetramer **14** with a tetrahedral geometry (Fig. 10) [61]. Different metal ions Zn(II) and Au(III) were inserted into two porphyrin centers of the dimer to modulate the electron transfer rate, where, first, one Au(III) ion was inserted to synthesize mono-metalated complex and then Zn(II) ion was inserted. The electron transfer reaction was investigated by transient absorption measurement by using 30-ps pulse laser in dimethylformamide (DMF). The rate of forward electron transfer reaction from Zn(II) porphyrin to Au(III) porphyrin in tetrameric complex **14** was accelerated by 18 times relative to the uncomplexed dimer due to the operation of superexchange mechanism of the intervening Cu(I) phenanthroline complex.

2,9-Diphenyl-1,10-phenanthroline derivatives have also been used to synthesize catenanes and rotaxanes [16–19]. Compound **19**, the first report of porphyrin-containing catenane, is represented in Figure 11 [62]. The first ring **17**

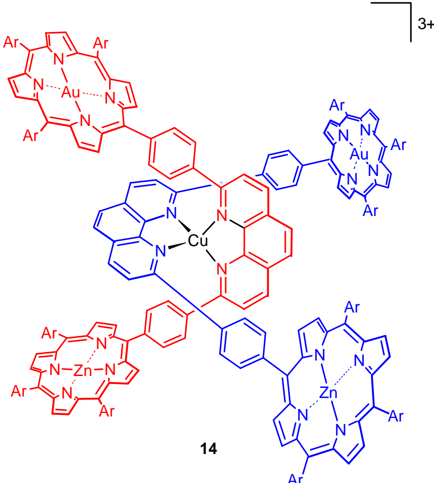


Figure 10. Porphyrin tetramer **14** by coordination of 1,10-phenanthroline, which bridges Zn(II)-porphyrinyl and Au(III)-porphyrinyl to Cu(I) ion [61].

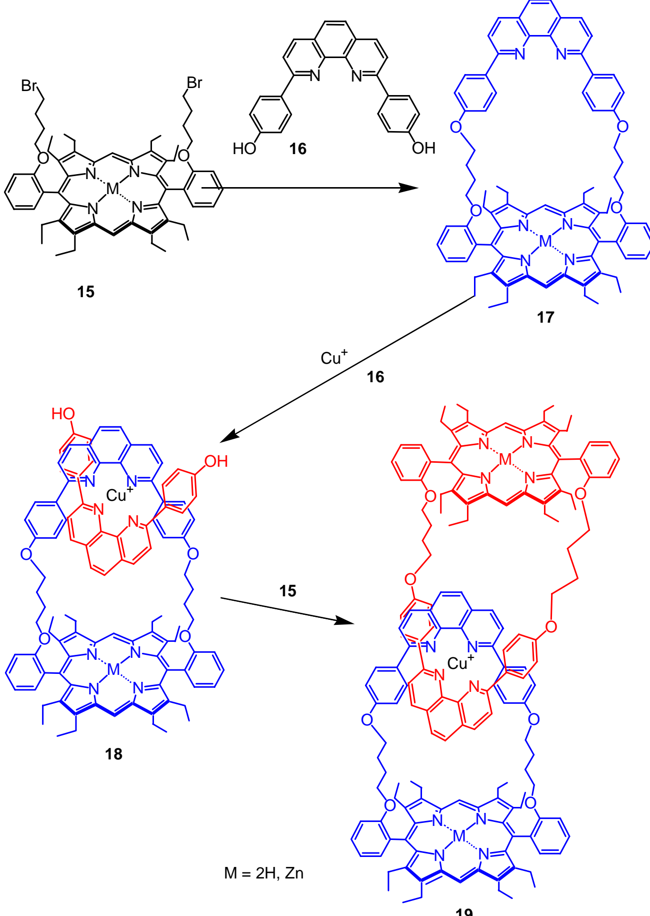


Figure 11. Synthetic route of catenane **19** [62].

was incorporated with the second open chain component **16** through the coordination interaction with Cu(I) ion to give a precursor **18**. Further coupling of compounds **18** with **15** gave rise to the formation of catenane **19** in a 33% yield, in which two porphyrin units were interlocked by complexation of two phenanthroline moieties with Cu(I) ion. The absorption spectra of **19** resembled those of monomeric species and indicated the absence of electronic interactions between two porphyrins of a large spatial separation distance.

Sauvage has reported several asymmetric catenanes and rotaxanes containing a Zn(II) porphyrin unit as an electron donor and an Au(III) porphyrin unit as an electron acceptor to investigate long-range electron transfer reactions. Catenane **24**, which consisted of Au(III) porphyrin and Zn(II) porphyrin units, was synthesized according to a strategy similar to the case of **19** with modified connection units, as shown in Figure 12 [63]. In this case, too, no significant interaction between two porphyrins was observed in the absorption spectrum, but highly efficient fluorescence quenching of Zn(II) porphyrin was observed. This quenching may be responsible for the electron transfer from Zn(II) porphyrin to Au(III) porphyrin, although the detailed mechanism has not been reported yet.

The construction of rotaxane-like threaded porphyrin assembly **26** was reported by the same author and co-workers (Fig. 13) [64]. Reaction of an equimolar mixture of

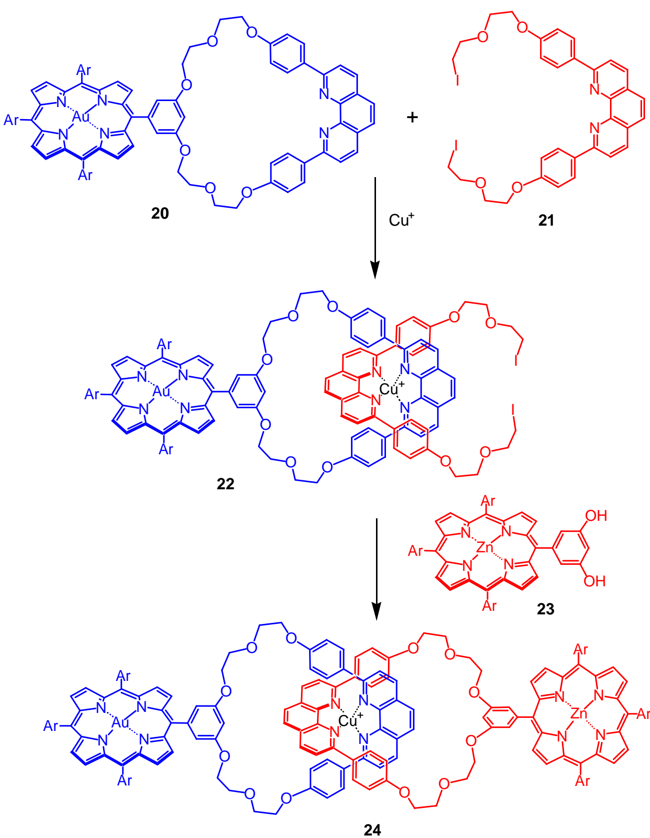


Figure 12. Synthetic route of catenane **24** [63].

Zn(II) and Au(III) porphyrins each bearing phenanthroline moiety in a closed ring, and string component **25** gave a mixture of three porphyrin species: Zn(II)-Zn(II), Au(III)-Au(III), and Zn(II)-Au(III) in the ratio of 2:1:1. The desirable hetero-complex Zn(II)-Au(III) **26** was separated by chromatography and no scrambling was observed in a solution at room temperature. When threading was undertaken for Zn(II) porphyrins, the isolated yields of Zn(II)-Zn(II) complex depended much on the structure of the connecting group L of **25**, being 6, 60, and 95% for trimethylene, 1,2-phenylene, and 4,4′-diphenylene ether, respectively.

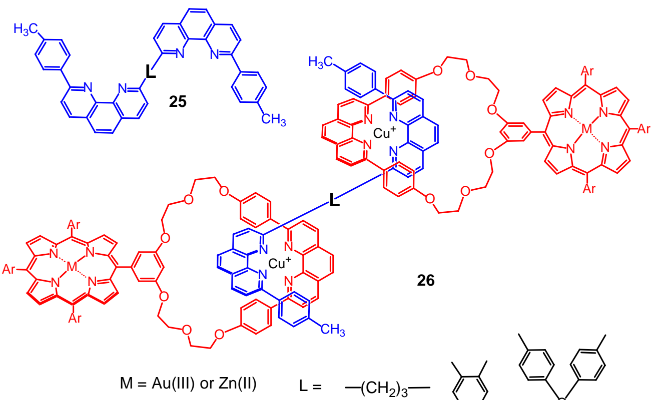


Figure 13. Structure of threaded self-assembly **26** [64].

The same author and co-workers have reported a real rotaxane **29** without fixation by metal coordination [65]. Two terminals of the string were connected to Zn(II) porphyrins, and a ring part, which was threaded by the string, had a Au(III) porphyrin, as shown in Figure 14. Reaction of precursor **27** having two formyl groups with benzaldehyde and dipyrromethane followed by zinc insertion gave **28**. Finally, **29** was obtained by the removal of Cu(I) ion from **28** by treatment with KCN. Steady-state fluorescence measurement showed strong fluorescence quenching of the Zn(II) porphyrin units by 87% indicating electron transfer from the Zn(II) porphyrin to the Au(III) porphyrin. The electron transfer rate was estimated roughly as 3.5×10^9 s^{-1}, which was slower than that of the covalently linked Au(III) and Zn(II) bisporphyrin by 2,9-diphenyl-1,10-phenanthroline [66] due to the lack of covalent bonds between porphyrins in **29**. In the case of rotaxane **30**, the distance between donor porphyrins and acceptor porphyrins was controlled by reversible complexation and decomplexation reactions at the phenanthroline units to generate long distance and close proximity orientations, as shown in Figure 15 [67]. The conformation of **30** was confirmed by observation of NOE between H5,6 and Hpy showing a proximity between the rear part of the 1,10-phenanthroline moiety bridging two zinc porphyrin units and the gold porphyrin

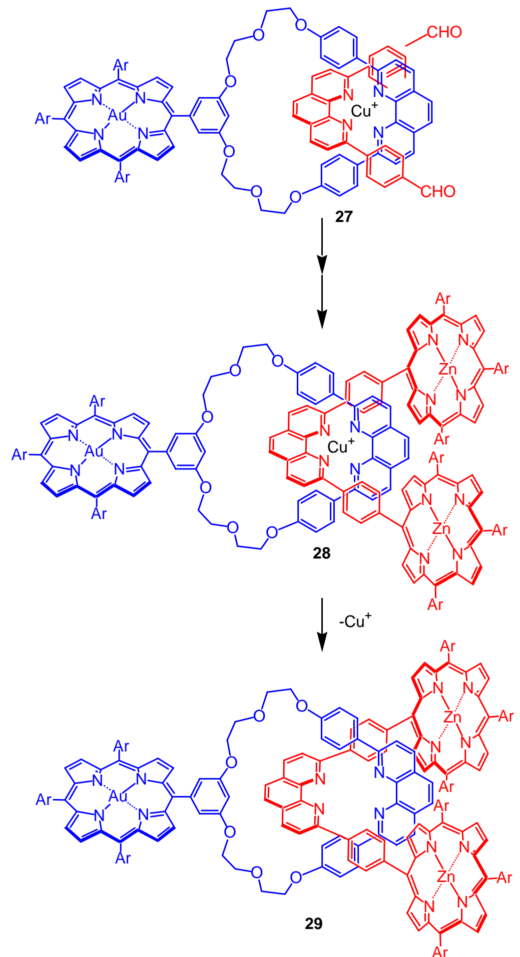


Figure 14. Synthetic route of rotaxane **29** [65].

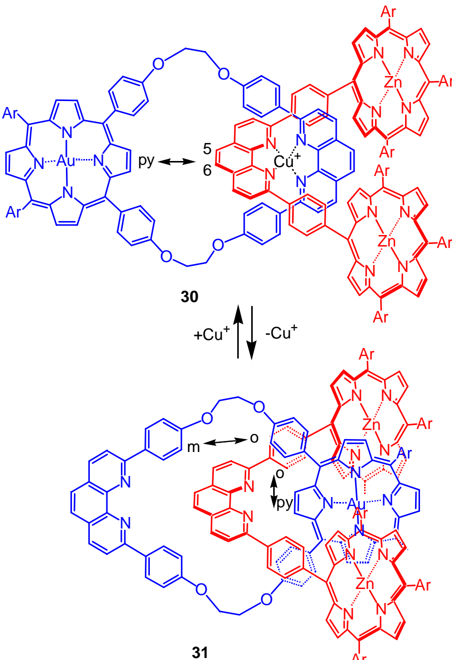


Figure 15. Control of the arrangement between gold porphyrin and zinc porphyrins by complexation and decomplexation at phenanthroline sites [67].

Figure 16. Linear porphyrin dimers assembled by chelate complexation of 1,10-phenanthrolines (**32** [68] and **33** [69]) and terpyridyl (**34** [70–72]) to outer metal ions.

part. On the other hand, NOE measurement of the decomplexed form **31** demonstrated the close proximity between Hm and Ho as well as Hpy and Ho′, as drawn schematically in Figure 15. The estimated center-to-center (Au-Zn) separations from space-filling model are 19 and 7 Å for **30** and **31,** respectively. Interesting photochemical properties are expected, but not examined.

Another example of self-assembly via chelate formation of 1,10-phenanthroline with tetrahedral Cu(I) or Zn(II) ion to form porphyrin dimers (**32, 33**) is shown in Figure 16. [68, 69]. In the case of **33** [69], different metal ions such as Ni(II), Cu(I/II), Zn(II), and Li(I) were employed for the complexation of the phenanthroline moiety and the changes in the absorption spectra were measured. The binding of all metal ions at the phenanthrolinic moiety perturbed the porphyrin π-system to give red-shifts of all peaks. The Soret and Q bands[1] of Zn(II) and Cu(II) porphyrins were much more sensitive to the variation of external metal ion bindings than those of free base porphyrin.

Terpyridyl is a tridentate ligand with three pyridyl nitrogens and is known to form very stable complexes with hexacoordinating metal ions such as Ru(II). Sauvage has reported three chromophore complexes containing free base or Zn(II) porphyrin, bis(terpyridine)Ru(II), and Au(III) porphyrin parts [70–72]. For example, triad **34** was synthesized to investigate the energy transfer reaction between them, as shown in Figure 16 [72]. Terpyridyl substituted Au(III) porphyrin, por(Au), was reacted with $RuCl_3$ followed by $AgBF_4$ to form a por(Au)-Ru(terpyridyl)Cl_2 complex, which was then converted to **34** by the reaction with terpyridyl substituted free base porphyrin. Photoinduced energy transfer from the triplet state of Au(III) porphyrin moiety to the free base porphyrin site (center-to-center distance of 2.1 nm) occurred in glassy media at low temperature with a rate constant of 2.5×10^7 s^{-1}. The energy transfer process was mediated by the central bis(terpyridine)Ru(II) moiety and was explained by an electron exchange mechanism [72].

Oxine is known to form extremely stable, neutral, and rigid complexes with metal ions such as Al(III) and Ga(III) by assembling three oxine units [73, 74]. *Meso*-oxinyl substituted porphyrin **35** afforded rigid gallium-tris(oxinyl) complex **37**, where three porphyrins were arranged like a pinwheel, as shown in Figure 17 [75, 76]. This minimum unit was extended to a dendritic structure **38** by introducing another oxinyl group at the facing *meso*-position. Molecular mechanics calculation (with MM+ force field) deduced the most stable configuration and the center-to-center separation distances between two porphyrin units were estimated as 15, 17 and 18.5 Å for *a-b*, *b-c*, and *c-a*, respectively. Interestingly, the fluorescence intensities of both Ga complexes **37** and **38** were increased significantly compared with each monomeric component **35** and **36**. At the same time, [efficient excitation energy transfer was found to occur between component porphyrins through the analysis of fluorescence anisotropy decay]. A rapid component of ~10 ps appearing in the complex was assigned to the energy transfer time constant between the nearest porphyrin units in the complex according to the Förster model. The excitation energy was stored in the complex by repeating energy transfer roughly

[1] Porphyrin has electronic transitions absorbing the visible light. The relatively weak transition from ground state (S_0) to first excited state (S_1), which corresponds to the HOMO-LUMO bandgap, and the strong one from S_0 to second excited state (S_2), are called Q band and Soret band (or B band), respectively (see Fig. 20).

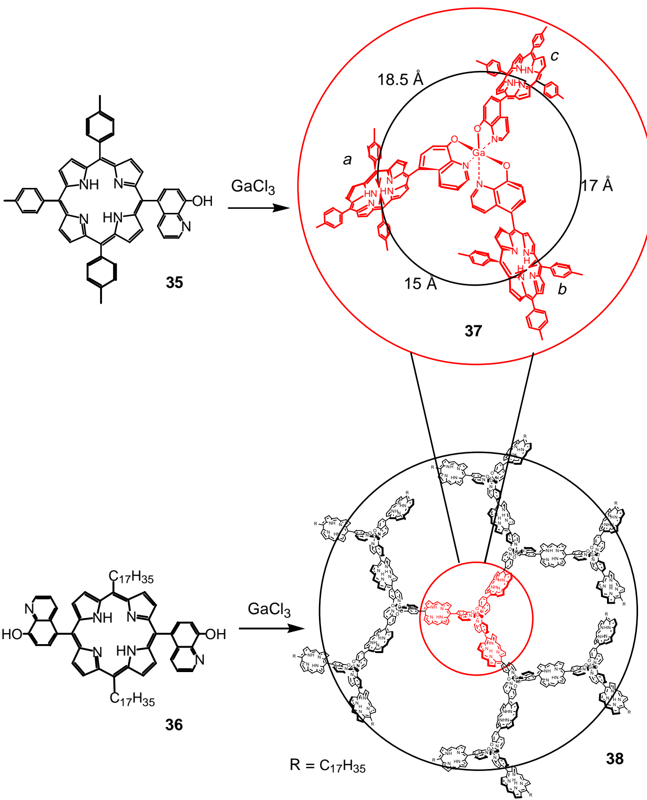


Figure 17. Mono(oxinyl)porphyrin **35**, bis(oxinyl)porphyrin **36**, and their tris(oxinato)Ga(III) complexes **37** and **38**.

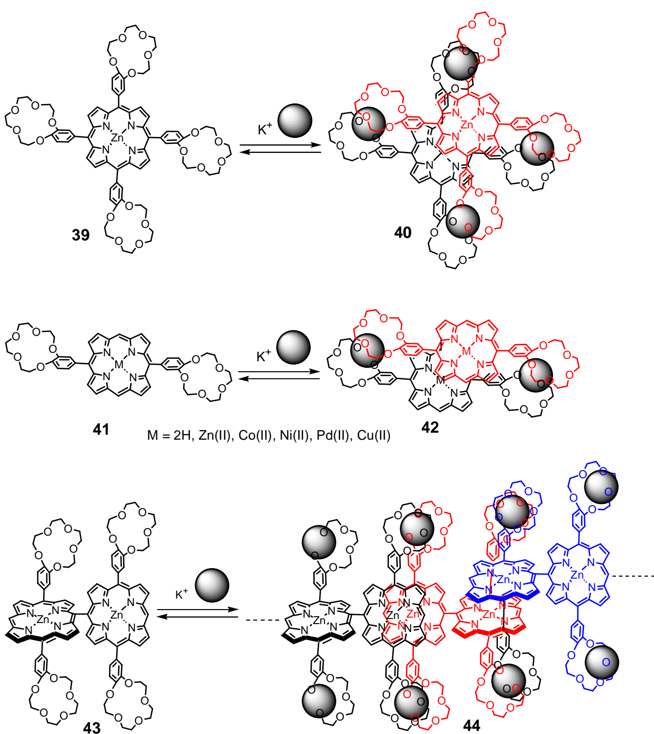


Figure 18. Self-assembled face-to-face porphyrin dimers (**40** [77] and **42** [80]) and polymer (**44** [80, 81]) by complexation of benzo-15-crown-5 substituents with potassium ion.

~800 times within its lifetime of 8.6 ns among three porphyrins in the complex **37** behaving like a photosynthetic light-harvesting complex [75].

2.3. Self-Assembly by Metal Ions with Crown Ether

5,10,15,20-Tetrakis(benzo-15-crown-5)porphyrinatozinc **39** forms face-to-face porphyrin dimer **40** by sandwich-type complexation of eight benzo-15-crown-5 substituents with four metal ions such as sodium, potassium, magnesium, calcium, and barium, as shown in Figure 18 [77–79]. Particularly, reaction with potassium ion gave the most stable face-to-face dimeric complex with a stability constant of 10^{23} M^{-5} in $CHCl_3$-MeOH (1:1) and the interplanar distance between two porphyrin rings was estimated as 4.2 Å. The complexation of 5,10,15-tris(benzo-15-crown-5)porphyrinatozinc with potassium ion exhibited much lower stability constant of ~10^{16} M^{-4}. Recently, the stability constants of 5,15-bis(benzo-15-crown-5)porphyrins **41** with potassium ion in $CHCl_3$/MeCN (2:1) have been determined as follows: 1.9, 2.2, 3.3, 3.5, 1.7, and 18.0 × 10^{18} M^{-3} for M = 2H, Zn(II), Co(II), Ni(II), Pd(II), and Cu(II), respectively [80]. Further, this dimerization induced by potassium ion has been applied to *meso-meso* linked bis(porphyrinatozinc) **43**, each porphyrin bearing two benzo-15-crown-5 substituents [80, 81]. Bisporphyrin **43** was linked by complexation with potassium ion to form polymer **44** in a mixed solvent of $CHCl_3$ and MeCN (2:1). The mean molecular weight and aggregation number were determined as ca. 10^5 Da and 50, respectively, by light-scattering measurement.

3. SELF-ASSEMBLED PORPHYRINS BY AXIAL COORDINATION

Porphyrin can incorporate many elements into the center as shown in Figure 1A [82]. In many cases, the central metal or nonmetal atom can accept one or two axial coordination(s) from outer ligand(s) (L_n, in Fig. 1B). The stability constant K_n and dynamic behavior of complexation between porphyrin center and axial ligand depends much on the nature of the central atom and electron-donating properties of the ligand. There are many possible combinations between central metal and ligand, and the nanostructure with most desirable optical/electronic functions can be obtained by careful choice of the best combination among these. Further, it is interesting that axial coordination of the outer ligand to the central element changes the spectroscopic and electrochemical properties of porphyrin. Such spectroscopic change in the visible light region varies the color of porphyrin and may be applicable for chemical sensors [83]. Chlorophylls both in the antenna complex [4–6] and in the special pair [7–12] are also fixed by membrane proteins through axial coordination of histidine-imidazole to their central Mg ions. These arrangements successfully drive the important photosynthetic functions. In this section, several examples of supramolecular architecture formation by axial coordination to porphyrin center will be presented.

3.1. Self-Assembled Porphyrin Dimer

Self-assembled porphyrin dimers by coordination of *meso*-substituted ligand to another porphyrin center are summarized with schematical drawing, as shown in Figure 19. In these cases, coordination directions of *meso*-substituted ligands such as 1-methyl-2-imidazolyl [84], 2-pyridyl [85, 86], 2-hydroxyphenyl [87, 88], and pyridyl groups connecting through arylamide spacer [89, 90], are orientated to almost perpendicular with respect to the porphyrin plane. As for central metal ions, it should be commented here that most transition-metal ions quench the excitation energy by the presence of odd spin or facilitate effective singlet-triplet intersystem crossing and no fluorescence is generally observable. Zn(II) ion is almost a sole exception and behaves similarly as Mg(II) in chlorophylls. Zn(II) is easily inserted into the porphyrin framework and the product is stable toward heat and acid. These facts cover almost completely the drawbacks of a little less efficiency in energy/electron transfer reactions since Mg(II) is difficult to incorporate and labile toward the action of heat and acid. Therefore, varieties of photosynthetic models have been constructed based on porphyrin Zn(II) complexes. In 1996, Zn(II) analogs of bacteriochlorophylls were found to be a main component of photosynthetic reaction center in the *acidiphilium bacteria* living under acidic conditions (pH 2.5–6.0) [91]. The principal nature of chlorophylls maintained in these Zn(II) analogs. Although interest in supramolecular structure formation sometimes introduces hexacoordinating metal ions such as Ru, Co, Mn, Fe, and others, these metalloporphyrins cannot act as functional mimics of photosynthetic energy/electron transfer reactions, but these metal ions may play important roles in electric conducting materials by using oxidation-reduction activities.

In the case of 1-methyl-2-imidazolyl substituent, porphyrin dimer **45** showed a large stability constant of over 10^{10} M^{-1} in chloroform [84]. ^1H NMR showed that a half of the porphyrin ring was affected by the strong ring current effect of the facing porphyrin, and the UV/visible absorption spectrum gave split Soret bands with splitting energy of 1035 cm^{-1} arising from exciton interactions of slipped cofacial arrangement of two porphyrin units in a solution (Fig. 20) [84, 92]. The large splitting comes from blue and red shifts by exciton interactions of face-to-face and head-to-tail transition dipoles, respectively (see Section 5). Since zinc ion embedded in porphyrin framework accepts only one

Figure 19. Dimerization of porphyrins by coordination of *meso*-substituted ligand to another porphyrin center [84–90]. Substituents of porphyrins have been omitted for clarity.

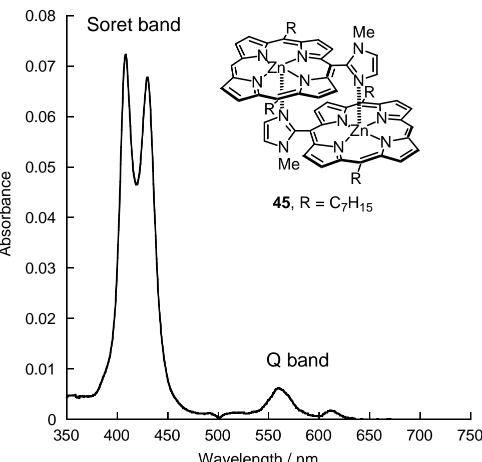


Figure 20. UV/visible absorption spectrum of complementary coordination dimer **45** in $CHCl_3$, where substituent R is *n*-heptyl group.

axial coordination, further organization cannot proceed and its chemistry becomes simplified to a great deal. From X-ray crystallographic study [93], the interplanar distance and the metal-to-metal distance were determined as 3.23 and 6.13 Å, respectively, which resembled those of dimeric structure of the native special pair in the bacterial photosynthetic reaction center [7–12]. These data suggested strong interactions between two chromophores which might have caused changes in optical and electrochemical properties. Similar distances were also found in 2-pyridyl-substituted porphyrinatozinc dimer **46** as 3.30 (3.31) and 5.955 (5.859) Å, respectively, determined by the X-ray study [85]. In this case, the split of Soret bands with splitting energy of ca. 1040 cm^{-1} was observed too.

[Since Ru(II) ion is inert for ligand exchange reaction compared with Zn(II) ion] and can accept hexacoordination with high stability, 2-pyridyl-Ru(II)-porphyrin dimer **47** with sixth ligation of CO or pyridyl was synthesized and their spectroscopic and electrochemical properties were reported [86]. Cyclic voltammogram measurements showed stepwise oxidations for the first oxidation potential of porphyrin rings, and for the Ru(II/III) ions of **47** due to strong interactions between two porphyrins. Absorption spectrum showed one broadened Soret band due to exciton interactions, but did not split in contrast to the zinc dimer complexes **45** and **46**.

2-Phenolate substituent in basic media could coordinate to trivalent central metal ions, Fe(III) and Mn(III), to afford neutral dimer complexes **48** [87] and **49** [88]. However, the crystal structure of Fe(III) complex **48** showed that two porphyrin planes of the dimer were not parallel but inclined with a dihedral angle of 13.5° and the average separation between the two porphyrin rings was 4.09 Å, indicating modest π-π interaction between two chromophores. Hunter and co-workers have reported several self-assembled porphyrins by using pyridyl pendant groups and reported also two types of dimer **50–52** exhibiting the stability constants K in the range 10^6–10^8 M^{-1} in chloroform [89, 90]. Self-assembled dimers by using pyridyl substituted porphyrin at β-position have been reported (Fig. 21) [85, 94]. ^1H NMR spectrum of dimer **53** [85] in $CDCl_3$ was broadened suggesting the exchange reaction with monomeric species at

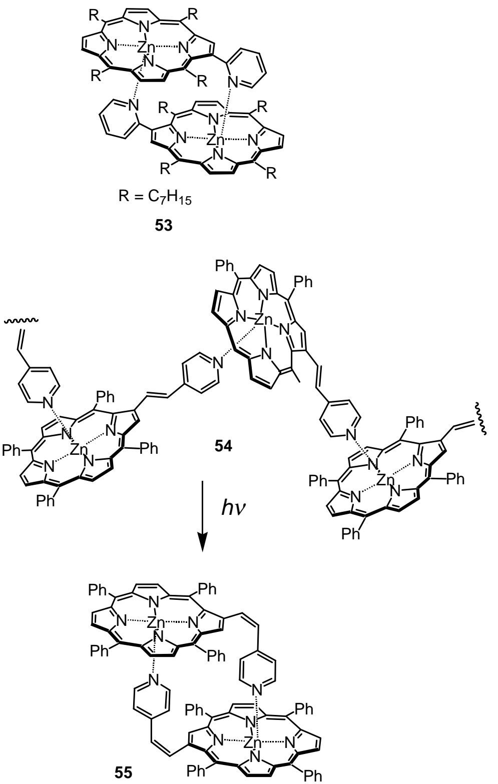


Figure 21. Self-assembled porphyrin dimers **53** [85] and **55** [94] by coordination of β-substituted ligand to another porphyrin center.

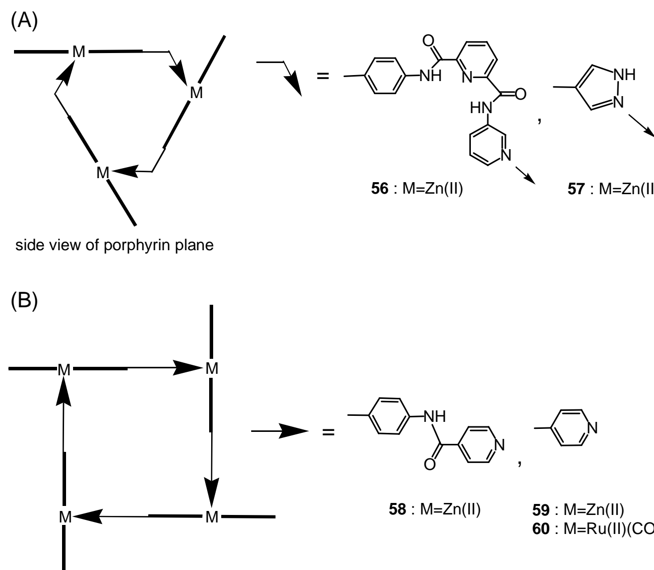


Figure 22. Cyclic trimer (A) [95–97] and tetramer (B) [95, 98–100] of porphyrins by coordination of *meso*-substituted ligand to another porphyrin center. The bold bar represents the side view of porphyrin plane.

room temperature. The spectrum was sharpened at −35 °C with a single set of dimer signals. The complexation reaction was certainly concentration dependent and measurements of absorption spectra undertaken in dilute CH_2Cl_2 solution (∼10^{-4} M) required lowering the temperature to approximately −40 °C to observe the complete formation of dimer **53**. 5,10,15,20-Tetraphenyl porphyrin (TPP) derivative instead of alkyl substituent did not give dimeric species due to significant intermolecular phenyl-phenyl contact. Burrell and co-workers [94] synthesized β-substituted 4-pyridyl-vinyl-TPP with *trans* (62%) and *cis* (20%) alkene linkages. Two isomers were different in the coordination direction of pyridyl nitrogen. Therefore, these zinc complexes gave polymeric species **54** and dimer **55** for *trans* and *cis* isomers, respectively. Both assembled structures were determined by single-crystal X-ray analyses. Two porphyrin planes of dimer **55** were not parallel with a dihedral angle of 32° due to geometrical requirements of the *cis* alkene linkage. The polymer **54** was converted into dimer **55** by irradiation with UV light.

3.2. Cyclic Porphyrin Oligomer

When 3-pyridyl group was introduced into the side arm instead of 4-pyridyl in **50**, the angle between coordination direction and porphyrin plane changed from 90° to near 150° to give porphyrin trimer **56**, as shown in Figure 22A [95, 96]. The ^{1}H NMR spectrum of **56** in $CDCl_3$ was concentration independent. The pyridyl protons were shifted upfield to a great deal at around −6.4 to −0.9 ppm due to the porphyrin ring current. The molecular weight corresponding to trimer was obtained from VPO measurements in $CHCl_3$ at the concentrations from 3 to 20 mM. The absorption spectrum of **56** in CH_2Cl_2 depended on the concentration; at lower concentration of 10^{-7} M, the spectrum corresponding to monomeric species was obtained, but a split of the Soret band by 9 nm in same intensity was observed at higher concentration of 10^{-4} M. The split of the Soret band at higher concentrations indicated the exciton coupling interaction [92] among three porphyrins. The stability constant of trimer **56** was determined as 5×10^{12} M^{-2}. A split of the Soret band of self-assembled porphyrin trimer **57** by coordination of *meso*-substituted pyrazolyl ligand was also reported [97]. In this case, the trimeric structure was cooperatively stabilized by coordinations of pyrazolyl to central metal ion and hydrogen bonds between pyrazolyl NH and carbonyl oxygen of another *meso*-substituent, as shown in Figure 23,

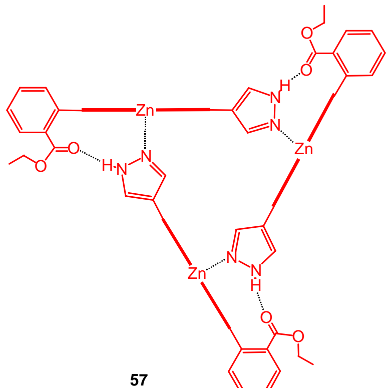


Figure 23. Cooperative stabilization of porphyrin trimer **57** by coordinations of pyrazolyl to central zinc and hydrogen bonds [97].

exhibiting the stability constant of 6×10^{13} M^{-2} at 22 °C. The trimeric structure was supported by ^1H NMR, ESI-MS, and VPO measurements.

Further, in order to obtain a square tetramer **58**, the angle was turned to 180° by changing the *meso*-substituent to 4-(isonicotinamide)phenyl, as shown in Figure 22B [95, 96]. The stability constant of tetramer **58** was determined as ca. 10^{13} M^{-3}. The tetrameric structure was also characterized by ^1H NMR, VPO, and absorption spectrum measurements. An attempt for construction of cyclic tetramer by *meso*-4-pyridyl porphyrinatozinc **59** was unsuccessful by the evidence of X-ray crystallographic study, which revealed the polymeric structure of $(\text{4-pyridyl-porphyrinatozinc})_n$ [98]. However, by using inert Ru(II) ion instead of labile Zn(II) as the central metal ion, the formation of tetramer **60** in solution was confirmed by NMR and FAB-mass spectrometries [99]. Variable temperature NMR study revealed that 4-pyridylporphyrinatozinc produced the polymeric species at room temperature. The result may be compatible with this X-ray study. However, cyclic tetramer **59** could form as a major product only by lowering the temperature to −40 °C [99]. Finally, the crystal structure of cyclic tetramer **59**, which was obtained from a benzene solution, was reported [100].

The natural light-harvesting antenna complexes in photosynthetic bacteria [4–6] have a higher-ordered macrocyclic ring structure composed of many chlorophyll molecules. In the first step of the photosynthesis reaction, the antenna complexes play important roles to absorb light and to transfer the absorbed light energy into the reaction center. Therefore, many chemists have desired to construct the porphyrin macromolecule as artificial antenna complexes by covalent as well as noncovalent approaches. Hunter and co-workers have synthesized monomeric cobalt-porphyrin having two side arms different in coordination direction and length, which may facilitate the coordination with a bend angle of about 30° between two porphyrin planes, to afford a large cyclic complex with approximately 12 porphyrin units, as shown in Figure 24 [101]. Here, the mass spectrometric study (FAB, ES, and MALDI) to detect the oligomeric species **61** was unsuccessful and only monomer peak was observed. The structural estimation was obtained from GPC (gel permeation chromatography) data, which were calibrated by using Co-porphyrin compounds as the reference, and estimated the number of monomer units to be 11.5.

Recently, a large macrocyclic antenna mimic by connecting imidazolyl-zinc coordination units, as explained in dimer **45**, and *m*-phenylene bridging parts has been successfully synthesized [102]. As shown in Figure 25, zinc insertion into free base gable porphyrin **62** gave a mixture of polymeric species **63**, macrocyclic pentamer **64**, and hexamer **65** by the evidence of GPC analysis (Fig. 26, thin line). Although polymeric mixture **63** with broad distributions of relatively higher molecular weight was observed, two separate peaks were detected at longer retention times corresponding to macrorings **64** and **65**. After the reorganization experiment, where the mixture was diluted in $CHCl_3$:MeOH = 7:3 to 1:1 mixture (down to 3.5 μM) to dissociate the coordinate bond and then the solvent was slowly evaporated at 25 ± 1 °C, the higher-molecular-weight part disappeared and converged to the two peaks (Fig. 26, bold line). Atomic force microscope (AFM) observation of the converged sample confirmed the presence of round-shaped particles with uniform height of ca. 1.5 nm, as shown in Figure 27. Small-angle X-ray scattering measurements were undertaken to determine the molecular size of **65**, which was isolated from the converged mixture by preparative GPC separation. The Guinier analysis of the scattering data gave diameters of 42.36 and 40.26 Å for sphere and cylinder approximations, respectively, according well with the outer diameter of cyclic hexamer ca. 41 Å estimated from molecular mechanics calculation. The fluorescence quantum yield of **65** was determined as 0.51 relative to that of the monomeric bis-zinc gable-porphyrin. Similar relations have been observed between the natural light-harvesting complex and its dimeric subunit [103]. This porphyrin macroring is regarded as the first successful antenna mimic to investigate the light-harvesting in photosynthesis.

3.3. One-Dimensionally Self-Assembled Porphyrin Polymer

One-dimensional multi-porphyrin array is one of the candidates for single-molecular electric wire. Therefore, several interesting researches on construction of such a molecular wire by noncovalent as well as covalent approaches have

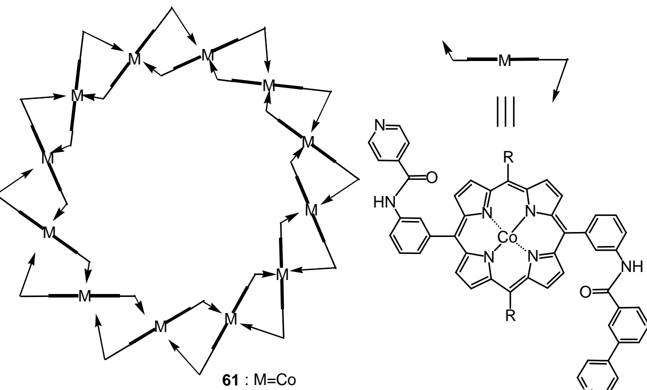


Figure 24. Cyclic porphyrin oligomer assembled by two different substituents inducing a bend angle of 30° between porphyrin planes [101].

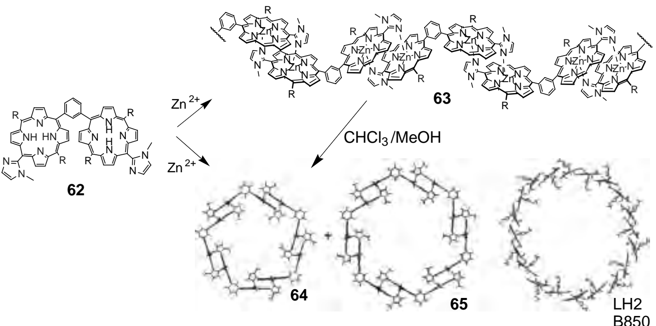


Figure 25. Formation of porphyrin macrorings **64** and **65** as light-harvesting antenna mimic [102].

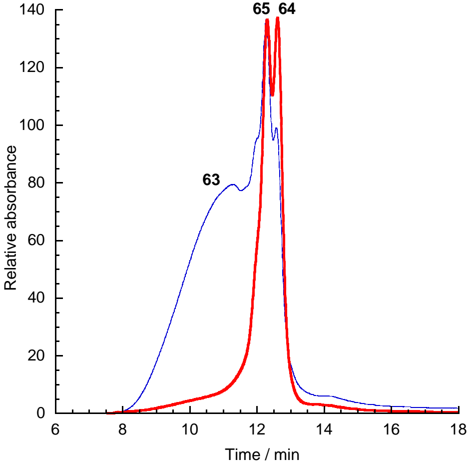


Figure 26. GPC charts of polymeric mixture **63** (thin line) and a converged mixture of **64** and **65** (bold line). Exclusion limit of column was 7×10^4 Daltons and eluent was chloroform.

been reported in recent years. In this section, one-dimensional multi-porphyrin arrays by coordination of *meso*-substituted ligand to another porphyrin center will be reviewed.

One of the examples is porphyrin aggregate by the complementary coordination of imidazolyl to Mg ion (**66**) with strong interaction between porphyrin rings (Fig. 28) [104–106]. As described in Section 3.2, zinc ion accepts only one axial coordination to provide dimer **45** and further organization of *trans*-bis(1-methyl-2-imidazolyl)porphyrinatoZn could not proceed. Therefore, Mg(II), which could accept the hexacoordination, was introduced into *trans*-bis(1-methyl-2-imidazolyl)porphyrin. NMR study of the self-organized complex of Mg porphyrin **66** in $CDCl_3$ showed a growth of the coordination structure by the continuous

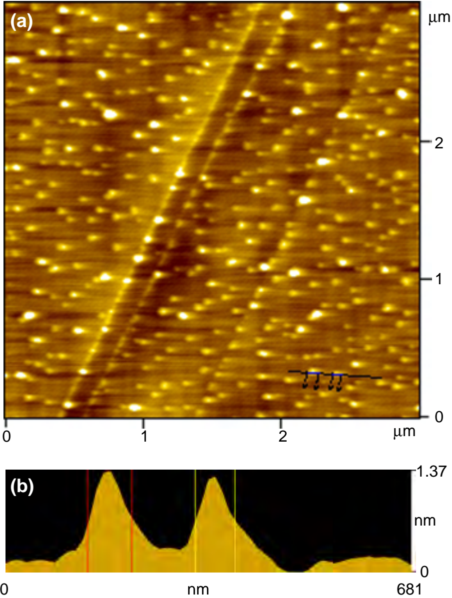


Figure 27. (a) AFM image of self-assembled porphyrin macroring spin-coated on a mica plate. (b) Cross section along the line in (a).

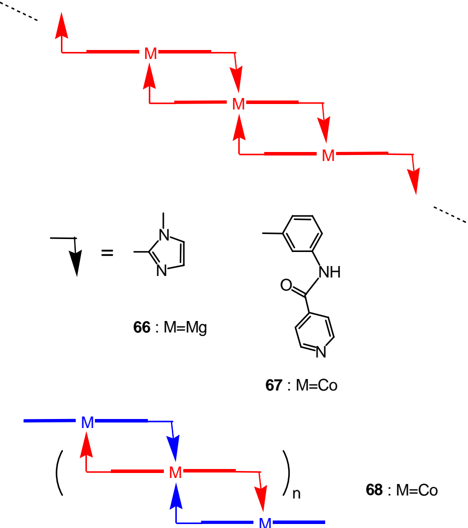


Figure 28. Linear porphyrin polymers **66** [104–106] and **67** [90] by coordination of *meso*-substituted ligand to other porphyrin centers, and oligomer **68** [90] terminated by mono-substituted chain stopper.

repetition of hexacoordination and was suggested by significant up-field shifts of protons due to the strong ring current effect and split Soret bands as well. The evidence of the elongation of organized structure could be obtained from electrospray ionization mass spectrum, which showed the existence of the assembly up to heptameric species [106].

Hydroquinone was introduced into a *meso*-position to obtain another type of one-dimensionally propagated array of Mg porphyrins, as shown in Figure 29 [107]. Figure 30 shows ^{1}H NMR spectra of free base **69** (A) and Mg complex **70** (B). The spectrum of **70** showed larger splits of β-pyrrole protons and tolyl protons than **69** due to increasing asymmetry by complexation. The protons assigned to the hydroquinone part shifted more characteristically from 7.21–7.45 to 6.58–6.38 ppm. Temperature-dependent NMR study down to −40 °C elucidated considerable upfield shifts for all protons in the following order:

$$\mathrm{Ha} \approx \mathrm{Hc} > \mathrm{Hb} \approx \mathrm{H}\beta' > \mathrm{H}\beta'' \approx \mathrm{Hi}$$
$$\approx \mathrm{Ho} > \mathrm{Hi}' > \mathrm{Ho}' > \mathrm{H}\beta$$

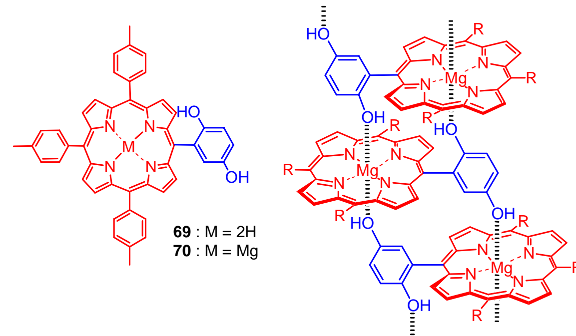


Figure 29. A proposed structure of mono(hydroquinonyl)-porphyrin Mg complex **70** (R = *p*-tolyl) [107].

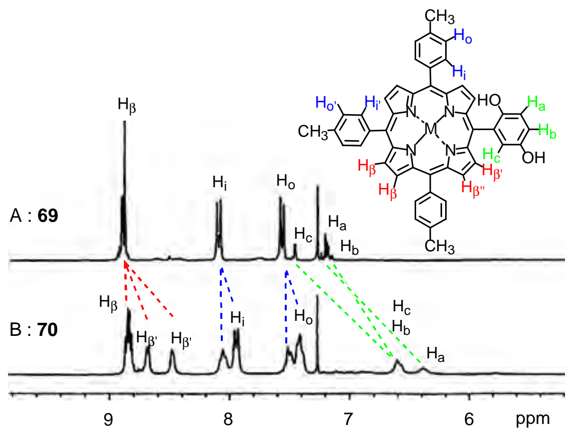


Figure 30. ^{1}H NMR spectra of free base **69** (A) and Mg complex **70** (B) in $CDCl_3$ (8 mM, 20 °C) [107].

The largest shifts were observed for hydroquinonyl peaks Hc and Ha suggesting the molecular organization of Mg porphyrin in a slipped cofacial arrangement.

Bis(*m*-(isonicotinamide)phenyl)-porphyrinatoCo allowed the formation of polymer **67** by coordination of pyridyl ligand to the central cobalt ion (Fig. 28) [90]. Cobalt ion can also accept the sixth coordination to afford the construction of a linear array in a way similar to the case of **66**. The pulse-gradient spin-echo NMR diffusion experiments, which give the information of relative molecular sizes, showed that the diffusion coefficient of **67** was much smaller than that of monomeric Co porphyrins, suggesting a large polymer. GPC analyses of **67** revealed the degree of polymerization of 95 at 7 mM concentration. A mixture of **67** and mono-substituted porphyrinatoCo **52** (Fig. 19), which acted as a chain stopper, gave oligomeric mixture **68** in solution. The length of the oligomer **68** decreased with the use of increasing amounts of the chain stopper **52**.

We succeeded in constructing one-dimensionally self-assembled giant porphyrin array **72** (Fig. 31) by coupling of the complementary coordination unit of imidazolyl to central zinc ion like **45** [108]. Here, the functional target was photosynthetic energy/electron transfer reactions and the central metal ion was limited to Zn(II) due to the reason already discussed. Molecular weight of **72** was determined by GPC analysis. Figure 32 illustrates chromatograms of **72** along with those of polystyrene standard mixtures. The distribution peak maximum appeared at around 1×10^5 Da corresponding to 160 porphyrin units of **71** and molecular length of 110 nm. Since elution of **72** started even before the exclusion limit of MW 5×10^5, the largest molecular weight was at least 5×10^5 Da corresponding to 800 porphyrin units and a 550 nm molecular length. AFM image of **72** dispersed on a mica plate from a chloroform solution in 2.5×2.5 μm range revealed that wirelike structures propagated as a single molecule several hundreds nanometers in length and around 1.3 nm in thickness, corresponding to the single-molecular chain of polyporphyrin (Fig. 33). Even such a stable structure forming polyporphyrin **72** could be dissociated into bisporphyrin unit **71** by the addition of MeOH or pyridine (see Section 5). A mixture of $(\mathbf{73})_2$ and **72** dissolved in a 1:5 molar ratio in chloroform was analyzed by GPC with the column having an exclusion limit of MW 7×10^4 (Fig. 34, line (a)). Two isolated peaks at 8 min (exclusion limit) and 14 min were detected indicating no scrambling between these components. The same mixture was dissolved in chloroform-methanol (1:1), followed by evaporation to reorganize. Then, a series of oligomers **74** marked by 1, 2, 3, and more (Fig. 34, line (b)) was observed. This reorganization process becomes important in the systems such as photocurrent generation [109, 110] and nonlinear optics [43, 110], where specific functional group R′ of **73** is required for facilitating electron transfer and molecular polarization, as well as the construction of large macrocyclic porphyrin arrays **64** and **65** as light-harvesting antenna complex model as discussed in Section 3.2. The photocurrent generation and NLO property will be discussed in Section 6.

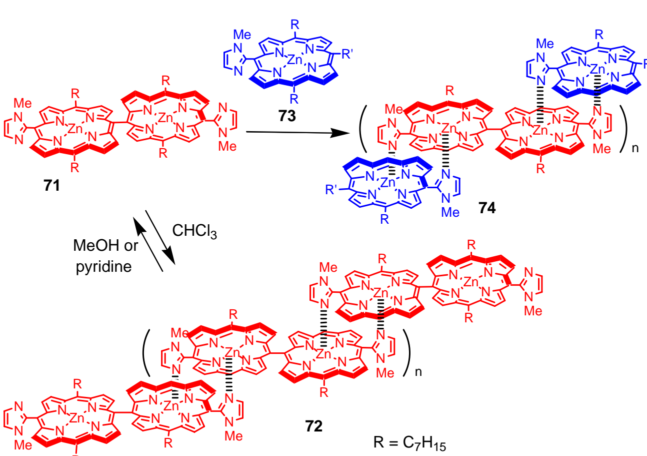


Figure 31. Formation of one-dimensional giant porphyrin array **72** and terminated oligomer **74** by the complementary coordination of imidazolyl to the central zinc(II) ion [108].

3.4. Other Porphyrin Assemblies by Axial Coordination

4,4′-Bipyridyl, 1,4-diazabicyclo[2.2.2]octane (DABCO), and mono- to tetra-pyridyl-appended porphyrins have been employed widely to construct a variety of unique porphyrin

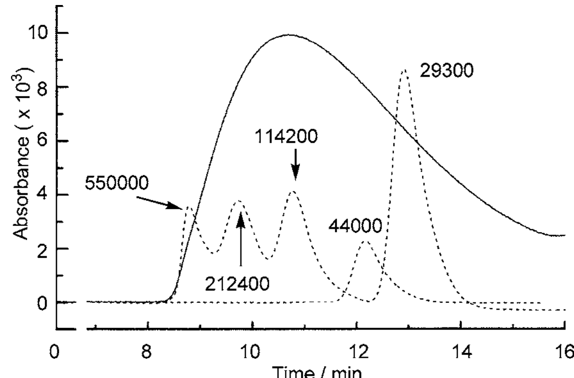


Figure 32. GPC chart of **72** (bold line) and mixtures of polystyrene standard (dotted line, the numbers indicate molecular weight) with a column of exclusion limit 5×10^5. The eluent was EtOH-free $CHCl_3$.

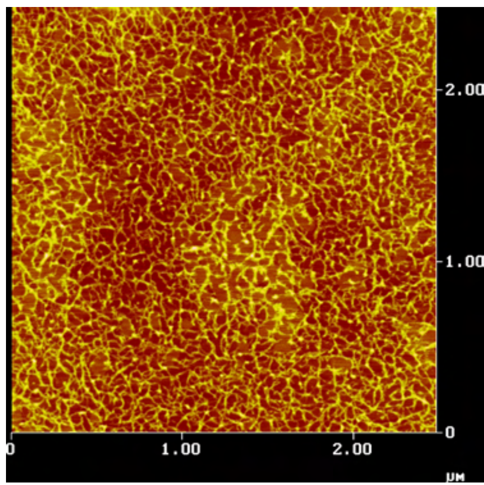


Figure 33. AFM image of linear multi-porphyrin array 72 dispersed on a mica plate.

architectures by axial coordination to other porphyrin centers. In many cases, Zn(II) was selected as the central metal ion to accept the axial coordination. Its relatively strong affinity constants especially to pyridyl gave various examples. Simple coordination chemistry by the acceptance of only one axial ligand, and its light-emitting property, are their characteristics. Many examples have been reported to push forward such noncovalent chemistry for controlling nanostructures and will be introduced in this section (some of these are already introduced in the discussion as seen in Fig. 5 and Sections 3.1–3.3).

One of the early examples conducted by Anderson et al. is shown in Figure 35 [111, 112]. Two zinc porphyrins connected by flexible alkoxycarbonylalkyl chains containing pyromellitdiimide, which is a strong electron acceptor, were synthesized to incorporate some guest molecules in the cavity. 5,15-Bis(4-pyridyl)porphyrin was bound by axial coordination of 4-pyridyl group to two zinc porphyrins as well as by π-π interactions between free base porphyrin and pyromellitdiimide units to give relatively large association constant

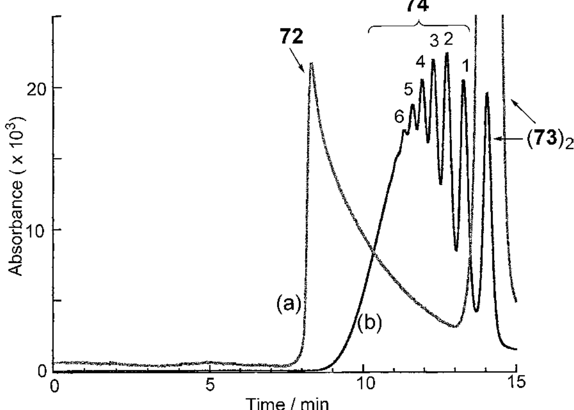


Figure 34. GPC chart of a mixture of 72 and $(73)_2$ in a 5:1 molar ratio (a) and the same sample mixture once dissolved in $CHCl_3$-MeOH (1:1), then evaporated, and redissolved in $CHCl_3$ (b). The exclusion limit of the column employed was 7×10^4.

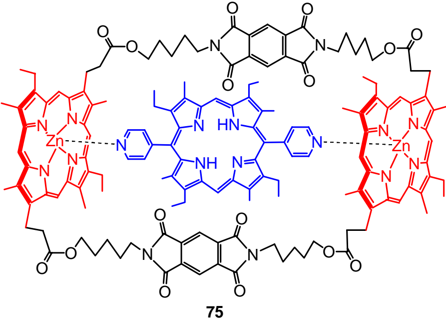


Figure 35. Self-assembled porphyrin trimer 75 stabilized by two axial coordinations and π-π interactions between free base porphyrin and pyromellitdiimide [111, 112].

of 2.6×10^5 M^{-1}. The fluorescence of free base porphyrin in 75 was quenched due to electron transfer to the pyromellitdiimide part. Chernook et al. synthesized self-assembled porphyrin systems 76 and 77 (Fig. 36), in which *trans*-bis(3-pyridyl)- and *cis*-bis(4-pyridyl)-substituted free base porphyrins bound to coplanar bis(porphyrinatozinc) by 2-fold coordination, respectively [113]. The complexation constants of both systems were in the range 10^6 to 10^7 M^{-1} in CH_2Cl_2 determined by absorption and fluorescence titration experiments. The emission from zinc porphyrin units within the complex was quenched and that from the free base increased, indicating the occurrence of singlet-singlet energy transfer between the components.

Hunter reported a self-assembled porphyrin trimer containing a naphthalenediimide bridging group as the electron acceptor 78 and its reference compound without a diimide grouping 79 [114]. Free base porphyrin was loaded onto bis(porphyrinatozinc) by two pyridyl arms (Fig. 37). These structures were confirmed by VPO, ^1H NMR , and absorption spectra showing a discrete and stable 1:1 complex at the concentration range of 10^{-2} to 10^{-7} M. The stability constant of 78 was determined as 3×10^8 M^{-1} in CH_2Cl_2 by using fluorescence titration technique. Photoinduced electron transfer to diimide quenched 70% fluorescence of free base porphyrin. The weak fluorescence from uncomplexed zinc porphyrin parts was not changed by the acceptance of a

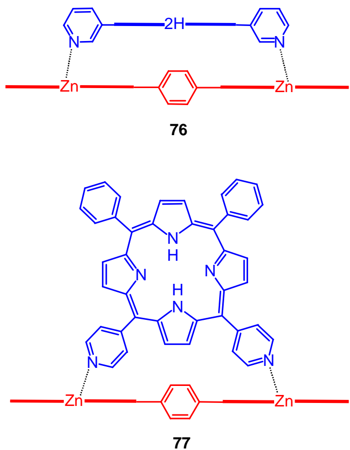


Figure 36. Self-assembled porphyrin trimers 76 and 77 [113].

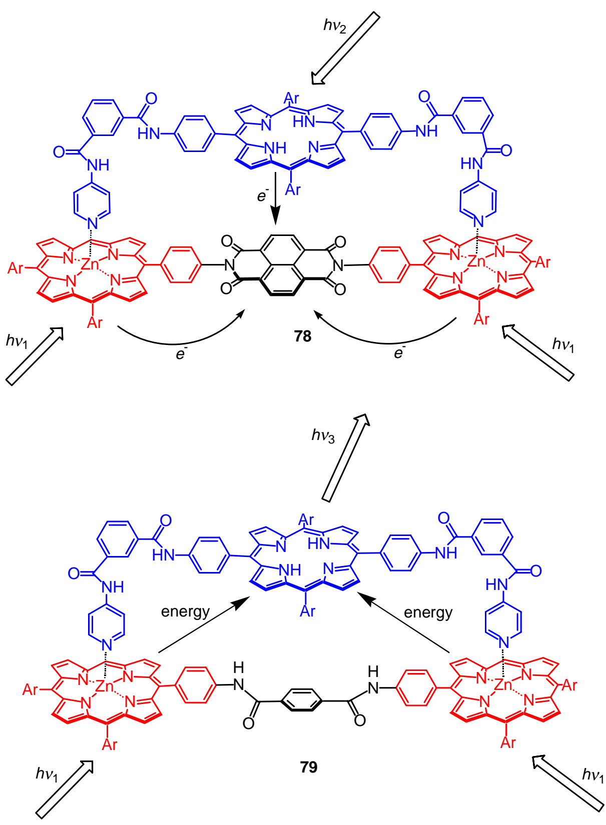


Figure 37. Self-assembled porphyrin trimer containing naphthalenediimide as an electron acceptor **78** and no acceptor **79** to control electron and energy transfer processes [114].

free base porphyrin as a guest. This indicated that fast electron transfer from zinc porphyrin to diimide units occurred concurrently in **78**. On the other hand, complex **79**, having no acceptor, showed energy transfer from zinc porphyrin to free base porphyrin. The electron and energy transfer processes are summarized schematically in Figure 37 [114].

Anderson has reported double-strand [42, 115, 116] and triple-strand [117] porphyrin ladders (**80** and **81**), as shown in Figure 38. The linear porphyrin arrays are conjugated by diacetylene linkages. Bridging reagents employed were 4,4′-bipyridyl and DABCO for **80**, and diacetylene linked free base 5,5′,15,15′-tetrakis(4-pyridyl)-bisporphyrin for **81**. In all cases, ^{1}H NMR spectra showed that all the protons of the bridging ligand were shielded by porphyrin ring current effect to give large up-field shifts. In the case of conjugated dimer **80** ($n = 2$) with DABCO, aggregation became much stronger with an equilibrium constant of 1.2×10^{7} M^{-1} for the formation of bimolecular porphyrin aggregate relative to that of monomer (3.8×10^{2} M^{-1}) in CH_2Cl_2 [115]. The stability constant for 2:2 complex was determined as 4×10^{21} M^{-3} in CH_2Cl_2. The double-strand conjugated polymer with 4,4′-bipyridyl exhibited nine times larger third-order optical nonlinearity than the single-strand one by holding the π-conjugation in a coplanar form [42]. The absorption spectrum of triple-strand **81** was not a simple sum of those of the components and the stability

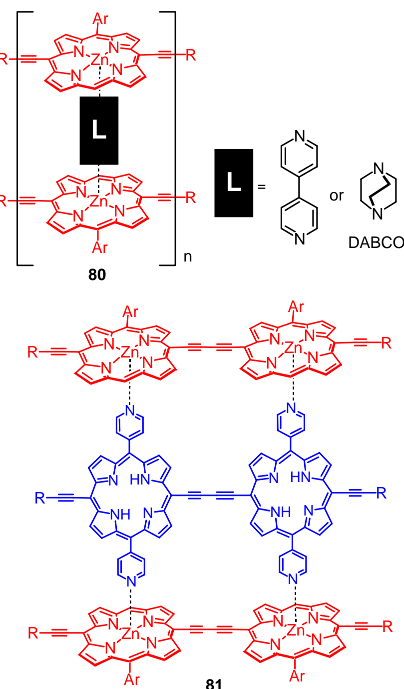


Figure 38. Double-strand conjugated porphyrin ladders **80** [42, 115, 116] and triple-strand **81** [117].

constant could not be determined accurately in CH_2Cl_2 (10^{16}–10^{19} M^{-2}) [117]. The spectral changes by the formation of **80** and **81** will be discussed in Section 5.

Alessio and co-workers synthesized a dimeric zinc porphyrin building block possessing the structure similar to compound **2** shown in Figure 3 by using Ru(II) instead of Pd(II) ion. Higher-order assemblies **82** and **83** were constructed from this building block by axial coordination of bridging ligand 4,4'-bipyridyl and 5,15-bis(4-pyridyl)porphyrin, respectively, as shown in Figure 39 [54]. One-dimensional, two-dimensional, and

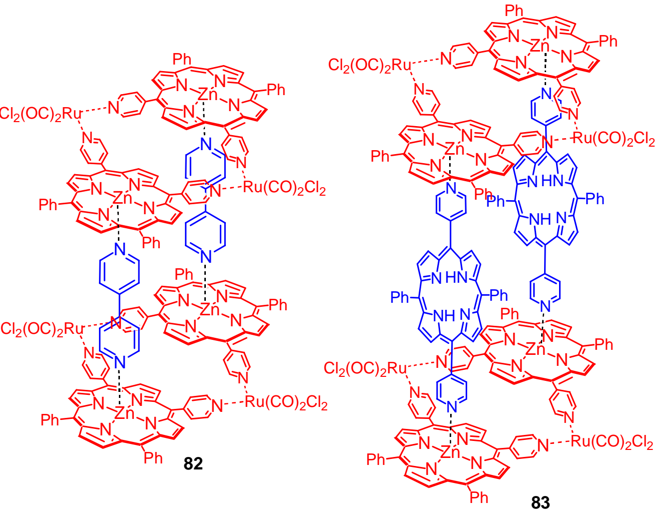


Figure 39. Molecular sandwiches **82** and **83** constructed by axial coordination of bridging ligand to Ru(II)-assisted dimeric porphyrin building block [54].

variable-temperature NMR investigations indicated assembly formation of sandwich-like 2:2 complex quantitatively in a $CDCl_3$ solution. Single-crystal X-ray analyses showed that **83** was a discrete assembly of two dimers connected by bridging free base porphyrins. The two bridging porphyrins oriented in a coplanar arrangement with a distance of ca. 11.4 Å.

The same authors have reported a discrete porphyrin tetramer **85** constructed by two-point self-coordinations of Ru(II)-assisted bis[5,10-bis(4-pyridyl)porphyrinatozinc(II)] (Fig. 40) [118]. Free base bisporphyrin **84**, which was complexed by $Ru(CO_2)Cl_2$ at each 4-pyridyl substituent and had one unbound 4-pyridyl residue on each porphyrin, as a building block was expected to form the tetramer **85** by insertion of Zn(II) ion into two porphyrins. ^{1}H NMR spectrum of the crude product showed sharp signals corresponding to discrete self-assembly as a main species and minor broad peaks attributed to oligomer. The pure **85** was isolated by silica gel chromatography and crystallized by slow diffusion of *n*-hexane into $CHCl_3$ solution. The single-crystal X-ray analysis revealed a propeller-like arrangement consisting of four porphyrins with a S4 symmetry along the rotation axis passing through two Ru(II) ions. The cooperative interaction of four axial coordinations of pyridyl substituent, which was arranged in a *cis* form (substituted at 5 and 10 positions of porphyrin), gave a discrete assembly by recognizing the building block.

At the same time, *meso-meso* linked bis(5-(4-pyridyl)porphyrinatozinc) **86** and its cyclic tetramer **87** have been reported (Fig. 41) [100]. The boxlike assembly **87** has a cavity of ca. $10 \times 10 \times 8$ Å^3. ^{1}H NMR spectrum gave sharp peaks and pyridyl protons were upfield-shifted by the effect of porphyrin ring current, suggesting the existence of a self-assembled cyclic tetramer **87** from **86** in $CDCl_3$. A GPC analysis of **87** showed a sharp elution peak with

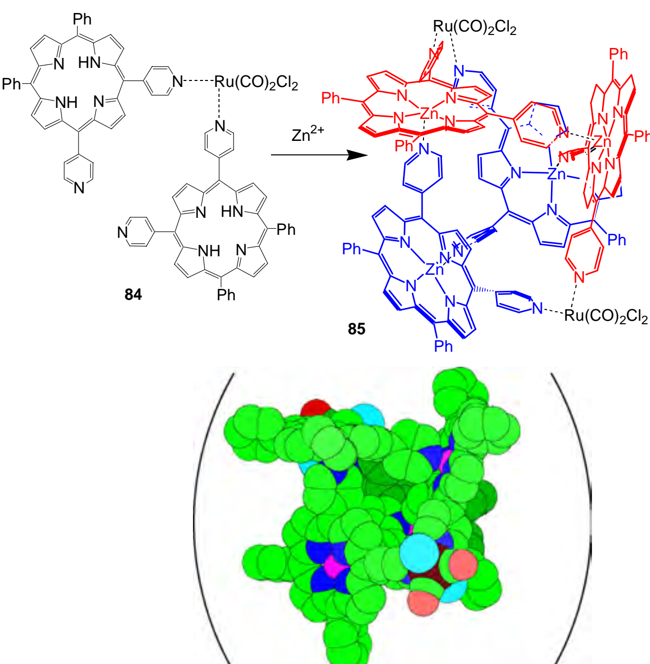


Figure 40. Dimer **84** and cyclic tetramer **85** [118].

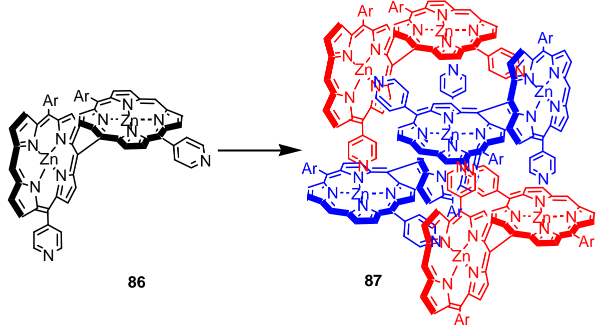


Figure 41. *Meso-meso* linked bis(5-(4-pyridyl)porphyrinatozinc) **86** and its cyclic tetramer **87** [100].

a shorter retention time than that of monomer indicating discrete cyclic assembly. The CSI-MS measurement in a THF solution observed the parent ion of tetramer **87** as the major peak with a much smaller peak corresponding to monomer. The association constant was estimated as at least 10^{25} M^{-3} in $CHCl_3$ from fluorescence spectrum, which did not depend on the concentration down to 1.6×10^{-8} M.

Imamura and Alessio and co-workers have developed a self-assembled porphyrin system by using Ru(II) or Os(II) ion as the central metal ion to construct kinetically stable architectures. Figure 42 represents a series of

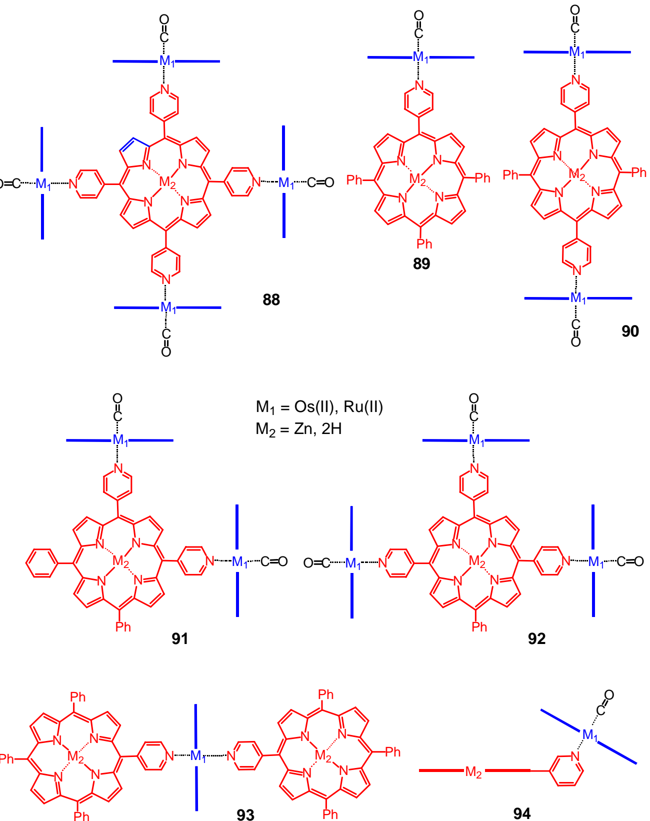


Figure 42. Schematical representation of systematic porphyrin assemblies by using coordination of *meso*-substituted pyridyl to Ru(II) or Os(II) center [119–124].

porphyrin arrays constructed by using coordination of *meso*-substituted pyridyls to central Ru(II) or Os(II) [119–124]. Alessio et al. prepared pentamer **88** by combination of four Ru(CO)-TPP walls with one central free base or zinc tetrapyridyl porphyrin as a bridging ligand [119]. Reaction of Ru(CO)(EtOH)-TPP with tetrakis(4-pyridyl)porphyrin was carried out in chloroform at room temperature to afford **88** rapidly and quantitatively as confirmed by ^{1}H NMR measurement. Imamura and co-workers reported a series of porphyrin assemblies **88–93** by using Os(CO)- or Ru(CO)-porphyrins as walls [120–122]. In these cases, Os(CO)(EtOH)- or Ru(CO)(MeOH)-porphyrins were used as the precursor of the wall and reacted with pyridyl porphyrin in toluene under Ar atmosphere and reflux conditions. The organizations were also confirmed by ^{1}H NMR measurement. The cyclic voltammogram and electronic spectral measurements of the oligomers indicated that there was no distinct interaction within the macrocycles. Two authors also reported porphyrin assemblies **94** by using an axial coordination of *meso*-substituted 3-pyridyl-porphyrin to Os(II) or Ru(II) center [123, 124].

Imamura and co-workers have reported porphyrin tetramers **95**, **96** [86], and octamer **97** [125], as shown in Figures 43 and 44, respectively. The substitution of axial CO ligands in dimer **47** (Fig. 19 in Section 3.1) and tetramer **60** (Fig. 22 in Section 3.2) with free base 4-pyridyl- or 3-pyridyl-porphyrin by the photolysis of CO by medium-pressure mercury lamp in toluene followed by pyridine coordination gave the desirable higher-ordered assemblies, whose structures were analyzed by elemental analyses and FAB- or ESI-MS measurements. ^{1}H NMR spectra also revealed the tetrameric and octameric structures. Integral intensities of NMR signals agreed with the 1:1 composition of Ru(II) and free base porphyrins (2:2 for tetramers **95**, **96** and 4:4 for octamer **97**). The pyridyl and β-pyrrolic protons of the newly introduced axial porphyrin unit were shifted to up-field due to the ring current effect of the ruthenium porphyrin part.

Sanders and co-workers have also reported self-assembled porphyrin systems by using axial coordination to Ru(II) and other metal ions. Figure 45 illustrates dendritic multi-porphyrin arrays **98** and **99** [126]. Compound **98** was synthesized in two steps, in which monomeric Ru(II) porphyrin Ru(CO)(solvent)-por and 1 eq. of free base trimer $(por)_3$ were reacted in benzene at room temperature to form $[Ru(CO)\text{-}por\cdot(por)_3]$ and then $[Ru(CO)\text{-}por\cdot(por)_3]$ and $(por)_3$ were coupled by photoirradiation to give $[(por)_3Ru\text{-}por\cdot(por)_3]$ **98**. Free base trimer having three 4-bipyridyl substituents was treated with monomeric Ru(CO)(solvent)-por in chloroform to give rise to the formation of **99**. The NMR measurements supported these structures. Crystals of **99** suitable for single-crystal X-ray analysis were not obtained, but tripyridyltriazine(Py_3T)-assisted trimer **100** could be crystallized by slow diffusion of methanol into a

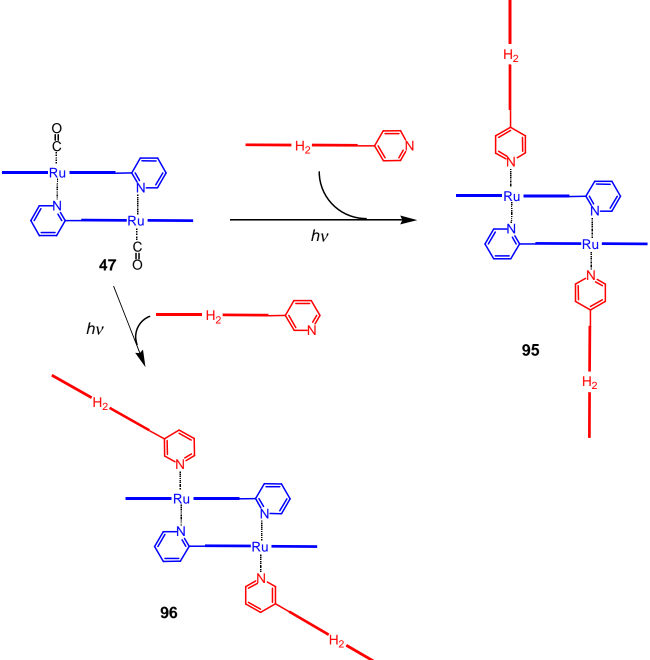


Figure 43. Formation of higher-ordered organization from dimer **47** by photoirradiation [86].

Figure 44. Formation of porphyrin octamer from tetramer by photoirradiation [125].

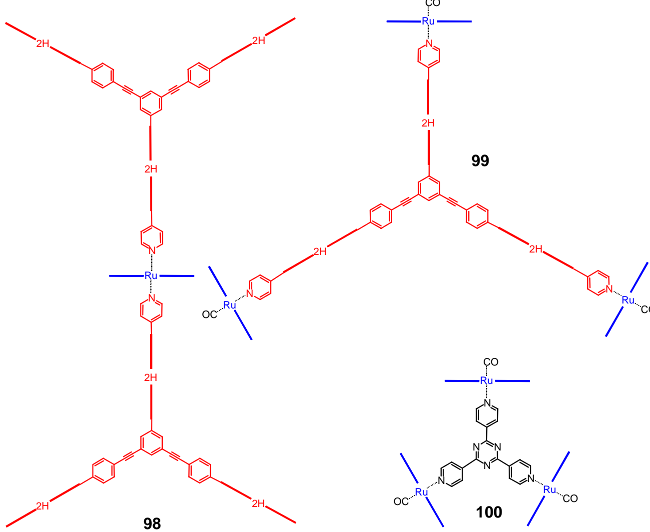


Figure 45. Dendric multi-porphyrin arrays based on Ru(II) [126].

toluene solution. The X-ray structure of **100** showed that the angles of Py_3T nitrogens and porphyrin plane were ranged from 85° to 90° and the distances between Ru(II) and nitrogens of Py_3T were 2.15 to 2.22 Å.

Self-assembled heterometallic oligoporphyrins **101** and **102** constructed by cooperative Zn(II) —N, Ru(II) —N, and Sn—O coordination bonds have been reported by the same group, as shown in Figure 46 [127]. The heterometallic porphyrin oligomers were designed to exhibit complementary geometries and cooperative binding properties by using different affinities toward ligand species. Here again, Ru(II) ion forms inert complexes with nitrogen ligands with hexacoordination, while Zn(II) porphyrin prefers nitrogen atoms with kinetically labile pentacoordination. On the other hand, Sn(IV) porphyrin prefers oxygen atoms with a hexacoordinate octahedral geometry and exhibits relatively slow ligand exchange rates. Therefore, carboxylate groups extended from a *meso*-phenylene position of Zn(II) or Ru(II) porphyrin bound to the Sn(IV) center of bis(4-pyridyl)porphyrin and at the same time two pyridyls of the Sn(IV) porphyrin coordinated to Zn(II) or Ru(II) centers in a cooperative fashion. These structures were determined by 1D and 2D (COSY and NOESY) NMR techniques.

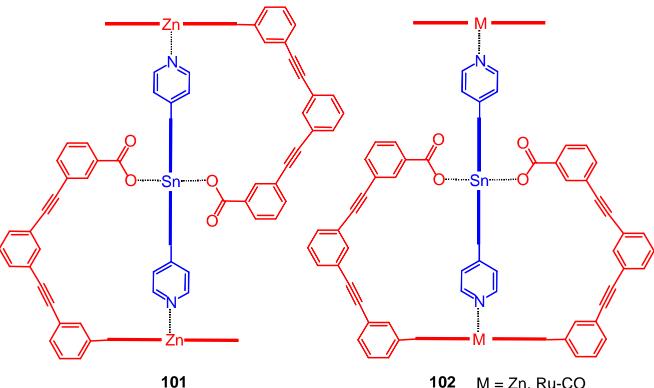


Figure 46. Self-assembled heterometallic oligoporphyrins constructed by cooperative Zn(II)—N, Ru(II)—N, and Sn—O coordination bonds [127].

Recently, Rh(III) porphyrin has been employed for building blocks of nanoarchitectures [128, 129]. Reaction of Rh(III) porphyrin monomer, whose axial positions were coordinated first by an iodide ion and a MeOH molecule, with 4,4′-bipyridine, DABCO, 4,4′-bipyrimidine, 2,7-diazapyrene, and 5,5′-dicyano-2,2′-bipyridine could give dimeric species **103** with these bridging diaza ligands by substitution of MeOH (Fig. 47). X-ray crystal structures of dimeric complexes with DABCO, 2,7-diazapyrene, and 5,5′-dicyano-2,2′-bipyridine were analyzed successfully. Porphyrin trimer **104** and pentamer **105** were synthesized by using a combination of carboxylate coordination to Sn(IV) and pyridine coordination to Rh(III) porphyrins. The ligands bridging Sn(IV) and Rh(III) porphyrins were isonicotinate and free base 5-(4-pyridyl)-15-(4-benzoate)porphyrin for **104** and **105**, respectively. X-ray crystal analysis of **104** showed that two Rh porphyrins

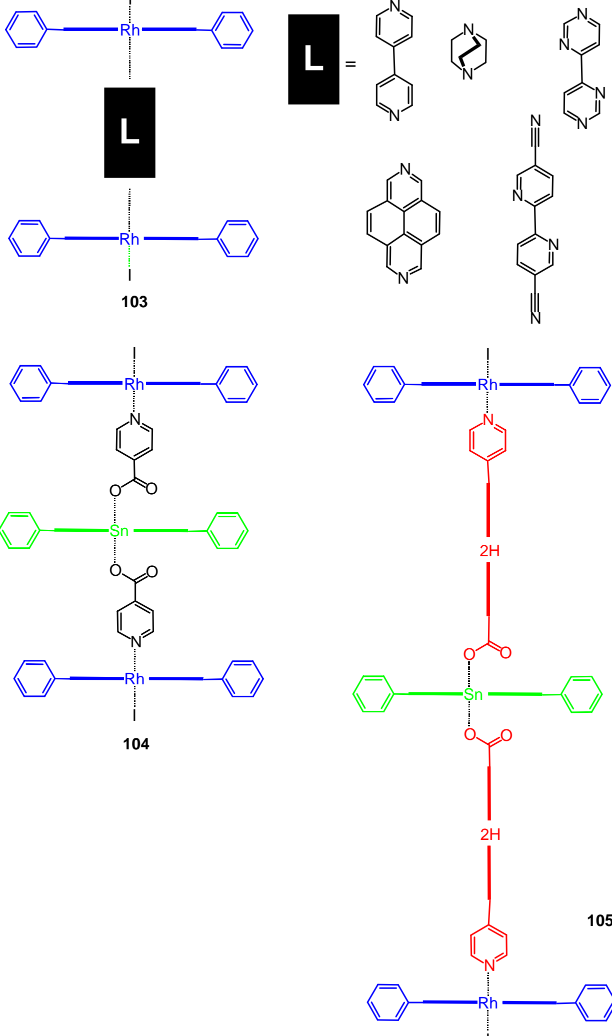


Figure 47. Supermolecules from Rh(III) and Sn(IV) porphyrins by using axial coordinations [128].

were equivalent through an inversion center of Sn atom and almost planar. The isonicotinate molecules were also planar and directed perpendicularly to the Rh porphyrin. An eleven-porphyrin array **106** was constructed by using three different metalloporphyrins: Sn(IV), Rh(III), Ni(II), and free base (Fig. 48) [128]. A Sn(IV) porphyrin acted as a core which was coordinated by two carboxylate substituted porphyrin trimers. Four Rh(III) porphyrins were bound to pyridyl groups at the four terminals of this array. The porphyrin array **106** was analyzed by 2D NMR techniques (COSY, NOESY, and TOCSY) to obtain spatial information. A cyclic porphyrin tetramer **107**, consisting of two bis(diphenylphosphine)-substituted zinc(II) porphyrins and two Rh(III)TPP units, was also constructed (Fig. 49) [129]. When Rh(III)-TPP was mixed with a same equivalent of bis(diphenylphosphine)-substituted Zn(II) porphyrin in $CDCl_3$, ^{1}H and ^{31}P NMR spectra suggested the formation of mixtures of several different species. When 0.5 equivalent of 4,4′-bipyridyl was added to the solution, cyclic porphyrin tetramer **107** was formed quantitatively. This process was monitored by ^{1}H NMR and all the information from ^{1}H, COSY, NOESY, and ^{31}P NMR spectra proved the existence of **107** as a single species.

Anderson et al. have developed a template-directed synthesis of cyclic porphyrin oligomers [130–132]. The synthesis of cyclic porphyrins by the use of appropriate templates

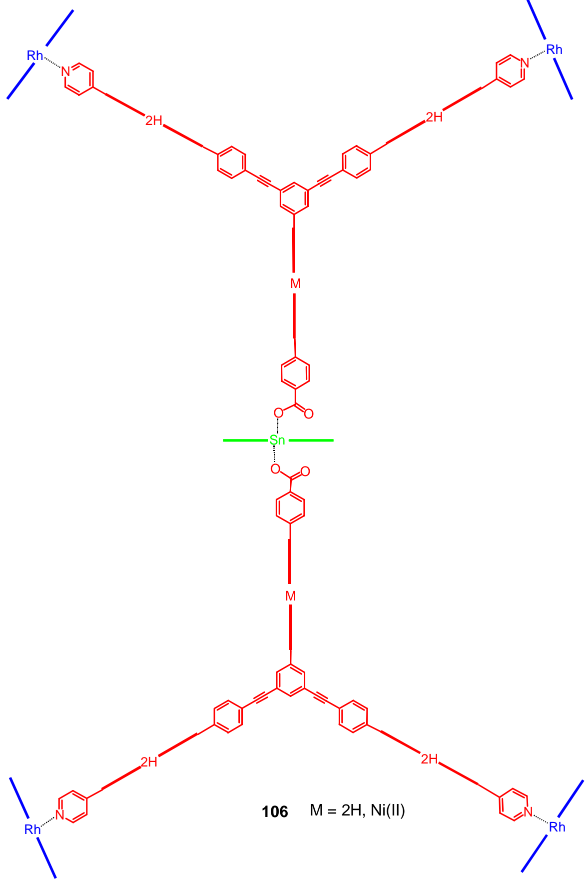


Figure 48. Self-assembled porphyrin 11-mer constructed from Rh(III) and Sn(IV) porphyrin building blocks [128].

Figure 49. Self-assembled cyclic porphyrin tetramer **107** by coordination of phosphine to Rh(III) center and bipyridyl bridging [129].

enabled to increase the yield significantly. The key step was the first process to self-assemble the substrates in a juxtaposition for the reaction, and then the desirable porphyrin macromolecule was produced by intramolecular covalent bond formation. As shown in Figure 50, 4,4′-bipyridyl was added to two zinc porphyrins to facilitate the closed conformation. The Glaser coupling reaction at two terminal acetylene groups of linear porphyrin dimer **108** produced cyclic dimer **109** in a 83% yield. The untemplated reaction gave only 21% of **109** with large amounts of other higher cyclic oligomers. On the other hand, a larger template reagent, tetrapyridyl-porphyrin, effectively promoted the formation of cyclic tetramer **110** containing one free base porphyrin with a high yield of 77%. In a similar way, the coupling reaction of monomer **111** in the presence of tridentate ligand Py_3T as the template gave a symmetric cyclic porphyrin trimer **112** in a 55% yield (Fig. 51).

4. SELF-ASSEMBLED PORPHYRINS BY HYDROGEN BONDS

Hydrogen bond formed between hydrogen atom adjacent to electronegative element and lone pair electron of mainly oxygen, nitrogen, and sulfur atoms is relatively weak interaction of 20 kJ/mol at most. However, stable aggregations can be formed through cooperative hydrogen bonds at multiple sites as seen in biological systems. For example, a double helix of DNA is built up by a lot of hydrogen bonds between nucleic acid base (nucleobase) pairings [13–15]. Sessler and co-workers have introduced cytosine and guanine derivatives, which can form a stable nucleobase pair of C-G by

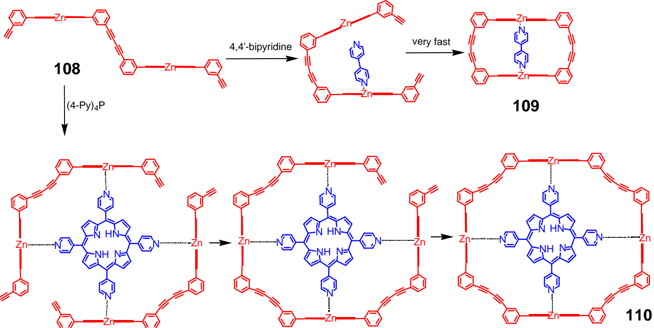


Figure 50. Template syntheses of cyclic dimer **109** and tetramer **110** assisted by 4,4′bipyridyl and tetrapyridyl-porphyrin.

Figure 51. Template synthesis of cyclic trimer **112** by using Py_3T [130–132].

triple hydrogen bonds, into each *meso*-position of respective porphyrins, as shown in Figure 52 [133]. Association constant of dimer **113·114** was determined to be 210 ± 40 $M^{-1}$by fluorescence lifetime measurement in a $CHCl_3$ solution. However, the photoinduced singlet-singlet energy transfer reaction within the hydrogen-bonded complex was not observed because of intracomplex diffusional encounter quenching between donor **113** and acceptor **114**. In order to suppress the flexibility, more rigid supramolecules have been developed by the combination of **115–117** [134]. In these systems, the donor and acceptor molecules were connected through *para*-phenylene linkages to hydrogen bonding units of guanosine (G) and cytidine (C), respectively.

Figure 52. Self-assembled porphyrins by triple hydrogen bonds in a nucleobase pairing [133, 134].

Association constant of dimer **115·116**, which was arranged side-by-side with an interplanar angle of ca. 90° and with the center-to-center distance of ca. 22.5 Å, was determined as 22,000 ± 2000 M^{-1} by ^{1}H NMR study in a CD_2Cl_2 solution. The photoinduced singlet-singlet energy transfer occurred within both dimer **115·116** and trimer **116·117·116**, and quantum yields were determined as 0.57–0.64.

Uracil derivatives have also been used in the construction of a triply hydrogen bonded self-assembly. Two 5,15-bis(5-uracil)porphyrinatoZn, in which two uracil substituents orientated in the *cis* arrangement as shown in Figure 53, formed a face-to-face dimer **118** by cumulative 12 hydrogen bonds bridged by two triaminopyrimidine molecules [135]. The interplanar distance between two porphyrins was estimated to be 10 Å providing a cage cavity to incorporate a host molecule such as 4,4′-bipyridine. Recently, five 2,6-diacetamido-4-pyridyl and four 6-uracil *meso*-substituted porphyrins have been synthesized for building up various types of multi-porphyrin supramolecules by triple hydrogen bonds between 2,6-diacetamido-4-pyridyl and uracil [136]. Two equivalents each of 5,10-bis(2,6-diacetamido-4-pyridyl)porphyrin and 5,10-bis(6-uracil)porphyrin in THF gave square tetramer **119** (Fig. 53). Formation of the tetramer was confirmed by ^{1}H NMR, dynamic light-scattering, ESI MS, and VPO measurements. ^{1}H NMR study showed the association constant of 6×10^{12} M^{-3} in THF-d_8.

Supramolecular multi-porphyrin macrocycle **120** by alternating disposition of three 1,3,5-triazine-bridged bis-porphyrin units and three barbituric acid units by eighteen

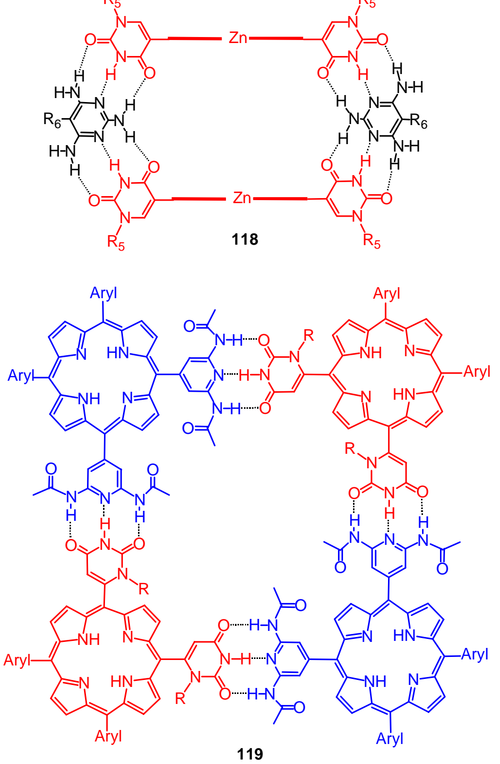


Figure 53. Self-assembled porphyrins of uracil derivatives **118** [135] and **119** [136] by triple hydrogen bonds formation.

N—H···N and N—H···O hydrogen bonds was synthesized (Fig. 54) [137]. ESI-MS measurement of a 1:1 mixture of bisporphyrin and barbituric acid in CH_2Cl_2 afforded the dicationic parent peak of **120** and VPO measurement suggested the molecular weight consistent with the assembly of three bisporphyrin and three barbituric units. ^1H NMR and NOE experiments indicated the formation of a discrete supramolecular structure rather than polymeric species and diffusion-ordered 2D NMR, which displays 1D chemical shift spectrum with a diffusion spectrum in the other dimension and gives information on the molecular radius [138], showed that over 90% of the two components were assembled to form assemblies of an approximately 2.5 nm size.

Two porphyrins bearing 4-pyrazolyl and 3,5-dimethyl-4-pyrazolyl substituent have been synthesized [139]. 3,5-Dimethyl-4-pyrazolyl-substituted porphyrin provided cyclic tetramer **122** as shown in Figure 55 with an association constant of 9.3×10^3 M^{-3} in $CDCl_3$. On the other hand, 4-pyrazolyl derivative without a methyl group favored the formation of dimeric species **121** and association constant was determined as 39 M^{-1}. The structural characterizations in both cases were performed by ^1H NMR, IR, and ESI-MS measurements. In the case of **121** without the methyl group in the pyrazolyl group, the rotational barrier between pyrazolyl and porphyrin plane was relatively low, and therefore steric hindrance between methyl groups at other *meso*-substituted tolyl groups might have prevented the formation of higher-ordered oligomer.

Self-assembled porphyrin dimer **123** through hydrogen bonds between two carboxyl groups with an association constant of 10^7 M^{-1} has been reported (Fig. 56) [140]. This dimer fabricated a cavity to incorporate a pyrazine molecule by coordination to two facing porphyrin centers of **123**. This molecular recognition has been applied to construct a multi-porphyrin system [141]. Thus, the free base porphyrin bearing four pyrazine arms **124** and dimer **123** allowed the formation of nonameric porphyrin assembly **128** with each

Figure 55. Two modes of intermolecular hydrogen bondings between pyrazolyl substituents with and without methyl groups giving porphyrin dimer **121** and tetramer **122**, respectively [139].

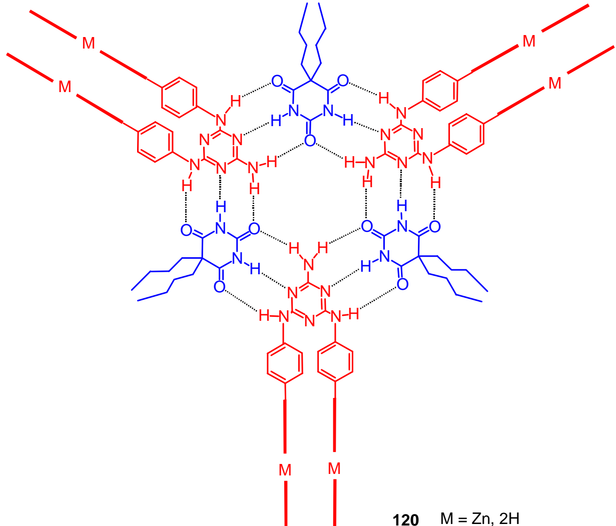


Figure 54. Supramolecular multi-porphyrin macrocycle **120** consisting of three triazine-bridged bisporphyrin units and three barbituric acid units [137].

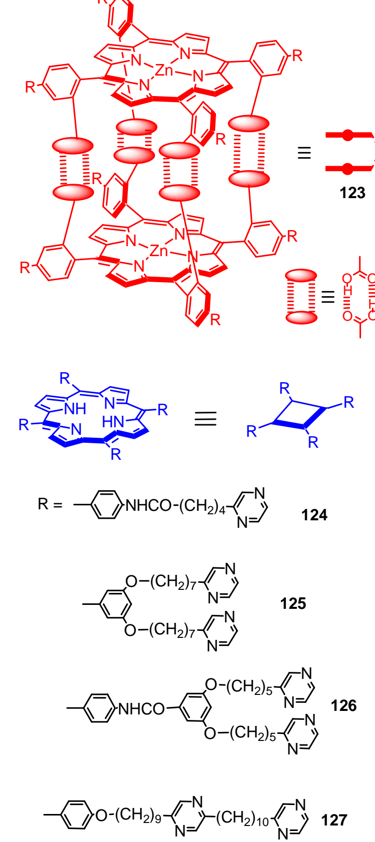


Figure 56. Schematic structure of dimer **123** organized by hydrogen bonds between carboxyl groups and free base porphyrin **124–127** bearing pyrazine pendants at *meso*-positions [140–142].

binding constant of 4×10^7 M^{-1}, as shown in Figure 57. Within the resulting complex, photoinduced energy transfer from zinc porphyrin moiety to free base porphyrin was observed with the energy transfer efficiency of 82%. The same authors synthesized three multi-porphyrin arrays assembling 16 porphyrin units **129–131** from three types of free base porphyrin having eight pyrazine substituents **125–127** by the same methodology [142]. The assemblies **129** and **130** were in a parallel configuration, while **131** was in a serial arrangement. All titration experiments for the formation of these assemblies showed similar complexation behavior and the binding constants of the inner and the outer pyrazine could not be separated. The existence of free **123** in solution was estimated to be less than 8% considering the binding constant. The energy transfer efficiencies were calculated from the fluorescence quenching of zinc porphyrin moieties in the complex as 61, 80, and 55% for **129**, **130**, and **131**, respectively. The reasons for the lowest efficiency observed in **131** having a serial arrangement of antenna units were considered as (i) the single-strand orientation of a side chain by two bulky antenna units constrained the farthest antenna units far from the central free base porphyrin, (ii) the energy transfer between inner and outer antenna units interfered with that from the antenna to the central porphyrin.

Two monomeric Cu(II) porphyrin units of **132** bearing four ammonium cations were tied by four bis-*p*-phenylene-[34]crown-10 molecules to give cofacially arranged supramolecular Cu(II) porphyrin dimer **133**, as shown in Figure 58 [143]. The X-ray analysis of single crystals of **133** with eight PF_6^- from an acetonitrile solution revealed a dimeric structure. The mean interplanar separations between two porphyrins and Cu-Cu separation were 3.65 Å and 4.76 Å, respectively, showing π-π stacking in a slipped cofacial orientation. In addition to the π-π interaction between porphyrins, the supramolecular assembly was stabilized by [N^+—H···O] and [C—H···O] hydrogen bonds between the dibenzylammonium sites of porphyrinsubstituent and the oxygen atoms of macrocyclic polyether linkages.

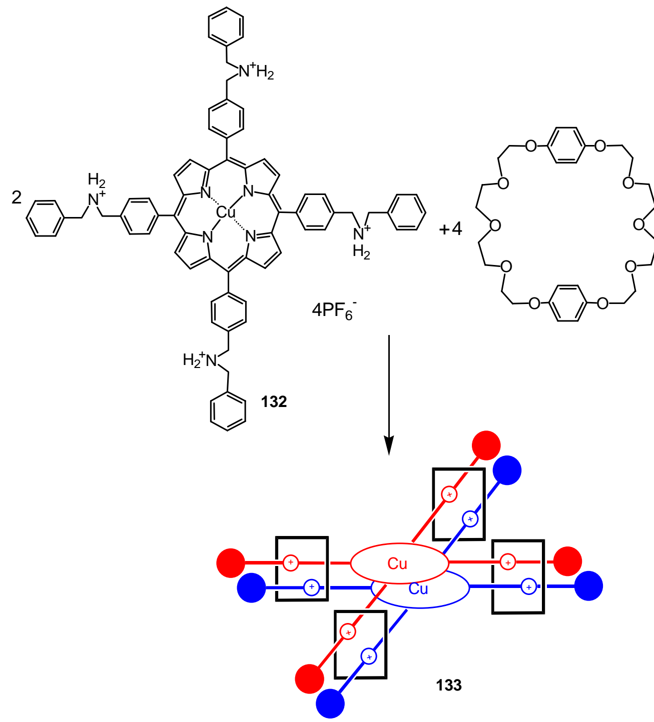


Figure 58. Formation of porphyrin dimer **133** tied by four bis-*p*-phenylene-[34]crown-10 molecules [143].

Supramolecular assembly by imidazolyl-imidazolyl hydrogen bonds afforded stacks of porphyrins [144, 145]. Bis(4(5)-imidazolyl)porphyrin **134** was synthesized to investigate the formation of cooperatively hydrogen-bonded aggregation (Fig. 59A) [144]. Porphyrin-porphyrin interaction in

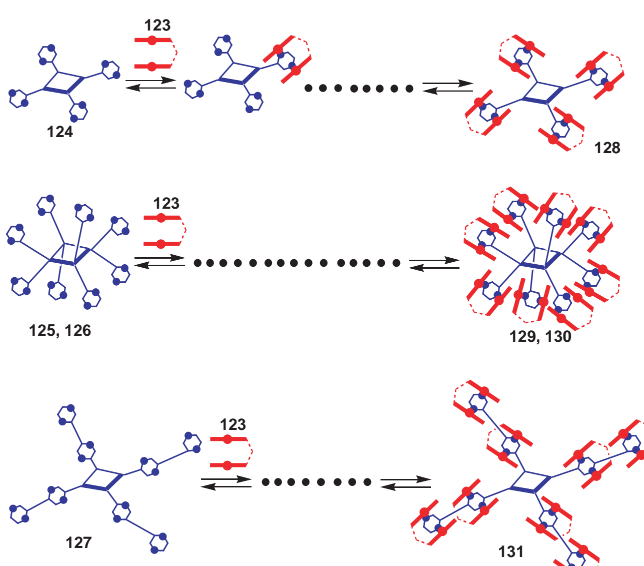


Figure 57. Formation of a porphyrin nonamer and 17-mers by stepwise complexation of **123** with **124–127** [141, 142].

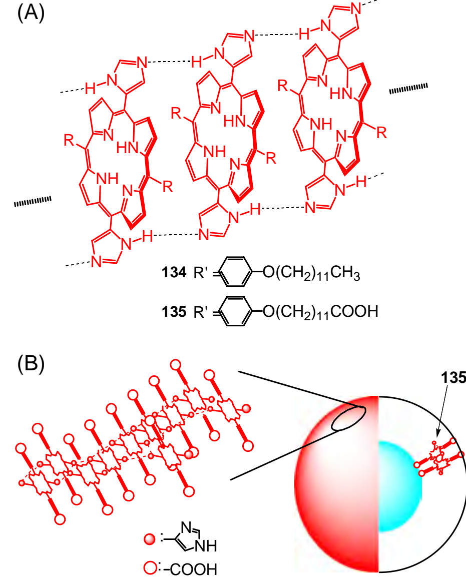


Figure 59. Schematic drawings of one-dimensional aggregation by imidazolyl-imidazolyl hydrogen bonds (A) [144], and of the liposome structure consisted of **135** in an aqueous solution (B) [145].

the resulting assembly of **134** reflected sensitively in their UV spectra. In nonpolar solvent such as toluene and cyclohexane, significant broadening of the Soret band was observed. This observation can be explained by the exciton coupling theory for two porphyrin chromophores. Since mono- and non-imidazolyl substituted porphyrins did not show such peak broadenings, cooperative hydrogen bonds and π-stacking interactions were necessary to form this supramolecular structure. ^{1}H NMR spectrum of **134** in $CDCl_3$ gave broad peaks, which were sharpened gradually by the addition of MeOD. Figure 60 shows the titration behavior of chemical shifts of imidazolyl, pyrrolic, and aromatic protons. The down-field shift with increasing CD_3OD amount was observed indicating gradual breakage of hydrogen bonds between imidazolyl groups. Energy and/or electron transfer reactions were examined from the supramolecule to the externally added acceptor, such as benzoquinone or Mn(III)porphyrin. The increased efficiency by a fact observed for the supramolecule compared with monomeric species was ascribed to delocalization of excitation energy over the whole space of the supramolecule.

This idea of supramolecular organization by imidazolyl-imidazolyl hydrogen bonds has been applied to form liposome by their own intermolecular interactions alone without assistance of any other lipid components [145]. Therefore, amphiphilic bis(imidazolyl)porphyrin **135** having two *meso*-ω-carboxyalkoxyphenyl substituents as hydrophilic head groups was synthesized. Compound **135** was dispersed in water by sonication and subjected to a gel filtration column separation to isolate the fraction corresponding to small unilamellar vesicles (Fig. 59B). The formation of liposome with a mean diameter of 27 ± 8 nm was supported by dynamic light-scattering (DLS) measurements. Figure 61a shows a top view of the AFM image and Figure 61b the side view along the line in Figure 61a. The observed width at the baseline level was 26 nm. A TEM image of the liposome dispersions showed many particles of diameters in the range 20–30 nm corresponding to the size distribution obtained by the DLS measurement, as shown in Figure 61c. The liposome possessed a cavity containing an interior aqueous phase and could keep polar solutes inside the liposome.

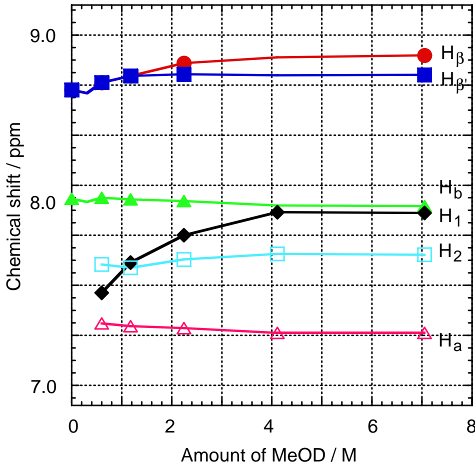


Figure 60. NMR titration of **134** in $CDCl_3$ with CD_3OD.

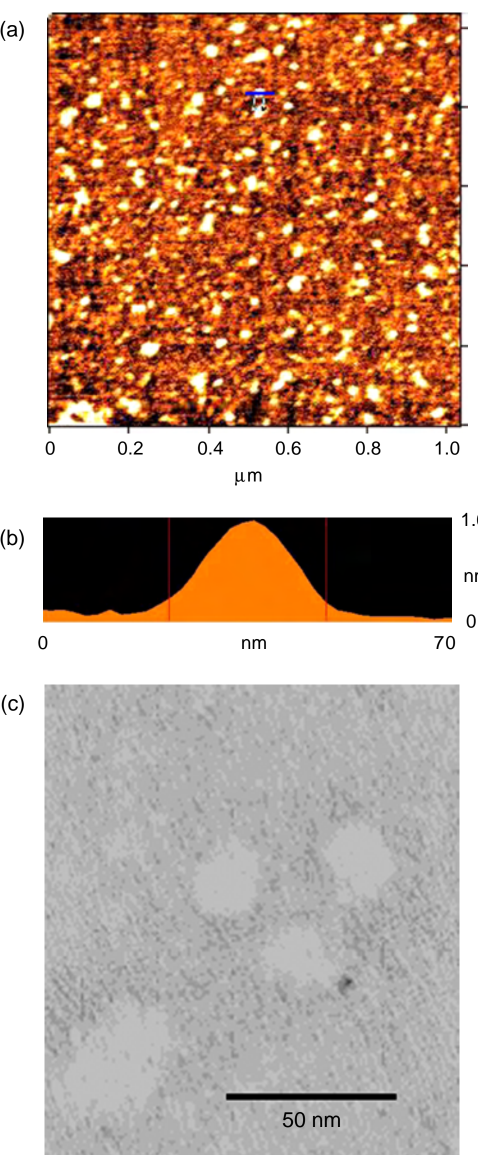


Figure 61. AFM image of dispersion prepared from **135** developed on mica. (a) A top view and (b) a cross section along the line in (a), (c) TEM image of negatively stained samples of dispersion from **135** with uranyl acetate.

The liposome obtained solely from porphyrin amphiphiles is regarded as a huge light-harvesting antenna, which could be regulated in a nanometer-scaled structure.

5. SPECTROSCOPIC CHARACTERIZATION OF SELF-ASSEMBLED PORPHYRIN ARRAYS

If the interaction between porphyrin chromophores in the resulting multi-porphyrin array is strong enough to induce the exiton couplings [92], its absorption spectrum is changed from that of a monomer or subunit, depending on the distance and orientation between porphyrin chromophores. In the self-assembled systems, porphyrins in the array are

arranged in various orientations to each other and investigation on spectral changes allows structural information of the assembly. Furthermore, the absorption and fluorescence spectra shift to longer wavelengths by fifth and sixth axial ligation to metalloporphyrin [146]. Here, we introduce two examples of our self-assembled porphyrin arrays to explain the exciton interaction between porphyrin chromophores.

Figure 62 shows absorption spectra of slipped cofacial dimer **45** (solid line) and its free base monomeric porphyrin **136** (dashed line) in $CHCl_3$ with a schematic representation of two perpendicular transition dipoles, M_x and M_y. In the case of monomeric porphyrin, these transitions are almost degenerated and therefore a single Soret band appears at 414 nm. On the other hand, the Soret band of dimer **45** split into two peaks to appear at 408 and 430 nm, as a result of blue- and red-shifts from the degenerated band, respectively. This splitting behavior can be explained by the exciton theory [92]. As shown in Figure 63, the interaction between two dipoles of M_x orienting along the head-to-tail direction induces two exciton states and the in-phase transition (in the same direction) is allowed, but the out-of-phase transition (in the opposite direction) is forbidden, resulting in the red-shift. In a similar way, the interaction between two M_y components in a face-to-face arrangement causes a blue-shift of the Soret band. These explanations accord with the observed spectrum and may apply to other self-assembled systems arranged in face-to-face or head-to-tail geometry.

In Section 3.3, we discussed the multi-porphyrin array **72** prepared by coupling of the complementary coordination unit of imidazolyl to central zinc **45**. The one-dimensional polymer **72** has a huge molecular weight in chloroform and exhibits large split of the Soret band appearing at around 410 and 490 nm; splitting width was approximately 80 nm, as shown in Figure 64. The splitting behavior of **72** can be explained again by using the exciton interactions caused by *meso-meso* coupling and coordination. Covalently linked porphyrin array at *meso*-positions has been developed first by Osuka and co-workers and its splitting behavior of the Soret band was explained by exciton interaction between

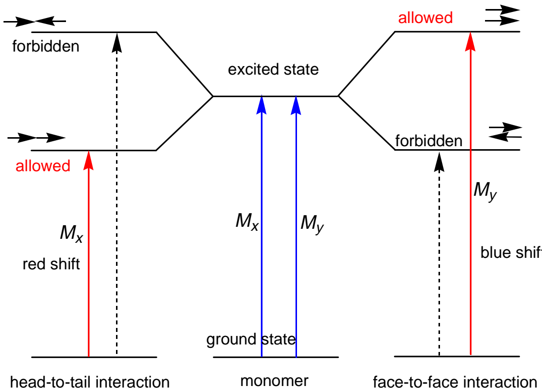


Figure 63. Energy level diagram for splitting behavior of the Soret band of dimer **45** in Figure 62.

M_x components inducing a large red-shift, with no interaction for M_y and M_z due to the orthogonality [147–149]. For example, the Soret band of *meso-meso* coupled free base porphyrin dimer (demetallated **71**) was split into 415 nm for M_y and M_z components and 453 nm for the M_z component; splitting width was 38 nm [108]. The complementary coordination in **72** induces another exciton interaction in a slipped cofacial arrangement. As mentioned earlier, the interaction of M_y's or M_z's components in a face-to-face orientation gives a blue-shift of the transition to 410 nm. On the other hand, M_x's components are shifted to a longer wavelength up to 490 nm by interaction with those of the neighboring porphyrin units. The shift at longer wavelength should depend on the degree of coordination. Therefore, the effect of addition of MeOH, which is a competing ligand to break the zinc-imidazolyl coordination bond, to a $CHCl_3$ solution of **72** was examined by monitoring the absorption spectrum (Fig. 64). The splitting width of the Soret band decreased with the addition of MeOH, resulting in an obvious blue-shift of the longer Soret band passing through a

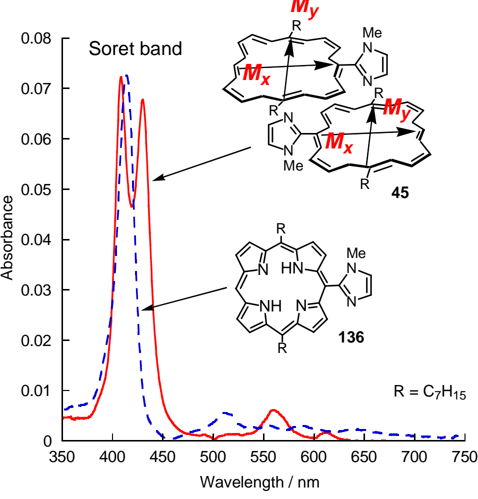


Figure 62. Absorption spectra of complementary coordination dimer of **45** (solid line) and its free base monomer **136** (dashed line) in $CHCl_3$.

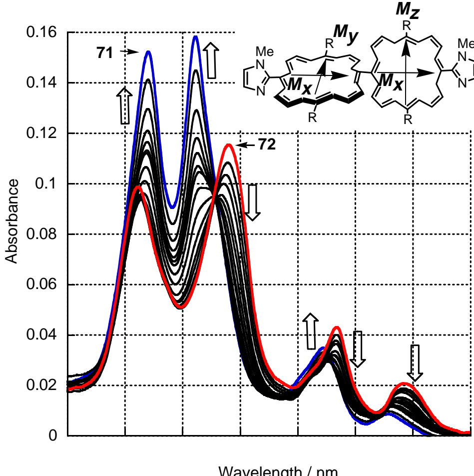


Figure 64. Change of absorption spectra by the addition of MeOH to a $CHCl_3$ solution of **72**. Each spectrum was measured after the addition of MeOH in the amounts 0, 1, 2, 3, 4, 5, 7, 9, 10, 12, 14, 16, 20, 25, and 30 volumes to 100 volume of the $CHCl_3$ according to the order of the arrow indicated.

clear isosbestic point at 478 nm and a smaller red-shift of the shorter Soret band. On the addition up to 30% MeOH, the shift reached almost the saturation point with the splitting of 41 nm corresponding to that of isolated **71**. These results clearly show that the addition of competing ligand such as MeOH can simply and reversibly control the length of self-assembled porphyrin array.

Conjugation of diacetylene linked porphyrin array conducted by Anderson could be controlled by the axial ligand and to construct and to break ladder complexes **80** and **81** (see Fig. 38) [42, 115–117]. The ladder formation was expected to force coplanar arrangement in neighboring porphyrin units, and might increase the conjugation between porphyrins inducing red-shift in absorption spectrum due to the decrease of HOMO-LUMO bandgap energy. As shown in Figure 65, the ladder complex **80** (n = 6 and L = DABCO) exhibited red-shifted Q-band at around 840 nm and Soret bands of a larger split width [116]. The spectral features were changed by the addition of excess DABCO breaking the ladder complex. The blue-shifted Soret band due to weak face-to-face interaction in the ladder complex shifted to longer wavelengths indicating the dissociation of the stack. At the same time, the longest Q band in the ladder shifted to shorter wavelengths because of the lack of conjugation by free rotation about the diacetylene bonds. When 4,4′-bipyridyl was chosen as the bridging ligand L, Q band became sharper and was red-shifted by 16 nm relative to that of the complex with DABCO, indicating more planar arrangement due to the elimination of the steric hindrance between the interstrand aryl groups. Figure 66 compares the absorption spectrum of free base dimer **138** (a), zinc dimer-pyridine complex **137** · (pyridine)$_2$ (b), and triple-strand **81** (c) in CH_2Cl_2. The spectrum (c) was not represented by the simple sum of spectra (a) and (b) indicating that the conformations of **137** and **138** were changed by the formation of the triple-strand **81**. The splitting width of the Soret band increased and the Q band shifted to longer wavelength showing expansion of conjugation due to increase of the planarity too.

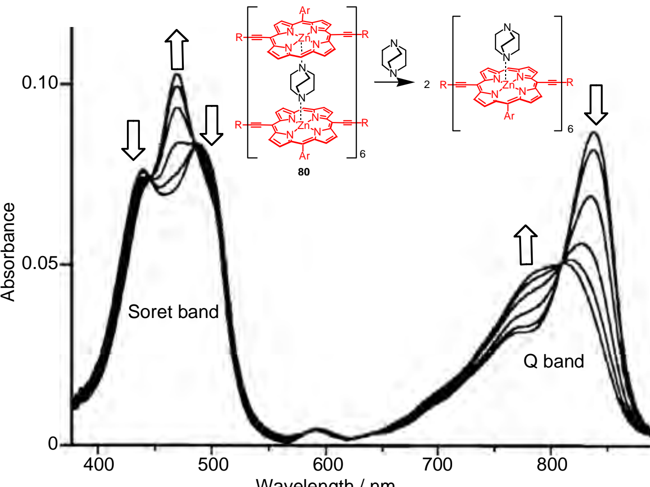


Figure 65. Titration of ladder complex **80** (n = 6 and L = DABCO) with excess DABCO in toluene. Reprinted with permission from [116], P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.* 121, 11538 (1999). © 1999, American Chemical Society.

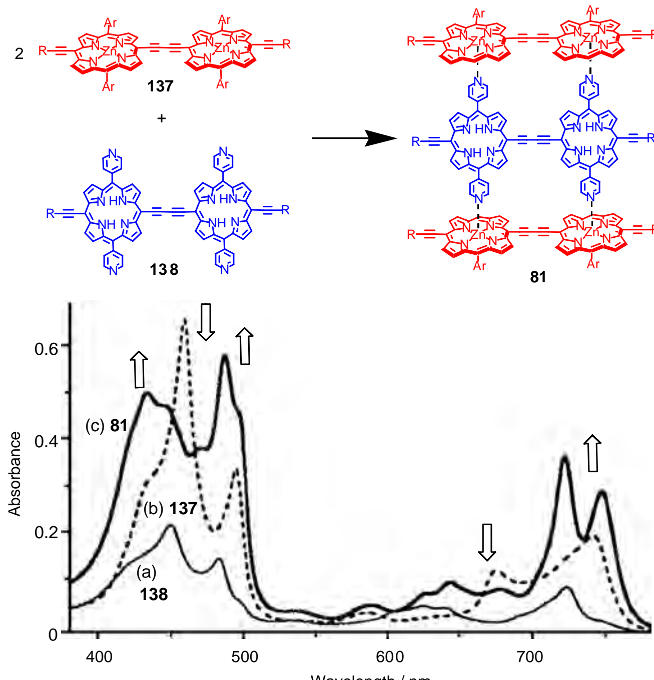


Figure 66. Absorption spectra of (a) **138**, (b) **137**·(pyridine)$_2$ in the presence of excess pyridine and (c) **81** in CH_2Cl_2. Reprinted with permission from [117], G. S. Wilson and H. L. Anderson, *Chem. Commun.* 1539 (1999). © 1999, Royal Chemical Society.

6. PROPERTIES OF SELF-ASSEMBLED PORPHYRIN ARRAYS

One of the goals of research on self-assembled porphyrin array is to develop practically useful materials in the molecular electronic and photonic devices as well as artificial photosynthetic system to convert light energy to electric power. Various types of self-assembled porphyrin arrays fabricated in nanometer-scale have been reported and introduced in this chapter. However, the technology to realize the molecular-sized electronic devices has not been achieved yet by using the multi-porphyrin array as well as other conductive organic molecules. On the other hand, NLO property and photocurrent generation by using multi-porphyrin arrays have been studied extensively in recent years. In this section, we will introduce the importance of noncovalent assembly of porphyrins to develop the NLO materials and highly efficient photocurrent generation system.

Third-order NLO materials with a large real part of susceptibility $\chi^{(3)}_{real}$ are useful for ultrafast optical switching and light modulation device to realize the all-optical networking [29–31]. Furthermore, two-photon absorption (TPA) is one of the third-order NLO phenomena corresponding to the imaginary part $\chi^{(3)}_{im}$ and is a nonlinear absorption process depending on an incident light intensity [150]. TPA materials could find useful applications for 3D optical memory, photodynamic therapy, optical power limiting, 3D microfabrication, and fluorescence microscopy. The strong optical nonlinearities are given in principle by large π-conjugated molecules. Additionally, introduction of donor/acceptor groups at the molecular terminals increases

its polarizability and induces much larger NLO response. As mentioned in Sections 3.4 and 5, the third-order NLO properties of π-conjugated multi-porphyrin arrays of diacetylene linkage have been studied [40–42]. Double-strand polymer **80** (L = 4, 4′-bipyridyl) exhibiting larger conjugation due to planarity between porphyrin units than that of single-strand polymer **139** (Fig. 67) was used for the NLO study to test the effect of increased conjugation on the optical nonlinearity by degenerate four-wave mixing measurement with 50-ps pulse laser at 1064 nm [42]. The number average molecular weights (M_n) of **139** and **80** were determined as 2.1×10^4 and 5.6×10^4 Da, respectively, by GPC analysis. The real part of molecular polarizability per porphyrin unit $|\gamma_{xyyx}|$/N of **80** was determined as 7.6×10^{-45} m^5 V^{-2}, which was nine times amplified relative to that of **139** (0.84×10^{-45} m^5 V^{-2}) by the formation of ladder structure, where SI units of m^5 V^{-2} may be converted to esu unit by multiplying $(4\pi/9) \times 10^{-8}$ [41]. Further, a large two-photon absorption cross section was observed at 1064 nm for **80** as 5×10^4 GM, which was also seven times enhanced by ladder structure formation. These observations were explained by a reasonable assumption of three-level model [40–42] and showed that the self-assembly of double-strand formation might be useful for nonlinear optics.

In addition to expansion of π-conjugation, introduction of donor/acceptor substituents into the array structure to induce higher polarizability is another important factor to increase the NLO property. The self-assembled linear porphyrin array **72** (Fig. 31) can scramble with a specific terminator porphyrin **73** in $CHCl_3$/MeOH or pyridine to afford oligomers **74**. Since the molecular polarizability of the oligomer is expected to increase by choosing the acceptor groups at both molecular ends, two series of porphyrin oligomers were simply synthesized by the reorganization process. One of the series had no acceptor (R′ = n-heptyl) **140–142** and the other was terminated with porphyrins as an acceptor **143–145** (Fig. 68) [43]. Table 1 summarizes results of optical Kerr effect (OKE) measurement with 150-fs pulse laser at an off-resonant wavelength of 800 nm. The $|\gamma_{yyyy}|$

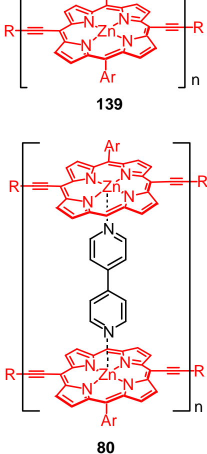


Figure 67. Structures of conjugated porphyrin polymer **139** and double-strand ladder **80** for NLO measurements by DFWM with 50-ps pulse laser [42].

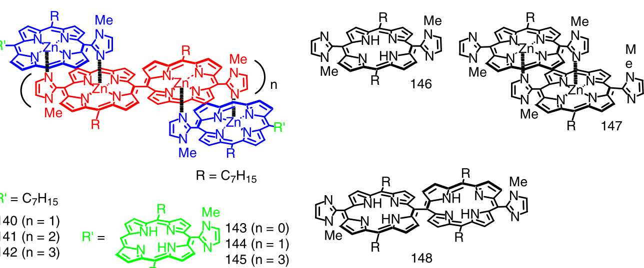


Figure 68. Structures of porphyrins for optical Kerr effect measurements with 150-fs laser [43].

values of oligomers **140–142** were in the range from 10^{-31} to 10^{-30} esu, while the $|\gamma_{yyyy}|$ of the series of compounds **143–145** were extremely large, ranging from 10^{-30} to 10^{-29} esu, 10 times larger than those of **140–142**. The large enhancements of the $|\gamma_{yyyy}|$ values were observed when free base porphyrins were attached at the terminal positions of core zinc porphyrin arrays. The terminal free base should have acted as an acceptor with respect to inner zinc porphyrin. The complementary coordination structure must also have contributed to such a large increment of nonlinear response. There are several reports on NLO materials, but the values are difficult to compare, because methods and conditions are different in each experiment. Here, NLO properties of organic molecules measured under similar conditions to the above, that is, employing an OKE technique with femtosecond pulse laser at off-resonant wave length, will be compared [151, 152]. The γ values, corresponding to $|\gamma_{yyyy}|$, of C_{60} derivative with electron donor, -$(NH_2)_2$CNCN, was determined to be 3.5×10^{-32} esu as the largest among C_{60} derivatives [151]. A donor-acceptor conjugated copolymer consisting of 2,7-diethynylfluorene as the acceptor unit and tetraphenyldiaminobiphenyl as the donor site with average molecular weight (M_w) of 17,600 (M_w/M_n = 2.8) showed the γ value of 450×10^{-32} esu [152]. These values may be comparable to those of the porphyrin system summarized in Table 1. There is a possibility to obtain large nonlinearities by covalent approach to connect donor/acceptor substituents to porphyrin, but it means a difficult synthetic target to develop such systems. The methodology of simple

Table 1. NLO properties of porphyrins by optical Kerr effect measurements.

Porphyrins (no. of units)	$\lvert\gamma_{yyyy}\rvert$ (10^{-32} esu)	γ_{real}/unit (10^{-32} esu)	$\lvert\chi^{(3)}_{yyyy}\rvert$ (10^{-15} esu)	$\chi^{(3)}_{real}$ (10^{-15} esu)	Conc. (10^{-6} M)
140 (4)	13		55		200
141 (6)	61		38		30
142 (8)	120		39		15
143 (4)	180	55	50	57	13
144 (6)	950	120	165	130	8.3
145 (10)	1300	150	163	150	5.0
146 (1)	14		91		320
147 (2)	8		24		150
148 (2)	7		4		25

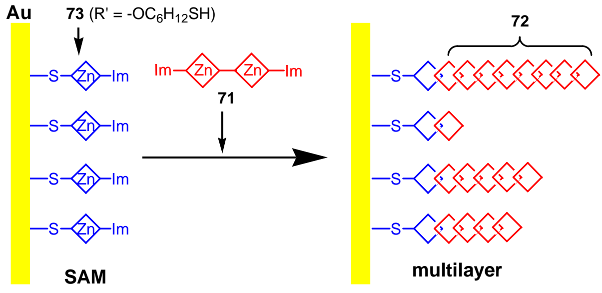


Figure 69. Schematic representation of self-assembled monolayer (SAM) of thiolate and imidazolyl appended monomeric porphyrin and multilayer prepared by accumulation process with bis(imidazolyl-porphyrinatoZn) **71** [109].

reorganization process in the self-assembled system presented here may afford a facile structural modification to gain even further large nonlinearities. This optical nonlinearity observed in a femtosecond time scale is a candidate for the application to ultrafast optical switching.

The self-assembled porphyrin system by complementary zinc-imidazolyl coordination has been applied to photocurrent generation [109]. Thiol group, –SH, is well known to be attached to gold surface to produce stable —S—Au bonds. Terminal monomeric porphyrin **73** bearing an ω-mercaptoalkyl substituent (R′ = -$OC_6H_{12}SH$) was attached onto gold electrodes to obtain a self-assembled monolayer, SAM (Fig. 69). Multi-porphyrin array **72** then propagated on this SAM simply by accumulating bisporphyrin **71** as follows: an aliquot of **72** dissolved in a 1:1 mixture of pyridine/nitrobenzene in order to dissociate the array structure into monomeric **71** was applied dropwise to the electrode and pyridine was evaporated to develop the complementary coordination of imidazolyl to the central zinc. Another aliquot of porphyrin was applied again and the accumulation procedure was repeated. Absorption spectrum of this multilayered electrode showed an increase of light absorption depending on the repeated number of the deposition cycles

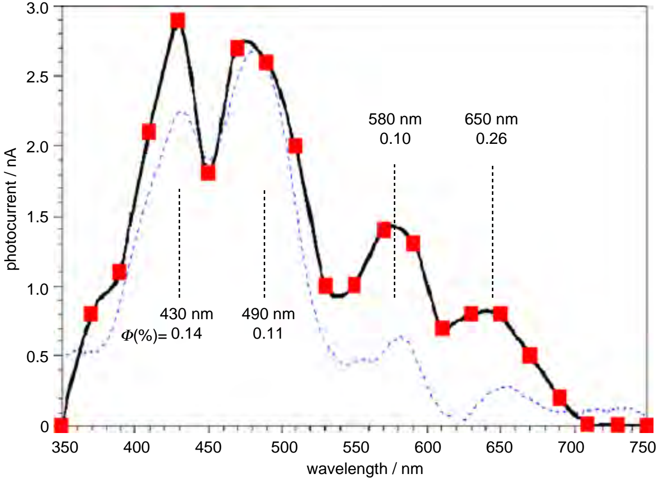


Figure 71. The action spectrum (square plot) with absorption spectrum (dotted line) of multilayered electrode.

in the range of 350 to 750 nm [109, 110]. AFM measurement confirmed long porphyrin assemblies extending from the surface of the electrode (Fig. 70). The heights were not homogeneous with a broad distribution and some assemblies reached a scale of a few hundred nanometers. Irradiation of the multilayered electrode generated photocurrent more efficiently compared with simple SAM. The action spectrum, which was obtained by plotting photocurrent as a function of the irradiating wavelength, resembled the absorption spectra of porphyrin-assembled electrodes in the whole range of 350 to 750 nm (Fig. 71). Photocurrent quantum efficiencies [153] at the respective absorption peaks were estimated as 0.14% (430 nm), 0.11% (490 nm), 0.10% (580 nm), and 0.26% (650 nm). Photocurrent values of SAM and multilayered electrodes (a)–(d) are plotted as a function of the absorption area ($A\lambda$) integrated over the range of 350 to 750 nm in Figure 72. The photocurrent density linearly increased with

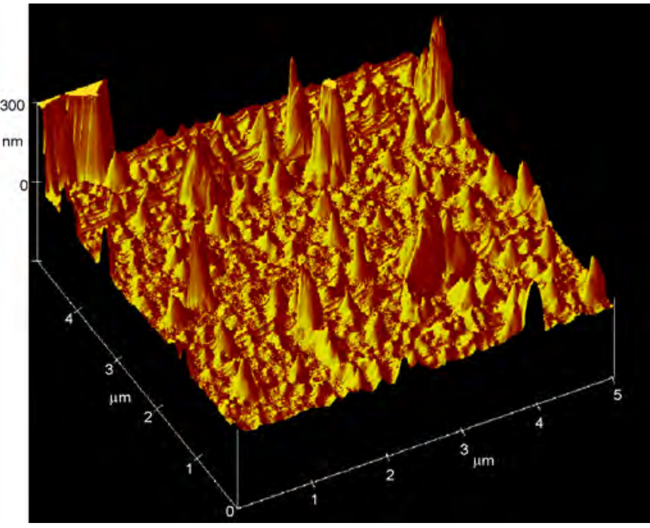


Figure 70. AFM image of multi-porphyrin array propagated from SAM on gold electrode.

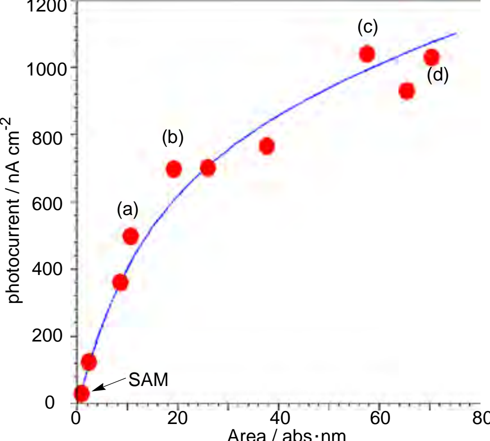


Figure 72. Plot of photocurrent density of SAM and other electrodes as a function of the absorption area integrated over the range 350–750 nm. Concentration and repetition cycles are (a) 1 mM and once (n = 2), (b) 0.1 mM and three times (n = 4), (c) 1 mM and three times (n = 13), and (d) 1 mM and four times (n = 16), respectively.

the increase of the absorption area especially in the range 0–20 $A\lambda$. This increase came from the improved efficiency in absorbing light energy and the excitation energy transfer by perpendicularly accumulated porphyrins from the surface. Further, effective charge separation and electron/hole transfer along the multi-porphyrin chains may be contributing to the photocurrent generation. Although the SAM technique has opened an excellent way to lead realistic materials, the single-molecular layer on the surface is limited by the low efficiency of light absorption. The process by self-assembly accumulating the chromophore to multiply light absorption may propose a strategy for solving the problem.

7. SUMMARY

In the past few years, much attention has been focused on nanostructure generation and nanotechnology, and large bodies of porphyrin architectures have been constructed by using covalent and noncovalent linking methodologies. In this review, we have presented self-assembled porphyrin systems toward photosynthetic mimics and single molecular devices constructed by noncovalent interactions of hydrogen bond, π-π interaction, coordinations of *meso*-substituted ligand to outer metal and other porphyrin centers. In order to build desirable nanostructures with specific function for the practical use, there are several important points on designing porphyrin assemblies: (1) Direction to connect porphyrins is determined by the coordination direction of the *meso*-substituted ligand and coordination geometry around the metal ion. The same is true for multiple hydrogen bonding interactions. (2) Stability of the porphyrin-porphyrin connection depends on the number of interactions, that is, cooperatively multiple bindings are more effective than single binding. (3) If transition metal ions with unoccupied d-orbitals are chosen as the porphyrin center, singlet excitation energy is quenched and therefore fluorescence and photocurrent generation will not be expected. Instead, variable oxidation states of the transition metal are thought to be advantageous for electric conduction and switching of the coordination properties. (4) Strength of the interaction between neighboring porphyrins increases with decrease of the interplanar distance, and then strong interaction causes change in their spectroscopic features and reduction-oxidation properties.

Synthesis and purification of porphyrins are not easy tasks considering their preparative methods associated with low yields and large amounts of by-products. Even so, the porphyrin building block fascinates researchers because of its high π-conjugation system assembled in one nanometer-sized building block. Porphyrins accommodate various metal ions into the center acting as molecular recognition site and allow the introduction of various functional groups at *meso*- and β-positions to construct wonderful nanoarchitectures.

GLOSSARY

Atomic force microscopy (AFM) One of the scanning probe microscopies by detecting the local force operating between sample surface and probe tip. The spatial resolution is usually a few nm.

Association constant Equilibrium constant of forward complexation reaction; $L + M = LM$, $K = [LM]/([L] \times [M])$. Synonymous with stability constant. Reciprocal of dissociation constant.

Chelate Metal complexation by multidentate ligand. The word "chelate" came from craw of crab in Greek.

Chemical shift In the NMR spectroscopy, the resonant frequency depends on the effective magnetic field around the nuclei, which are affected by induced magnetic field by electron. The shielding effect increases with an electron density. The chemical shift is proportional to strength of the magnetic field and represented as the ratio of frequency with that of reference in ppm.

Cyclic voltammetry Electrochemical method to determine the oxidation/reduction potential.

Dendrimer Well-regulated highly branched polymer.

Dynamic light scattering (DLS) Characterization of particle size from diffusion coefficient given by the light scattering measurement.

Fluorescence Emission of light from excited singlet state after the absorption of light energy.

Fluorescence anisotropy decay The measurement of anisotropy as a function of time by laser pulses provides the rate of the excited energy transfer or migration.

Fluorescence quantum yield A ration of the number of emitted photons to the number of absorbed photons.

Förster model Energy transfer theory explained as dipole-dipole interaction between donor and acceptor. In this model, the emission band of donor must overlap with the absorption wavelength of acceptor.

Gel permeation chromatography (GPC) Analytical or preparative method to separate a mixture based on difference in molecular size or shape. GPC analysis gives an information on molecular weight by comparison with appropriate standard samples.

Hard-soft-acid-base rule (HSAB rule) One of the measures for relative strengths of acids and bases. An acid, usually a metal ion, prefers the bases having similar softness or hardness.

Heavy-atom effect See the singlet-triplet intersystem crossing.

Highest occupied molecular orbital and lowest unoccupied molecular orbital (Homo-Lumo) Usually, HOMO and LUMO correspond to the ground state (S_0) and the lowest excited state (S_1), respectively.

Hydrophobic interaction Hydrophobic solutes strongly interact with each other in hydrophilic media such as in water.

Liposome Spherical bilayer membrane structure entrapping inner aqueous phase with a size of around a few nm to a few μm.

Mass spectroscopy Determination of molecular weight by detecting the positively or negatively charged molecular ion. There are several ionization methods such as FAB (fast-atom-bombardment), ESI (electrospray ionization), MALDI (matrix-assisted-laser-desorption-ionization), and so on.

Molecular mechanics calculation A calculation method giving optimized-structure and energy for conformations of

molecules based on the empirical bond lengths and angles, strain, torsion, steric repulsion, van der Waals interaction, and electrostatic interaction.

Nuclear magnetic resonance spectroscopy (NMR) NMR gives information on the chemical environment around nuclei. The nuclei having no spin such as ^{12}C exhibits no magnetic resonance, but if net spin is not zero as seen in ^{1}H and ^{13}C, we can observe NMR signals allowing information about chemical structure such as chemical shift, spin coupling, and integration.

Optical kerr effect (OKE) Refractive index of molecule or media is proportional to the incident light intensity.

π-π interaction π-Stacking interaction due to overlapping of π-electron cloud between aromatic groups.

Peptide Amide bond, -NH-CO-, formed from amino group of an amino acid and carboxy group of another amino acid.

Ring current effect The shielding effect, which comes from the magnetic field generated by π-electron system. Effective magnetic field of nuclei is shielded by this effect to reduce the chemical shift.

Singlet-triplet intersystem crossing Transition from single/triplet state to triplet/single state. In principle, single-triplet transition can not occur due to spin-forbidden. However, we can observe the triplet state transition from the singlet state mostly by the presence of spin-orbit coupling. Facile intersystem crossing can be expected when a molecule contains a heavy atom because of large spin-orbit coupling (heavy atom effect).

Small-angle X-ray scattering measurements (SAXS) SAXS provides information on the size and shape of macromolecule or colloid in the spatial range from 1 to 300 nm.

Space-filling model Same with CPK model developed by Corey-Pauling-Koltun. The tool for molecular graphics and modeling based on van der Waals radii.

Superexchange mechanism Electron transfer reaction, which occurred through chemical bond allowing long-range electron transfer even at low temperature. Molecular orbital of bridging or spacer group might be used as the electron path from donor to acceptor.

TEM (transmission electron microscopy) TEM works like a optical microscope by using electrons instead of light. The shade of transmitted electrons through the sample was detected by fluorescence screen attaining a possible spatial resolution higher than 1 nm.

Transient absorption measurement Time resolved absorption spectrum measurements providing information on an excited state of a molecule and a lifetime of excited state or charge separated state by using laser pulses as the excitation light.

van der Waals interactions Interactions between instantaneous electrical polarization induced by electron movement in the molecule, called van der Waals dispersion forces.

Vapor phase osmometry (VPO) Determination of a mean molecular weight by vapor-pressure and osmometric measurements of a sample solution.

X-ray crystallography Powerful tool for determination of three-dimensional structure of molecule by measuring X-ray diffraction of a single crystal. Precision atomic coordinates, thermal parameters, and symmetric properties can be obtained.

REFERENCES

1. K. M. Smith, Ed., "Porphyrins and Metalloporphyrins." Elsevier, Amsterdam, 1976.
2. D. Dolphin, Ed., "The Porphyrins." Academic Press, New York, 1978.
3. K. Kadish, K. M. Smith, and R. Guilard, Eds., "The Porphyrin Handbook." Academic Press, New York, 1999.
4. G. McDermott, S. M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz, R. J. Cogdell, and N. W. Isaacs, *Nature* 374, 517 (1995).
5. J. Koepke, X. Hu, C. Muenke, K. Schulten, and H. Michel, *Structure* 4, 581 (1996).
6. K. McLuskey, S. M. Prince, R. J. Cogdell, and N. W. Isaacs, *Biochemistry* 40, 8783 (2001).
7. J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel, *J. Mol. Biol*. 180, 385 (1984).
8. J. P. Allen, G. Feher, T. O. Yeates, D. C. Rees, J. Deisenhofer, H. Michel, and R. Huber, *Proc. Natl. Acad. Sci. USA* 83, 8589 (1986).
9. J. P. Allen, G. Feher, T. O. Yeates, H. Komiya, and D. C. Rees, *Proc. Natl. Acad. Sci. USA* 84, 5730 (1987).
10. J. Deisenhofer, O. Epp, I. Sinning, and H. Michel, *J. Mol. Biol*. 246, 429 (1995).
11. A. Zouni, H.-T. Witt, J. Kern, P. Fromme, N. Krauss, W. Saenger, and P. Orth, *Nature* 409, 739 (2001).
12. P. Jordan, P. Fromme, H. T. Witt, O. Kukas, W. Saenger, and N. Krauss, *Nature* 411, 909 (2001).
13. L. Stryer, "Biochemistry," 4th ed. W. H. Freeman and Campany, New York, 1995.
14. D. Voet and J. G. Voet, "Biochemistry," 3rd ed. Wiley, New York, 2002.
15. S. J. Lippard and J. M. Berg, "Principles of Bioinorganic Chemistry." University Science Books, California (1994).
16. J.-M. Lehn, "Supramolecular Chemistry: Concepts and Perspectives." VCH, Weinheim, 1995.
17. J. L. Atwood, J. E. D. Davies, D. D. Macnicol, and F. Vögtle, Eds., "Comprehensive Supramolecular Chemistry." Pergamon, Oxford, 1996.
18. H. J. Schneider and A. K. Yatsimirsky, Eds., "Principles and Methods in Supramolecular Chemistry." Wiley, Chichester, 2000.
19. W. Jones and C. N. R. Rao, Eds., "Supramolecular Organization and Materials Design." Cambridge University Press, Cambridge, 2002.
20. J.-C. Chambron, V. Heitz, and J.-P. Sauvage, in "The Porphyrin Handbook" (K. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 6, Chap. 40, p. 1. Academic Press, New York, 1999.
21. A. Aviram and M. A. Ratner, *Chem. Phys. Lett.* 29, 277 (1974).
22. A. Aviram and M. A. Ratner, Eds., "Molecular Electronics—Science and Technology." Annals of the New York Academy of Sciences, New York, 1998.
23. F. L. Carter, Ed., "Molecular Electronic Devices II." Marcel Dekker, New York, 1987.
24. R. W. Keyes, *Phys. Today* 45, 42 (1992).
25. Y. Wada, in "Molecular Electronics—Science and Technology" (A. Aviram and M. A. Ratner, Eds.), Vol. 852, p. 257. Annals of the New York Academy of Sciences, New York, 1998.
26. H. Dai, J. Kong, C. Zhou, N. Franklin, T. Tombler, A. Cassell, S. Fan, and M. Chapline, *J. Phys. Chem. B* 103, 11246 (1999).
27. Z. Yao, H. W. Ch. Postma, L. Balents, and C. Dekker, *Nature* 402, 273 (1999).

28. J.-H. Chou, M. E. Kosal, H. S. Nalwa, N. A. Rakow, and K. S. Suslick, in "The Porphyrin Handbook" (K. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 6, Chap. 41, p. 43. Academic Press, New York, 1999.
29. P. N. Prasad and D. J. Williams, Eds., "Introduction to Nonlinear Optical Effects in Molecules and Polymers." Wiley, New York, 1991.
30. J. Zyss, Ed., "Molecular Nonlinear Optics: Materials, Physics and Devices." Academic Press, Boston, 1994.
31. C. Bosshard, K. Sutter, P. Prêtre, J. Hulliger, M. Flörsheimer, P. Kaatz, and P. Günter, "Organic Nonlinear Optical Materials." Gordon and Breach, Amsterdam, 1995.
32. G. Chen and S. Mukamel, *J. Phys. Chem*. 100, 11080 (1996).
33. R. R. Tykwinski, U. Gubler, R. E. Martin, F. Diederich, C. Bosshard, and P. Günter, *J. Phys. Chem*. *B* 102, 4451 (1998).
34. S. Hahn, D. Kim, and M. Cho, *J. Phys. Chem. B* 103, 8221 (1999) and references therein.
35. C. Maloney, H. Byrne, W. M. Dennis, W. Blau, and J. M. Kelly, *Chem. Phys.* 121, 21 (1988).
36. T. Sakaguchi, Y. Shimizu, M. Miya, T. Fukumi, K. Ohta, and A. Nagata, *Chem. Lett.* 1992, 281.
37. S. Guha, K. Kang, P. Porter, J. F. Roach, D. E. Remy, F. J. Aranda, and D. V. G. L. N. Rao, *Opt. Lett.* 17, 264 (1992).
38. F. Z. Henari, W. J. Blau, L. R. Milgrom, G. Yahioglu, D. Phillips, and J. A. Lacey, *Chem. Phys. Lett.* 267, 229 (1997).
39. M. Terazima, H. Shimizu, and A. Osuka, *J. Appl. Phys*. 81, 2946 (1997).
40. J. R. G. Thorne, S. M. Kuebler, R. G. Denning, I. M. Blake, P. N. Taylor, and H. L. Anderson, *Chem. Phys.* 248, 181 (1999).
41. S. M. Kuebler, R. G. Denning, and H. L. Anderson, *J. Am. Chem. Soc.* 122, 339 (2000).
42. T. E. O. Screen, J. R. G. Thorne, R. G. Denning, D. G. Bucknall, and H. L. Anderson, *J. Am. Chem. Soc.* 124, 9712 (2002).
43. K. Ogawa, T. Zhang, K. Yoshihara, and Y. Kobuke, *J. Am. Chem. Soc.* 124, 22 (2002).
44. M. Fujita, J. Yazaki, and K. Ogura, *J. Am. Chem. Soc.* 112, 5645 (1990).
45. M. Fujita, F. Ibukuro, H. Hagihara, and K. Ogura, *Nature* 367, 720 (1994).
46. M. Fujita, *Chem. Soc. Rev.* 27, 417 (1998).
47. C. M. Drain and J.-M. Lehn, *J. Chem. Soc., Chem. Commun.* 2313 (1994).
48. R. V. Slone and J. T. Hupp, *Inorg. Chem.* 36, 5422 (1997).
49. C. M. Drain, F. Nifiatis, A. Vasenko, and J. D. Batteas, *Angew. Chem. Int. Ed.* 37, 2344 (1998).
50. P. J. Stang, J. Fan, and B. Olenyuk, *Chem. Commun.* 1453 (1997).
51. J. Fan, J. A. Whiteford, B. Olenyuk, M. D. Levin, P. J. Stang, and E. B. Fleischer, *J. Am. Chem. Soc.* 121, 2741 (1999).
52. E. Alessio, E. Ciani, E. Iengo, V. Yu. Kukushkin, and L. G. Marzilli, *Inorg. Chem.* 39, 1434 (2000).
53. K. E. Splan, M. H. Keefe, A. M. Massari, K. A. Walters, and J. T. Hupp, *Inorg. Chem*. 41, 619 (2002).
54. E. Iengo, E. Zangrando, R. Minatel, and E. Alessio, *J. Am. Chem. Soc*. 124, 1003 (2002).
55. H. Yuan, L. Thomas, and L. K. Woo, *Inorg. Chem.* 35, 2808 (1996).
56. A. Prodi, C. J. Kleverlaan, M. T. Indelli, F. Scandola, E. Alessio, and E. Iengo, *Inorg. Chem.* 40, 3498 (2001).
57. E. Iengo, R. Minatel, B. Milani, L. G. Marzilli, and E. Alessio, *Eur. J. Inorg. Chem.* 609 (2001).
58. E. Iengo, B. Milani, E. Zangrando, S. Geremia, and E. Alessio, *Angew. Chem., Int. Ed.* 39, 1096 (2000).
59. N. Fujita, K. Biradha, M. Fujita, S. Sakamoto, and K. Yamaguchi, *Angew. Chem. Int. Ed.* 40, 1718 (2001).
60. Y. Tomohiro, A. Satake, and Y. Kobuke, *J. Org. Chem.* 66, 8442 (2001).
61. A. M. Brun, S. J. Atherton, A. Harriman, V. Heitz and J.-P. Sauvage, *J. Am. Chem. Soc*. 114, 4632 (1992).
62. M. Momenteau, F. Le Bras, and B. Loock, *Tetrahedron Lett.* 35, 3289 (1994).
63. D. B. Amabilino and J.-P. Sauvage, *New J. Chem*. 22, 395 (1998).
64. D. B. Amabilino, C. O. Dietrich-Buchecker, and J.-P. Sauvage, *J. Am. Chem. Soc.* 118, 3285 (1996).
65. M. Linke, J.-C. Chambron, V. Heitz, and J.-P. Sauvage, *J. Am. Chem. Soc.* 119, 11329 (1997).
66. A. M. Brun, A. Harriman, V. Heitz, and J.-P. Sauvage, *J. Am. Chem. Soc*. 113, 8657 (1991).
67. M. Linke, J.-C. Chambron, V. Heitz, J.-P. Sauvage, and V. Semetey, *Chem. Commun*. 2469 (1998).
68. M. J. Crossley, P. L. Burn, S. J. Langford, and J. K. Prashar, *J. Chem. Soc., Chem. Commun*. 1921 (1995).
69. T. A. Vannelli and T. B. Karpishin, *Inorg. Chem.* 39, 340 (2000).
70. A. Harriman, F. Odobel, and J.-P. Sauvage, *J. Am. Chem. Soc.* 116, 5481 (1994).
71. A. Harriman, F. Odobel, and J.-P. Sauvage, *J. Am. Chem. Soc.* 117, 9461 (1995).
72. L. Flamigni, F. Barigelletti, N. Armaroli, B. Ventura, J.-P. Collin, J.-P. Sauvage, and J. A. G. Williams, *Inorg. Chem.* 38, 661 (1999).
73. R. Ballardini, G. Varani, M. T. Indelli, and F. Scandola, *Inorg. Chem.* 25, 3858 (1986).
74. H. Schmidbaur, J. Lettenbauer, D. L. Wilkinson, G. Müller, and O. Kumberger, *Z. Naturforsch.* 46b, 901 (1991).
75. I. V. Rubtsov, Y. Kobuke, H. Miyaji, and K. Yoshihara, *Chem. Phys. Lett*. 308, 323 (1999).
76. Y. Kobuke, H. Miyaji, and K. Ogawa, *Supramol. Chem.* 14, 159 (2002).
77. V. Thanabal and V. Krishnan, *J. Am. Chem. Soc.* 104, 3643 (1982).
78. V. Thanabal and V. Krishnan, *Inorg. Chem.* 21, 3606 (1982).
79. B. Maiya and V. Krishnan, *Inorg. Chem.* 24, 3253 (1985).
80. H. Shinmori, Y. Yasuda, and A. Osuka, *Eur. J. Org. Chem.* 1197 (2002).
81. H. Shinmori and A. Osuka, *Tetrahedron Lett*. 41, 8527 (2000).
82. J. K. M. Sanders, N. Bampos, Z. Clyde-Watson, S. L. Darling, J. C. Hawley, H.-J. Kim, C. C. Mak, and S. J. Webb, in "The Porphyrin Handbook" (K. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 3, Chap. 15, p. 1. Academic Press, New York, 1999.
83. N. A. Rakow and K. S. Suslick, *Nature* 406, 710 (2000).
84. Y. Kobuke and H. Miyaji, *J. Am. Chem. Soc*. 116, 4111 (1994).
85. R. T. Stibrany, J. Vasudevan, S. Knapp, J. A. Potenza, T. Emge, and H. J. Schugar, *J. Am. Chem. Soc.* 118, 3980 (1996).
86. K. Funatsu, T. Imamura, A. Ichimura, and Y. Sasaki, *Inorg. Chem.* 37, 4986 (1998).
87. H. M. Goff, E. T. Shimomura, Y. J. Lee, and W. R. Scheidt, *Inorg. Chem.* 23, 315 (1984).
88. G. M. Godziela, D. Tilotta, and H. M. Goff, *Inorg. Chem.* 25, 2142 (1986).
89. C. A. Hunter and L. D. Sarson, *Angew. Chem., Int. Ed. Engl*. 33, 2313 (1994).
90. U. Michelsen and C. A. Hunter, *Angew. Chem., Int. Ed.* 39, 764 (2000).
91. N. Wakao, N. Yokoi, N. Isoyama, A. Hiraishi, K. Shimada, M. Kobayashi, H. Kise, M. Iwaki, S. Itoh, S. Takaichi, and Y. Sakurai, *Plant Cell Physiol.* 37, 889 (1996).
92. M. Kasha, H. R. Rawls, and M. A. El-bayyoumi, *Pure Appl. Chem.* 11, 371 (1965).
93. Y. Inaba, A. Nomoto, and Y. Kobuke, unpublished work.
94. A. K. Burrell, D. L. Officer, D. C. W. Reid, and K. Y. Wild, *Angew. Chem., Int. Ed.* 37, 114 (1998).
95. X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, and L. D. Sarson, *J. Chem. Soc., Chem. Commun.* 2563 (1995).
96. X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, and L. D. Sarson, *J. Chem. Soc., Chem. Commun.* 2567 (1995).
97. C. Ikeda, Y. Tanaka, T. Fujihara, Y. Ishii, T. Ushiyama, K. Yamamoto, N. Yoshioka, and H. Inoue, *Inorg. Chem.* 40, 3395 (2001).

98. E. B. Fleischer and A. M. Shachter, *Inorg. Chem.* 30, 3763 (1991).
99. K. Funatsu, T. Imamura, A. Ichimura, and Y. Sasaki, *Inorg. Chem.* 37, 1798 (1998).
100. A. Tsuda, T. Nakamura, S. Sakamoto, K. Yamaguchi, and A. Osuka, *Angew. Chem., Int. Ed.* 41, 2817 (2002).
101. R. A. Haycock, C. A. Hunter, D. A. James, U. Michelsen, and L. R. Sutton, *Org. Lett.* 2, 2436 (2000).
102. R. Takahashi and Y. Kobuke, *J. Am. Chem. Soc.*, in press (2003).
103. M. C. Chang, P. M. Callahan, P. S. Parkes-Loach, T. M. Cotton, and P. A. Loach, *Biochemistry* 29, 421 (1990).
104. Y. Kobuke and H. Miyaji, *Bull. Chem. Soc. Jpn*. 69, 3563 (1996).
105. H. Miyaji and Y. Kobuke, *Mol. Cryst. Liq. Cryst*. 276, 283 (1996).
106. Y. Kobuke and N. Nagata, *Mol. Cryst. Liq. Cryst*. 342, 51 (2000).
107. H. Miyaji, Y. Kobuke, and J. Kondo, *Chem. Lett.* 497 (1996).
108. K. Ogawa and Y. Kobuke, *Angew. Chem., Int. Ed*. 39, 4070 (2000).
109. A. Nomoto and Y. Kobuke, *Chem. Commun.* 1104 (2002).
110. Y. Kobuke and K. Ogawa, *Bull. Chem. Soc. Jpn.*, in press (2003).
111. H. L. Anderson, C. A. Hunter, and J. K. M. Sanders, *J. Chem. Soc., Chem. Commun.* 226 (1989).
112. H. L. Anderson, C. A. Hunter, M. N. Meah, and J. K. M. Sanders, *J. Am. Chem. Soc.* 112, 5780 (1990).
113. A. V. Chernook, A. M. Shulga, E. I. Zenkevich, U. Rempel, and C. von Borczyskowski, *J. Phys. Chem.* 100, 1918 (1996).
114. C. A. Hunter and R. K. Hyde, *Angew. Chem., Int. Ed. Engl.* 35, 1936 (1996).
115. H. L. Anderson, *Inorg. Chem*. 33, 972 (1994).
116. P. N. Taylor and H. L. Anderson,*J. Am. Chem. Soc.* 121, 11538 (1999).
117. G. S. Wilson and H. L. Anderson, *Chem. Commun.* 1539 (1999).
118. E. Iengo, E. Zangrando, S. Geremia, R. Graff, B. Kieffer, and E. Alessio, *Chem. Eur. J*. 8, 4670 (2002).
119. E. Alessio, M. and B. G. Maiya, *J. Chem. Soc., Chem. Commun.* 1411 (1996).
120. N. Kariya, T. Imamura, and Y. Sasaki, *Inorg. Chem.* 36, 833 (1997).
121. A. Kimura, K. Funatsu, T. Imamura, H. Kido, and Y. Sasaki, *Chem. Lett.* 207 (1995).
122. K. Funatsu, A. Kimura, T. Imamura, A. Ichimura, and Y. Sasaki, *Inorg. Chem.* 36, 1625 (1997).
123. N. Kariya, T. Imamura, and Y. Sasaki, *Inorg. Chem.* 36, 1658 (1998).
124. E. Alessio, S. Geremia, S. Mestroni, E. Iengo, I. Srnova, and M. Slouf, *Inorg. Chem.* 38, 869 (1999).
125. A. Okumura, K. Funatsu, Y. Sasaki, and T. Imamura, *Chem. Lett.* 779 (1999).
126. S. L. Darling, C. C. Mak, N. Bampos, N. Feeder, S. J. Teat, and J. K. M. Sanders, *New J. Chem.* 23, 359 (1999).
127. H.-J. Kim, N. Bampos, and J. K. M. Sanders, *J. Am. Chem. Soc.* 121, 8120 (1999).
128. J. E. Redman, N. Feeder, S. J. Teat, and J. K. M. Sanders, *Inorg. Chem.* 40, 2486 (2001).
129. E. Stulz, Y.-F. Ng, S. M. Scott, and J. K. M. Sanders, *Chem. Commun.* 524 (2002).
130. S. Anderson, H. L. Anderson, and J. K. M. Sanders, *Acc. Chem. Res*. 26, 469 (1993).
131. S. Anderson, H. L. Anderson, and J. K. M. Sanders, *J. Chem. Soc., Perkin Trans.* 1 2247 (1995).
132. S. Anderson, H. L. Anderson, and J. K. M. Sanders, *J. Chem. Soc., Perkin Trans.* 1 2255 (1995).
133. A. Harriman, D. J. Magda, and J. L. Sessler, *J . Phys. Chem.* 95, 1530 (1991).
134. J. L. Sessler, B. Wangt, and A. Harriman, *J. Am. Chem. Soc.* 117, 704 (1995).
135. C. M. Drain, R. Fischer, E. G. Nolen, and J.-M. Lehn, *J. Chem. Soc., Chem. Commun.* 243 (1993).
136. X. Shi, K. M. Barkigia, J. Fajer, and C. M. Drain, *J. Org. Chem.* 66, 6513 (2001).
137. C. M. Drain, K. C. Russell, and J.-M. Lehn, *J. Chem. Soc., Chem. Commun.* 337 (1996).
138. K. F. Morris and C. S. Johnson, Jr., *J. Am. Chem. Soc.* 115, 4291 (1993).
139. C. Ikeda, N. Nagahara, E. Motegi, N. Yoshioka, and H. Inoue, *Chem Commun.* 1759 (1999).
140. Y. Kuroda, A. Kawashima, Y. Hayashi, and H. Ogoshi, *J. Am. Chem. Soc.* 119, 4929 (1997).
141. Y. Kuroda, K. Sugou, and K. Sasaki, *J. Am. Chem. Soc.* 122, 7833 (2000).
142. K. Sugou, K. Sasaki, K. Kitajima, T. Iwaki, and Y. Kuroda, *J. Am. Chem. Soc.* 124, 1182 (2002).
143. M. C. Feiters, M. C. T. Fyfe, M.-V. Martínez-Díaz, S. Menzer, R. J. M. Nolte, J. F. Stoddart, P. J. M. van Kan, and D. J. Williams, *J. Am. Chem. Soc.* 119, 8119 (1997).
144. N. Nagata, S. Kugimiya, and Y. Kobuke, *Chem. Commun.* 1389 (2000).
145. N. Nagata, S. Kugimiya, and Y. Kobuke, *Chem. Commun.* 689 (2001).
146. R. Humphry-Baker and K. Kalyanasundaram, *J. Photochem.* 31, 105 (1985).
147. A. Osuka and H. Shimidzu, *Angew. Chem. Int. Ed. Engl.* 36, 135 (1997).
148. A. Nakano, A. Osuka, I. Yamazaki, T. Yamazaki, and Y. Nishimura, *Angew. Chem. Int. Ed.* 37, 3023 (1998).
149. T. Ogawa, Y. Nishimoto, N. Yoshida, N. Ono, and A. Osuka, *Angew. Chem. Int. Ed.* 38, 176 (1999).
150. S. Kershaw, in "Characterization Techniques and Tabulations for Organic Nonlinear Optical Materials" (M. G. Kuzyk and C. W. Dirk, Eds.), Chap. 7. Marcel Dekker, New York, 1998.
151. S. Wang, W. Huang, R. Liang, and Q. Gong, *Phys. Rev. B* 63, 153408 (2001).
152. X. Zhan, Y. Liu, D. Zhu, W. Huang, and Q. Gong, *Chem. Mater.* 13, 1540 (2001).
153. A. Aoki, Y. Abe, and T. Miyashita, *Langmuir* 15, 1463 (1999).

www.aspbs.com/enn

Self-Assembled Porphyrinic Nanoarchitectures

Xin Chen

City University of New York, New York, USA

Charles Michael Drain

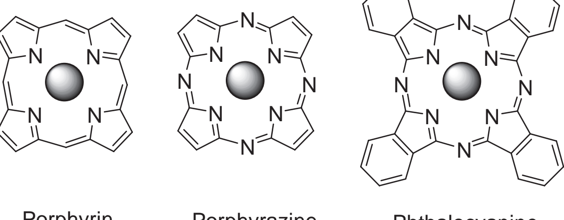
City University of New York, New York, USA and
The Rockefeller University, New York, New York, USA

CONTENTS

1. Introduction
2. Self-Assembled and Self-Organized Porphyrinic Materials—Nomenclature and Descriptors
3. Assemblies Mediated by Coordination Chemistry
4. Assemblies Mediated by Hydrogen Bonding
5. Assemblies Mediated by Electrostatics
6. Porphyrins on Surfaces
7. Miscellaneous Multi-Porphyrin Assemblies and Aggregates
8. Summary
Glossary
References

1. INTRODUCTION

The manifold applications of porphyrins, porphyrazines, and phthalocyanines derive from their photophysical and electrochemical properties, their remarkable stability, as well as their predictable and rigid structure. These applications range from oxidative catalysts, to photonic materials, to therapeutics. The properties of all three pigments are modulated by appending various chemical moieties onto the macrocycles, by choice of metallo derivative, and by the choice of environment. In materials containing multichromophoric



systems, furthermore, the relative orientation of the chromophores, the nature of the linker, and the size of the system also dictate the properties and functions. Nanoscaled assemblies, aggregates, or crystals composed of these dyes display photonic properties unobtainable either by single molecules or by bulk materials. With these concepts in mind, there has been considerable effort in the design and synthesis of chromophores that self-assemble or self-aggregate into predictable, precisely positioned structures. Although this chapter focuses on porphyrinic systems, many of the same concepts and methodologies are applicable for the latter two pigments.

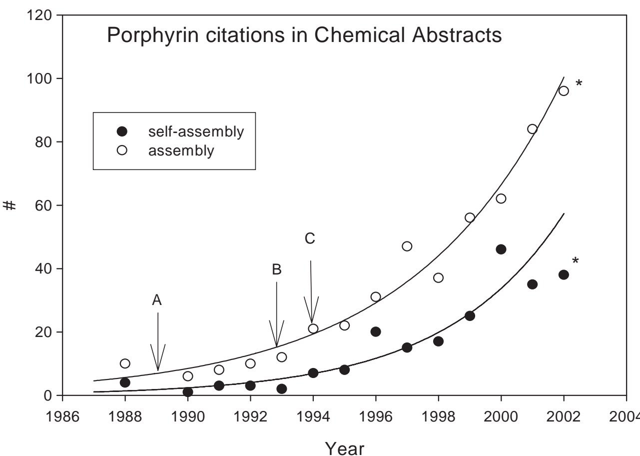
1.1. Porphyrinoids

Porphyrins are tetrapyrrole macrocycles that are found in nature in systems such as hemoglobin and various photosynthetic systems [1], but a very large number of derivatives have been synthesized in the laboratory. Historically, the phthalocyanines were discovered early on in synthetic dye industry, and the porphyrazines came later. Though all three types of

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

macrocycle have related photophysical properties, the porphyrins have been the primary focus of research because of synthetic accessibility and greater solubility in organic solvents. While much of the research in the past decades on multiporphyrin systems has been directed toward understanding electron/energy transfer processes in biology [2–8] and to a lesser extent in polymers [9–14], more recently, some of the focus has shifted toward other applications such as components of molecular electronic and photonic materials [3, 15–19]. Other applications exploit both the photophysical properties and the rigid structure of porphyrins to make materials such as molecular sieves [20–22], catalysts [23], sensors [24], actuators [25], and nonlinear optics [26]. Much elegant organic synthetic work has yielded discrete multiporphyrinic systems held together by covalent bonds, especially with acetylene and phenyl acetylene linkers [13, 27–46], and by direct bonding or fusion of the porphyrin macrocycles [29, 47–52]. The physical characterization of these types of molecules has provided insights into the complexities of electron/energy transfer processes [28, 32, 33, 53]. The Achilles' heel of most discrete molecular systems containing more than a few porphyrins is that the overall yield is too low to be commercially viable. Similarly, the synthesis of porphyrin polymers, either by themselves or grafted onto/into a variety of well-established polymeric systems, has also yielded interesting materials that have a distribution of properties due to polydispersity [9–12]. In most cases, the precise positions of the photoactive moieties in polymers are neither predictable nor precise. Both the relative positions of the macrocycles and the nanoscaled size of the system are necessary for the function of many photonic materials [54]. Thus, many labs have exploited supramolecular chemistry to self-assemble and self-organize porphyrins into a large variety of nanoscaled arrays, aggregates, or crystals that serve as the active components of materials with diverse functions [19, 35, 55–58]. Since the supramolecular chemistry of porphyrins is by far the most developed, the focus of this chapter is on the recent developments in functional, supramolecular, porphyrinic materials. However, studies on the assemblies of porphyrazines and phthalocyanines have yielded interesting results [59].

The opportunity to provide a perspective on self-assembling and self-organizing porphyrinic materials is both exciting because of the tremendous progress in the field (Graph 1), and daunting because of the limited time and space to accomplish the task. Thus, boundaries must be set. The primary focus of this chapter will be on the developments of the last few years since there are several reviews and chapters that cover the large body of work on the chemistry, the properties, and application of porphyrins up to early 2000 [35, 56–61]. There are also a variety of reviews on both self-assembly and multiporphyrin systems [55, 56, 61–68]. Unfortunately, the excellent work on covalently bound multiporphyrin arrays, dendrimers [33, 69–72], and polymers such as phenylacetylene linked arrays of Lindsey [27, 28, 31–34, 37, 38], the directly linked and fused systems of Osuka and others [29, 47–51, 73], the acetylene linked arrays of Therien [39–46] and Anderson [74, 75] or the polymeric materials of Jones [9–12, 76, 77], and the multichromophoric photosynthetic mimics of Gust [78, 79] among others, is



Graph 1. The number of papers cited in chemical abstracts on self-assembling porphyrins is increasing exponentially. Key papers include the electrostatic assembly of an ion-conducting chain (A) [257], a discrete dimer by H-bonding (B) [237], and discrete arrays by metal ion coordination (C) [140]. * = Projected.

also beyond the scope of the present review.[1] There are also numerous reports on the incorporation/encapsulation of porphyrins into gels and other matrices [80]. All of these covalent or aggregated systems have provided a wealth of knowledge on the chemistry, photophysics, and potential applications of porphyrins to modern materials. The design and synthesis of supramolecular porphyrinic materials build upon our understanding of the biological and covalent analogues, in the hope of understanding or mimicking the former, and improving or finding new properties compared to the latter.

1.2. Biological Inspiration and Current Uses

Any discussion of porphyrins must acknowledge that the initial inspiration for research in the field arises from the diverse functions of these molecules in various biological processes. Photosynthesis is the source of most of the free energy that makes life on earth possible—both in terms of stored energy and in the production of dioxygen. Porphyrins [1] are the chief arbitrator of solar energy in photosynthesis; therefore life depends on these molecules. A variety of diverse organisms exploit porphyrins in extraordinarily different ways because of their stability, their photo- and electrochemical properties, and the ability to fine-tune these functions by modifications of either the tetrapyrrole macrocycle or environmental factors. These functions include: (1) chlorophylls are used in the collection and conversion of solar energy into chemical potential in photosynthesis [6], (2) iron porphyrins are used as oxidation catalysts such as in cytochromes P450 [81], (3) iron porphyrins serve as electrical conduits or shuttles such as in cytochromes C, (4) nickel porphyrinoids are used as reducing agents such as in methyl-coenzyme-M reductase [82], (5) iron porphyrins act as the transporter of small molecules such as in hemoglobin and myoglobin. It is not surprising, then, that the chemistry,

[1] Our aim is to derive general themes from the "mis-shapen chaos of well-seeming forms!" (W. Shakespeare, *Romeo and Juliet I*, 1, 184).

structure, and function of porphyrins continue to be of interest even after over 150 years of research [1].

Porphyrins also form the basis of many commercial products that serve, among other things, as catalysts, sensors, and therapeutics. The synthesis and application of porphyrins and porphyrinic materials to *circa* 2000 has been extensively reviewed [60]. At present the porphyrins in these products are either monomers or short conformationally dynamic oligomers, and the function arises from the aforementioned diverse photo- and electrochemical properties of the macrocycle as well as their robustness. In contrast to the commercial products, the function of biological photosynthetic systems relies on arrays of several to hundreds of porphyrins in a precise structural arrangement [2, 6, 83–86]. These nanoscaled solar energy converters are nearly quantum efficient! Inspired partly by biology and partly by the need for technological innovation to make materials with new functions, there has been substantial effort to synthesize multiporphyrinic systems by both covalent, *vide supra*, and non-covalent bonds. For reviews to 2000, see [35, 55, 56, 58, 60, 61, 65, 67, 87–93]. Both routes have advantages and disadvantages. Covalently linked multichromophore arrays (see, e.g., [19, 28, 29, 37, 49, 50, 61, 94, 95]) are discrete molecules and can be purified, thus facilitating characterization of both structure and function. These covalent systems have been pivotal in our understanding of the photo and electronic (photonic) properties of large dye-containing molecules. Most large, complex multiporphyrinic structures are difficult to make in appreciable quantities and yields via synthetic organic chemistry, and this is a major impediment to their use in commercially viable products.

1.3. Self-Assembly and Self-Organization

Supramolecular chemistry uses designed intermolecular interactions (non-covalent bonds) and shape complementarities to self-assemble or self-organize simpler molecules into more complex structures that are often unobtainable by synthetic organic chemistry (see, e.g., [89, 96, 97]). Intermolecular forces are heuristically divided into several sometimes-overlapping types, but all arise from the distribution of electrons on the molecule or atom. The self-assembly and self-organization of molecules into ever more complex structures relies on the molecular information programmed into the component molecules in terms of specific and nonspecific intermolecular interactions [97]. Conceptually, specific intermolecular interactions (such as metal ion coordination and hydrogen bonding) tend to be directional and more predictable than nonspecific intermolecular interactions (such as van der Waals and other electrostatic forces) which tend to be nondirectional and have a broader range of interaction energies and distances [54]. In general, self-assembly refers to the synthesis of discrete or highly ordered systems wherein there is little tolerance for error. In contrast, self-organization tends to entail the organization of many copies of molecular or supramolecular subunits, and is more tolerant of error [98]. Both self-processes are stochastic and dynamic in the sense that the molecules search conformational, dynamic, and associative space as they evolve into a structure or structures that represent a thermodynamic minimum. Intermediate kinetic products with different structures may sometimes be observed and isolated. An elegant essay by Lehn [97] further discusses molecular informatics. Since supramolecular entities are usually thermodynamic products formed from equilibrating solutions, the problem with this approach is that the structure often changes as conditions vary. However, the photonic properties of several robust supramolecular porphyrinic systems have been reported (see, e.g., [54, 77, 99–103]).

Supramolecular materials and devices have hierarchical structure, and in analogy to the terms used in structural biology, these can be classified as the primary, secondary, tertiary, and quaternary structure of a system (Fig. 1) [98, 100, 104]. The primary structure is the individual molecular components. The secondary structure is the designed, supramolecular entity formed by self-assembly processes. Though there is much to be discovered, synthetic chemistry and synthetic supramolecular chemistry are well-developed fields. The tertiary structure then describes the structural relationship of the supermolecule to its immediate environment, as in crystal packing and aggregation where the supermolecules self-associate, or in another modality, how the supramolecular system positions into/onto substrates. Though there has been notable progress especially with porphyrin tectons for crystal engineering [20, 21, 105–110], there is much to be learned in this arena, as we are unable to *a priori* predict the packing of unit cells into crystals [111–113] or to control aggregation such that the product is monodispersed. Finally, the quaternary structure describes how the self-assembled or self-organized material is incorporated into a device via another self-process (as opposed to, e.g., manipulation of individual structures with a scanning probe microscope tip). The interconnections to the macroscopic world may also be a part of the quaternary structure. This last structural descriptor is especially important for molecular electronics and nanoscaled devices, and is perhaps the least understood and developed.

In many ways the self-assembly of supramolecular porphyrinic materials has the opposite benefits and deficiencies as those made by synthetic methods—ease of synthesis from smaller components in greater yields, but added difficulties in characterization of both structure and function. Since

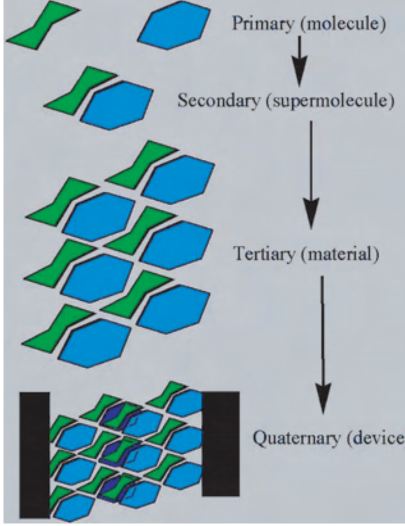


Figure 1. Levels of organization of functional materials.

supramolecular entities are usually thermodynamic products resulting from oftentimes-complex equilibria, characterization of self-assembled and self-organized systems can be difficult. Structural characterization of many supramolecular systems relies heavily on crystallography [20, 21, 105–113], but the structure in the crystal may or may not be relevant to the structure in solution. Several mass spectral techniques have been particularly useful in solution-phase characterization of self-assembled structures [114]. However, since most noncatalytic applications require solid-state porphyrinic materials such as crystals, colloids, or solids on surfaces, the solid state is of greatest concern. Recent developments in supramolecular porphyrin chemistry have begun to address this issue by correlating the structures in both liquid and solid phases [20, 99, 100]. Though generally not specifically addressed by practitioners of supramolecular chemistry, if the material must be crystalline in order to function as designed, the yield of crystals should also be considered. An approach used by many researchers is to design, synthesize, and study smaller supramolecular systems and use these results as a guide to form more complex structures with the same or similar molecular building blocks.

The practical applications of any nanoscaled photonic material based on organic substituents must thoroughly address three major stability issues: thermal fluctuations, oxidation/reduction reactions, and sensitivity to dioxygen. The structural fidelity of self-assembled and/or self-organized materials both in terms of deposition onto surfaces and in terms of thermal fluctuations is a major design criterion that has only recently begun to be addressed [115]. Porphyrins, phthalocyanines, and porphyrazines can be exploited to address all of these issues [60]. More recently, there has been a shift from synthetic supramolecular chemistry towards the design of nanoscaled systems that perform functions that rely on supramolecular structure [54, 55, 116]. For porphyrins [104], these applications include molecular electronics [16, 18, 19, 79, 117, 118], sensors [119–121] or sieves [20, 68, 106, 108, 122], receptors [123, 124], magnets [125–127], and molecular capsules for systems such as fullerenes [128–130].

1.4. Nanoscaled Photonic Materials

It is well established that the function of many materials depends on the geometric order of the component molecules and/or atoms as well as the molecular dynamics [54, 55, 97, 115, 116]. This is especially true of materials—biological or man-made—that are used to conduct electrons and/or photons (photonic materials), since these functions exquisitely depend on the relative alignment and relative energies of the molecular orbitals of the component molecules. Beyond the choice of molecule and the solid-state structure, the photonic properties of a system can be fine-tuned by making components that are on the scale of tens to hundreds of nanometers. Indeed, for the case of photonic materials, nanoscaled devices have properties that can be obtained neither by the component molecule nor by the "bulk" material. The unique properties of nanoscaled photonic materials arise from the confinement of the electrons to a volume with dimensions much less than the wavelength of light. There are numerous reviews on this topic, for example, for systems assembled by coordination chemistry [65, 101, 104]. The applications of porphyrin-based nanoparticles, assemblies, and layers may make inroads first as component devices for highly specialized applications, but as the field of nanoscaled photonics matures they will find further applications that disrupt the current technologies [131]. There are two fundamentally different approaches to self-organization: one uses designed interactions between molecules/ions to build up a three-dimensional structure, and the other uses an external mold such as aerogels to dictate the aggregation of the building blocks, which potentially can synthesize materials over "all length scales" [132]. The second method has been used mostly for inorganic materials, but these new methods are incorporating more complex organic molecules. This latter method also affords geometries not possible by simple self-assembly processes.

2. SELF-ASSEMBLED AND SELF-ORGANIZED PORPHYRINIC MATERIALS—NOMENCLATURE AND DESCRIPTORS

Given the above, the goal of much of the research on supramolecular chromophoric materials is to design materials with functions and properties that are beyond the range of covalent arrays and polymers. The formation of nonlinear optical (NLO) materials is often cited as a goal because of the need for non-centrosymmetric crystals (see, e.g., [26, 74] for recent porphyrin-based examples). Other applications include shape-selective catalysts, molecular sieves, sensors, actuators, and molecular electronics. The focus of this perspective is to provide insights into the formation and utilization of self-assembled and self-organized porphyrinic materials with special emphasis on the relationship between structure and function. Furthermore, we have generally limited the scope to papers published in the last three years since there are several excellent reviews up to *circa* 2000 *vide supra*. The perspective is divided by the mode of assembly/organization (e.g., metal ion coordination, hydrogen bond, electrostatic) and into discrete, open, and polymeric systems.

We have attempted to group the structures found in the literature into several broad, and occasionally overlapping, topological categories in order to discern trends in synthetic methodologies, structure, and function. "Discrete" means that the self-assembling building blocks are designed to result in only one supramolecular product albeit with varying yields. "Closed" describes a system that encompasses an area or volume wherein each molecule is at least ditopic, and is a subset of discrete systems. Conversely, "open" indicates a linear or branched system with obvious ends that in principle can be polymeric or discrete. Generally, at least one component of topologically open systems is monotopic. Discrete systems tend to be assembled from specific intermolecular interactions and tend to be topologically saturated, but this is not a requirement. The mode of self-assembly via directed or specific (H-bond, metal coordination) and nonspecific (electrostatic, van der Waals, etc.) intermolecular interactions is also an essential feature not only of the structure

but of the material's function, *vide infra*. The stylized representation of the multiporphyrin systems in the figures will aid in the discussion, and the key is presented in Figure 2. There have been several attempts to develop a systematic nomenclature for self-assembled supermolecules—especially those assembled by coordination chemistry [63, 133], but simple schematic representations better serve the purposes of structural identification and this review. The choice of assembly method, metalloporphyrin, the directionality of the linkers or recognition moieties, and the exocyclic and/or auxiliary ligands are all used to design porphyrin arrays of varying degrees of complexity. Some supramolecular entities are constructed in more than one step or reaction, but in general this is not a hierarchical assembly process. The hierarchical organization of supramolecular arrays—tertiary structure—can be accomplished either in a second step by the addition of auxiliary chemical entities designed to stitch the supermolecules into some higher-order structure, by imbedding them in a matrix that organizes them into an aggregated structure, or *in-situ* by secondary self-processes. The packing of supramolecular entities into a crystal can belong to more than one of these overlapping categories depending on the design and nature of intermolecular interactions. In some cases, the hierarchical organization of porphyrinic arrays is concomitant with their incorporation into a device—quaternary structure.

The chemical and photophysical properties of porphyrins and metalloporphyrins can be adjusted to perform a variety of functions that can be fine-tuned in several ways. Substitutions on the macrocycle (such as *meso-tert*-butyl and *meso*-adamantyl) that distort the otherwise planar structure can modulate the ground-state absorption spectra as well as the spectral dynamics of the excited state by six orders of magnitude [134, 135]. Supramolecular assemblies of porphyrins with various quenchers have been used to study intermolecular electron transfer [8]. Transition metals with appropriate redox potentials can also modulate the porphyrin excited-state lifetime [28, 33] by transient intramolecular electron transfer reactions such that the complete quenching of luminescence is observed for Ni(II) porphyrins [134, 135]. Nonluminescent species are of interest for optical limiting devices [26, 52, 74]. Halogenation of the macrocycle and/or its substituents can modulate both spectral properties and the redox potential of ligated metals [60]. The desired properties of porphyrinic materials can thus be modified as well.

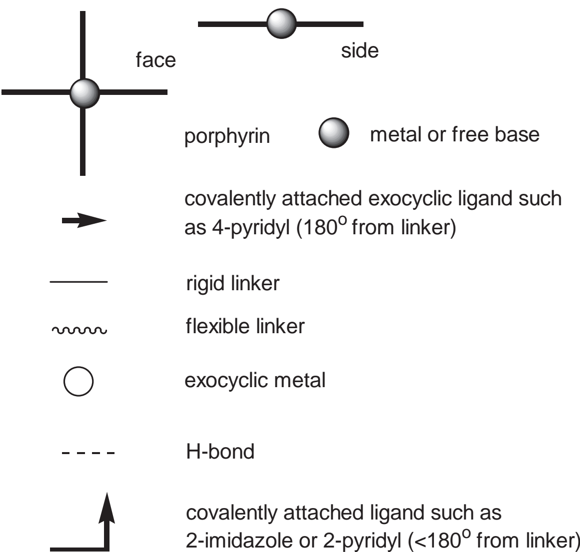


Figure 2. Key to symbols used in figures.

3. ASSEMBLIES MEDIATED BY COORDINATION CHEMISTRY

The well-understood coordination geometries of transition metals (e.g., [136]), coupled with multitopic ligands designed with complementary geometries, have yielded a very large number of discrete, cyclic nanostructures [88, 137]. Coordination chemistry is by far the most used method to assemble porphyrins into various structures—discrete and polymeric—and porphyrins bearing pyridyl groups are the basis of much of these [19, 25, 56, 61, 63, 65, 67, 87, 92, 93, 133, 138]. The square planar geometry of the macrocycle, plus the axial positions of metalloporphyrins, plus the coordination chemistry of the various transition metals that can serve as linkers, allow for the design of a very large number of self-assembled species. There are many variations on this general theme in the formation of crystalline solids wherein at least the unit cell is designed [20, 21, 106, 108, 110]. It should be noted that coordination polymers and materials have a long, rich history. Virtually all porphyrinic materials assembled by transition metals have different photophysical properties than the individual porphyrinic building blocks because of the heavy-atom effect [99, 115, 139]. Additionally, the metal ion linkers may have redox potentials that are appropriate to accept or donate electrons from the excited-state porphyrin. The potential applications of both porphyrin-based and other transition-metal assembled systems for optoelectronics have been well established [101].

3.1. Closed Multiporphryin Arrays

The formation, characterization, and exploitation of closed multiporphyrin systems mediated by metal ion coordination has been the central theme of several research groups for the last few years as these are generally quite easy to make and are the thermodynamically favored product. As initially reported, porphyrinic squares can be formed in two ways [140]: (1) when the porphyrins serve as the corners and metal ions as the sides (Fig. 3A), and (2) when the porphyrins constitute the sides and the metal ions the corners (Fig. 3B). One of the obvious applications of these closed systems is that they can serve as tailor-made hosts for a variety of guests. The size of the cavity can be tuned by a variety of ways, and the use of metalloporphyrins further diversifies the potential uses as sieves, sensors, and catalysts. The metal ion linkers have included Pt, Pd, Ru, and Re, but many others can also be used depending on their coordination geometry, ligand substitution patterns, and desired functional properties [23, 102, 104, 122, 139–151]. These structures are quite predictable and are almost unavoidable as long as labile *cis* or equatorial metal ion positions are

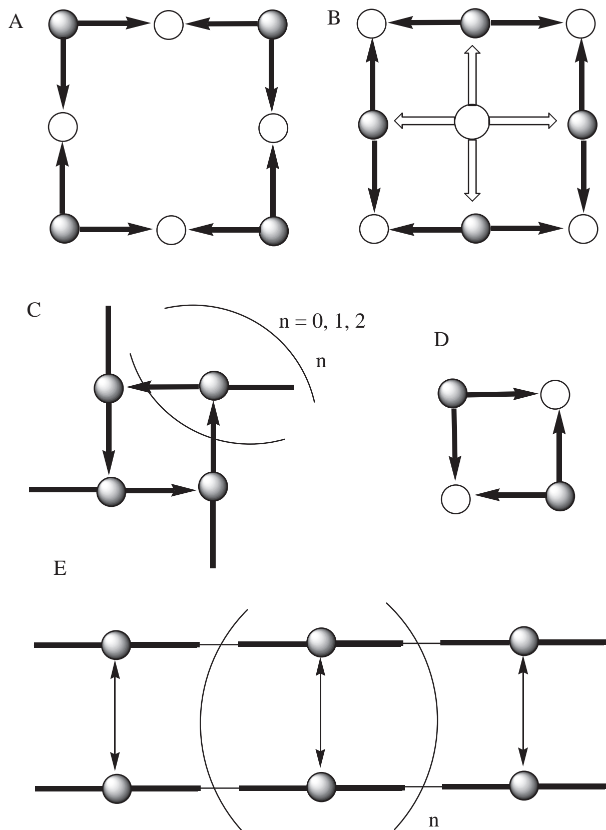


Figure 3. Closed arrays.

available for Figure 3B, and labile *trans* or axial metal ion positions are available for Figure 3A. The porphyrins in supramolecular arrays such as Figure 3A are less conformationally dynamic and are on average co-planar because of the 90° coordination topology of the macrocycle. These features make Figure 3A-type metalloporphyrin squares ideal candidates for secondary self-assembly steps mediated by 180° ditopic ligands such as 4,4'-bipyridine, DABCO, etc. to form layered structures wherein each layer is in register (exactly aligned) [140]. Alternatively, secondary self-organizational processes, such as pi-stacking, can result in nanoscaled aggregates [115]. Self-organization of 3A-type or 3B-type assemblies is also possible by appending long-chain hydrocarbons to the periphery, which allows thin films of these materials to be deposited onto surfaces (see Section 6.2.3).

The porphyrins in squares of the type found in Figure 3B can rotate to some extent, presenting a face toward the inside of the array, and thus enclose a greater volume which can be exploited to bind guests for molecular recognition or catalysis [23, 102, 104, 122, 145–151]. These reports have shown that 3B-type assemblies can be incorporated into devices or perform a function. Microporous films containing 3B assemblies serve as a chemical sensor, the assumption being that the analyte resides in the host, and the signal transduction mechanism relies on the diffraction of visible light because of refractive index changes. Porphyrinic squares such as Figure 3B have been incorporated into other films and devices that demonstrate that they can act as molecular sieves. When the sides of Figure 3B are metalloporphyrins, di- tri- and tetratopic guests of appropriate size (such as other porphyrins shown in outline in Fig. 3B) are readily incorporated [122] as would be expected from crystal structures of similarly arranged lattices of porphyrins. Tetra- and tritopic porphyrin guests will be perpendicular to the sides and will enhance the structural integrity of the square and curtail the conformational dynamics. Since the guest porphyrin can have a redox active metal, this arrangement could be used as an oxidation catalyst that would lessen the self-oxidation reactions that hamper the applications of catalysts based on single porphyrin molecules. The large $\sim 2.5\text{-nm}^2$ area on each face precludes regio-specific reactions. Conversely, 180° ditopic guests can rotate such that the porphyrin square host affords two smaller cavities and regio-specific reactions are favored. This approach to the design of new catalysts is facile because each of the building blocks can be systematically altered to optimize the activity or selectivity.

Figure 3B-type squares assembled using transition metals bearing chiral BINAP ligands afford a means to modulate the function of the array [88, 92, 93, 152–154]. Even though the chiral moieties are outside the central cavity, substantial circular dichroism effects in the porphyrin region are observed, indicating that the photonic properties are substantially different than the nonchiral nanoscaled material. The added feature of this approach is that the supramolecular species is more soluble in organic solvents. A variant of the Figure 3A motif that extends the cavity size utilizes the organometallic 1,4-di-platinate of benzene [152]. The luminescent transition metals used as linking ions afford yet another way to modulate the functionality. The co-crystallization of C_{60} with Figure 3B-type arrays has yielded pillared systems [130].

Figure 3C-type squares represent a simple square self-assembled from one building block linked by axial coordination of exocyclic ligand [155, 156]. These and similar systems are self-complementary because of their shape and ligand direction. For rigid, exocyclic 90° ditopic ligands, squares are the major product (given the usual caveats of self-assembly thermodynamics), but with more flexible external ligands, oftentimes there are significant amounts of the open polymer or supramolecular macrocycles with different numbers of subunits $n \neq 1$, depending on the thermodynamics of the intermolecular bonds, rotation energetics, etc. [157–159]. A 3C square using ruthenium metalloporphyrins leaves the sixth axial site open for further derivatization, and when four equivalents of a free base monopyridylporphyrin are added, which bind the four available axial sites, an assembly containing eight porphyrins is obtained [160]. A variant of this mode of assembly uses ~90° covalently linked metalloporphyrin dimers with axial coordinating ligands at the corner of the link such that self-dimerization results in a closed capsule with four porphyrins [161].

Simple dimers composed of two porphyrins and two metal ions (such as in Fig. 3D) can be considered closed in that they enclose a space [140], but these types of self-assembled porphyrinic systems can also be considered the first generation of linear polymers (see Section 3.2). These dimers are useful as a basis for characterization of larger structures and as building blocks to form hierarchical structures

[141–143, 152, 153, 162]. The exocyclic metal ions need not be only structural, but can add to the functionality of the system such as the modulation of redox properties [141]. Assemblies containing emissive rhenium linkers [150] also modulate the photophysics [139]. The linkers may also serve as a means to dictate the packing of unit cells in a crystal lattice [20, 106, 108]. Another capsule is formed by the dimerization of porphyrins bearing ligands 120° from the porphyrin plane and by four linking metal ions to give a structure with cofacial porphyrins [163]. As an example of guest-assisted assembly, a porphyrin trimer that encapsulates a prism-shaped space has been reported [164]. When two metalloporphyrins are covalently linked with an appropriate spacer, they can be coordinated by a porphyrin bearing two exocyclic ligands [118,165].

The initial report on the self-assembled dimers and squares of types 3A, 3B, and 3D [140] also described a secondary self-assembly step. Two 180°-linked-type dimers with zinc metalloporphyrins were assembled into a square structure (or a two-rung ladder) by the addition of 4,4’-bipyridine (Fig. 4A). Similarly, two 3B-type supramolecular squares were assembled into a cubic structure (Fig. 4B) by the addition of the same ditopic ligand [140] to make a supramolecular species with eight porphyrins—each in an exactly defined geometry. Other later examples of this strategy include ruthenium linked dimers [162].

Another mode of forming porphyrinic squares is to rigidly bond two metalloporphyrins via a variety of organic moieties and then link two of these molecules by a variety of 180° ditopic ligands ($n = 0$ in Fig. 3E) [166]. The ditopic ligands can be chosen for their binding strength to the metal center of the porphyrin, and include 4,4′-bipyridine, DABCO, pyrazine, and terephthalic acid. When n is greater than 0, porphyrin ladders are formed where the binding of the sides increases as the number of rungs increases [74, 75]. Two zinc porphyrins covalently attached by diacetylene can form a square 3E-type where $n = 0$ upon addition of two equivalents of a rigid 180° ditopic bipyridyl acetylene derivative, or form open oligomers when the bipyridyl linkers are flexible [167].

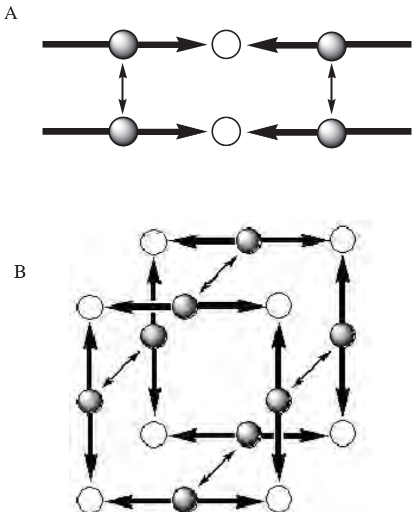


Figure 4. 180° ditopic ligands can can be used in secondary self-assembly steps to link two one-dimensional arrays to form a ladder (A), and two two-dimensional arrays to form a cube.

It is well established that the synthetic yield of large cyclic molecules increases dramatically by incorporation of a template into the reaction mixture. This can be considered an assembly process because the guest directs the synthesis of the host. The methodology has been employed to make macrocycles containing two to six porphyrins (Fig. 5A, B, C) [13, 168–170].

By taking advantage of well-known differences in the ligand affinities of different metals, it is possible to generate more complex closed nanostructures such as Figure 5D in a combinatorial [171] process, wherein the differences between zinc and rhenium metalloporphyrins for pyridyl and phosphine ligands preferentially form a symmetric cyclic structure. This differential metal coordination strategy has been used to form open arrays as well (see Section 3.2).

A 21-component porphyrin nonamer (Scheme 1) can be assembled from three types of porphyrins (4 corners, 4 sides, 1 center) and 12 $PdCl_2$in good yields [99, 115]. This is the next generation up from the squares, but it may represent the ultimate discrete, closed array that can be made by this method because many attempts to form the next generation array have resulted in poor yields of what is likely the discrete array of 16 porphyrins (4 corners, 8 sides, 4 centers). In the same secondary self-assembly method used for the dimers and squares, 4,4’-bipyridine can be used to assemble two nonamers to yield a sandwich of 18 porphyrins—all in a precise location in the supramolecular architecture. Without the bipyridine, these nonamers can self-organize into aggregates that have properties unique to this scale (see Section 7.2).

3.2. Open Multiporphyrin Arrays

Discrete open systems generally contain one or two types of porphyrin building block, but occasionally higher-order oligomers are formed. The latter case is usually accomplished by capping the open ligands of a central porphyrin or by a multistep supramolecular synthesis, which combines dimers and/or trimers with different metal ligand affinity

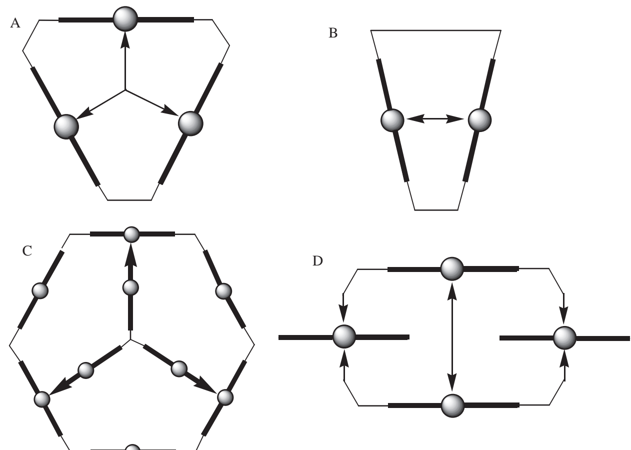
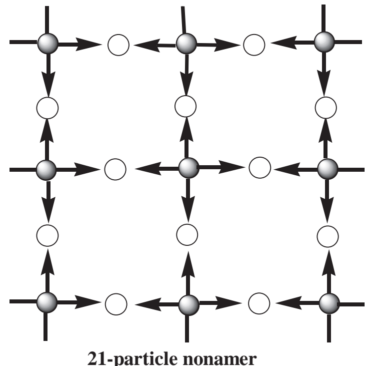


Figure 5. Template-assisted formation of multiporphyrin macrocycles.



Scheme 1.

strategies [56, 63, 88, 133]. Supermolecules with more than one type of porphyrin and/or metal ion linker can exhibit a broader range of functions [8, 101]. Some of the solid-state structures fall in this category [20, 108, 109]. This limited range of discrete open self-assembled systems results from the fact that merely adding "stopping" units still results in statistical mixtures where the average length is dictated by the thermodynamics of the intermolecular interactions. Most of these systems are conformationally dynamic because of one or more monotopic interactions (Fig. 6).

Conceptually, the discrete open arrays that assemble more than one porphyrin around a central metal ion (Fig. 6A) are similar to the simple dimers using the same strategy using Pt(II) or Pd(II) metal ions [140]. In addition to other reports using these square planar metals [92, 152, 173], ruthenium has been widely exploited (see, e.g., [139, 144, 174]) to form dimers with porphyrins at 90° or 180° angles. These three metals have also been used to form tetrameric 6A-type arrays (e.g., [173]). There are a variety of ligands other than pyridine that can be appended to the porphyrin macrocycle

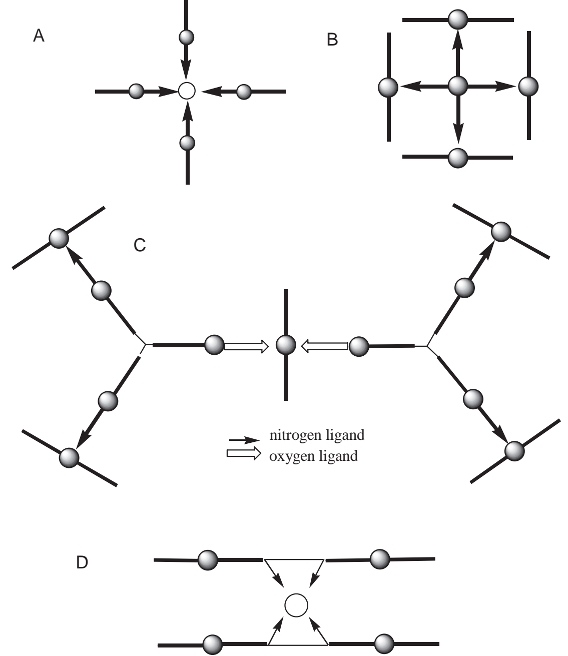


Figure 6. Open arrays.

that facilitate self-assembly mediated by metal ion coordination, and one of the most common is the terpyridyl moiety. Thus, two porphyrins bearing a single terpyridyl group can be dimerized by addition of a wide variety of metals [175–178].

In a similar approach, coordinating metalloporphyrins around a central chemical entity that bears more than one ligand capable of coordinating in the axial position can make discrete open multimers of porphyrins. If the central multitopic ligand is tetra-4-pyridylporphyrin (TPyP), then four ruthenium porphyrins can coordinate to make a pentamer (Fig. 6B) [139, 144, 179, 180]. The pyridyl groups can also be in the 3-position to give a pentamer with two porphyrins on each side of the central macrocycle [181]. Covalently linked metalloporphyrin dimers can be assembled with a free base porphyrin bearing flexible amine ligands [182]. The advantage of this strategy is that the free base porphyrin can act as an electron/energy trap and the peripheral porphyrins as antennae. Other multitopic central ligands have included 4,4′-bipyridine, DABCO [179, 183], and recently several cross-conjugated macrocycles bearing pyridyl moieties at a 180° angle [184] have been reported. These latter ligands are much larger supramolecular building blocks than the former. 2,2′-Bipyridines bearing porphyrins can be used to assemble arrays when the bipyridine binds a metal ion. Thus, when each half of the bipyridine is linked to a porphyrin via vinyl groups [185], or each half is directly attached to either the 4 and 4′ [186, 187] or the 5 and 5′ positions [187] (Fig. 6D), arrays of four porphyrins are formed upon the 2,2′-bipyridyl subunit binding to tetracoordinate metals. When the porphyrins are on the 4,4′-positions and a hexacoordinate metal is used, in principle a trimer bearing six porphyrins would be formed. Terpyridines bearing a 4′-pyridyl moiety have been used; in this case, when the terpyridyl unit binds a metal ion and the pyridyl units bind ruthenium porphyrins, a dimer is formed [188], and the same type of strategy works with bipyridyl linkers [189].

Figure 6B-type arrays are the parent structures for both lattices and linear polymers in which the metalloporphyrin is hexacoordinate and/or only the TPyP is used, *vide infra*. Several metalloporphyrins are hexacoordinate (e.g., Fe, Cr) and in principle the Ru porphyrins in the cited examples can be made hexacoordinate by photolytic removal of the CO ligand. These discrete systems have been used to examine the effects of heavy atoms and the assembly conformation on the photophysical properties [139, 141, 144, 162, 180] of ruthenium porphyrin oligomers. Other examples of this mode of assembly include a tripyridylphosphine [190] that is itself a ligand for a rhodium carbonyl catalyst.

Another route to discrete but open porphyrin architectures uses essentially three supramolecular synthetic steps and knowledge of the concept of hard-soft ligand-metal binding interactions. The propensity of tin porphyrins to be hexacoordinate and their preference for hard ligands such as oxygen, can be used in conjunction with metalloporphyrins with metals (such as ruthenium, zinc, or rhodium) that prefer softer ligands (such as pyridyl), to build complex multiporphyrinic structures. An elegant example can be found in Figure 6C, wherein two building blocks bearing three covalently bound porphyrins (two with a pyridyl group and one with a benzoic acid group) are dimerized by the coordination of benzoate to a tin porphyrin, and the four terminal pyridyl

groups on this dimer then axially coordinate to four metalloporphyrins [183]. Complex multiporphyrin arrays, such as shown in Figure 6C, can be built in reverse order as well by capping the two exocyclic pyridyl groups on a covalently bound trimer with Ru porphyrins to form an intermediate with a carboxyphenyl group free. In a second step, stitching two of these together via coordination of both axial sites on a Sn(IV) porphyrin results in a system with 11 porphyrins of four different types. 4-Carboxypyridine can be used with one tin porphyrin and two rhodium porphyrins to form a linear trimer [183, 191–193].

3.3. Porphyrin Polymers and Aggregates

What distinguishes the simple open multimers from the polymers discussed in this section is largely the fact that two or more of the components of the self-assembling system are at least ditopic. Though they have a long, rich history, coordination polymers are of renewed interest in nanoscaled materials (see, e.g., [194]). The simplest cases, (Fig. 7A, B, C) are merely polymeric forms of the dimers discussed earlier, but more complex polymers can be formed (such as in Fig. 7D and E). Again, since the photonic properties of multiporphyrinic systems depend exquisitely on the relative size, orientation, and mode of assembly, each of the linear polymers is expected—and observed—to have different properties. It is well known, for example, that face-to-face aggregates have different optical properties than the side-to-side aggregates. The parent dimers to 7A and C were reported along with the porphyrin squares [140], and subsequent papers described tape 7C [99]. Tapes such as 7D (e.g., [195]) and 7E have been found to have large third-order nonlinear optical properties [26, 196]. These latter modes of self-assembly may be somewhat related to the autocatalyzed self-aggregation of certain bacteriochlorophylls in nonpolar solvents where a secondary alcohol on a pyrrole axially coordinates the magnesium (e.g., [197]). Many of the structures of these types of systems (Fig. 7A–E) have been characterized by single crystal X-ray analysis and other solid-state methods [21, 109, 198–200]. A 7C-type polymer that forms a two-dimensional network of hexacoordinated zinc porphyrins was recently reported [201]. Many of the earliest publications on porphyrin crystal engineering relied on metallotetrapyridylporphyrin self-coordination or on free base tetrapyridylporphyrin coordination to external metal ion linkers [20, 21, 106, 108, 110]. Hexacoordinate metallotetraphenylporphyrin derivatives with flexible pyridyl ligands at the 3-position of the 5, 15 phenyl groups can self-organize into 7D-type linear polymers, which under the right conditions can close to form a ring [202]. More rigid 5,15-bis(imidazolyl)-10,20-(4-dodecyloxyphenyl)-porphyrinato zinc molecules self-assemble into rigid 7D-type polymers [4]. The self-complementary coordination of 2-aminophenyl-porphyrinato zinc subunits to form a 7D-type n=0 dimer is in equilibrium with the open oligomers [203]. Hydrothermal processing continues to provide crystals with more rigid structures, likely the thermodynamic products, of porphyrins with a variety of metal ions. A recent example uses tetracarboxyphenyl groups and Ca(II) to make an interpenetrating network which is further organized by strong hydrogen bonds [22]. This illustrates the fact that even in solid-state crystals there may be many different polymorphic structures using the same or similar supramolecular design principles.

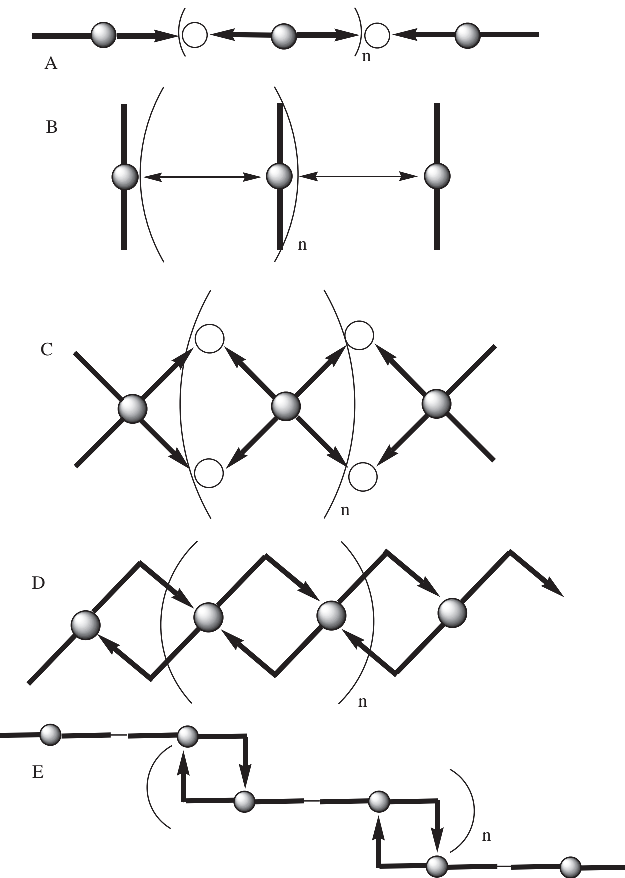


Figure 7. Polymeric arrays.

One interesting method to make polymeric multiporphyrin arrays combines both the synthesis of covalently bound polymers of tin metalloporphyrins, and the axial coordination of these subunits by porphyrins bearing a single benzoate group (Fig. 8A) [193]. The hybrid approach of combining second types of synthetic porphyrin oligomers followed by coordination of these oligomers by other porphyrins [179, 204] or other chromophores promises to greatly expand the repertoire of porphyrin self-assembly, as it bridges several levels of self-assembly and self-organization [97, 98]. Polymers of type 8B and 8C are readily accessible as well. Recently reported is a method for making porphyrin derivatives with the coordinating exocyclic ligand built directly on (a part of) the macrocycle [205–208]. This allows for the formation of coplanar dimers and multimers where the macrocycles are "conjugated" via the metal ion coordinating bonds (Fig. 8C). These dimers, for example, have photonic and electrochemical properties not seen by other means of self-assembly. This approach, combined with other modes of self-assembly, may afford interesting materials with complementary photonic properties. An 8C-type structure is an intermediate in the synthetic pathway to [2]porphyracene (a fused porphyrin dimer) [52, 209].

Mesogenic systems formed from self-assembled porphyrin dimers (Fig. 7A, $n = 0$) appended with long-chain

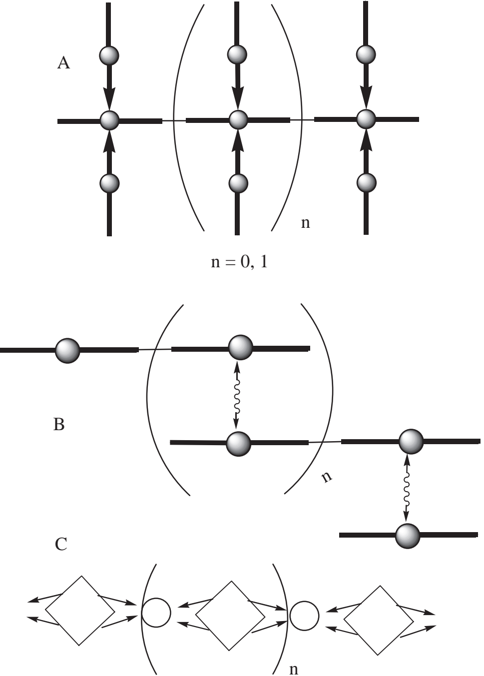


Figure 8. Other polymers.

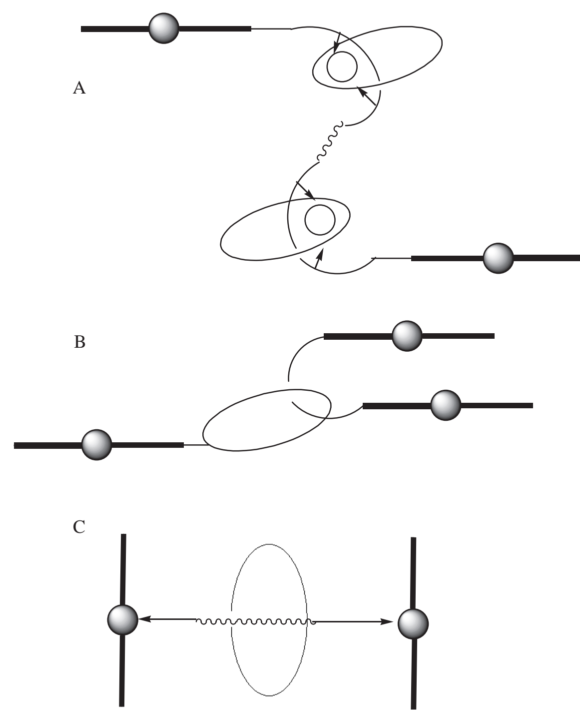


Figure 9. Arrays made with mechanical bonds.

hydrocarbons have been found to form ringlike domains of greatly varying size, thickness, and height [210–212]. A variety of mesogenic systems have been reviewed [66, 67, 213–215], and a few will be discussed in Section 6.2.3.

3.4. Porphyrins Linked by Mechanical Bonds

Rotaxanes and catenanes are chemical entities where a mechanical bond joins two different molecules; minimally, the former involves a linear guest and a ring-shaped host (Fig. 9), while the latter links two rings as in a chain [216]. Most of the self-assembled multiporphyrinic systems to date are rotaxanes, and coordination chemistry is generally the method to string the linear guest molecule through the host macrocycle to result in systems like Figure 9A and B [159, 175, 217–221]. Alternatively the ring may be formed around a template. In addition to examining electron/energy transfer in systems held together by mechanical bonds, the porphyrins act as a probe of the molecular and mechanical dynamics of these systems [87, 217, 222]. The rearrangement of these molecules represents prototypical molecular machines in that they can serve as nanoscaled “ball and socket” joints, “universal” joints, and wheel and axles [216]. Rotaxane systems have been widely proclaimed to be part of the initial foray into molecular based nanoscaled electronic devices [15, 16, 55, 223, 224], but much remains to be solved—especially the interconnections between devices. Another mode of rotaxane formation is to use coordination chemistry to stopper the rotaxanes (Fig. 9C) [225, 226]. The 9C-type rotaxanes afford an avenue toward the synthesis of more complex multiporphyrinic arrays by exploiting the differential metal ion coordination strategy, or the stoppers themselves may be arrays of porphyrins.

3.5. Themes and Trends

These structures rely on the coordination chemistry of the metalloporphyrin and the coordination of external metal ions by exocyclic ligands as separate but designable building blocks that assemble in essentially one step. For the closed multiporphyrin arrays, one of the most common themes is that these types of structures are excellent hosts to a variety of guests. The size of the internal cavity can be fine-tuned by many means, including the size (length) of the exocyclic ligand, secondary self-assembly steps, and the formation of structures that are larger than the most common square tetramer. Metal mediated self-assembly of porphyrin arrays yields structures that are more robust, and the design of specific structures for a given function is more predictable than for other means of self-assembly or self-organization. Fullerenes are an example of large, chromophoric guests for which self-assembled arrays and dimers (Fig. 10) have been tailor-made [128–130] or incorporated into self-assembled structures [227]. Therefore, it is likely that this mode of assembly will continue to dominate the organization of molecules by both selfprocesses. In addition to tailoring the cavity to the size and shape of target guests, the various photonic properties of these nanoscaled materials can be exploited for redox catalysts, sensors, and sieves.

4. ASSEMBLIES MEDIATED BY HYDROGEN BONDING

One of the overarching themes in nature is the organization of matter into ever more complex structures that, when properly arranged, function as viruses, cells, and organisms. One of the underlying principles of this organization is

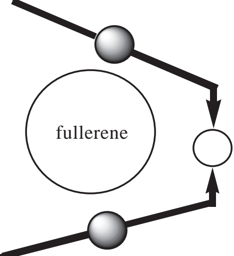


Figure 10. Self-assembled host binding a guest.

shape and intermolecular bond complementarity—the information contained in the molecule is ultimately sufficient for molecular self-assembly. Though more complex structures often require enzymes or templates, the fundamental building blocks do not. Of the specific intermolecular interactions, hydrogen bonding is by far the most exploited for both defining structure and storing information. The role of hydrogen bonds in these functions arises from their number, energetics, and their directionality. For example, the complex topological structure and function of proteins rely heavily on hydrogen bonds, as do the base pairs in nucleic acids DNA and RNA. These structures form and are stabilized via the cooperative interactions of many complementary hydrogen bonds (see, e.g., [228]). Nature relies on the fact that these interactions are reversible so that mistakes in structure can be reversed and corrected. It is from biology that scientists take their initial inspiration in the design and exploitation of complementary hydrogen-bonding entities to self-assemble molecules into various nanoscaled materials. The structure and function of these materials arise from the molecular and supramolecular structure as well as the mode of assembly. The design, strength, and uses of hydrogen bonds in supramolecular chemistry have been well reviewed [111, 229–234].

The rigid structure of the porphyrins make them ideal scaffolds to append hydrogen-bonding groups that allow self-assembly upon addition of the complementary hydrogen bonding motif. Functional biomimetics continues to drive much of the research on porphyrin materials and receptors (see, e.g., [235, 236]). The nanoscaled structures resulting from the self-assembly and/or self-organization of multiporphyrinic arrays and materials mediated by complementary hydrogen-bonding groups, like the metal ion mediated counterparts, can be closed, discrete, open, or polymeric. There are many fewer reports on self-assembling porphyrin arrays mediated by H-bonds because these intermolecular interactions are much weaker, which hampers characterization in solution. Also, the solid-state structures may be different than those in solution. (Thus many of the reports of porphyrin assembly via hydrogen bonds are solid-state structures [20, 21, 107, 108] because of the difficulties in characterization.) Hydrogen-bonding motifs are directional, but they may also be ditopic in that more than one complementary group can bond to the same moiety. When the complementary molecule is a porphyrin, multiporphyrin arrays result.

4.1. Closed Multiporphyrin Arrays

H-bonds are directional and the cooperative formation of closed self-assembled arrays can result in sometimes surprisingly stable materials (Fig. 11). These relatively weak intermolecular interactions can be exploited in materials wherein responses to thermal or solvent polarity are desired. Other than temperature and solvent as factors in the assembly, the concentration range for the closed structures is limited. As is well known in biology, open polymeric structures are favored at high concentrations, and of course at low concentrations monomers or partially assembled species are present. Another feature of many hydrogen-bonding groups is that there are specific sequences of hydrogen bond donors (D) and acceptors (A); for example the DAD triple H-bonding potential of a triaminotriazine is complementary to that of the ADA motif of barbituric acids. But because of the reversibility of H-bonding interactions, slipping these two complementary groups by two hydrogen bonds results in a system whereby the two entities remain bound, but with only hydrogen bond. The consequence of this is that even the closed arrays are still dynamic in solution under the best of circumstances. Limited interporphyrin charge and energy transfer has been observed in squares assembled by H-bonds (Fig. 11A), but in general the porphyrin photophysics is more similar to the monomers than is found for the coordination arrays. The early examples of these

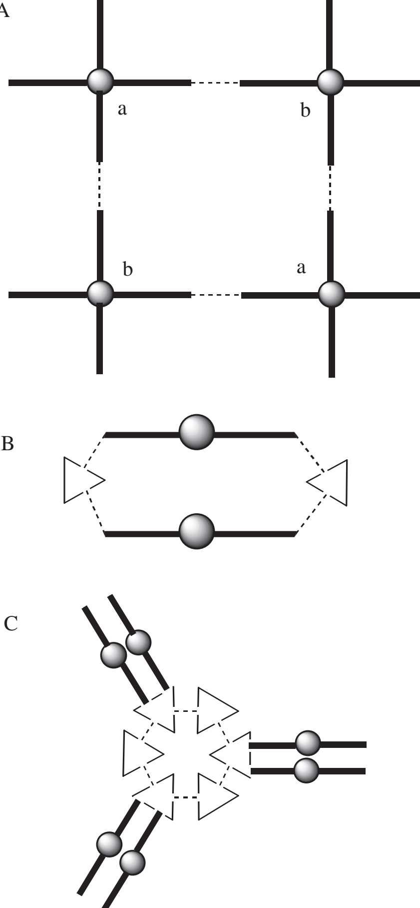


Figure 11. Closed arrays.

closed arrays used uracylic and triazine moieties along with a nonporphyrin complementary hydrogen-bonding group (triangles) to form a supramolecular cage (Fig. 11B) [237]. A rosette with three co-facial porphyrin pairs (Fig. 11C) [238] is formed from a bisporphyrintriazine and a variety of barbituric acid derivatives. It is possible to design closed square arrays of porphyrins (Fig. 11A) by placing two copies of one type of a monotopic hydrogen-bonding moiety at 90° on one porphyrin (a in Fig. 11A) and two copies of the complementary monotopic hydrogen-bonding group at 90° on another porphyrin (b in Fig. 11A) [239]. Since the hydrogen-bonding groups in 11B and C are ditopic, several possible structures can result, but conditions are chosen so that the predominant form of the supermolecule is closed. It is also possible to exploit the aforementioned dynamics of complementary hydrogen-bonding systems to design and synthesize squares that are formed via self-complementary interactions (Fig. 11A, where $a = b$) [172]. In this particular case the intermolecular forces are about an order of magnitude weaker than the square made via heterocomplementary H-bond groups [239]. Comparing the thermodynamics of forming the closed arrays (90° ditopic porphyrin) with the thermodynamics of forming the open arrays (180° ditopic porphyrin) mediated by the same complementary H-bonding groups allows the assessment of the role of cooperativity in the formation of closed structures without changing recognition motifs or molecular dynamics. There are other reports of closed 11C-type structures [240]. More robust 11B-type cage structures are formed by the α^4 rotamer (all directed towards the same face) of tetra-(2-carboxyphenyl)porphyrins wherein the carboxy moieties bind the two porphyrins with four sets of self-complementary double H-bonds [241, 242]. When four equivalents of these metalloporphyrin cages are combined with a porphyrin bearing four pyrazines flexibly attached to four meso phenyl groups, a secondary self-assembly step occurs wherein each pyrazine group coordinates both of the zinc centers of the cage. Encapsulation of all four meso-appended pyrazine derivatives results in a nonameric array [243], and the antenna effect on energy transfer for these nonameric arrays was studied. The porphyrins used for 11B-type cages generally have two or more rotameric conformations (atropimers), but the supramolecular cage traps the conformation with the hydrogen-bonding groups on the same face [237]. While closed structures 11A and 11C are relevant to electron/energy transfer across hydrogen bonds, 11B-type cages may be exploited for guest recognition or catalysis. H-bonds and ion-pairing can be used to make structures with porphyrins and iron terpyridine [244, 245]. A square can be formed from two 180° covalently linked zinc porphyrins when 4(3-H)pyrimidone is added, where the nitrogen on the heterocycle binds axially to the zincs and the amido positions serve as self-complementary H-bond units. This is an H-bond version of a 3E-type square ($n = 0$) [246].

4.2. Open Multiporphyrin Arrays

As mentioned earlier, the same set of porphyrins assembled by hydrogen bonding can be monomeric, polymeric, or closed depending on conditions. Other than simple dimers [239, 240], there are few examples of open multiporphyrinic systems that are discrete. The problem with forming open multiporphyrinic systems via hydrogen bonding is that the intermolecular interactions are so weak, that tapes such as a trimer (Fig. 12A, $n = 1$) are formed in low yields and are only one of several supramolecular entities in the solution. In the solid state, X-ray studies reveal that a variety of other nonspecific intermolecular forces (pi stacking, van der Waals, etc.) reinforce the structure mediated by hydrogen bonds (see, e.g., [22, 68, 107, 229, 233, 239, 247–249]).

4.3. Porphyrin Polymers and Aggregates

Because sugars have a variety of recognition functions in living systems, porphyrins appended with a variety of sugars have been the focus of several research groups. The multiple hydroxy groups on sugars make them ideal candidates for the formation of porphyrin aggregates and polymers (e.g., [250]). Several reports on the formation of gels and fibers from porphyrins bearing sugar groups (Fig. 12B) have appeared [251, 252]. Imidazole-imidazole hydrogen bonding has been used to make 12C-type polymers [196, 253, 254].

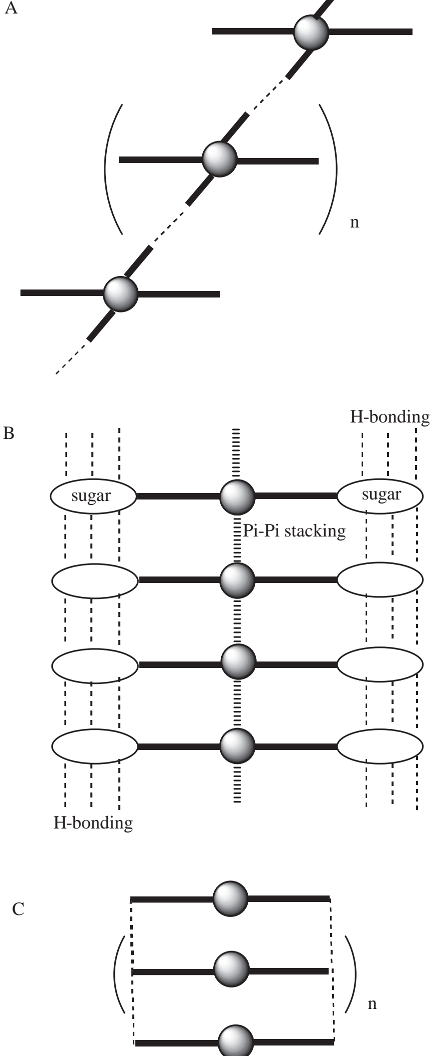


Figure 12. Polymers by H-bonds.

There are many solid-state examples of porphyrins assembled by hydrogen bonding. Though the crystals are not nanoscopic, the specific intermolecular interactions that are exploited to place the porphyrins in a designed geometry and order often result in nanoscaled channels, or cavities that can be used to recognize guests (sensors), act as molecular sieves, or as shape and size selective catalysts [20, 67, 68, 105–108, 249]. In addition to the one-dimensional structures in Figure 12, there are a variety of structures with designed two- and three-dimensional interactions such as those in Figure 13 [105, 106]. The crystal structure of 12C-type columnar aggregates reveals remarkably stable channels [249].

4.4. Themes and Trends

Though each individual hydrogen bond is only worth a few kcal/mol, there are many complementary organic groups that are capable of more than three H-bonding interactions (e.g., [255]). The >8-kcal/mol total interaction energy between these latter groups in nonpolar solvents is more than sufficient to make robust self-assembled materials, though complementary groups with more than three H-bonding sites have not been generally applied to porphyrin materials chemistry. The weak intermolecular interactions can be exploited as structural sensors to environmental factors such as temperature and solvent, wherein the resultant modulation of interchromophore distances alters the photonic properties which report on the status of the assembly. The aggregates of porphyrins are also relevant to the formation and activities of photosynthetic antenna complexes, which may or may not use a protein scaffold to maintain structural integrity of the chlorophyll arrays.

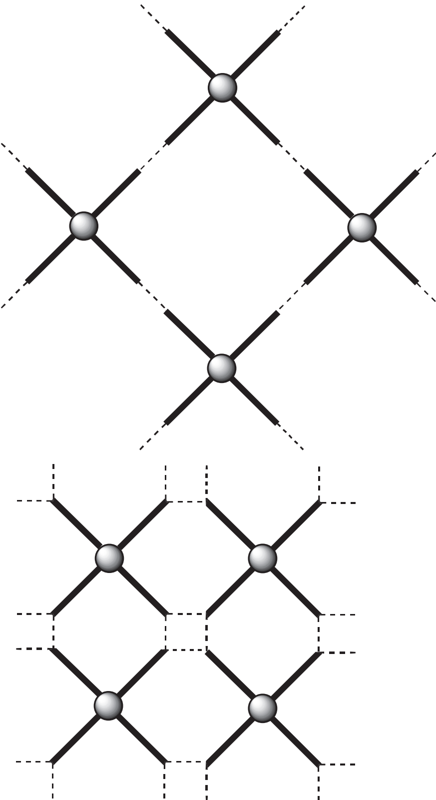


Figure 13. Lattices and arrays by H-bonds.

5. ASSEMBLIES MEDIATED BY ELECTROSTATICS

For the purpose of discussion in this review, electrostatic assembly is defined as arising from the interactions of full or partial charges (ions and dipoles) on the component molecules, though other intermolecular forces such as hydrogen bonding and metal ion coordination certainly have electrostatic contributions. Charge-charge interactions are usually neither specific nor directional, but dipolar interactions are directional. The specific placement of charged groups in proteins, lipids, and other biological molecules or biopolymers is an important aspect of both tertiary and quaternary structure, as well as of function. Phosphatidylcholine lipid bilayers, for example, are formed by a variety of nonspecific intermolecular electrostatic interactions [256]: (1) electrostatic interactions in the charged head groups, (2) dipolar interactions in the ester region, (3) and van der Waals interactions amongst the hydrocarbon tails. These nonspecific interactions are necessary in membranes to assure that they are structurally dynamic, that is, fluid.

5.1. Porphyrin Polymers and Aggregates

There have been numerous papers on the mixing of positively charged porphyrins with those bearing negative charges, or with other molecules with complementary charge. One of the first examples of porphyrin self-organization (as opposed to uncontrolled aggregation or crystallization) by any means used transiently formed porphyrin cations with lipophilic anions such as tetraphenylborate to create “linear” chains that conducted ions across a lipid bilayer [257, 258]. This was an early example of a working molecular electronic device, and it performed as a photogated ion-conducting transistor (see Section 7.1) [259]. Charged porphyrins include the anionic forms tetra-carboxyphenylporphyrins (TCPP), tetra-phosphonatophenylporphyrins (TPPP), and tetra-sulfonatophenylporphyrins (TSPP), the cationic tetra-N-methyl-pyridiniumporphyrins (TMePyP), and the tetrakis-tri-alkylanalyniumporphyrins (TAAPP). Mixing solutions of TSPP and TMePyP along with a surfactant results in specific types of aggregates [260], where the size depends on surfactant, solvent, temperature, and concentration, just as it does for other self-assembled aggregates such as those made by hydrogen bonding. Without the surfactant, these species aggregate as well to form what are likely more ordered structures in solution [261, 262]. TSPP has been found to interact with phthalocyanine dyes [263]. The addition of an amino acid such as phenylalanine to a mixture of cationic and anionic porphyrins induces them to form chiral aggregates [264]. The anion-controlled assembly of porphyrin-bicyclic guanidine conjugates, which are much less symmetric than the porphyrins described earlier, results in chiral aggregates [265]. This latter system is noteworthy because there is more information designed into the molecule, so more structural information is read out and expressed in the aggregate structure. Also reported, is the formation of chiral aggregates

of less symmetric charged porphyrins by vortex mixing [266, 267], where the vortex is somewhat akin to a chiral mold or template such as porphyrin-DNA assemblies [268, 269]. Sulfonated C_{60} has been mixed with TMePyP to form electrostatic assemblies on chitosan films [270]. Several solid-state materials have been reported where the structures of the system arise at least in part by electrostatic interactions, for example, between porphyrins with carboxylate anions and potassium anions [20]. Irregular, polydispersed aggregates are also formed from TSPP in solution [261, 262], and even from neutral porphyrins such as TPP and TPyP [271–273]. In addition to porphyrins bearing charges on the peripheral groups, the protonated diacids of some porphyrins in solution [274] undergo structural changes that result in the meso aryl groups possessing a smaller dihedral angle with the macrocycle, which allows aggregation of zwitterionic derivatives. The electrostatic association of TSPP with cationic polyamines forms pH-dependent species at lower pH values [275], and the charged polymer has been used to adsorb this porphyrin onto glass substrates (see Section 6.2.4).

5.2. Themes and Trends

The electrostatic association of porphyrins to other dye molecules, to polymers, and to charged particles continues to be a fruitful strategy to make nanoscaled materials. However, the relative orientation of the chromophores, the size of the particles/aggregates, and the polydispersity of these systems are difficult to control and make them difficult to incorporate into nanoscaled devices. A related method is the layer-by-layer deposition of charged porphyrins, metalloporphyrins, and charged polymers. These systems are remarkably robust and afford a greater degree of control of molecular position (see Section 6.2.4).

6. PORPHYRINS ON SURFACES

Despite the advantages of using the processes of self-assembly and self-organization to construct materials that can serve as components of nanoscaled devices, several significant obstacles must be overcome. The advantages include efficient use of chemicals, and the ability to organize molecules in a precise architecture. But because of the reversibility of the intermolecular interactions that allow self-assembly and self-organization to result in thermodynamic products in high yields, the major obstacles to employing such materials are issues of stability. Device stability to thermal fluctuations, to redox chemistry, to environmental changes such as humidity, and to dioxygen is crucial. But, all of these factors can affect the equilibrium of self-assembled systems, thus the structure and function. Another issue is that most structures formed by self-processes take place in solution, and the deposition of these structures onto surfaces with high structural fidelity is not necessarily straightforward. Most applications require the removal of the solvent, and the inherent concentration changes often perturb the equilibrium, which results in structural rearrangements of the supramolecular system. Scanning probe microscopies, such as atomic force microscopy (AFM) and scanning tunneling microscopy (STM), are some of the best methods to evaluate structures on surfaces [276]. There are a few examples of self-assembled (discrete, monodispersed) systems that can be deposited onto surfaces with high structural fidelity and are stable at room temperature. As one might expect, these successful systems are largely assembled via metal ion coordination, as hydrogen bonding is generally too weak. There is a large body of work on forming self-organizing systems on surfaces, including monolayers, liquid crystals, layer-by-layer electrostatically formed films, and chemisorbed materials. The synthesis and characterization of electropolymerized thin films of tetra(4-aminophenyl)porphyrin from an aqueous solution onto electrodes to form an emeraldine-type polymer was reported [277] though the morphology of the film was not well defined.

6.1. Self-Assembled Structures

One of the first examples of self-assembled porphyrin arrays that address the above issues is a self-assembled nonamer. The 3×3 tessellation of nine porphyrins of three topologies (+, T, and L-shaped) is mediated by the *trans* coordination geometry of Pd(II) in a variety of organic solvents (Scheme 1). These nonamers (the next generation of a 3A square) can then self-organize into columnar stacks of 1 to ~ 20 layers to form 6.5 nm $\times$ 0.5 nm to 10 nm nanoscaled aggregates [99]. The length of the columnar stack can be fine-tuned by choice of appended alkane, by choice of solvent, by choice of metal, and by kinetics [100, 115]. These stacks can be deposited onto glass surfaces without any observable change in structure or function because of the weak interactions between the aggregate and the substrate and the strength of the intermolecular forces holding the nonamer and the stack together. The individual (nonaggregated) nonamer can be deposited on a Au 111 surface—an example of a purely self-assembled structure deposited on a surface (a 14G-type of arrangement). Most importantly for applications, all of these constructs, and their photonic functions, are stable in air at room temperature for years [100, 104, 115]! The next step will be to self-organize the self-assembled nonamers into two-dimensional arrays on surfaces. An elegant use of electrochemical potential of an electrode surface that modulates the surface mobility of adsorbed molecules allows for the self-assembly porphyrin arrays has been reported [278].

6.2. Self-Organized Structures

Because self-organized systems are generally less ordered and more tolerant of defects than self-assembled systems, both in solution and on surfaces, these types of systems are generally easier to make and characterize. Most of these systems can be divided into three general categories: (1) the macrocycles can be chemically bound to the surface, such as porphyrins bearing hydrocarbons terminated with thiols; (2) the macrocycles can be bound to the surface by coordination chemistry, such as when ligands are attached to the surface either chemically or by adsorption; and (3) self-assembly by direct adsorption on the surface under controlled conditions or by using moieties that impart liquid-crystal properties to the molecules. Some of the applications for these types of layered films are to modify electrodes to make them selective for various analytes, and to modify fibers for detection of analytes by optical absorption or emission. The placement

of long-chain hydrocarbons allows for the two-dimensional organization of porphyrins [279] and phthalocyanines [103, 280, 281] on surfaces that can be characterized by scanning probe microscopy [276, 282].

6.2.1. Chemically Bound

Though commonly referred to as self-assembled monolayers (SAMs), these materials are more accurately described as self-organizing systems (Fig. 14A, B, E, F, where 'x' = S or other reactive group). There has been considerable research on SAMs and there are numerous reports on porphyrin containing monolayers. One of the most common avenues into these materials is to chemically bind (Fig. 14A) a hydrocarbon-bearing thiol group to a gold surface [95, 283–288]. Other substrates such as indium-tin oxide (ITO) [287] and polypyrrole [289] have been used. The nature of the hydrocarbon linker modulates the physical and photonic activity of the monolayer. More complex (structurally, redox, and optically) photonic triple-decker molecules containing porphyrins, phthalocyanines, and both porphyrins and phthalocyanines have been used to form monolayers with more complex redox and optical properties [36, 95, 290–295]. The increased physicochemical complexity affords an opportunity to design and characterize devices that act as multistate counters for molecular electronics applications. Thiol bound monolayers terminating with a porphyrin bearing an imidazole facing away from the surface can serve as a template to construct 7E-type multilayers, but as with other multilayers, the length of the self-coordinating polymer varies, the dispersity increasing as each layer is added [296].

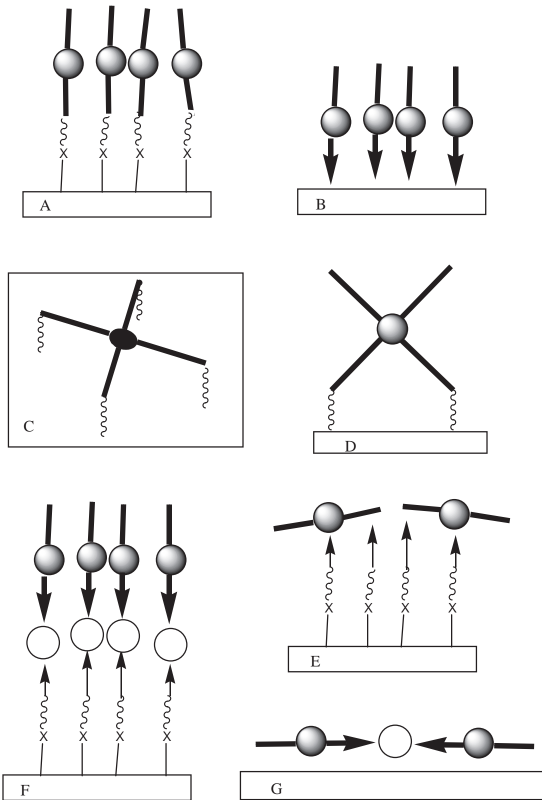


Figure 14. Self-assembled porphyrinic structures on surfaces.

The interaction of carboxylate moieties with surfaces (Fig. 14B, C, D) has also been used to form monolayers on gold [200, 279, 297, 298], mica [299], and at the air-water interface [300]. The weaker substrate-molecule interactions of these systems allow more facile modification of the surface. When porphyrins bear reactive groups (poised to form supramolecular or covalent bonds) on the side or edge facing away from the surface, more layers may be added with varying degrees of structural fidelity to the initial layer. Characterization of multilayer systems is often more difficult. A hybrid system with octacarboxlyate porphyrins adsorbed face-down onto a gold surface surrounded by a bolaamphophile or other monolayers has been reported [301–303]. The motivation behind this work is to make nanoscaled test tubes on the surface for further reaction or sensing applications.

6.2.2. Coordination

There are several examples of 14E-type monolayers where a ligand is covalently attached to the substrate and this binds to an axial position of a metalloporphyrin [304–310]. This method to make monolayers on surfaces generally orients the plane of the porphyrin parallel to the plane of the surface, but there is a lot of room for the macrocycle to tilt so that there is a range of orientations. This method allows for studies by surface plasmon resonance, surface-enhanced Raman scattering, and surface-enhanced infrared absorption. Again, one of the potential applications of these systems is the use of surface-modified electrodes for analytical chemistry.

As mentioned earlier, the relative orientation of the porphyrin rings with respect to each other and with respect to the surface is of paramount importance to the photonic properties of the system. An elegant way to use coordination chemistry to orient metalloporphyrins perpendicular to the surface is to make monolayers with flexible linkers and ligands that generally are poised to bind 6-coordinate metalloporphyrins in both axial positions, resulting in layers with the plane of the macrocycle perpendicular to the surface [311, 312].

6.2.3. Thin Films Using Liquid-Crystal-Type Interactions

One of the early successes in the self-organization of a porphyrin-based device utilized liquid-crystal-type interactions to form thin films of a photoconducting zinc porphyrin with eight dodecyloxy groups on the pyrroles [117, 313, 314]. Electron-hole pairs are generated upon irradiation of a section of a device that has this porphyrin derivative placed between two optically transparent electrodes with an applied electric field. The device then performs as a high-density nanosecond charge trap that can be used as an optical memory device. Similar porphyrins were used to make liquid-crystal thin films between ITO coated glass slides, which displayed electric-field modulated near-field photoluminescence [315]. These studies show that there is a correlation between the device's physical properties and the morphology of the self-organized, photoactive, organic layer. Other liquid-crystal-forming porphyrins with properties dependent on the nature and position of the hydrocarbon have also

been reported [316–318]. There are several examples of discrete multiporphyrin arrays (such as 3A-type, 3B-type without the guest, and 6A-type) that have long-chain hydrocarbons appended on the periphery that form monolayers by the Langmuir–Blodgett method (e.g., [319]), which then can be transferred to surfaces such as glass with reasonable structural integrity [320]. These films are an early example of using a secondary procedure to self-organize self-assembled supermolecules. A good example of an alternating-current light-emitting device based on Langmuir–Blodgett films of porphyrins was reported [321], where the transient character of electroluminescence is explained in terms of a space-charge assisted electron injection.

Solid vesicle membranes made of bixin-porphyrin derivatives have been reported to form stable colloidal solutions, which can be photopolymerized to yield 30- to 120-nm spheres [322] which can then be cast onto surfaces. Self-assembled fibrillar networks are formed from porphyrins bearing two lipid groups on the same side [323].

6.2.4. Layer-by-Layer Deposited Films

Only a few examples of films containing porphyrins made by the layer-by-layer technique are cited, because to date the degree of organization after the first layer is somewhat limited. Briefly, the process involves dipping a charged substrate into a solution containing an organic molecule or polymer of the opposite charge. After annealing and rinsing, it is dipped into a second solution that has molecules or polymers with the opposite charge as the first. This can be repeated, but the structural fidelity generally decreases with increasing number of layers. The density of defects in each layer is a major issue for most layer-by-layer systems, but is not generally reported. For porphyrins, an early example used TSPP (tetra anionic) and TPyP appended with four cationic $RuBiPy_2$ units [324]. Two-dimensional arrays formed from TPyP and cadmium ions made by Langmuir–Blodgett methods and deposited on surfaces [325], or the equivalent palladium ion linked array made with the layer-by-layer method [326], are two of the best examples of self-organized two-dimensional layers using specific intermolecular interactions. The stepwise self-organization of a bis-bipyridinium–porphyrin dyad and gold nanoparticles on a conductive ITO surface results in a system where the redox properties are significantly different than the component molecules [327]. A cationic polyamine adsorbed onto glass promotes adsorption of anionic TSPP [275] and similar systems [328]. The association of tetrapyridyl cobalt porphyrin coordinated to four $Ru(BiPy)_2$ associates with anionic Zn(II)TSPP in layer-by-layer films on electrodes has been used to examine nitrite oxidation as an analytical application [329]. ITO electrodes modified with porphyrin-fullerene dyads as the active layer gives rise to improved light-harvesting properties in the visible region [330].

6.3. Themes and Trends

The ability to deposit specifically designed self-assembled (discrete) supramolecular structures onto surfaces without change in structure or function remains a significant challenge for the exploitation of this methodology in real-world applications and devices. Self-organized monolayer systems, on the other hand, are more accessible because of the tolerance for local defects, but the horizontal dimensions of these are generally quite large. The layer-by-layer method has both advantages (ease of preparation) and disadvantages (greater defect density than SAMs), but may find numerous applications as sensors, and other systems requiring surface modification. The adsorption of a $La(III)(TPyP)_2^-$ sandwich compound onto a liquid Hg surface can provide access to two-dimensional arrays on this and other surfaces when they are chemically cross-linked with a dibromo hydrocarbon [331, 332]. The overall structure of these systems is likely a grid, but since the cross-linking step results in kinetic rather than thermodynamic products, the degree or order is likely not comparable to self-assembled/organized systems. This work is notable since it is an attempt to organize porphyrin sandwich compounds on surfaces with greater structural stability and arrangement unobtainable by self-assembled monolayers of similar compounds (see Section 6.2.1).

7. MISCELLANEOUS MULTIPORPHYRIN ASSEMBLIES AND AGGREGATES

There are several applications for nanoscaled porphyrin aggregates where the architectural order of the macrocycles is of little consequence or indeed undesirable. For example, nanoparticles are used for drug delivery, or as nonspecific sensor elements in cross-reactive or stochastic sensor arrays. A recent report describes the formation of 10- to 75-nm radius porphyrin nanoparticles using methods similar to those used to make inorganic nanoparticles [333], specifically, by injecting an organic solution containing the porphyrin and a small percentage of a short polyethylene glycol (PEG) into water (Scheme 3B). The solvent ratio, the specific PEG, and the rate of mixing, all affect the particle size. Dynamic light scattering and the optical spectra in solution characterized the particles. But more importantly for device applications, many of the particles can be deposited on glass surfaces without agglomeration, and without disrupting the nanoparticles. It was shown that iron porphyrin nanoparticles catalyze the oxidation of cyclohexene with about 1000 times the turnover of the individual porphyrin (normalized to the same number of porphyrins). The product ratios were close to those of the individual porphyrin, but it remains to be seen if the nanoparticle catalysts exhibit regioselectivity on more complex substrates.

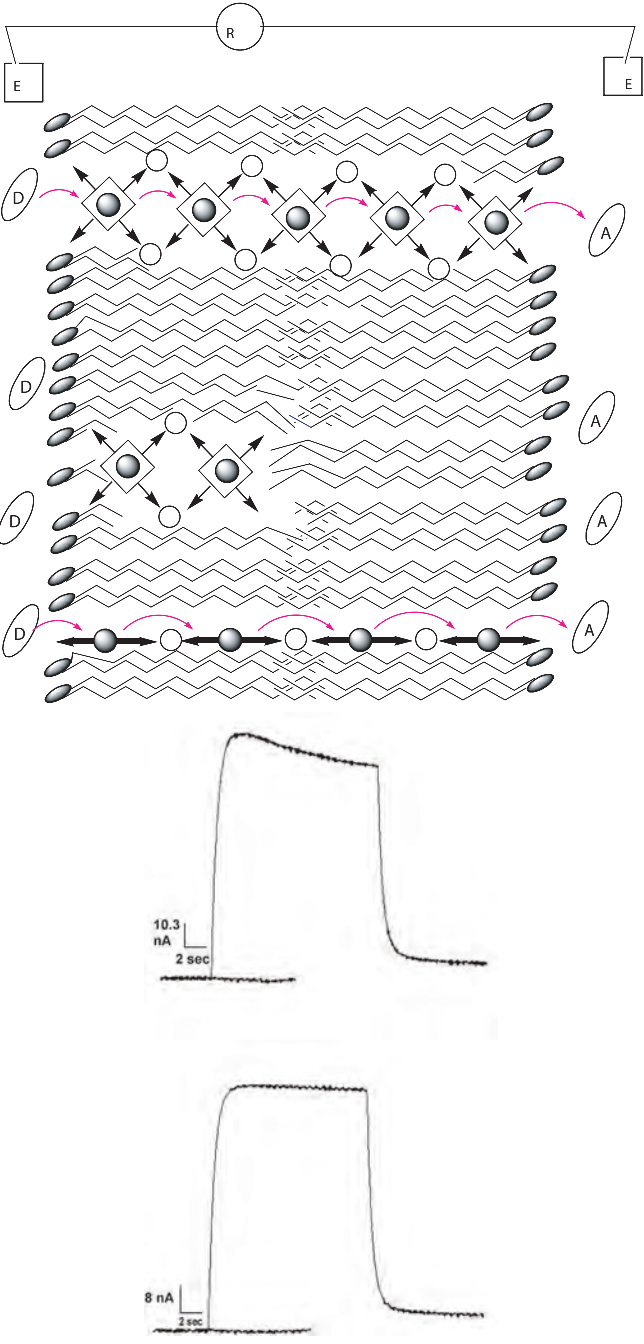
A porphyrin appended with four meso perylene imides also aggregates to form nanoparticles, but in this case there is likely a much greater degree of organization than in the previous system [104, 334]. These particles display photonic properties that depend on the aggregation state as well as the component molecule. Porphyrins have also been used as a means to organize Au nanoparticles [335].

7.1. Membranes: Molecular Electronics

Molecular ionic transistors. Lipid bilayer membranes and vesicles are self-organizing systems that mimic cell and organelle walls. The trans-membrane distance is ~8 nm. Again, mostly because of the relevance to photosynthesis, the reactivity and function of porphyrins in bilayer membranes and vesicles have been studied for more than 30 years

[336, 337], especially in terms of interfacial electron transfer, trans-membrane electronic and ionic currents, and redox chemistry. Early on it was realized that porphyrins partition into the bilayers with some degree of order—into the polar, ester group region of the head-group and perpendicular to the bilayer-water interface. The somewhat serendipitous discovery of a photo-gated *ion-conducting* transistor in the late 1980s [257–259] dovetailed well with the electrical circuit models of lipid bilayer function first proposed almost two decades earlier (e.g., [336, 337]). The charge-sensitive ion-conducting device (CSIC) was composed of lipophilic anions and porphyrins in a lipid bilayer. In the dark, with an applied potential of 40 mV, there is a space-charge limited conductance of the lipophilic anions [256]. Upon illumination with laser light in the absorption region of the macrocycle or white light, interfacial electron transfers from the lattice of porphyrins on both sides of the bilayer to aqueous electron acceptors resulted in two phenomena that allowed the conductance of the system to increase by several orders of magnitude. (1) The resultant porphyrin cation lattices (one on each side) significantly reduced the space-charge that limited the conductance of the lipophilic anions across the bilayer; and (2) the porphyrin cations now paired with the lipophilic anions to form an 'ion chain' which further reduced the barrier for ions crossing the hydrocarbon core of the membrane [259, 338]. Of course the stability of this system is limited by the stability of the membrane, and the response time is limited by the dynamics of assembling the ion chain. The inclusion of a protonophore results in a system that can photochemically pump protons to form a potential gradient [339, 340] across the membrane—one of the first self-organizing mimics of the essential outcome of photosynthesis which drives the synthesis of ATP. Other bilayer conducting systems with C_{60} and C_{70} fullerenes also serve as a basis for molecular electronics designs [341].

A molecular electronic transistor. The inclusion of a self-assembled membrane-spanning tape of porphyrins results in a significant improvement of the above system (Scheme 2) [98], and represents a molecular electronic device with all four levels of structure described in the introduction. By pre-assembling the conducting supermolecule, the gating time is significantly reduced, and the system need not rely on the movement of large ions across the viscous bilayer. Thus, when an electron donor is placed on one side of the bilayer and an electron acceptor on the opposite side, there is a chemical potential defined by the redox potentials of the donor and acceptor. Applying a voltage to this system results in no measurable conductance. Upon illumination with white light, the *electronic* conductance across the membrane increases by 50- to 1000-fold over controls without the donor or without the acceptor and with no applied voltage. This is an example of a self-organized, molecule-based photo-gated transistor, in this case driven by the electrochemical potential between the donor and acceptor. The self-assembled tapes are formed either by complementary H-bonds or by metal ion coordination. In addition to the advantages and limitations stated in the introduction, the benefit of self-assembly in this case is that the length of the tape is dictated by the width of the membrane. The water, phosphates, and esters compete for the hydrogen bonds and destabilize the coordination bonds. The electron donor and acceptor serve as the source and the drain, respectively, and the system is gated by light. This system also demonstrates that the interconnections to nanoscaled devices need not be nanoscaled *a priori*.



Scheme 2. Type-7C self-assembled linear polymers span an ~8 nm lipid bilayer parallel to the hydrocarbon chains (top and bottom). Not all supermolecules span the bilayer (center). Upon illumination with white light, the excited-state zinc porphyrin transfers an electron to an aqueous electron acceptor anthraquinone sulfate (A). The hole then hops across the supermolecule until quenched by an electron donor $K_4Fe(II)(CN)_6$ (D). Saturated calomel electrodes (E) reside in 0.1 M KCl solutions on each side of the bilayer, and an operational amplifier (R) measures the conductance. Conductance from 7A-type tape (top), and 7C-type tape (bottom). The baseline traces represent controls with no acceptor.

7.2. Vesicles

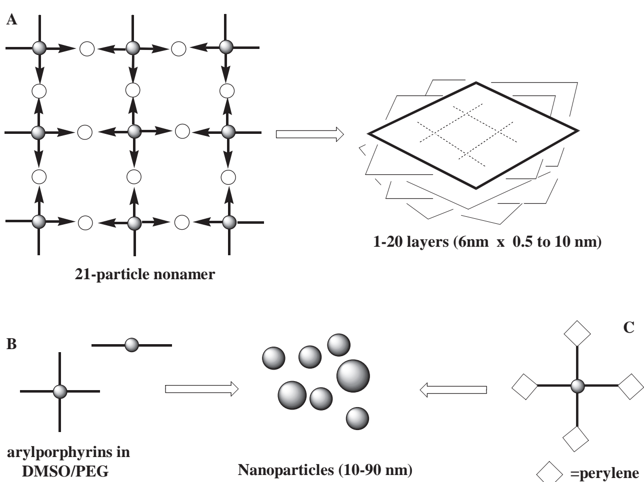
Another elegant example of the formation of potential gradients across membranes of vesicles uses a covalently bound porphyrin-vinyl-fullerene triad and similar molecules. Though these are molecular entities, they can be self-organized in bilayers of vesicles so that they mediate trans-membrane electrical currents upon illumination. The current then generates a potential that can drive the synthesis of ATP [78]. This is a dramatic example of the combined power of organic synthesis (primary structure) coupled with the self-organizing properties of lipids. Through the years there have been many other reports on the activity of porphyrins in membranes such as cytochrome P450 mimics (e.g., [81, 211]). A porphyrin bearing eight dendrons terminated with eight acetylenes can be self-encapsulated upon polymerization of the acetylenes using a Grubb's catalyst. The removal of the porphyrin core leaves a hollow dendrimer capable of recognizing a variety of porphyrins [71]. This remarkably effective example of molecular imprinting works because of the three-dimensional molecular organization of the dendrimer. But, the focus here is on self-assembled multiporphyrinic systems. Liposomes composed of porphyrins bearing long-chain hydrocarbon-substituted imidazoles [253, 342] have been reported.

7.3. Aggregates of Supermolecules

Remarkably, the self-assembly of 3 × 3 nonameric arrays can be accomplished by the stoichiometric addition of 12 Pd(II) ions to four "L"-shaped porphyrins that serve as corners, four "T"-shaped porphyrins that constitute the sides, and one "+"-shaped central porphyrin (see Section 3.1) [99]. These nonameric arrays can then stack to form columnar aggregates (Scheme 3A) that are ~7 nm in diameter by ~1 nm tall [100, 115]. These columnar stacks exhibit properties unique to this scale and degree of supramolecular order in that they are remarkably luminescent [104]. There are numerous papers on the aggregation of various porphyrin derivatives (neutral, negatively, and positively charged) where the quantum yield and lifetimes are ~10 times less than the individual porphyrin (e.g., [80, 262, 271–273]). This is in marked contrast to the self-assembled nonamers and columnar stacks of nonamers where both the heavy atom effect and aggregations should lead to quite meager luminescence, but instead the luminescence is only quenched by ~ 50%. This observation further indicates a high degree of order in the columnar stacks. These nanoscaled materials are stable in air at room temperature for years without detectable changes in their optical or morphological character [104].

7.4. Themes and Trends

Porphyrins have been known to aggregate since their first isolation and synthesis. The efficiency in harvesting solar energy, for example, arises from the specific alignment and organization of chlorophyll pigments in antenna complexes and reaction centers. Since many nanotechnological applications of porphyrins require narrowly dispersed nanoaggregates, there is a need to predictably and reliably control the size of the particles. Encapsulation of aggregates is an obvious method, but this may not be appropriate for all applications. There have been a few recent reports on the controlled formation of porphyrin nanoparticles with predictable size without resorting to an external matrix.



Scheme 3. Other porphyrin aggregates.

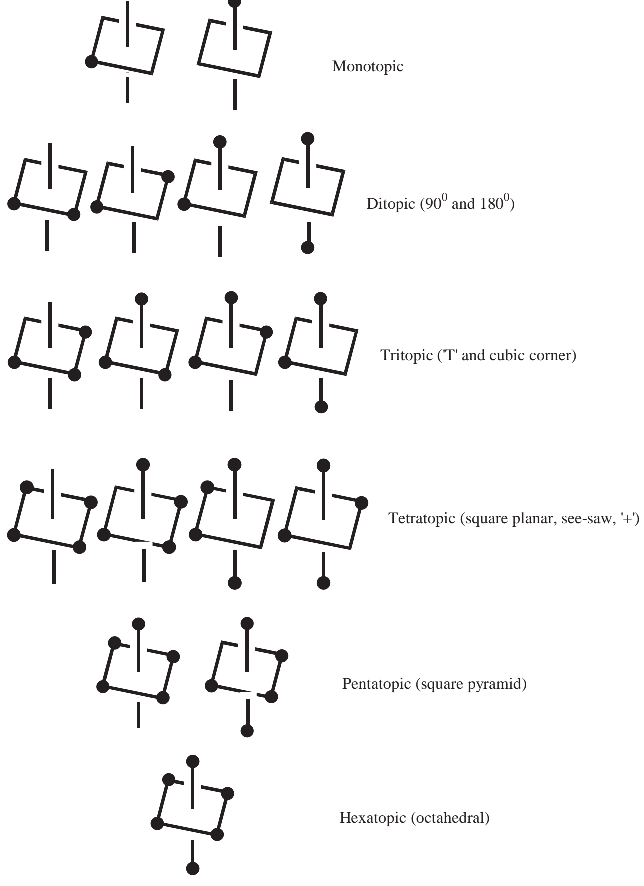


Figure 15. Topologies of a metalloporphyrin bearing four intermolecular binding motifs at the meso positions.

8. SUMMARY

After nearly 40 years of intense research, synthetic supramolecular chemistry has developed a wide variety of tools, methods, and strategies to form a staggering number of self-assembled and self-organized nanoscaled entities with diverse functions. Because porphyrins are rigid molecules, they have predictable shapes and orientations; therefore, these molecules are very often used as techtons. Porphyrins can act as monotopic caps, as linear linkers, as 90° turns, as bifurcating "T"-shaped crossroads, as square planar linkers or centers, and metalloporphyrins can serve as square pyramidal components, or as octahedral components (Fig. 15). Obviously, incorporating more complex moieties capable of more than one intermolecular interaction (such as meso 3,5-dicarboxyphenyl- and 3,5-dipyridylphenyl-) or attaching these molecular recognition moieties to the pyrrole positions will significantly increase the number of topologies available for the design of supramolecular porphyrinic systems. Coupling these porphyrin topologies with the various coordination geometries of transition metals and/or the various geometries of other multitopic ligands can result in supramolecular systems that are only limited by one's imagination. The requirements of the application dictate the design of self-assembled or self-organized systems such as porphyrins, but because of the rich supramolecular chemistry of porphyrins these requirements can be met. The resultant nanoscaled materials can have a wide variety of properties as photonics, catalysts, sensors, and sieves—each of these functions can be fine-tuned by modifications of either the macrocycle or the bound metal ions.

Recent endeavors into the formation of dynamic combinatorial libraries may afford a new means to find multiporphyrin arrays that have unimagined or optimal structures, with unimagined or optimal functions [171, 343]. Materials that switch their structure as a response to some environmental factor or a chemical entity [97, 344] can also be formed with porphyrinoids, which would then serve as a reporter of the structural change. Finally, methods of porphyrin synthesis have facilitated their use both in the lab and for commercial applications [94, 345].

GLOSSARY

Photonic The interplay between photophysical properties of a molecule or material and the electronic properties of the material, such as in photoinduced electron or energy transfer.

Self-assembly The spontaneous association of molecules and/or ions to form a specific supramolecular entity.

Self-organization The spontaneous association of molecules and/or ions to form polymeric materials, or dynamic two-dimensional systems such as liquid crystals.

Supramolecular Referring to a discrete entity containing two or more molecules/ions associated by intermolecular forces (such as hydrogen bonding, coordination chemistry, and electrostatic interactions) in a specific geometric arrangement.

ACKNOWLEDGMENTS

We would like to thank Sandy-Anne Noel and Yanilda Garcia for assistance with the references, and Diana Samaroo and Garry Hanan for careful readings of the manuscript. Funding from The National Science Foundation (CHE-0135509, IGERT DGE-9972892), the Israel-US Binational Science Foundation (1999082), and the National Institutes of Health (SCORE) is gratefully acknowledged. Dedicated to Marie M. K. Drain on the occasion of her 70th Birthday.

REFERENCES

1. D. C. Mauzerall, *Clin. Dermat.* 16, 195 (1998).
2. M. Freemantle, *C & EN* Oct. 26, 37 (1998).
3. F. T. Hong, *J. Mol. Elect.* 5, 163 (1989).
4. Y. Kobuke and N. Nagata, *Mol. Cryst. Liq. Cryst.* 342, 51 (2000).
5. M. R. Wasielewski, *Chem. Rev.* 92, 435 (1992).
6. M. R. Wasielewski, in "The Photosynthetic Reaction Center," Vol. II (J. Deisenhofer and J. Norris, Eds.), Vol. 11, p. 465. Academic Press, New York, 1993.
7. V. K. Yachandra, K. Sauer, and M. P. Klein, *Chem. Rev.* 96, 2927 (1996).
8. I. Willner, E. Kaganer, E. Joselevich, H. Durr, E. David, M. J. Günter, and M. R. Johnston, *Coord. Chem. Rev.* 171, 261 (1998).
9. B. Jiang, S.-W. Yang, and W. E. Jones, Jr., *Chem. Mater.* 9, 2031 (1997).
10. B. Jiang, R. Niver, S.-W. Yang, and W. E. Jones, Jr., *Syn. Met.* 94, 205 (1997).
11. B. Jiang, S. W. Yang, S. L. Bailey, L. G. Hermans, R. A. Niver, M. A. Bolcar, and W. E. Jones, Jr., *Coord. Chem. Rev.* 171, 365 (1998).
12. B. Jiang, S.-W. Yang, P. T. Lam, and W. E. Jones, Jr., *Mat. Res. Soc. Symp. Proc.* 488, 323 (1998).
13. S. Anderson, H. L. Anderson, A. Bashall, M. McPartlin, and J. K. M. Sanders, *Angew. Chem. Int. Ed.* 34, 1096 (1995).
14. S. Hecht, H. Ihre, and J. M. J. Frechet, *J. Am. Chem. Soc.* 121, 9239 (1999).
15. K. S. Kwok and J. C. Ellenbogen, *Mater. Today* 5, 28 (2002).
16. J. C. Ellenbogen and J. C. Love, *Proc. IEEE* 88, 386 (2000).
17. J. M. Tour, *Acc. Chem. Res.* 33, 791 (2000).
18. C. S. Lent, *Science* 288, 1597 (2000).
19. A. K. Burrell and M. R. Wasielewski, *J. Porph. Phthal.* 4, 401 (2000).
20. Y. Diskin-Posner, S. Dahal, and I. Goldberg, *Angew. Chem. Int. Ed.* 39, 1288 (2000).
21. R. K. Kumar, Y. Diskin-Posner, and I. Goldberg, *J. Incl. Phenom. Macrocyclic Chem.* 37, 219 (2000).
22. M. E. Kosal, J.-H. Chou, and K. S. Suslick, *J. Porph. Phthal.* 6, 377 (2002).
23. M. L. Merlau, M. d. P. Mejia, S. T. Nguyen, and J. T. Hupp, *Angew. Chem. Int. Ed.* 40, 4239 (2001).
24. V. V. Borovkov, J. M. Lintuluoto, H. Sugeta, M. Fujiki, R. Arakawa, and Y. Inoue, *J. Am. Chem. Soc.* 124, 2993 (2002).
25. A. Robertson and S. Shinkai, *Coord. Chem. Rev.* 205, 157 (2000).
26. K. Ogawa, T. Zhang, K. Yoshihara, and Y. Kobuke, *J. Am. Chem. Soc.* 124, 22 (2002).
27. A. A. Bothner-By, J. Dadok, T. E. Johnson, and J. S. Lindsey, *J. Phys. Chem.* 100, 17551 (1996).
28. D. Holten, D. F. Bocian, and J. S. Lindsey, *Acc. Chem. Res.* 35, 57 (2002).
29. A. Osuka, N. Tanabe, R. Zhang, and K. Maruyama, *Chem. Lett.* 1505 (1993).
30. G. S. Wilson and H. L. Anderson, *Chem. Commun.* 1539 (1999).
31. A. Ambroise, J. Li, L. Yu, and J. S. Lindsey, *Org. Lett.* 2, 2563 (2000).

32. A. Ambroise, R. W. Wagner, P. D. Rao, J. A. Riggs, P. Hascoat, J. R. Diers, J. Seth, R. K. Lammi, D. F. Bocian, D. Holten, and J. S. Lindsey, *Chem. Mater.* 13, 1023 (2001).
33. M. d. R. Benites, T. E. Johnson, S. Weghorn, L. Yu, P. D. Rao, J. R. Diers, S. I. Yang, C. Kirmaier, D. F. Bocian, D. Holten, and J. S. Lindsey, *J. Mater. Chem.* 12, 65 (2002).
34. J. Li and J. S. Lindsey, *J. Org. Chem.* 64, 9101 (1999).
35. J. S. Lindsey, *New J. Chem.* 15, 153 (1991).
36. K. M. Roth, D. T. Gryko, C. Clausen, J. Li, J. S. Lindsey, W. G. Kuhr, and D. F. Bocian, *J. Phys. Chem. B* 106, 8639 (2002).
37. R. W. Wagner, J. S. Lindsey, J. Seth, V. Palaniappan, and D. F. Bocian, *J. Am. Chem. Soc.* 118, 3996 (1996).
38. L. Yu and J. S. Lindsey, *J. Org. Chem.* 66, 7402 (2001).
39. J. T. Fletcher and M. Therien, *Inorg. Chem.* 41, 331 (2002).
40. J. T. Fletcher and M. J. Therien, *J. Am. Chem. Soc.* 124, 4298 (2002).
41. A. G. Hyslop, M. A. Kellet, P. M. Iovine, and M. J. Therien, *J. Am. Chem. Soc.* 120, 12676 (1998).
42. R. Kumble, S. Palese, V. S.-Y. Lin, M. J. Therien, and R. M. Hochstrasser, *J. Am. Chem. Soc.* 120, 11489 (1998).
43. S. M. LeCours, H.-W. Guan, S. G. DiMagno, C. H. Wang, and M. J. Therien, *J. Am. Chem. Soc.* 118, 1497 (1996).
44. V. S.-Y. Lin, S. G. DiMagno, and M. J. Therien, *Science* 264, 1105 (1994).
45. K. T. Moore, J. T. Fletcher, and M. J. Therien, *J. Am. Chem. Soc.* 121, 5196 (1999).
46. P. J. F. de Rege, S. A. Williams, and M. J. Therien, *Science* 269, 1409 (1995).
47. M. G. H. Vicente, L. Jaquinod, and K. M. Smith, *Chem. Commun.* 1771 (1999).
48. N. Aratani and A. Osuka, *Org. Lett.* 3, 4213 (2001).
49. N. Aratani and A. Osuka, *Macromol Rapid Commun.* 22, 725 (2001).
50. N. Aratani, A. Osuka, Y. H. Kim, D. H. Jeong, and D. Kim, *Angew. Chem. Int. Ed.* 39, 1458 (2000).
51. A. Nakano, A. Osuka, T. Yamazaki, Y. Nishimura, S. Akimoto, I. Yamazaki, A. Itaya, M. Murakami, and H. Miyasaka, *Chem. Eur. J.* 7, 3134 (2001).
52. A. Tsuda, H. Furuta, and A. Osuka, *J. Am. Chem. Soc.* 123, 10304 (2001).
53. H. A. M. Biemans, A. E. Rowan, A. Verhoeven, P. Vanoppen, L. Latterini, J. Foekema, A. P. H. J. Schenning, E. W. Meijer, F. C. de Schryver, and R. J. M. Nolte, *J. Am. Chem. Soc.* 120, 11054 (1998).
54. A. P. Alivisatos, P. F. Barbara, A. W. Castleman, J. Chang, D. A. Dixon, M. L. Klein, G. L. McLendon, J. S. Miller, M. A. Ratner, P. J. Rossky, S. I. Stupp, and M. E. Thompson, *Adv. Mater.* 10, 1297 (1998).
55. D. Philp and J. F. Stoddart, *Angew. Chem. Int. Ed.* 35, 1154 (1996).
56. T. Imamura and K. Fukushima, *Coord. Chem. Rev.* 198, 133 (2000).
57. G. F. Swiegers and T. J. Malefetse, *Coord. Chem. Rev.* 225, 91 (2002).
58. D. S. Lawrence, T. Jiang, and M. Levett, *Chem. Rev.* 95, 2229 (1995).
59. T.-H. Tran-Thi, *Coord. Chem. Rev.* 160, 53 (1997).
60. K. M. Kadish, K. M. Smith, and R. Guilard, "The Porphyrin Handbook," Vols. 1–12. Academic Press, New York, 2000.
61. A. K. Burrell, D. L. Officer, P. G. Plieger, and D. C. W. Reid, *Chem. Rev.* 101, 2751 (2001).
62. W. Sliwa and B. Mianowska, *Transition Met. Chem.* 25, 491 (2000).
63. G. F. Swiegers and T. J. Malefetse, *Chem. Rev.* 100, 3483 (2000).
64. H. E. Toma and K. Araki, *Coord. Chem. Rev.* 196, 307 (2000).
65. J. Wojaczýnski and L. Latos-Grażýnski, *Coord. Chem. Rev.* 204, 113 (2000).
66. C.-T. Chen and K. S. Suslick, *Coord. Chem. Rev.* 128, 293 (1993).
67. J.-H. Chou, M. E. Kosal, H. S. Nalwa, N. A. Rakow, and K. S. Suslick, in "The Porphyrin Handbook" (K. M. Kadish, K. M. Smith, and R. Guilard Eds.), Vol. 6, p. 43. Academic Press, New York, 2000.
68. M. E. Kosal and K. S. Suslick, *J. Solid State Chem.* 152, 87 (2000).
69. F. Diederich and B. Felber, *Proc. Natl. Acad. Sci. USA* 99, 4778 (2002).
70. K. Inoue, *Progr. Polym. Sci.* 25, 453 (2000).
71. S. C. Zimmerman, M. S. Wendland, N. A. Rakow, I. Zharov, and K. S. Suslick, *Nature* 418, 399 (2002).
72. E. K. L. Yeow, K. P. Ghiggino, J. N. H. Reek, M. J. Crossley, A. W. Bosman, A. P. H. J. Schenning, and E. W. Meijer, *J. Phys. Chem. B* 104, 2596 (2000).
73. A. Naoki, T. Akihiko, and O. Atsuhiro, *Syn. Lett.* 11, 1663 (2001).
74. T. E. O. Screen, J. R. G. Thorne, R. G. Denning, D. G. Bucknall, and H. L. Anderson, *J. Am. Chem. Soc.* 124, 9712 (2002).
75. P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.* 121, 11538 (1999).
76. D. M. Sarno, L. J. Matienzo, and W. E. Jones, Jr., *Inorg. Chem.* 40, 6308 (2001).
77. D. M. Sarno, B. Jiang, D. Grosfeld, J. O. Afriyie, L. J. Matienzo, and W. E. Jones, Jr., *Langmuir* 16, 6191 (2000).
78. D. Gust, T. A. Moore, and A. L. Moore, *Acc. Chem. Res.* 34, 40 (2001).
79. P. A. Liddell, G. Kodis, A. L. Moore, T. A. Moore, and D. Gust, *J. Am. Chem. Soc.* 124, 7668 (2002).
80. W. Xu, H. Guo, and D. L. Akins, *J. Phys. Chem. B* 105, 1543 (2001).
81. S. B. Ungashe and J. T. Groves, *Adv. Inorg. Biochem.* 9, 317 (1994).
82. C. M. Drain, D. B. Sable, and B. B. Corden, *Inorg. Chem.* 27, 2396 (1988).
83. H. J. Dutton, *Photosynth. Res.* 52, 175 (1997).
84. N. Krauss, W.-D. Schubert, O. Klukas, P. Fromme, H. T. Witt, and W. Saenger, *Nature Struct. Biol.* 3, 973 (1996).
85. D. Mauzerall and N. L. Greenbaum, *Biochim. Biophys. Acta* 974, 119 (1989).
86. G. McDermott, S. M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz, R. J. Cogdell, and N. W. Isaacs, *Nature* 374, 517 (1995).
87. J.-C. Chambron, V. Heitz, and J.-P. Sauvage, in "The Porphyrin Handbook" (K. M. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 6, p. 1. Academic Press, New York, 2000.
88. S. Leininger, B. Olenyuk, and P. J. Stang, *Chem. Rev.* 100, 853 (2000).
89. J.-M. Lehn, *Pure Appl. Chem.* 66, 1961 (1994).
90. M. Fujita, *Chem. Soc. Rev.* 27, 417 (1998).
91. G. M. Whitesides, E. E. Simanek, J. P. Mathias, C. T. Seto, D. N. Chin, M. Mammen, and D. M. Gordon, *Acc. Chem. Res.* 28, 37 (1995).
92. P. J. Stang and B. Olenyuk, *Acc. Chem. Res.* 30, 502 (1997).
93. P. J. Stang, *Chem. Eur. J.* 4, 19 (1998).
94. J. S. Lindsey, in "The Porphyrin Handbook" (K. M. Kadish, K. M. Smith, and R. Guilard, Eds.), Vol. 1, p. 45. Academic Press, New York, 2000.
95. K. M. Roth, N. Dontha, R. B. Dabke, D. T. Gryko, C. Clausen, J. S. Lindsey, D. F. Bocian, and W. G. Kuhr, *J. Vac. Sci. Technol. B* 18, 2359 (2000).
96. J.-M. Lehn, *Angew. Chem. Int. Ed.* 29, 1304 (1990).
97. J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* 99, 4763 (2002).
98. C. M. Drain, *Proc. Natl. Acad. Sci. USA* 99, 5178 (2002).
99. C. M. Drain, F. Nifiatis, A. Vasenko, and J. D. Batteas, *Angew. Chem. Int. Ed.* 37, 2344 (1998).
100. C. M. Drain, J. D. Batteas, G. W. Flynn, T. Milic, N. Chi, D. G. Yablon, and H. Sommers, *Proc. Natl. Acad. Sci. USA* 99, 6498 (2002).
101. K. Kalyanasundaram and M. Grätzel, *Coord. Chem. Rev.* 77, 347 (1998).
102. G. A. Mines, B.-C. Tzeng, K. J. Stevenson, J. Li, and J. T. Hupp, *Angew. Chem. Int. Ed.* 41, 154 (2002).

103. S. B. Lei, C. Wang, S. X. Yin, H. N. Wang, F. Xi, H. W. Liu, B. Xu, L. J. Wan, and C. L. Bai, *J. Phys. Chem. B* 105, 10838 (2001).
104. C. M. Drain, J. T. Hupp, K. S. Suslick, M. R. Wasielewski, and X. Chen, *J. Porph. Phthal.* 6, 241 (2002).
105. Y. Diskin-Posner, R. K. Kumar, and I. Goldberg, *New. J. Chem.* 23, 885 (1999).
106. Y. Diskin-Posner, G. K. Patra, and I. Goldberg, *Eur. J. Inorg. Chem.* 2515 (2001).
107. Y. Diskin-Posner, G. K. Patra, and I. Goldberg, *Acta Cryst.* m344 (2002).
108. I. Goldberg, *Cryst. Eng. Comm.* 4, 109 (2002).
109. R. K. Kumar and I. Goldberg, *Angew. Chem. Int. Ed.* 37, 3027 (1998).
110. R. K. Kumar, S. Balasubramanian, and I. Goldberg, *Inorg. Chem.* 37, 541 (1998).
111. S. Subramanian and M. J. Zaworotko, *Coord. Chem. Rev.* 137, 357 (1994).
112. B. Moulton and M. J. Zaworotko, *Curr. Opinion Solid State Mater. Sci.* 6, 117 (2002).
113. B. Moulton and M. J. Zaworotko, *Chem. Rev.* 101, 1629 (2001).
114. C. A. Schalley, *Int. J. Mass Spectrom*. 194, 11 (2000).
115. T. N. Milic, N. Chi, D. G. Yablon, G. W. Flynn, J. D. Batteas, and C. M. Drain, *Angew. Chem. Int. Ed.* 41, 2117 (2002).
116. S. I. Stupp, M. U. Pralle, G. N. Tew, L. Li, M. Sayar, and E. R. Zubarev, *MRS Bull.* 25, 42 (2000).
117. C.-Y. Liu, H.-L. Pan, M. A. Fox, and A. J. Bard, *Science* 261, 897 (1993).
118. E. I. Zenkevich, C. v. Borczyskowski, A. M. Shulga, S. Bachilo, U. Rempel, and A. Willert, *Chem. Phys.* 275, 185 (2002).
119. J. H. Wosnick and T. M. Swager, *Curr. Opinion Chem. Biol.* 4, 715 (2000).
120. M. Biesaga, K. Pyrzynska, and M. Trojanowicz, *Talanta* 51, 209 (2000).
121. R. Purrello, S. Gurrieri, and R. Lauceri, *Coord. Chem. Rev.* 190–192, 683 (1999).
122. R. V. Slone and J. T. Hupp, *Inorg. Chem.* 36, 5422 (1997).
123. B. Linton and A. D. Hamilton, *Chem. Rev.* 97, 1669 (1997).
124. H. Ogoshi, T. Mizutani, T. Hayashi, and Y. Kuroda, "Porphyrins and Metalloporphyrins as Receptor Models in Molecular Recognition," Vol. 6. Academic Press, New York, 2000.
125. A. J. Epstein, *MRS Bull.* 25, 33 (2000).
126. V. Laget, C. Hornick, P. Rabu, M. Drillon, and R. Ziessel, *Coord. Chem. Rev.* 178–180, 1533 (1998).
127. J. S. Miller, *Polyhedron* 20, 1723 (2001).
128. V. Georgakilas, F. Pellarini, M. Prato, D. M. Guldi, M. Melle-Franco, and F. Zerbetto, *Proc. Natl. Acad. Sci. USA* 99, 5075 (2002).
129. D. Sun, F. S. Tham, C. A. Reed, L. Chaker, and P. D. W. Boid, *J. Am. Chem. Soc.* 124, 6604 (2002).
130. D. Sun, F. S. Tham, C. A. Reed, and P. D. W. Boyd, *Proc. Natl. Acad. Sci. USA* 99, 5088 (2002).
131. C. M. Christensen, *MRS Bull.* 27, 278 (2002).
132. G. A. Ozin, *Chem. Commun.* 419 (2000).
133. G. F. Swiegers and T. J. Malefetse, *Coord. Chem. Rev.* 225, 91 (2002).
134. C. M. Drain, C. Kirmaier, C. J. Medforth, D. J. Nurco, K. M. Smith, and D. Holten, *J. Phys. Chem.* 100, 11984 (1996).
135. C. M. Drain, S. Gentemann, J. A. Roberts, N. Y. Nelson, C. J. Medforth, S. Jia, M. C. Simpson, K. M. Smith, J. Fajer, J. A. Shelnutt, and D. Holten, *J. Am. Chem. Soc.* 120, 3781 (1998).
136. L. Fabbrizzi, M. Licchelli, and P. Pallavicini, *Acc. Chem. Res.* 32, 846 (1999).
137. B. J. Holliday and C. A. Mirkin, *Angew. Chem. Int. Ed.* 40, 2022 (2001).
138. S.-S. Sun and A. J. Lees, *Coord. Chem. Rev.* 230, 170 (2002).
139. A. Prodi, C. J. Kleverlaan, M. T. Indelli, F. Scandola, E. Alessio, and E. Iengo, *Inorg. Chem.* 40, 3498 (2001).
140. C. M. Drain and J.-M. Lehn, *Chem. Commun.* 2313 (1994).
141. E. Iengo, R. Minatel, B. Milani, L. G. Marzilli, and E. Alessio, *Eur. J. Inorg. Chem.* 609 (2001).
142. E. Alessio, E. Iengo, and L. G. Marzilli, *Supramol. Chem.* 14, 103 (2002).
143. E. Iengo, B. Milani, E. Zangrando, S. Geremia, and E. Alessio, *Angew. Chem. Int. Ed.* 39, 1096 (2000).
144. E. Alessio, E. Ciani, E. Iengo, V. Y. Kukushkin, and L. G. Marzilli, *Inorg. Chem.* 39, 1434 (2000).
145. S. Belanger and J. T. Hupp, *Angew. Chem. Int. Ed.* 38, 2222 (1999).
146. S. Belanger, M. H. Keefe, J. L. Welch, and J. T. Hupp, *Coord. Chem. Rev.* 190–192, 29 (1999).
147. P. H. Dinolfo and J. T. Hupp, *Chem. Mater.* 13, 3113 (2001).
148. M. H. Keefe, K. D. Benkstein, and J. T. Hupp, *Coord. Chem. Rev.* 205, 201 (2000).
149. S. T. Nguyen, D. L. Gin, J. T. Hupp, and X. Zhang, *Proc. Natl. Acad. Sci. USA* 98, 11849 (2001).
150. K. E. Splan, M. H. Keefe, A. M. Massari, K. A. Walters, and J. T. Hupp, *Inorg. Chem.* 41, 619 (2002).
151. M. E. Williams and J. T. Hupp, *J. Phys. Chem. B* 105, 8944 (2001).
152. J. Fan, J. A. Whiteford, B. Olenyuk, M. D. Levin, P. J. Stang, and E. B. Fleischer, *J. Am. Chem. Soc.* 121, 2741 (1999).
153. M. Schmitz, S. Leininger, J. Fan, A. M. Arif, and P. J. Stang, *Organometallics* 18, 4817 (1999).
154. F. M. Tabellion, S. R. Seidel, A. M. Arif, and P. J. Stang, *J. Am. Chem. Soc.* 123, 7740 (2001).
155. M. M. Yatskou, R. B. M. Koehorst, A. van Hoek, H. Donker, and T. J. Schaafsma, *J. Phys. Chem. A* 105, 11432 (2001).
156. G. Ercolani, M. Ioele, and D. Monti, *New J. Chem.* 25, 783 (2001).
157. C. Ikeda, Y. Tanaka, T. Fujihara, Y. Ishii, T. Ushiyama, K. Yamamoto, N. Yoshioka, and H. Inoue, *Inorg. Chem.* 40, 3395 (2001).
158. X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, and L. D. Sarson, *Chem. Commun.* 2563 (1995).
159. C. A. Hunter, C. M. R. Low, M. J. Packer, S. E. Spey, J. G. Vinter, M. O. Vysotsky, and C. Zonta, *Angew. Chem. Int. Ed.* 40, 2678 (2001).
160. A. Okumura, K. Funatsu, Y. Sasaki, and T. Imamura, *Chem. Lett.* 779 (1999).
161. M. R. Johnston, M. J. Latter, and R. N. Warrener, *Org. Lett.* 4, 2165 (2002).
162. E. Iengo, E. Zangrando, R. Minatel, and E. Alessio, *J. Am. Chem. Soc.* 124, 1003 (2002).
163. A. Ikeda, M. Ayabe, S. Shinkai, S. Sakamoto, and K. Yamaguchi, *Org. Lett.* 2, 3707 (2000).
164. N. Fujita, K. Biradha, M. Fujita, S. Sakamoto, and K. Yamaguchi, *Angew. Chem. Int. Ed.* 40, 1718 (2001).
165. A. Willert, S. Bachilo, U. Rempel, A. Shulga, E. Zenkevich, and C. von Borczyskowski, *J. Photochem. Photobiol. A: Chem.* 126, 99 (1999).
166. C. A. Hunter and R. Tregonning, *Tetrahedron* 58, 691 (2002).
167. L. J. Twyman and A. S. H. King, *Chem. Commun.* 910 (2002).
168. D. W. J. McCallien and J. K. M. Sanders, *J. Am. Chem. Soc.* 117, 6611 (1995).
169. S. Rucareanu, O. Mongin, A. Schuwey, N. Hoyler, A. Gossauer, W. Amrein and H.-u. Hediger *J. Org. Chem.* 66, 4973 (2001).
170. M. Nakash and J. K. M. Sanders, *Perkin Trans.* 2, 2189 (2001).
171. E. Stulz, Y.-F. Ng, S. M. Scott, and J. K. M. Sanders, *Chem. Commun.* 524 (2002).
172. C. M. Drain, X. Shi, T. Milic, and F. Nifiatis, *Chem. Commun.* 287 (2001).
173. H. Yuan, L. Thomas, and L. K. Woo, *Inorg. Chem.* 35, 2808 (1996).
174. E. Alessio, M. Macchi, S. L. Heath, and L. G. Marzilli, *Inorg. Chem.* 36, 5614 (1997).
175. M. Andersson, M. Linke, J.-C. Chambron, J. Davidsson, V. Heitz, L. Hammarstrom, and J.-P. Sauvage, *J. Am. Chem. Soc.* 124, 4347 (2002).

176. I. M. Dixon, J.-P. Collin, J.-P. Sauvage, and L. Flamigni, *Inorg. Chem.* 40, 5507 (2001).
177. L. Flamigni, F. Barigelletti, N. Armaroli, J.-P. Collin, I. M. Dixon, J.-P. Sauvage, and J. A. G. Williams, *Coord. Chem. Rev.* 190–192, 671 (1999).
178. L. Flamigni, I. M. Dixon, J.-P. Collin, and J.-P. Sauvage, *Chem. Commun.* 2479 (2000).
179. C. C. Mak, N. Bampos, and J. K. M. Sanders, *Chem. Commun.* 1085 (1999).
180. A. Prodi, M. T. Indelli, C. J. Kleverlaan, E. Alessio, and F. Scandola, *Coord. Chem. Rev.* 229, 51 (2002).
181. E. Alessio, S. Geremia, S. Mestroni, I. Srnova, M. Slouf, T. Gianferrara, and A. Prodi, *Inorg. Chem.* 38, 2527 (1999).
182. R. A. Haycock, A. Yartsev, U. Michelsen, V. Sundstrom, and C. A. Hunter, *Angew. Chem. Int. Ed.* 39, 3616 (2000).
183. J. E. Redman, N. Feeder, S. J. Teat, and J. K. M. Sanders, *Inorg. Chem.* 40, 2486 (2001).
184. K. Campbell, R. McDonald, and R. R. Tykwinski, *J. Org. Chem.* 67, 1133 (2002).
185. J. L. Allwood, A. K. Burrell, D. L. Officer, S. M. Scott, K. Y. Wild, and K. C. Gordon, *Chem. Commun.* 747 (2000).
186. Y. Tomohiro, A. Satake, and Y. Kobuke, *J. Org. Chem.* 66, 8422 (2001).
187. C. M. Drain, K. F. Cheng, and K. Grohmann, *Inorg. Chem.* 42, 2075 (2003).
188. K. Chichak and N. R. Branda, *Chem. Commun.* 523 (1999).
189. K. Chichak and N. R. Branda, *Chem. Commun.* 1211 (2000).
190. V. F. Slagt, J. N. H. Reek, P. C. J. Kamer, and P. W. N. M. van Leeuwen, *Angew. Chem. Int. Ed.* 40, 4271 (2001).
191. G. D. Fallon, S. J. Langford, M. A.-P. Lee, and E. Lygris, *Inorg. Chem. Commun.* 5, 715 (2002).
192. B. G. Maiya, N. Bampos, A. A. Kumar, N. Feeder, and J. K. M. Sanders, *New J. Chem.* 25, 797 (2001).
193. A. A. Kumar, L. Giribabu, D. R. Reddy, and B. G. Maiya, *Inorg. Chem.* 40, 6757 (2001).
194. S. R. Batten, *Curr. Opinion Solid State Mater. Sci.* 5, 107 (2001).
195. U. Michelsen and C. A. Hunter, *Angew. Chem. Int. Ed.* 39, 764 (2000).
196. K. Ogawa and Y. Kobuke, *Angew. Chem. Int. Ed.* 39, 4070 (2000).
197. T. S. Balaban, J. Leitich, A. R. Holzwarth, and K. Schaffner, *J. Phys. Chem. B* 104, 1362 (2000).
198. C. V. K. Sharma, G. A. Broker, J. G. Huddleston, J. W. Baldwin, R. M. Metzger, and R. D. Rogers, *J. Am. Chem. Soc.* 121, 1137 (1999).
199. C. V. K. Sharma, G. A. Broker, and R. D. Rogers, *J. Solid State Chem.* 152, 253 (2000).
200. C. V. K. Sharma, G. A. Broker, G. J. Szulczewski, and R. D. Rogers, *Chem. Commun.* 1023 (2000).
201. K. M. Barkigia, P. Battioni, V. Riou, D. Mansuy, and J. Fajer, *Chem. Commun.* 956 (2002).
202. R. A. Haycock, C. A. Hunter, D. A. James, U. Michelsen, and L. R. Sutton, *Org. Lett.* 2, 2435 (2000).
203. M. Gardner, A. J. Guerin, C. A. Hunter, U. Michelsen, and C. Rotger, *New J. Chem.* 309 (1999).
204. C. C. Mak, N. Bampos, S. L. Darling, M. Montalti, L. Prodi, and J. K. M. Sanders, *J. Org. Chem.* 66, 4476 (2001).
205. S. Richeter, C. Jeandon, R. Ruppert, and H. J. Callot, *Chem. Commun.* 91 (2001).
206. S. Richeter, C. Jeandon, J. Gisselbrecht, R. Ruppert, and H. J. Callot, *J. Am. Chem. Soc.* 124, 6168 (2002).
207. S. Richeter, C. Jeandon, R. Ruppert, and H. J. Callot, *Chem. Commun.* 266 (2002).
208. S. Richeter, C. Jeandon, C. Sauber, J.-P. Gisselbrecht, R. Ruppert, and H. J. Callot, *J. Porph. Phthal.* 6, 423 (2002).
209. K.-i. Sugiura, T. Matsumoto, S. Ohkouchi, Y. Naitoh, T. Kawai, Y. Takai, K. Ushiroda, and Y. Sakata, *Chem. Commun.* 1957 (1999).
210. L. Latterini, R. Blossey, J. Hofkens, P. Vanoppen, F. C. De Schryver, A. E. Rowan, and R. J. M. Nolte, *Langmuir* 15, 3582 (1999).
211. A. P. H. J. Schenning, D. H. W. Hubert, M. C. Feiters, and R. J. M. Nolte, *Langmuir* 12, 1572 (1996).
212. P. Foubert, P. Vanoppen, M. Martin, T. Gensch, J. Hofkens, A. Helser, A. Seeger, R. M. Taylor, A. E. Rowan, R. J. M. Nolte, and F. C. D. Schryver, *Nanotech*. 11, 16 (2000).
213. S. R. Collinson and D. W. Bruce, in "Transition Metals in Supramolecular Chemistry" (J.-P. Sauvage, Ed.), p. 285. John Wiley & Sons Ltd., New York, 1999.
214. B. Donnio and D. W. Bruce, in "Structure and Bonding," Vol. 95, p. 193. Springer Verlag, Berlin, 1999.
215. Y. Shimizu, J. Matsuno, M. Miya, and A. Nagata, *Chem. Commun.* 2411 (1994).
216. J.-P. Sauvage, *Acc. Chem. Res.* 31, 611 (1998).
217. M. Linke, J.-C. Chambron, V. Heitz, J.-P. Sauvage, and V. Semetey, *Chem. Commun.* 2469 (1998).
218. M. Linke, J.-C. Chambron, V. Heitz, J.-P. Sauvage, S. Encinas, F. Barigelletti, and L. Flamigni, *J. Am. Chem. Soc.* 122, 11834 (2000).
219. D. B. Amabilino, C. O. Dietrich-Buchecker, and J.-P. Sauvage, *J. Am. Chem. Soc.* 118, 3285 (1996).
220. N. Solladie, J.-C. Chambron, and J.-P. Sauvage, *J. Am. Chem. Soc.* 121, 3684 (1999).
221. M. Andersson, M. Linke, J.-C. Chambron, J. Davidsson, V. Heitz, J.-P. Sauvage, and L. Hammarstrom, *J. Am. Chem. Soc.* 122, 3526 (2000).
222. J.-C. Chambron, J.-P. Collin, J.-O. Dalbavie, C. O. Dietrich-Buchecker, V. Heitz, F. Odobel, N. Solladié, and J.-P. Sauvage, *Coord. Chem. Rev.* 178–180, 1299 (1998).
223. V. Balzani, M. Gomez-Lopez, and J. F. Stoddart, *Acc. Chem. Res.* 31, 405 (1998).
224. C. P. Collier, E. W. Wong, M. Belohradsky, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, and J. R. Heath, *Science* 285, 391 (1999).
225. M. J. Gunter, N. Bampos, K. D. Johnstone, and J. K. M. Sanders, *New J. Chem.* 25, 166 (2001).
226. K. Chichak, M. C. Walsh, and N. R. Branda, *Chem. Commun.* 847 (2000).
227. F. Hauke, A. Swartz, D. M. Guldi, and A. Hirsh, *J. Mater. Chem.* 12, 2088 (2002).
228. N. Kobko, L. Paraskevas, E. d. Rio, and J. J. Dannenberg, *J. Am. Chem. Soc.* 123, 4348 (2001).
229. C. B. Aakeroy and K. R. Seddon, *Chem. Soc. Rev.* 397 (1993).
230. E. A. Archer, H. Gong, and M. J. Krische, *Tetrahedron* 57, 1139 (2001).
231. G. R. Desiraju, *Acc. Chem. Res.* 35, 565 (2002).
232. L. J. Prins, D. N. Reinhoudt, and P. Timmerman, *Angew. Chem. Int. Ed.* 40, 2382 (2001).
233. T. Steiner, *Angew. Chem. Int. Ed.* 41, 48 (2002).
234. C. Schmuck and W. Wienand, *Angew. Chem. Int. Ed.* 40, 4363 (2001).
235. R. Fiammengo, M. Crego-Calama, and D. N. Reinhoudt, *Curr. Opinion Chem. Biol.* 5, 660 (2001).
236. J. M. J. Frechet, *Proc. Natl. Acad. Sci. USA* 99, 4782 (2002).
237. C. M. Drain, R. Fischer, E. Nolen, and J. M. Lehn, *Chem. Commun.* 243 (1993).
238. C. M. Drain, K. C. Russel, and J.-M. Lehn, *Chem. Commun.* 337 (1996).
239. X. Shi, K. M. Barkigia, J. Fajer, and C. M. Drain, *J. Org. Chem.* 66, 6513 (2001).
240. C. Ikeda, N. Nagahara, E. Motegi, N. Yoshioka, and H. Inoue, *Chem. Commun.* 1759 (1999).
241. Y. Kuroda, A. Kawashima, Y. Hayashi, and H. Ogoshi, *J. Am. Chem. Soc.* 119, 4929 (1997).

242. K. Sugou, K. Sasaki, K. Kitajima, T. Iwaki, and Y. Kuroda, *J. Am. Chem. Soc.* 124, 1182 (2002).
243. Y. Kuroda, K. Sugou, and K. Sasaki, *J. Am. Chem. Soc.* 122, 7833 (2000).
244. T. B. Norsten, K. Chichak, and N. R. Branda, *Chem. Commun.* 1794 (2001).
245. T. B. Norsten, K. Chichak, and N. R. Branda, *Tetrahedron* 58, 639 (2002).
246. M. J. Plater, S. Aiken, and G. Bourhill, *Tetrahedron Lett.* 42, 2225 (2001).
247. T. S. Balaban, A. Eichhofer, and J.-M. Lehn, *Eur. J. Org. Chem.* 24, 4047 (2000).
248. E. E. Simanek, L. Isaacs, X. Li, C. C. C. Wang, and G. M. Whitesides, *J. Org. Chem.* 62, 8994 (1997).
249. P. Bhyrappa, S. R. Wislon, and K. S. Suslick, *Supramol. Chem.* 9, 169 (1998).
250. P. Pasetto, X. Chen, C. M. Drain, and R. W. Franck, *Chem. Commun.* 81 (2001).
251. S.-I. Tamaru, M. Nakamura, M. Takeuchi, and S. Shinkai, *Org. Lett.* 3, 3631 (2001).
252. S.-I. Tamaru, S.-Y. Uchino, M. Takeuchi, M. Ikeda, T. Hatano, and S. Shinkai, *Tetrahedron Lett.* 43, 3751 (2002).
253. N. Nagata, S.-I. Kugimiya, and Y. Kobuke, *Chem. Commun.* 1389 (2000).
254. D. Paul, J. A. Wytko, M. Koepf, and J. Weiss, *Inorg. Chem.* 41, 3699 (2002).
255. R. P. Sijbesma, F. H. Beijer, L. Brunsveld, B. J. B. Folmer, J. J. K. K. Hirscbreg, R. F. M. Lange, J. K. L. M. Lowe, and E. W. Meijer, *Science* 278, 1601 (1997).
256. C. M. Drain and D. Mauzerall, *Biophys. J.* 63, 1544 (1992).
257. C. M. Drain, B. Christensen, and D. Mauzerall, *Proc. Natl. Acad. Sci. USA* 86, 6959 (1989).
258. C. M. Drain and D. Mauzerall, *Bioelectrochem. Bioenerg.* 24, 263 (1990).
259. C. M. Drain and D. C. Mauzerall, *Biophys. J.* 63, 1556 (1992).
260. N. C. Maiti, S. Mazumdar, and N. Periasamy, *J. Phys. Chem. B* 102, 1528 (1998).
261. N. Micali, A. Romeo, R. Lauceri, R. Purrello, F. Mallamace, and L. M. Scolaro, *J. Phys. Chem. B* 104, 9416 (2000).
262. N. Micali, F. Mallamace, A. Romeo, R. Purrello, and L. M. Scolaro, *J. Phys. Chem. B* 104, 5897 (2000).
263. A. V. Gusev and M. A. J. Rodgers, *J. Phys. Chem. A* 106, 1985 (2002).
264. R. Lauceri, A. Raudino, L. M. Scolaro, N. Micali, and R. Purrello, *J. Am. Chem. Soc.* 124, 894 (2002).
265. V. Kral, F. P. Schmidtchen, K. Lang, and M. Berger, *Org. Lett.* 4, 51 (2002).
266. J. M. Ribo, J. M. Bofill, J. Crusats, and R. Rubires, *Chem. Eur. J.* 7, 2733 (2001).
267. J. M. Ribo, J. Crusats, F. Sagues, J. Claret, and R. Rubires, *Science* 292, 2063 (2001).
268. R. F. Pasternack, E. J. Gibbs, P. J. Collings, J. C. dePaula, L. C. Turzo, and A. Terracina, *J. Am. Chem. Soc.* 120, 5873 (1998).
269. R. F. Pasternack, E. J. Gibbs, D. Bruzewicz, D. Stewart, and K. S. Engstrom, *J. Am. Chem. Soc.* 124, 3533 (2002).
270. H. Jiang, W. Su, J. Hazel, J. T. Grant, V. V. Tsukruk, T. M. Cooper, and T. J. Bunning, *Thin Solid Films* 372, 85 (2000).
271. R. F. Khairutdinov and N. Serpone, *J. Phys. Chem. B* 103, 761 (1999).
272. A. V. Udal'tsov, L. A. Kazarin, and A. A. Sweshnikov, *J. Mol. Struct.* 562, 227 (2001).
273. A. V. Udal'tsov, L. A. Kazarin, V. A. Sinani, and A. A. Sweshnikov, *J. Photochem. Photobiol. A: Chem.* 151, 105 (2002).
274. D. L. Akins, H.-R. Zhu, and C. Guo, *J. Phys. Chem.* 100, 5420 (1996).
275. M. A. Castriciano, A. Romeo, and L. M. Scolaro, *J. Porph. Phthal.* 6, 431 (2002).
276. L. C. Giancarlo and G. W. Flynn, *Annu. Rev. Phys. Chem.* 49, 297 (1998).
277. E. M. Bruti, M. Giannetto, G. Mori, and R. Seeber, *Electroanalysis* 11, 565 (1999).
278. Y. He, T. Ye, and E. Borguet, *J. Am. Chem. Soc.* 124, 11964 (2002).
279. Z. Zhang, N. Yoshida, T. Imae, Q. Xue, M. Bai, J. Jiang, and Z. Liu, *J. Coll. Inter. Sci.* 243, 382 (2001).
280. T. Ohshiro, T. Ito, P. Buhlmann, and Y. Umezawa, *Anal. Chem.* 73, 878 (2001).
281. X. Qiu, C. Wang, Q. Zeng, B. Xu, S. Yin, H. Wang, S. Xu, and C. Bai, *J. Am. Chem. Soc.* 122, 5550 (2000).
282. G.-y. Liu, S. Xu, and Y. Qian, *Acc. Chem. Res.* 33, 457 (2000).
283. A. Ishida and T. Majima, *Chem. Commun.* 1299 (1999).
284. A. Ishida and T. Majima, *Nanotech.* 10, 308 (1999).
285. A. Ishida and T. Majima, *Chem. Phys. Lett.* 322, 242 (2000).
286. N. Nishimura, M. Ooi, K. Shimazu, H. Fujii, and K. Uosaki, *J. Electroanal. Electrochem.* 473, 75 (1999).
287. H. Yamada, H. Imahori, Y. Nishimura, I. Yamazaki, and S. Fukuzumi, *Chem. Commun.* 1921 (2000).
288. H. Imahori, Y. Nishimura, H. Norieda, H. Karita, I. Yamazaki, Y. Sakata, and S. Fukuzumi, *Chem. Commun.* 661 (2000).
289. T. W. Hanks, B. Bergman, and P. Dillon, *Syn. Met.* 121, 1431 (2001).
290. C. Clausen, D. T. Gryko, R. B. Dabke, N. Dontha, D. F. Bocian, W. G. Kuhr, and J. S. Lindsey, *J. Org. Chem.* 65, 7363 (2000).
291. D. T. Gryko, C. Clausen, and J. S. Lindsey, *J. Org. Chem.* 64, 8635 (1999).
292. D. T. Gryko, F. Zhao, A. A. Yasseri, K. M. Roth, D. F. Bocian, W. G. Kuhr, and J. S. Lindsey, *J. Org. Chem.* 65, 7356 (2000).
293. J. Li, D. Gryko, R. B. Dabke, J. R. Diers, D. F. Bocian, W. G. Kuhr, and J. S. Lindsey, *J. Org. Chem.* 65, 7379 (2000).
294. K.-H. Schweikart, V. L. Malinovskii, J. R. Diers, A. A. Yasseri, D. F. Bocian, W. G. Kuhr, and J. S. Lindsey, *J. Mater. Chem.* 12, 808 (2002).
295. K. M. Roth, J. S. Lindsey, D. F. Bocian, and W. G. Kuhr, *Langmuir* 18, 4030 (2002).
296. A. Nomoto and Y. Kobuke, *Chem. Commun.* 1164 (2002).
297. Z. Zhang and T. Imae, *Stud. Surf. Sci. Catal.* 132, 585 (2001).
298. Z. Zhang and T. Imae, *Nano Lett.* 1, 241 (2000).
299. T. Imae, T. Niwa, and Z. Zhang, *J. Nanosci. Nanotech.* 2, 37 (2002).
300. B. Choudhury, A. C. Weedon, and J. R. Bolton, *Langmuir* 14, 6192 (1998).
301. M. Skupin, G. Li, W. Fudickar, J. Zimmermann, B. Roder, and J.-H. Fuhrhop, *J. Am. Chem. Soc.* 123, 3454 (2001).
302. W. Fudickar, J. Zimmermann, L. Ruhlmann, J. Schneider, B. Roder, U. Siggel, and J.-H. Fuhrhop, *J. Am. Chem. Soc.* 121, 9539 (1999).
303. G. Li and J.-H. Fuhrhop, *Langmuir* 18, 7740 (2002).
304. D. A. Offord, S. B. Sachs, M. S. Ennis, T. A. Eberspacher, J. H. Griffin, C. E. D. Chidsey, and J. P. Collman, *J. Am. Chem. Soc.* 120, 4478 (1998).
305. Z. Zhang, T. Imae, H. Sato, A. Watanabe, and Y. Ozaki, *Langmuir* 17, 4564 (2001).
306. M. S. Boeckl, A. L. Bramblett, K. D. Hauch, T. Sasaki, B. D. Ratner, and J. W. Rogers, Jr., *Langmuir* 16, 5644 (2000).
307. F. Armand, P.-A. Albouy, F. D. Cruz, M. Normand, V. Huc, and E. Goron, *Langmuir* 17, 3431 (2001).
308. F. Da Cruz, K. Driaf, C. Berthier, J.-M. Lameille, and F. Armand, *Thin Solid Films* 349, 155 (1999).
309. Z. Zhang, S. Hou, Z. Zhu, and Z. Liu, *Langmuir* 16, 537 (2000).
310. Z. Zhang, R. Hu, and Z. Liu, *Langmuir* 16, 1158 (2000).
311. G. Ashkenasy, G. Kalyuzhny, J. Libman, I. Rubinstein, and A. Shanzer, *Angew. Chem. Int. Ed.* 38, 1257 (1999).
312. G. Kalyuzhny, A. Vaskevich, G. Ashkenasy, A. Shanzer, and I. Rubinstein, *J. Phys. Chem. B.* 104, 8238 (2000).
313. M. A. Fox, *Acc. Chem. Res.* 32, 201 (1999).
314. C.-y. Liu, H.-l. Pan, M. A. Fox, and A. J. Bard, *Chem. Mater.* 9, 1422 (1997).

315. D. M. Adams, J. Kerimo, C.-Y. Liu, A. J. Bard, and P. F. Barbara, *J. Phys. Chem. B* 104, 6728 (2000).
316. B. R. Patel and K. S. Suslick, *J. Am. Chem. Soc.* 120, 11802 (1998).
317. M. Kimura, Y. Saito, K. Ohta, K. Hanabusa, H. Shirai, and N. Kobayashi, *J. Am. Chem. Soc.* 124, 5274 (2002).
318. H. D. Burrows, A. M. R. Gonsalves, M. L. P. Leitao, M. d. G. Miguel, and M. M. Pereira, *Supramol. Sci.* 4, 241 (1997).
319. T. Ogi, H. Ohkita, S. Ito, and M. Yamamoto, *Thin Solid Films* 415, 228 (2002).
320. J. Foekema, A. P. H. J. Schenning, D. M. Vriezema, F. B. G. Benneker, K. Norgaard, J. K. M. Kroon, T. Bjornholm, M. C. Feiters, A. E. Rowan, and R. J. M. Nolte, *J. Phys. Org. Chem.* 14, 501 (2001).
321. A. Chowdhury and A. J. Pal, *Thin Solid Films* 385, 266 (2001).
322. T. Komatsu, E. Tsuchida, C. Böttcher, D. Donner, C. Messerschmidt, U. Siggel, W. Stocker, J. P. Rabe, and J.-H. Fuhrhop, *J. Am. Chem. Soc.* 119, 11660 (1997).
323. T. Sagawa, S. Fukugawa, T. Yamada, and H. Ihara, *Langmuir* 18, 7223 (2002).
324. K. Araki, M. J. Wagner, and M. S. Wrighton, *Langmuir* 12, 5393 (1996).
325. D.-J. Qian, C. Nakamura, and J. Miyake, *Langmuir* 16, 9615 (2000).
326. D.-J. Qian, C. Nakamura, and J. Miyake, *Chem. Commun.* 2312 (2001).
327. M. Lahav, T. Gabriel, A. N. Shipway, and I. Willner, *J. Am. Chem. Soc.* 121, 258 (1999).
328. P. G. Van Patten, A. P. Shreve, and R. J. Donohoe, *J. Phys. Chem. B* 104, 5986 (2000).
329. J. R. C. d. Rocha, G. J.-F. Demets, M. Bertotti, K. Araki, and H. E. Toma, *J. Electroanal. Chem.* 526, 69 (2002).
330. D. M. Guldi, F. Pellarini, M. Prato, C. Granito, and L. Troisi, *Nano Lett.* 2, 965 (2002).
331. T. F. Magnera, L. M. Peslherbe, E. Körblová, and J. Michl, *J. Organomet. Chem.* 548, 83 (1997).
332. J. Michl and T. F. Magnera, *Proc. Natl. Acad. Sci. USA* 99, 4788 (2002).
333. X. Gong, T. Milic, C. Xu, J. D. Batteas, and C. M. Drain, *J. Am. Chem. Soc.* 124, 14290 (2002).
334. T. van der Boom, R. T. Hayes, Y. Zhao, P. J. Bushard, E. A. Weiss, and M. R. Wasielewski, *J. Am. Chem. Soc.* 124, 9582 (2002).
335. I. Sloufova-Srnova and B. Vlckova, *Nano Lett.* 2, 121 (2002).
336. F. T. Hong and D. Mauzerall, *Biochim. Biophys. Acta* 275, 479 (1972).
337. F. T. Hong and D. Mauzerall, *Proc. Natl. Acad. Sci. USA* 71, 1564 (1974).
338. K. Sun and D. Mauzerall, *Biophys. J.* 71, 295 (1996).
339. K. Sun and D. Mauzerall, *Proc. Natl. Acad. Sci. USA* 93, 10758 (1996).
340. K. Sun and D. Mauzerall, *Biophys. J.* 71, 309 (1996).
341. K. C. Hwang and D. Mauzerall, *Nature* 361, 138 (1993).
342. N. Nagata, S.-i. Kugimiya, and Y. Kobuke, *Chem. Commun.* 689 (2001).
343. J.-M. Lehn, *Chem. Eur. J.* 9, 2455 (1999).
344. P. N. W. Baxter, R. G. Khoury, J.-M. Lehn, G. Baum, and D. Fenske, *Chem. Eur. J.* 6, 4140 (2000).
345. C. M. Drain and X. Gong, *Chem. Commun.* 2117 (1997).

Self-Organization of Colloidal Nanoparticles

Joydeep Dutta

Asian Institute of Technology, Klong Luang, Pathumthani, Thailand

Heinrich Hofmann

Swiss Federal Institute of Technology, Lausanne, Switzerland

CONTENTS

1. Introduction
2. Self-Assembly and Self-Organization
3. Self-Organization of Nanoparticles Into Superlattice
4. Conclusions
Glossary
References

1. INTRODUCTION

The emerging fields of nanoscience and nanoengineering are leading to unprecedented understanding and control over the fundamental building blocks of all physical matter. Today, nanophase engineering expands in a rapidly growing number of structural and functional materials, both inorganic and organic, allowing the manipulation of mechanical, catalytic, electrical, magnetic, optical, and electronic functions. The synthesis of nanophase or cluster-assembled materials is usually based upon the creation of separated small clusters which are then fused into a bulk-like material or as a thin film or even by embedding into a solid matrix, materials like polymers or glass either in a disordered manner or in an ordered array.

There is a need to learn how to manipulate nanosized particles and form the ensemble into periodically ordered particulate materials for application into devices making use of the special physical properties that arise due to the nanometric crystallite sizes, like the quantum effects, for example. From recent developments in the cluster source technologies (thermal, laser vaporization, aerosol techniques, chemical vapor deposition, and sputtering) [1–4], it is now possible to produce intense cluster beams of any materials, even the most refractory or complex systems, for a wide range of sizes from a few atoms to a few thousand atoms [5].

Significant progress in the organization of semiconductor quantum dots (QD) has been made by employing a variety of microelectronic fabrication techniques [6–10], but the most widely used strategy employs the deposition of a narrower bandgap, compressively strained epilayer over a wider bandgap substrate, which results in the spontaneous nucleation (self-assembly) of coherently strained three-dimensional islands on top of a thin ($\sim$1 nm), two-dimensional "wetting layer." Subsequent deposition of substrate material to encapsulate the QD produces a fully three-dimensional confinement potential. This approach however, is, outside the scope of the present review where we will concentrate on the fabrication of ordered layers or structures from colloidal particles.

An elegant method to organize nanometric particles is by the so-called "self-assembly methods often inspired by nature (biomimetic processes). When self-assembly is used to organize nanomaterial into a useful form, irrespective of whether the structures so formed are points, lines, layers, or composites, they are going to be periodic in arrangements, since such arrangements minimize the free energy of the systems. Methods developed to produce three-dimensional, bulk-like hierarchical structures include biomimetic methods, which use polypeptides as building blocks, and amphiphile and colloidal templating, which use amphiphilic or colloidal mesophases as templates for inorganic mesoporous materials. Designing finite mesostructures with a given geometry is the future challenge to make use of the special properties arising due to the scaling of the particle sizes for applications. Ordered assemblies of nanometer-sized particles are thus an interesting class of nanomaterials that provide exceptional potentials to achieve one-two-and three-dimensional structures for a wide variety of applications ranging from photonic devices to memory devices, two-dimensional arrays of magnetic nanoparticles [11–13], or single-electron microelectronic devices, amongst others [14–16]. In order to make full use of this potential

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

for materials engineering, it is important to identify and gain control over the relevant growth and ordering parameters both during the formation of ordered structures as well as after the growth process, so that the structures so formed do not get distorted.

This is a field in its nascent stage and experimentalists are still struggling to build up on the ideas propagated by the theorists to develop the complex technology required to give two- or three-dimensional structures to a material on the nano or micron scale (in this case the primary materials are of nanoscopic length scales). More and more interesting structures, prepared by a variety of methods twinned around "self-assembly" methods, are being reported regularly in the literature. Self-assembly and self-organization are concepts that are broadly used over a whole range of fields including physics, chemistry, biology, and engineering. Nanoscale self-assembly is a concept that nature has been making use of since the beginning of life, and it is only recently that we have started realizing the potential of utilizing this phenomenon for the artificial control of matter [17].

Letting aside the self-assembly or self-organization processes, even nanotechnology did not emerge as an experimental science before the 1980s due to the unavailability of suitable tools to allow researchers to observe—let alone manipulate—individual atoms. It evolved dramatically with the invention of two new microscopy techniques: atomic force microscopy (AFM) and scanning tunneling microscopy (STM). Both these techniques were a radical departure from previous types of microscopy, which worked by reflecting either light (in the case of optical microscopes) or an electron beam (in the case of electron microscopes) off a surface and onto a lens. But no reflective microscope, not even the most powerful, could image an individual atom. The discovery of the STM got scientists and engineers to look into the self-organization processes of nanoscopic matter more closely, as STM brought in the possibility of visualizing at the nanoscopic levels and also the possibility of moving atoms to structure them into useful forms. Direct writing of nanoscopic features with atoms is a very slow process that is normally achieved, for example, by utilizing tools like the scanning probe microscope (SPM). Thus, a more practical method is albeit—the use of natural forces to structure nanoscopic features. With the discovery of other microscopic tools like AFM, and the improvement of the electron microscopy tools further developed the possibilities of visualizing and implementing the experiments designed by researchers which has given boosted efforts to explore the immense possibilities this field has in reserve.

Two approaches for the building up of nanoscopic features have been envisaged: the so-called "top-down" method and "bottom-up" methods. In the "top down" methods, the features are written directly onto a substrate, for example, by electron beams, and then by applying appropriate etching and deposition processes, the nanoscopic features are engraved. In the "bottom-up" approach, nanocomponents are made from precursors in the liquid, solid, or gas phase employing either chemical or physical deposition processes that are integrated into building blocks within the final material structure. The "bottom-up" approach, or simply the assembly from atoms, include wide-ranging methods encompassing supramolecular chemistry to biomimetic processes to template synthesis, amongst others. The aim of this article is to consider the "bottom-up" approach for building up nanostructures focusing on some of the self-organization methods that are used in structuring colloidal nanoparticles into regular films, arrays, or three-dimensional structures. This article will attempt to distinguish between the self-assembly processes (that is often misconceived in some experiments reported in the literature) from the self-organization of colloidal nanoparticles.

2. SELF-ASSEMBLY AND SELF-ORGANIZATION

Organization during self-assembly processes is driven mainly by competing molecular interactions like hydrophobic versus hydrophilic components, gravitational, van der Waals, or Coulombic interactions. Self-assembly is an assembly process in which only the constituents of the final structure take part, that is, get incorporated into the resulting structure [18]. Self-organization, in retrospect, is defined as a mechanism for building patterns, processes, and structures at a higher level through multiple interactions among the components at the lower level(s), where the components interact through local rules that do not explicitly code for the pattern [19–20]. In colloidal materials, self-assembly process has often been related to describe the growth of nanoparticles, for example, of controlling the size of the particles by self-organization of polymers on these growing particles and has to be clearly differentiated from the self-organization processes of the colloidal particles into either two- or three-dimensional structures.

The phenomena of pattern formation and self-organization occur as natural responses of complex systems to strong external stimulation. The collective interaction between system components under external driving forces, which drive the system far from equilibrium, results in the self-organization of its constituents [21–22]. Some of the salient features of this viewpoint can be found in the behavior of chemical [23–24], liquid crystal nematics [25–26], and fluid systems [27]. It has also been recognized that condensed matter systems show a rich variety of patterns and self-assembled microstructures under conditions as diverse as solidification [28], electro-chemical deposition [29], plastic deformation [30–31], surface modifications [32], rapid laser heating [33], irradiation by energetic particles [34], magnetic domains [35], and more recently, atom deposition into self-assembled monolayers [36]. All these phenomena are currently being revisited, as self-assembly is nothing new; for example, lustre decoration of medieval and renaissance pottery consists of self-assembled silver and copper nanocrystals within the glassy phase of the ceramic glaze. A peculiar property of this lustre was enumerated to arise due to the high density of pseudo-spherical nanocrystals and their locally homogeneous distribution in the ceramic [37–39]. Another historic example of self-assembly (or even self-organization) of nanoparticles is found in the Maya blue paints [40]. Studies of authentic samples from Jaina Island show that the material is made of needle-shaped palygorskite (a clay) crystals that form a superlattice with a period of 1.4 nm, with intercalates of amorphous silicate

substrate containing inclusions of metal (Mg) nanoparticles. The beautiful tone of the blue color is obtained only when both these nanoparticles and the superlattice are present together, as has been shown by the fabrication of synthetic samples.

2.1. The Self-Assembly Process

As explained earlier, self-assembly in a general sense, might be defined as the spontaneous formation of complex hierarchical structures from predesigned building blocks, typically involving multiple energy scales and multiple degrees of freedom. It is also a very general principle in nature, as seen in the formation for example, of membranes from lipid molecules or the living cell as probably the most important paradigm [41]. Self-assembly has been defined by Kuhn and Uhlman as a 'process in which supramolecular hierarchical organization is established in a complex system of interlocking components' [42]. Self-assembled monolayers (SAMs) have been studied in great details by surface scientists by using different types of molecules on a host of different substrates. Detailed description and discussion on the SAMs are outside the scope of this article, but a comprehensive report of SAMs as described by surface scientists have been recently published [43]. As an example, the self-assembly can be achieved through an alkane chain typically with 10–20 methylene units that can be given a head group with a strong preferential adsorption to a chosen substrate. For example, thiol (S-H) groups adsorb readily from a solution onto a gold surface creating a dense monolayer with the tail group pointing outwards from the surface. Numerous work, both calculations [44] and experiment [45], has been reported describing the self-assembly of thiol molecules to control gold nanoparticle sizes during synthesis [46] and/or during the assembly of these particles on a substrate including a milestone review recently published by Brust and Kiely [47].

2.2. Surface Modification of Nanoparticles

Particles in the nanometer-size range have a strong tendency to agglomerate due to van der Waals interactions. It is therefore important to develop synthetic methods by which the particles can be stabilized, that is, where repulsive forces between the particles can be provided to balance this attraction (Fig. 1(A)) [48]. Generally two types of stabilization are used to prevent agglomeration of nanoparticles, namely, electrostatic stabilization and steric stabilization by adsorbed molecules or steric hindrance [49–50]. Electrostatic stabilization involves the creation of an electrical double layer arising from ions adsorbed on the surface and associated counterions that surround the particle. Thus, if the electric potential associated with the double layer is sufficiently high, the Coulombic repulsion between the particles will prevent their agglomeration (Fig. 1(B) and Fig. 2) [51–52].

Steric stabilization can be achieved by the adsorption of large molecules such as polymers at the surface of the particles (Fig. 3). Indeed, the coil dimensions of polymers are usually larger than the range over which the attraction forces between colloidal particles are active. Two distinct effects describe this type of stabilization (the volume restriction contribution and the osmotic diffusion), and they both contribute to the interaction free energy [46, 53–54]. First, the fact that the adsorbed molecules are restricted in motion causes a decrease in the configurational entropic contribution to the free energy (Fig. 3). Second, the local increase in concentration of polymer chains between approaching particles results in an osmotic repulsion, since the solvent reestablishes equilibrium by diluting the polymer molecules and separating the particles. This can be described by the energy of free mixing of polymer segments and solvent molecules, calculated by the Flory–Krigbaum theory [55].

The conformation of adsorbed polymers tends to be controlled by the strength of segment/surface interactions [56], which may be described by the classical loop-train-tail model [57–58]. For an effective particle stabilization, it is important that the polymer form a complete, dense layer around the particle [59]. Then, the polymeric stabilizer must have sufficient tail length and adsorb uniformly enough to screen the attractive interaction between the particles.

Another requirement for good particle stabilization is the use of appropriate solvents. The stabilizing polymer has to possess high affinity with the solvent in order to solvate the particles and form an extended layer for screening the van der Waals attraction between the particles. It is well known, however, that adsorption is generally stronger when the affinity of the polymer to the solvent is low [60–61]. Therefore, block or graft copolymer micelles, in the presence of a selective solvent that solubilizes one of the blocks,

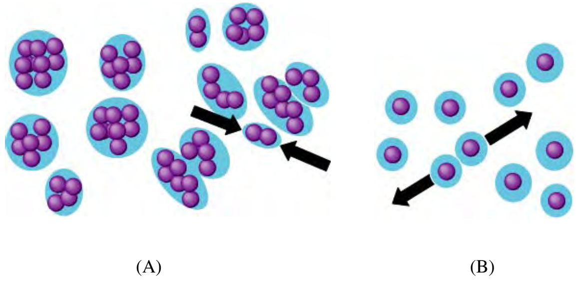


Figure 1. Particles in a colloid (A) uncharged particles are free to collide and agglomerate and (B) charged particles repel each other.

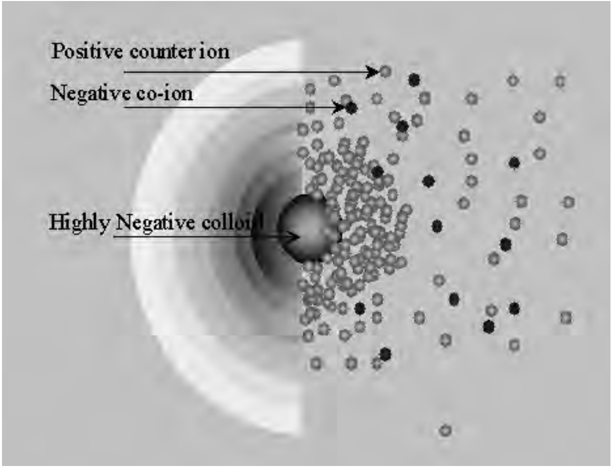


Figure 2. Schematic representation of the stabilization of a nanoparticle by charging the surface with counter ions.

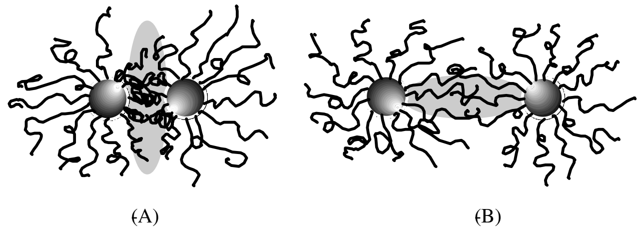


Figure 3. Steric stabilization of metal colloid particles by polymers: (A) entropic effect, (B) osmotic effect.

lead to expanded chains which occupy more surface area. In the case of polymers with strong anchoring groups, spontaneous adsorption occurs from good solvent [46]. De Gennes has discussed the density profile of polymers attached to the surface by one end and in contact with an appropriate solvent [62]. Another important advantage of covalently attached polymers is that the resulting interaction with the particle surface is stronger and desorption cannot occur by solvent effects or variations in temperature [63–64].

As previously mentioned, alkanethiols spontaneously self-assemble onto colloidal gold particles and this capping of gold particles by alkanethiols are sometimes termed three-dimensional self-assembly, which is often differentiated from the so-called two-dimensional self-assembly on gold thin films [65–66] or alkylsilanes on silica films. To differentiate between the arrangement of nanoparticles into ordered structures and the passivation of nanoparticles by polymeric entities [67], it is our belief the stabilization of nanoparticle growth or agglomeration by polymers is a *self-assembly process*. For example, the self-assembly on gold particles occurs through a strong thiolate bond between the surfactant and the gold surface which leads to the assembly of these surfactant molecules on the surface of gold nanoparticles [68]. There are numerous reports on the self-assembly of alkanethiols, aromatic thiols, and alkylamines onto gold and silver colloidal particles, which has been reviewed in great detail [47, 69–70]. Figure 4 illustrates an example of surface derivatization of gold and silver particles with three-dimensional SAMs of ω-functionalized thiol molecules.

Analogous to the attachment of organic stabilizers, biological components have also been used for stabilization of these colloids. Typically, wet-chemical preparation of the nanoparticles is carried out in the presence of stabilizing agents (often citrates, phosphanes, or thiols) which bind to the atoms exposed at the surface of the nanoparticles. This capping leads to a stabilization and prevents uncontrolled growth and aggregation of the nanoparticles. In the case of

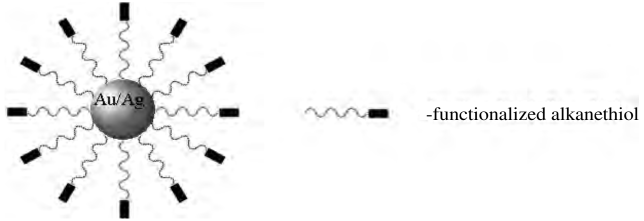


Figure 4. Schematic representation of the surface functionalization of a gold or silver nanoparticle by an alkanethiol.

a labile capping layer such as citrate, biomolecules can be linked directly with a metal particle by exchange reactions with stronger binding ligands. This method is usually applied in the coating of colloidal gold with thiol-containing proteins, for example, immunoglobulins (IgG) and serum albumins, which have cysteine residues that are accessible for the heterogeneous interphase coupling. If no such residues are available in the native proteins, thiol groups can be incorporated by chemical means [71] or by genetic engineering. For example, DNA molecules can be synthesized with alkylthiol groups, and n-alkylthiolated DNA has been used extensively in the preparation of DNA-functionalized gold and semiconductor nanoparticles [72–77]. As an interesting alternative, DNA oligonucleotides, which contain several adenosyl phosphothioate residues at their ends, have also been used to directly interact with the metal surface of nanoparticles [78]. The use of cyclic disulfide linkers [77–78] leads to nanoparticle capping, which are more stable towards ligand exchange than the corresponding conjugates prepared with the conventional reagents that contain a single thiol group or acyclic disulfide units. The greater stability is likely a result of the anchoring of the ligands to the nanoparticles through the two sulfur atoms [77].

Surface functionalized metal nanoparticles have been used for a variety of applications. Terms like "nanofactory" has been often cited, which basically study the reactivity's of functional groups anchored to gold colloidal particles [47, 79]. Another approach to use these self-assembled metal clusters has arisen due to the local field enhancement and nanoscale resonant behavior of metal clusters excited by an electromagnetic radiation. In metal particles, absorption in the UV-visible spectra (with a peak at 520 nm for 10 nm gold particles, for instance) is due to the excitation of a plasmon resonance which corresponds to a collective oscillation frequency of conduction electrons [80]. These absorption spectra can be theoretically explained by the Mie theory [81–82], in which the theoretical absorption spectrum of dilute spherical particles ("surface plasmon resonance") is related to their size, and their frequency-dependant dielectric constant, compared to the surrounding medium [83–84]. Immunogold labeling and immunogold chromatography as well as surface plasmon resonance transduction of metal cluster binding, which has been successfully commercialized using these surface stabilized nanoparticles, is outside the scope of this review [85]. Numerous other applications that include the deposition of these self-assembled nanoparticles have been reported, for example, local plasmon sensor [86], colorimetric DNA analysis [87], optically anisotropic windows [88–89], surface enhanced raman spectroscopy (SERS) substrates [90], amongst others. Several authoritative reviews are available dealing with the self-assembly and self-organization of gold nanoparticles [47, 91]. Self-assembled monolayer formation provides an easy route towards surface functionalization by organic molecules containing suitable functional groups like -SH, -CN, -COOH, -NH_2, and silanes on selected metallic (Au, Cu, Ag, Pd, Pt, Hg, and C) [92], as well as semiconducting surfaces (Si, GaAs, ZnS, CdS, etc.) [93–98]. Self-assembly inside a nanoscale cavity has also been reported where, for example, the reaction between divalent cadmium or zinc with sulfide or selenide is performed in a restricted environment in

the nanometer scale. Examples of nanoscale cavities include porous glasses and xerogels [99], reverse micelles [100], zeolites [101], membranes [102–103], and hollow proteins, amongst others [104]. Yeast and tomato virus, and likely other organisms, produce CdS quantum dots as a detoxification response to an overload of cadmium [105–106]; the CdS thus produced is coated with a particular peptide, which was used as the stabilizing thiol to make CdS quantum dots [107].

3. SELF-ORGANIZATION OF NANOPARTICLES INTO SUPERLATTICE

The general interest in the self-organization of spherical and non-spherical nanoparticles, of late, has arisen due to the possible applications in a wide range of domains [91, 108–110]. Periodic arrays are structures where nanoparticles are arranged in a regular fashion in either two or three dimensions [111]. These structures would be useful for displays, sensors, data storage, or photonic bandgap materials [91]. Monosized spheres of silica and polystyrene have most often been used for studying the self-organization processes, due to their easy availability. On the other side, QDs of semiconducting and metallic materials, due to their potential and already realized applications, as well as fundamental chemical and physical interest, have been studied in great details with some activities on the self-organization phenomena reported [112].

Nonlithographic bottom-up approaches based on self-assembly and self-organization are especially appealing because of their intrinsically low overhead for large-scale production. This approach has been useful in the self-organization of monolayer-protected metal nanoparticles into periodic two-dimensional (2D) arrays, with many of these assemblies demonstrating novel optical or electronic properties as a function of particle size or interparticle spacing [68, 113–116]. Numerous examples of 2D arrays comprised of small ($<$10 nm) gold nanoparticles have been reported, but details on well-ordered 2D arrays of larger gold nanoparticles are not available in the literature. Particles beyond a certain size (about 15 nm) tend to agglomerate into multilayer or three-dimensional aggregates rather than form 2D monoparticulate films [117]. This can be attributed to the large Hamaker constant for gold [118–119] and the rapid increase in van der Waals attraction between particles as a function of size [120], as well as the loss of surfactant chain mobility on the planar facets of the nanoparticles.

Polystyrene (PS) nanoparticles are commercially available and have been used for a wide range of applications ranging from applications in fields of photonics [121–125], nanotechnology [126–132], and life science [133–140] to fundamental studies in the behavior of colloidal suspensions, like the role of hydrodynamic [140–141] and entropic [142] forces and studies on phase transition [143] and crystallization [112] of "model atoms" in two or three dimensions. Irrespective of this plurality of applications, at the focus of these activities is the adsorption of the particles as a 2D layer onto a surface, where, depending on the specific aims of the subject, close-packed ordered arrays or irregular, less densely packed layers are selectively used. The schemes proposed so far for patterning PS particles range from purely physical concepts, such as micro-molding in capillaries (MIMIC) [144] and electrophoresis(to be discussed in a later section) to those combining physical interaction mechanisms like electrostatic or capillary forces with the selectivity of chemically patterned surfaces [145–146]. The self-organization techniques that have been used to build 2D or three-dimensional (3D)-ordered structures are schematically represented in Figure 5 and are discussed in greater details in the next few sections.

3.1. Self-Organization Through Capillary Forces

A comprehensive review of various types of capillary forces on particles has been reported by Kralchevsky and Denkov, which is summarized in Figure 6 [147]. If a droplet of the colloid suspension is dried slowly on an unpatterned polar surface (such as silica), the particles aggregate at the rim of the droplet because of attractive capillary forces between the particles when the water film thickness is in the dimension of the particle diameter, which is schematically shown in Figure 6 [147–148]. The mobility of the electrostatically attached particles is lost after complete drying, suggesting a three-step mechanism for the particle adsorption. First, positioning and adhesion in the suspension liquid are controlled by charge and polar interactions between the substrate and particle surfaces. In the second step, capillary forces between the particles and the surface laterally displace the particles during drying. After complete evaporation in the third stage, an irreversible reorganization of the particle–substrate interface occurs, which prevents resuspension or displacement when resubmerging the assembly into the adsorption liquid. A submonolayer of electrostatically adsorbed particles forms on the charged pattern surface in the bulk phase of the suspension. The submonolayer coverage is caused by Coulomb repulsion of the like-charged particles, which prevents dense packing in the aqueous medium. During evaporation of the suspension, the liquid front moves very slowly over the adsorbed particle layer, and a thin liquid meniscus remains over the particle assembly, which generates attractive capillary forces between the particles, as shown in Figure 6(c).

Because of the slow-moving, droplet front, particles from the suspension do have time to migrate into the thin liquid film at the drying front and aggregate into multilayer over the colloid assembly pattern driven by capillary forces (Fig. 7). Many more factors influencing the particle–surface interactions, such as surface topography and corrugation [149–151] or even nanoscopic air bubbles [152] might have to be taken into account in order to completely understand the ordering phenomena of colloidal nanoparticles through

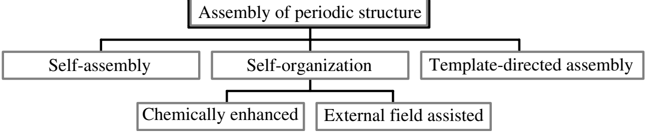


Figure 5. Diagram showing various self-organization techniques reported in the literature.

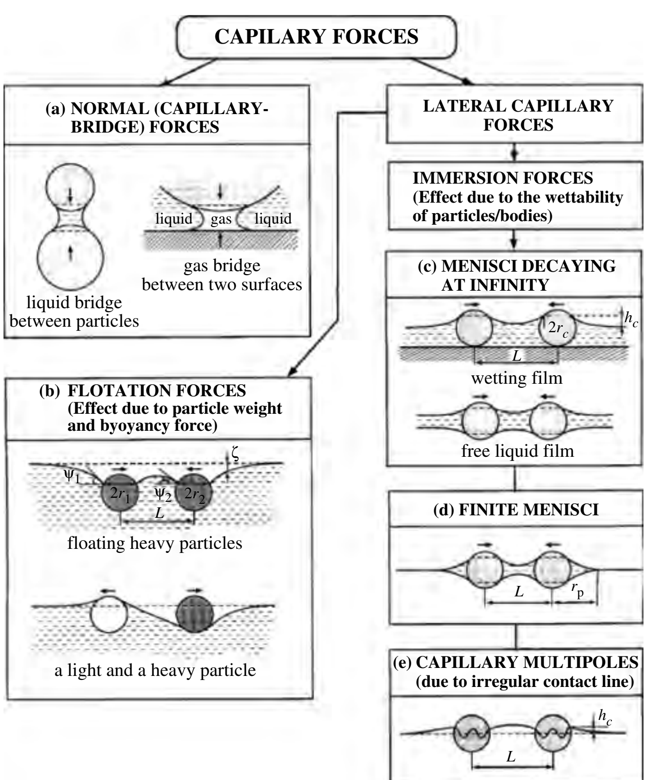


Figure 6. Types of capillary forces: (a) The *normal* capillary forces can be due to either liquid-in-gas or gas-in-liquid capillary bridges, which lead to particle-particle and particle-wall interactions, the force is directed normally to the contact line. In the case of *lateral* capillary forces (b, c, d, e), the force is parallel to the contact line. The interaction is due to the overlap of interfacial deformations created by the separate particles. (b) In the case of *flotation* force, the deformations are caused by the particle weight and buoyancy. In the case of *immersion* forces (c, d, e) the deformations are related to the wetting properties of the particle surface: position and shape of the contact line, and magnitude of the contact angle. When the deformation around an isolated particle is axisymmetric, we deal with "*capillary charges*," one can distinguish cases of infinite (c) and finite (d) menisci. (e) The forces between particles of undulated or irregular contact line can be described as interactions between "*capillary multipoles*," in analogy with electrostatics. Reprinted with permission from [147], P. A. Kralchevsky and N. D. Denkov, *Curr. Opin. Coll. Inter. Sci.*, 6, 383 (2001). © 2001, Elsevier Science.

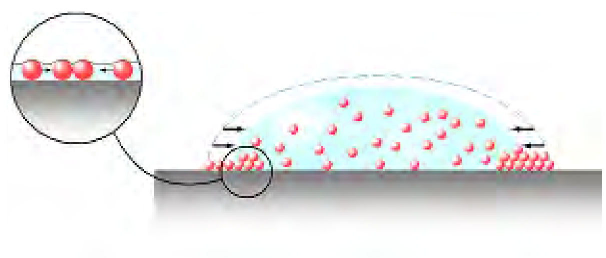


Figure 7. Schematic representation of a droplet of a colloid suspension dried slowly on an unpatterned polar surface (such as silica), the particles aggregate at the rim of the droplet because of attractive capillary forces between the particles when the water film thickness is in the dimension of the particle diameter.

capillary forces. Another important factor that should be considered is that during the drying process, the ionic strength of the colloid changes affecting the double-layer thickness and hence the final organization process.

3.2. Template-Assisted

In this section, we review only the approaches that combine templating with attractive capillary forces to assemble particles into complex aggregates with well-controlled sizes, shapes, and internal structures [153–154]. Surface confinement provided by liquid droplets has been used to assemble colloidal particles or microfabricated building blocks into spherical objects [155–157]. In this process, patterned arrays of templates are fabricated in thin films of photoresists (for example, spin-coated on glass, silicon wafer, or quartz substrates), using conventional photolithography techniques used in microelectronics or printing processes. Control over the number of particles in each such template is governed by the size of the particles and the size and shape of the templates [153]. Patterned arrays of relieves on solid substrates have also been exploited to grow colloidal crystals with unusual crystalline order and orientations [149, 158–159].

Templating against opaline arrays of colloidal spheres has also been successfully applied to the fabrication of 3D macroporous structures from a wide variety of materials, including organic polymers, ceramic materials, inorganic semiconductors, and metals (for example, as in Fig. 8) [161–176]. Fabrication based on this approach is remarkable for its simplicity, and for its fidelity in transferring the structure from the template to the replica. The size of the pores and the periodicity of the porous structures can be precisely controlled and readily tuned by changing the size of the colloidal spheres. There is no doubt that a similar approach is extendible to other materials. The only requirement seems to be the availability of a precursor that can infiltrate into the void spaces among colloidal spheres without significantly swelling or dissolving the template (usually made of

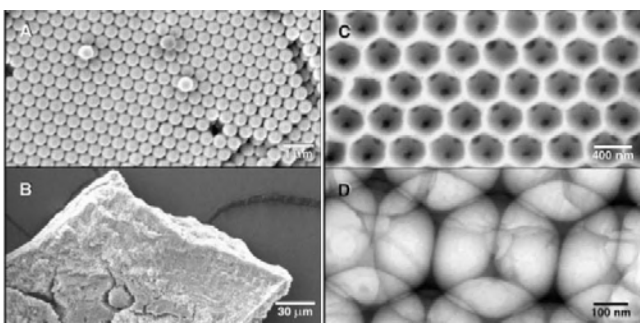


Figure 8. (A) Scanning electron micrograph (SEM) of a PS colloidal crystal. (B) Low magnification SEM of a single particle of 3DOM titania. (C) High magnification SEM of 3DOM silica showing close-packed macropores, which are interconnected through smaller windows. The white regions are walls of the first layer, the gray regions are walls of the second layer, and the black regions are voids. (D) Transmission electron micrograph (TEM) of 3DOM silica, which has amorphous walls. The dark regions are the silica walls. Reprinted with permission from [160], A. Stein and R. C. Schroeden, *Curr. Opin. Solid State Mater. Sci.* 5, 553 (2001). © 2001, Elsevier Science.

polystyrene beads or silica spheres). Some gaseous precursors have also been employed in this process, albeit the initial product deposited on the surfaces of samples might block the flow in of gaseous precursors [175]. Presently, the smallest colloidal spheres that have been successfully used in this method are ~35 nm in diameter [167]; the lower limit to the particle size that can be incorporated into this technique has not been completely established.

In another article, fabrication of germanium inverse opals by a combination of wet-chemical and gas-phase synthesis has been reported by Meseguer et al. [177]. Ge is transparent to infrared radiation (below 0.67 eV at room temperature) and it presents an extremely high refractive index contrast ($n = 4$). This makes Ge inverse opals one of the most promising PBG material made out of a semiconductor. It was reported that the photonic bandgap material formation by the classical infiltration method, using hydrolysis followed by a reduction and the use of chemical vapor deposition (CVD) technique, increased the homogeneity of the germanium layer. In the hydrolysis route, a germanium precursor ($Ge(OCH_3)_4$) was completely infiltrated in the opal voids (in a template made with 870 nm silica particles), followed by a hydrolysis process to obtain GeO_2. Figure 9(A) shows opal voids impregnated with small (~100 nm) GeO_2 microcrystallites. Ge was obtained by a reduction process at 550 °C. To increase the infiltration percentage, the opals are subjected to multiple cycles of hydrolysis followed by a reduction process. The Ge infiltrated opals were then chemically etched to remove silica particles with the same method used for the other inverse opals. In this way, we seek to remove the SiO_2 spheres from the composite and obtain Ge inverse opals. Figure 9(C) shows SEM images of the cleft edge of a Ge inverse lattice. In a further improvement of this classical method, the authors also reported the use of CVD for the infiltration of silica templates made from slightly larger spheres (1200 nm). The germanium precursor, Ge_2H_6 gas, was at first infiltrated into the template placed in a high vacuum cell (5×10^{-6} torr), and then, the sample cell was cooled by using liquid nitrogen, which gave rise to the condensation of the precursor into the opal void volume. Thereafter, upon heating the reactor up to the decomposition temperature of Ge_2H_6, microcrystalline germanium were formed in the pores, as shown in Figure 9. By this method, the authors claim that the Ge filling fraction can be controlled up to 100% of the template pore volume. The infiltrated samples were later chemically etched to obtain a germanium inverse opal with long-range FCC order of the original template.

Recently, Yang et al. demonstrated a simple and convenient method for fabricating hierarchically porous materials by combining colloidal arrays and block copolymers into a single system [176]. In summary, dewetting of colloidal dispersions from contoured surfaces, attractive capillary forces, and geometric templating can be combined to provide an attractive method of assembling colloidal particles into complex aggregates. The success of this organization approach depends on the control of a variety of parameters like the meniscus of the curvature of the rear edge of the liquid slug, concentration, and surface charges on the colloidal particles and the templates. However, though it has not been conclusively proved yet, this method may be limited to a few tens of nanometer regions due to the interaction of the Brownian motion and the capillary forces. One such example of the use of physical templating and capillary forces to assemble monodispersed spherical colloids into uniform aggregates with well-controlled sizes, shapes, and structures is shown in Figure 10 (and also Table 1). When an aqueous dispersion of colloidal particles is allowed to dewet from a patterned solid surface with appropriate relief structures, the particles get trapped and assemble into aggregates whose structures are determined by the geometric confinement provided by the templates. By this method, the capability and feasibility of assembling polystyrene beads and silica colloids (~150 nm in diameter) into complex aggregates which include polygonal or polyhedral clusters, linear or zigzag chains, and circular rings, have been demonstrated and a couple of examples are shown in Figures 11 and 12.

Immobilization of metal colloid particles on a suitable structured matrix using a self-assembly technique is one of the promising strategies for the construction of

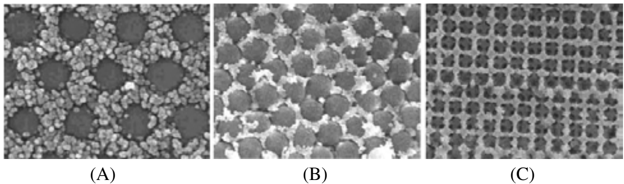


Figure 9. SEM images of an opal at the different stages of Ge infiltration. (A) shows a GeO_2-silica composite after hydrolysis reaction. (B) is the Ge-silica that results from the reduction process. After the etching process, one obtains the inverse lattice (C). The opal template is formed from 870 nm silica particles and then sintered at 950 °C for 3 h. Reprinted with permission from [177], F. Meseguer et al., *Coll. Surf. A.* © 2002, Elsevier Science.

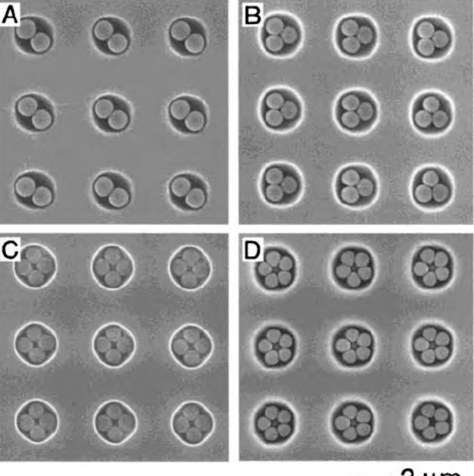


Figure 10. SEM images of polygonal aggregates formed by templating PS beads against 2D arrays of cylindrical holes (~2 μm diameter). (A) 2D array of dimers formed from 1.0 μm PS beads; (B) 2D array of trimers formed from 0.9 μm PS beads; (C) 2D array of square tetramers formed from 0.8 μm PS beads; and (D) 2D array of pentagons formed from 0.7 μm PS beads. Reprinted with permission from [178], Y. Yin et al. *J. Amer. Chem. Soc.* 123, 8718 (2001). © 2001, American Chemical Society.

Table 1. Control of the Self-Assembled Particles By Using Cylindrical Templates of Different Structure of Clusters.

Geometric shape of the template	Structures of the cluster	D/d
D	d	1.00-2.00
		2.00-2.15
		2.15-2.41
		2.41-2.70
		2.70-3.00
		3.00-3.30

Source: Reprinted with permission from [178], Y. Yin et al. *J. Amer. Chem. Soc.* 123, 8718 (2001). © 2001, American Chemical Society.

colloidal nanostructures [92, 179]. Assembling a homogeneous dispersion in a well-defined 3D nanostructure offers an alternative route to these stepwise 3D structure formations [180–181]. Zeolites and porous membranes (like

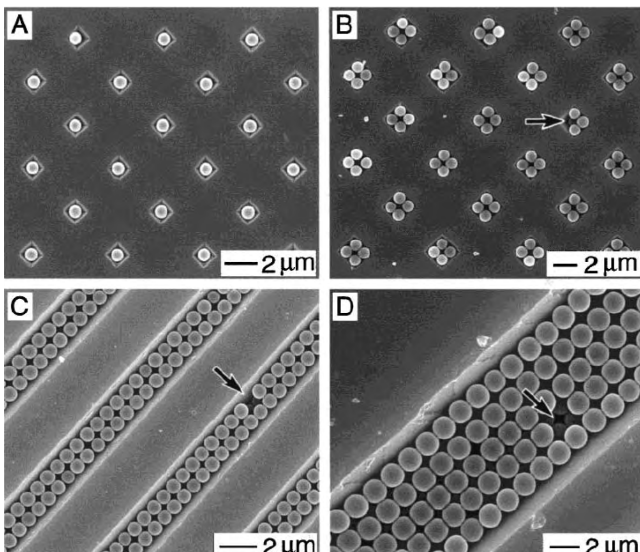


Figure 11. SEM images of 2D arrays of colloidal aggregates which were assembled under the confinement of templates etched in the surfaces of Si (100) substrates; (A) 800-nm PS beads in square pyramidal cavities 1.2 μm wide at the base; (B) 1.0 μm silica colloids in square pyramidal cavities 2.2 μm wide at the base; (C) 0.8 μm PS beads in V-shaped grooves 2.5 μm wide at the top; and (D) 1.6 μm PS beads in V-shaped grooves 10 μm wide at the top. Note that the use of V-shaped grooves as the templates also allowed one to control the orientation of the colloidal crystals. In (C) and (D), the face-center-cubic structures have a (100) orientation rather than (111), the one that is most commonly observed when spherical colloids are crystallized into 3D lattices. The arrows indicate defects, where one can also see the colloidal beads underneath the first layer of the structure. Reprinted with permission from [178], Y. Yin et al., *J. Amer. Chem. Soc.* 123, 8718 (2001). © 2001, American Chemical Society.

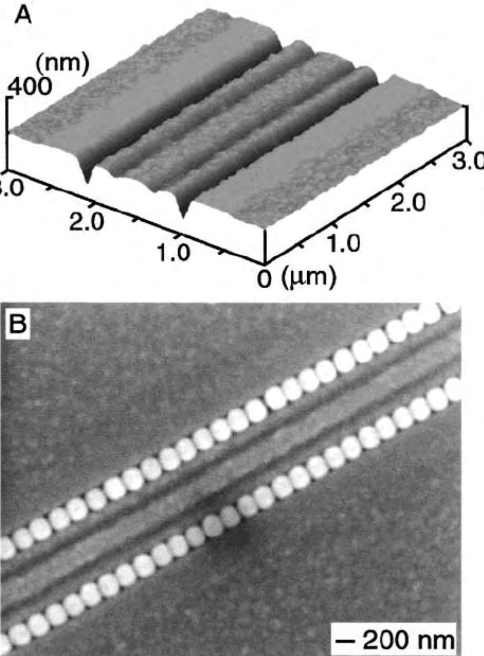


Figure 12. (A) AFM image of the template (a parallel 2D-array of trenches that were 150 and 150 nm in width and depth, respectively), which was fabricated using near-field optical lithography. (B) SEM image of the linear chains that were formed by templating 150-nm PS beads against the trenches shown in (A). These PS beads represent the smallest objects that have been successfully incorporated in the TASA process. Reprinted with permission from [178], Y. Yin et al., *J. Amer. Chem. Soc.* 123, 8718 (2001). © 2001, American Chemical Society.

porous silicon, anodized alumina, etc.) have also been used as templates for the growth of nanoparticles inside the pores [182]. One such example of three dimensionally arranged gold colloids has been prepared in the nanoscale pores of alumina membranes [181]. Immobilization of the colloid particles performed by vacuum incorporation were achieved by a modification of the inner walls of the porous alumina substrate with alkoxysilanes $Y\text{-}(CH_2)_x\text{–}Si(OR)_3$ and $Y\text{-}(CH_2)_x\text{–}SiR(OR)_2$ bearing suitable functional groups ($Y = NH_2$, SH), followed by anchoring of the colloids by self-assembly onto the modified surface. Well-washed membranes are dark-red in color upon incorporation of gold colloids, and a schematic description of an alumina membrane containing immobilized gold is shown in Figure 13. A part of the sample showing the particles located only near the pore walls is shown in Figure 14(A), which is schematically explained in Figure 14(B). The particles are closely packed, vertically overlapped, and generally observed along the walls.

3.3. Electrostatic Complexation of Nanoparticles with Charged Langmuir Monolayers

The air-water interface has been recognized to be an excellent media for organization of inorganic cations using charged amphiphilic monolayers (Langmuir films) [183].

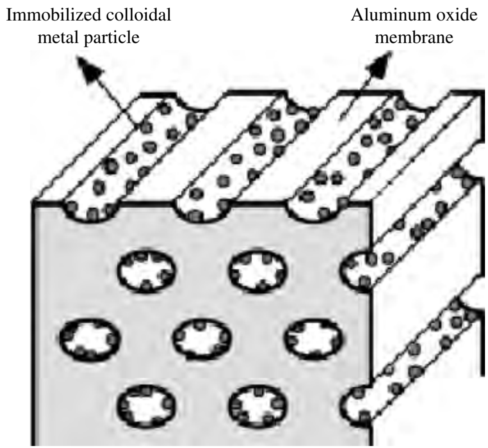


Figure 13. Model structure of a colloid-charged porous alumina membrane. Reprinted with permission from [181], T. Hanaoka et al., *Eur. J. Inorg. Chem.* 807 (1998). © 1998, Wiley-VCH.

This approach has been extended to assemble large inorganic anions [184–185] and biological macromolecules [186–187]. Colloidal particles can apparently be trapped at the liquid surface as a result of the electrostatic and surface tension forces [188–191]. The interaction between the ions/macromolecules in solution and charged lipid Langmuir monolayers drives the organization at the air-water interface. The nanoparticles organized in this fashion by the Langmuir–Blodgett (LB) method can then be transferred into appropriate substrates, for example, carboxylic-derivatized silver nanoparticles can be immobilized at octadecylamine (ODA) Langmuir monolayer surfaces [192–194].

The behavior of particle monolayers at liquid surfaces has also received considerable attention of late but was already studied some 30 years ago [195–196]. Kumaki reported an interesting work on monomolecular PS nanoparticles (~50 nm diameter) monolayers and discussed the possible significance of the surface pressures on the organization of these particles [197]. However, application of the LB technique to colloidal particles led to submonolayers with coverage less than 80% [198]. Only recently, Bardosova et al. [199] reported on the successful Langmuir–Schaefer transfer of monolayers of silica particles modified with 3-aminopropyl groups.

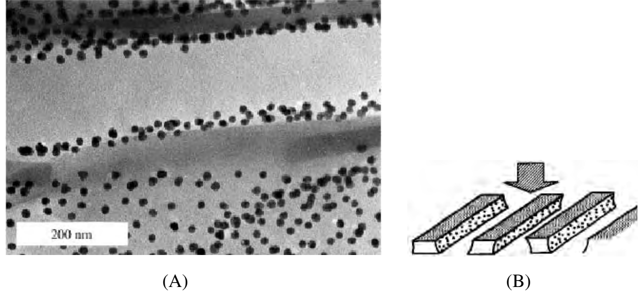


Figure 14. (A) TEM image of gold colloids in the pores. Both the upper and lower walls were removed by sectioning; (B) schematic representation illustrating the situation. Reprinted with permission from [181], T. Hanaoka et al., *Eur. J. Inorg. Chem.* 807 (1998). © 1998, Wiley-VCH.

The formation of 2D arrays by self-assembly on solid supports make use of numerous strategies [200–203], for example, a 2D array of colloidal spheres can be formed at the air-liquid interface which can then be subsequently transferred onto a surface of a solid substrate. The surfaces of these colloidal spheres have to be modified such that they will only be partially immersed into the surface of a liquid after they have been spread onto the air-liquid interface through a spreading agent (usually an alcohol) [189]. It is the strong attractive interactions (e.g., those between dipoles induced by the asymmetric interface) among the colloidal spheres that lead to the spontaneous formation of a 2D aggregate at the interface. The morphology of the aggregate usually exhibits fractal characteristics, but it can also be changed by varying a number of parameters such as the size, the number concentration, the surface hydrophobicity, or the charge density on the colloids, and the electrolytic properties of the underlying liquid [189]. In a recent demonstration, for example, Kondo et al. were able to generate 2D arrays of silica colloids (1 μm in diameter) with relatively large domain sizes by controlling the degree to which the silica colloids were immersed into the liquid surface [204]. Deckman et al. [205], Lenzmann et al. [206], and Fulda and Tieke [207] applied the LB film technique in this fashion to obtain polycrystalline 2D arrays of polymer latexes over areas as large as several square centimeters.More recently, Burmeister et al. also demonstrated a similar technique that was capable of forming ordered 2D arrays of colloidal spheres on various types of substrates [208].

Another method that has been largely explored by Nagayawa, Picard, Rakers, et al. [202–203, 209–213], uses the attractive capillary forces among colloidal spheres to organize them into a hexagonal 2D array in a thin film of liquid supported on a flat substrate. In a typical experiment, a liquid dispersion of colloidal spheres is spread onto the surface of a solid substrate. When the solvent slowly evaporates under a controlled condition, these colloidal spheres are self-assembled into a closely packed, hexagonal array, which has also been followed experimentally by using an optical microscope [202, 214]. They found that a nucleus—an ordered region that consists of a number of colloidal spheres—was first formed when the thickness of the liquid layer approached the diameter of the colloids. More colloids were driven toward this nucleus by a convective transport, and eventually organized around the nucleus due to the attractive capillary forces. A flat, clean, and chemically homogeneous surface has to be used in order to generate a highly ordered array with relatively large domain sizes. Solid substrates, such as glass slides or silicon wafers, have been used in this technique. Lazarov et al. have also explored the use of liquids such as prefluorinated oil (F-oil) or mercury as the substrates in forming highly ordered 2D arrays of colloidal spheres [215]. At the air-water interface under appropriate passivation conditions, mid-nanometer-sized gold particles can undergo self-organization into densely packed monoparticulate films [216]. These films, when transferred onto hydrophilic Formvar-coated copper

grids by vertical retraction of immersed substrates, produced monoparticulate films with variable extinction and reflectance properties, related to the hexagonal close-packed arrays that form in a micron length scale. The extinction bands of these arrays shifted by hundreds of nanometers to near-IR wavelengths and broadened enormously with increasing periodicity. Large particle arrays also demonstrated extremely high SERS, with enhancement factors $>10^7$, which were correlated with increasing periodicity.

A slight modification of this technique has been reported by Goldenberg et al. [217] following ideas of Aveyard et al. [218], which seems an interesting method for practical application in building up 2- and 3D lattices of spheres. This simple and fast technique utilizes the nonmiscibility of liquids and trapping particles between two liquids of different viscosity so as to make a monolayer of the film in the liquid interface, which can be suitably transferred onto an appropriate substrate. Goldenberg et al. utilized this technique to organize hexagonally ordered monolayers of hydrophilic micrometer size PS and silica particles at water/alkane interfaces, and to transfer them onto solid substrates. These monolayers showed good diffraction properties as observed by optical microscopy with a Bertrand lens [217]. A schematic representation of this method is shown in Figure 15(A), while a photo of actual latex particles floating on hexane is shown in Figure 15(B); a typical film so formed is shown in Figures 15(C) and 15(D) and the collected particles on the glass in Figures 16(A) and 16(B). The actual experimental condition of the PS particles floating in the water-hexane interface is shown in Figure 17(A), while 17(B) shows the long-range order in these monolayer films.

These methods or a supplement of other methods has been used to make 3D arrays of colloidal nanoparticles.

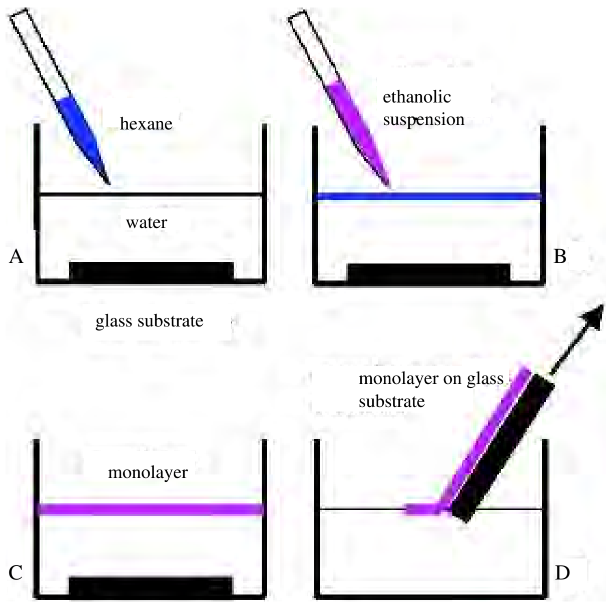
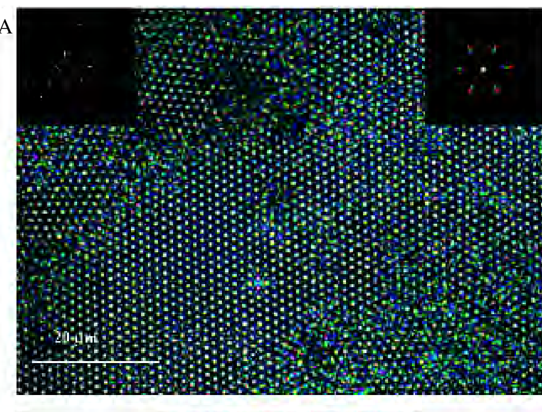


Figure 15. Scheme 1. Schematic description of the particle monolayer formation at the water-hexane interface: (A, B) spread of particle from ethanol dispersion on the water-alkane interface; (C) formation of the floating hexagonal arrays of particles: (D) transfer of particle onto solid substrate. Reprinted with permission from [217], L. M. Goldenberg et al., *Langmuir*, 18, 5627 (2002). © 2002, American Chemical Society.



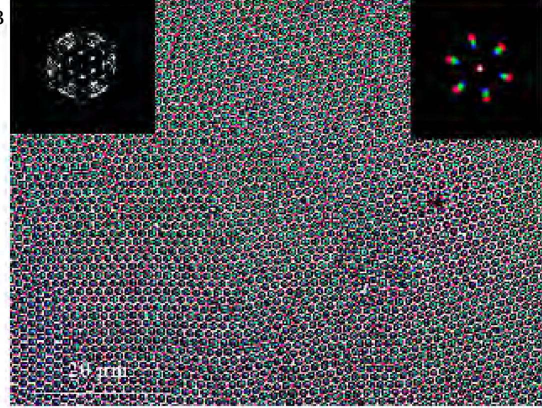


Figure 16. Optical microscopic image of a monolayer of 1.4 μm polystyrene-2,3-epoxypropyl methacrylate (PS-EPMA) (A), and (B) silica (1.7 μm) particles transferred from the water/hexane interface. Inset: corresponding diffraction images with Bertrand lens (right insert) and 2DFFT transformation (left insert, 512 × 512 pixels). Reprinted with permission from [217], L. M. Goldenberg et al., Langmuir 18, 5627 (2002). © 2002, American Chemical Society.

Sedimentation in a gravitational field is the simplest approach that has been used for the formation of 3D crystalline arrays from colloidal spheres [219–222]. Although it looks simple, the method involves several complex processes such as gravitational settling, translational diffusion (or Brownian motion), and crystallization (nucleation and growth). In addition, these force parameters need to be properly controlled as well as the size and density of the colloidal spheres and the sedimentation rate to succeed in making 3D colloidal lattices. The colloidal spheres can always

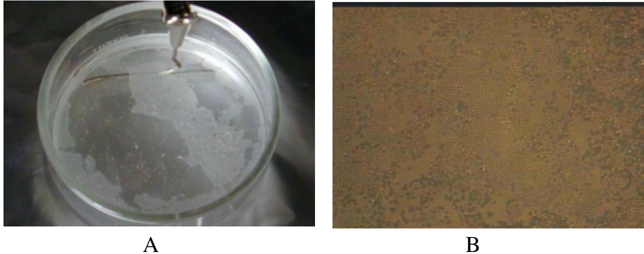


Figure 17. (A) Polystyrene particles spread on a water hexane interface, and (B) the successive film that can be collected from this partially covered layer on a glass slide.

settle completely to the bottom of a container as long as the size and density of these spheres are sufficiently high. Only when the sedimentation process is slow enough, the colloidal spheres concentrated at the bottom of the container will undergo a hard-sphere disorder-to-order phase transition to form a three dimensionally ordered lattice [223–224].

Monodispersed silica colloids are most commonly employed in sedimentation due to the high density of amorphous silica. Opalescent structures (usually referred to as synthetic or artificial opals) have been obtained from these colloidal materials under carefully controlled conditions [225–229]. Two types of crystalline arrays of colloidal spheres (or colloidal arrays) have been extensively studied: the first type includes fcc lattices formed from highly charged colloidal spheres and their volume fractions of colloids are often less than 10%; the second type is a Cubic-close-packed (ccp) structure (also a Face-centered-cubic (fcc) lattice) and its volume fraction of colloids is always close to 74%. It is generally accepted that the 3D crystalline arrays produced by this method have a ccp structure (or a fcc lattice with a packing density of ~74%) similar to that of a natural opal [229–230]. The preference of a ccp structure over a hexagonal-close-packed (hcp) one has been suggested to be a result of the difference in entropy between these two structures [231]. Recently, van Blaaderen et al. demonstrated the use of lithographically defined surfaces as templates to grow 3D crystalline arrays with desired spatial structures [232]. This process is the mesoscopic equivalent of epitaxial growth: highly ordered and well-controlled arrays as large as ~1 cm^3 could be generated. Jiang et al. also developed a layer-by-layer sedimentation method for fabricating ccp arrays of silica colloids; these arrays have a tightly controlled number of layers along the [111] direction [233].

3.4. Chemically Patterned Substrates

Stripes and channels with alternating wetability can also be produced by using wetting instabilities during the LB film transfer that has long been used for building up thin organic coatings of monomolecular layers (monolayers) onto solid substrates [234]. For example, Gleiche et al. [235] used monolayers of L-α-dipalmitoyl-phosphatidycholine (DPPC) on mica to generate a structured surface with a channel lattice exhibiting a high wetability contrast forming a lateral structure, which can be obtained by rapidly withdrawing a mica substrate (1,000 μms^{-1}) at a low monolayer surface pressure and constant temperature. Under these conditions, the film adsorption becomes unstable, leading to periodic interruptions in the molecular deposition which generate regular hydrophilic channels of about 200 nm in width and are separated by hydrophobic stripes of monomolecular height and latitude of about 800 nm. This structured surface can be used as a template to deposit materials along the channels with high adsorption selectivity by making use of the different wetting behavior of the hydrophilic channels and the hydrophobic stripes, for example, by either an anisotropic wetting/dewetting process (determined by the surface structure) or by using capillary (capillary filling) and electrostatic forces.

Gold clusters (Au55) stabilized by an organic ligand shell dissolved in 1-phenyloctane were dropped on the structured mica surface which filled the channels with Au55 clusters as shown in Figure 18. Mostly single, large-cluster aggregates (almost 4 nm in height) were lined up in the channels, but some two-layers were also noticed (see spike structure in Figure 18(B)). In contrast, the DPPC monolayer is virtually not wetted and only a few cluster aggregates are located on the top of the stripe region, resulting in a nearly perfect selective adsorption.

Self-organization on chemically patterned surfaces can be achieved in a rather straightforward way by the use of charged surfaces, as the nanoparticles in suspension are most often charged to prevent coagulation. Substrates are patterned by molecular layers in mesoscale dimensions. Substrates for such chemisorption are usually gold surfaces regiospecifically modified with functionalized thiol molecules by microcontact printing [146] or silica surfaces that have been photo-patterned by a preadsorbed cationic monomer (as shown in Figures 19 and 20) [236]. In these cases, the contrast between covered and uncovered areas, as well as the packing density within the covered regions, depends strongly on the electrochemical parameters of

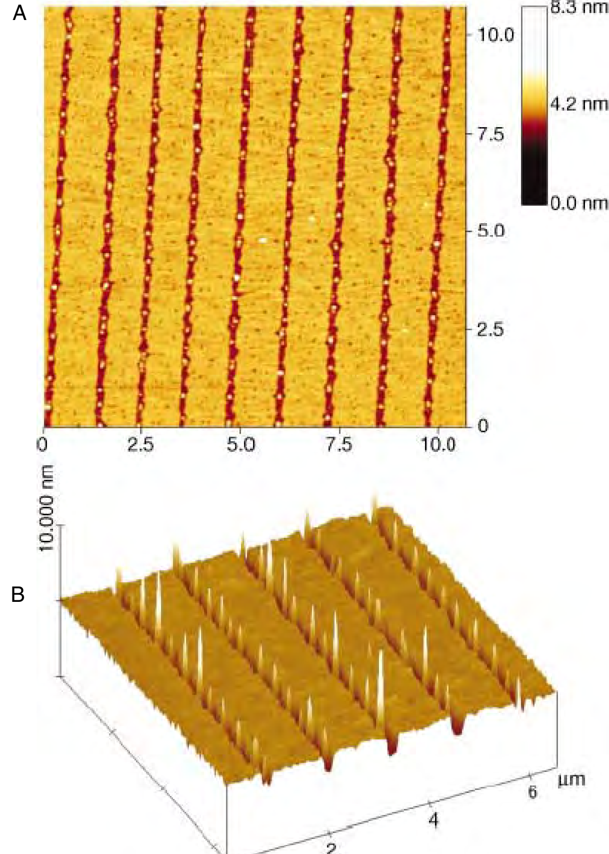


Figure 18. Dynamic scanning force microscopy (SFM) image of liquid-deposited Au55 metal cluster. (A) topography of the selectively adsorbed organic-ligand-stabilized metal cluster. Cluster aggregates (bright spots) were aligned along the channels even when scanning the previously wetted area. (B) the spikes depicted in the three dimensionally rendered selected area indicate the large height difference (image size 6.5 × 6.5 μm^2). Reprinted with permission from [235], M. Gleiche et al., *Nature* 403, 173 (2000). © 2000, Macmillan Magazines Ltd.

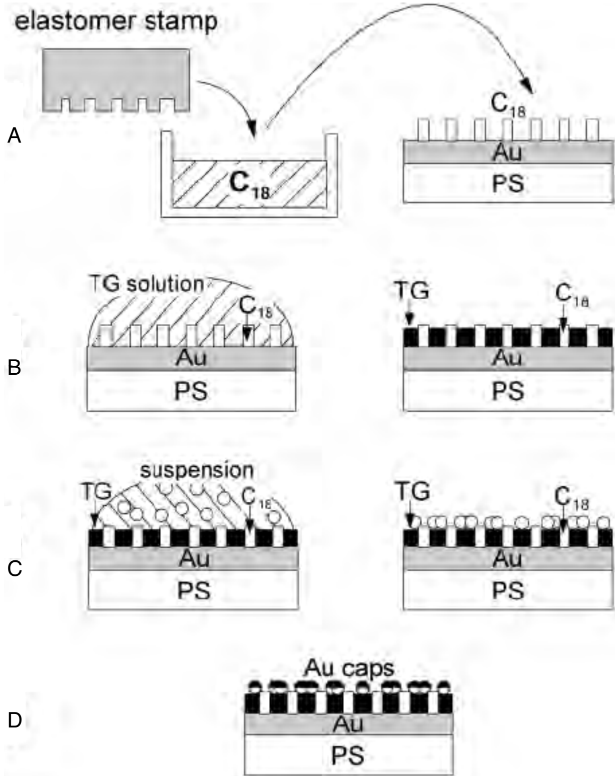


Figure 19. Schematic representation of a chemically induced patterning of PS particles via microcontact printing: (A) first, octadecanethiol (C18) is stamped onto a native gold surface via an elastomer stamp, which had been previously dipped into ethanolic C18 solution for a few seconds and dried in air; (B) next, the surface is covered with an aqueous thioglycolate (TG) solution for 1 h, to adsorb TG on the native gold areas; (C) after rinsing and drying of the chemical pattern on the gold surface, it is exposed to a suspension of PS nanoparticles, which had been prepared according to the standard procedure. After 1 h, the suspension is washed off with DI water and a patterned PS particle layer is obtained. (D) for analysis the layer was coated with 20 nm-thick gold film deposited via thermal evaporation. Reprinted with permission from [236], M. Himmelhaus and H. Takei, *Phys. Chem. Chem. Phys.* 4, 496 (2002). © 2002, Royal Society of Chemistry.

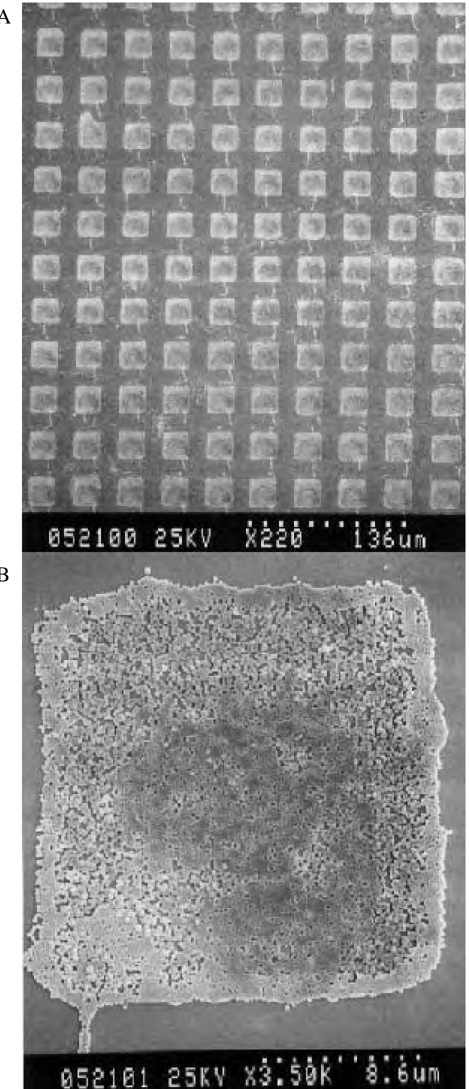


Figure 20. SEM micrographs of a patterning experiment by utilization of an elastomer stamp with a square pattern as structure for printing. (A) micrograph gives a survey, while that on (B) demonstrates the particle density within one square. Note the onset of regular close packing at several locations within the square. Dimensions of the pattern: squares (24 μm)^2, gaps 15 μm. Reprinted with permission from [236], M. Himmelhaus and H. Takei, *Phys. Chem. Chem. Phys.* 4, 496 (2002). © 2002, Royal Society of Chemistry.

the system, such as ionic strength, pH, and effective surface charge. Moreover, only moderate packing density is achieved due to the repulsive forces between the particles [237]. One way to overcome this problem is the application of cationic surfactants that increase the charge shielding and, due to adsorption onto the particle surface, introduce attractive forces such as van der Waals interaction. Recently, Chen et al. demonstrated the validity of this concept [145].

The site-selective assembly of colloidal polymer particles onto laterally patterned silane layers has also been extensively studied for the assembly of colloidal nanoparticles at mesoscale dimensions [145–146, 237–239]. The structured silane monolayers on silicon-oxide substrates are either fabricated by liquid- or gas-phase deposition or a combination of both methods, using different trialkoxysilanes with a photolithographic patterning technique. By using this method, various types of surface functionalization, such as regions with amino functions next to areas of the bare silica surface or positively charged regions of a quaternary ammonium silane surrounded by a hydrophobic octadecylsilane film have been reported (as shown in Figures 21 and 22). Often, photo-protected amino groups were, used which allowed direct photo-patterning after monolayer preparation, leading to free NH_2 groups at the irradiated regions. Depending on the assembly conditions (different surface functionalization, pH, and drying conditions), the particles can be directed selectively onto a specific surface region. The regiospecific assembly of colloidal particles onto laterally structured silane monolayers can generate complex structures. The required monolayer patterns can be obtained by a combination of the industrially well-established photolithography process with the deposition of silanes from the vapor and liquid phase onto technologically and highly

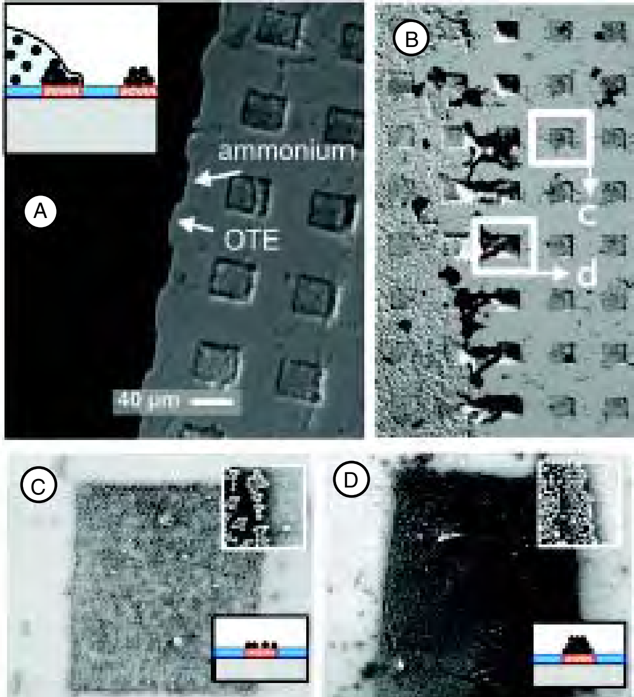


Figure 21. Optical micrographs (difference inter ference contrast) of the drying front of a latex suspension droplet on an ammonium silane (squares)-octadecylsilane monolayer pattern before (A) and after (B) removal of the liquid (inset) Schematic side view of a resting droplet on the substrate. The SEM images show a particle submonolayer on the ammonium pattern formed by electrostatic assembly in suspension (C after rapid removal of the suspension liquid) and a colloid multilayer on a similar pattern generated by capillary forces at the slow moving drying from (A) and schematics in C and D show the individual particles of the submonolayer and multilayer, respectively. octadecyltriethoxysilane (OTE). Reprinted with permission from [237], U. Jonas et al., PNAS, 99, 5034 (2002). © 2002, National Academy of Science.

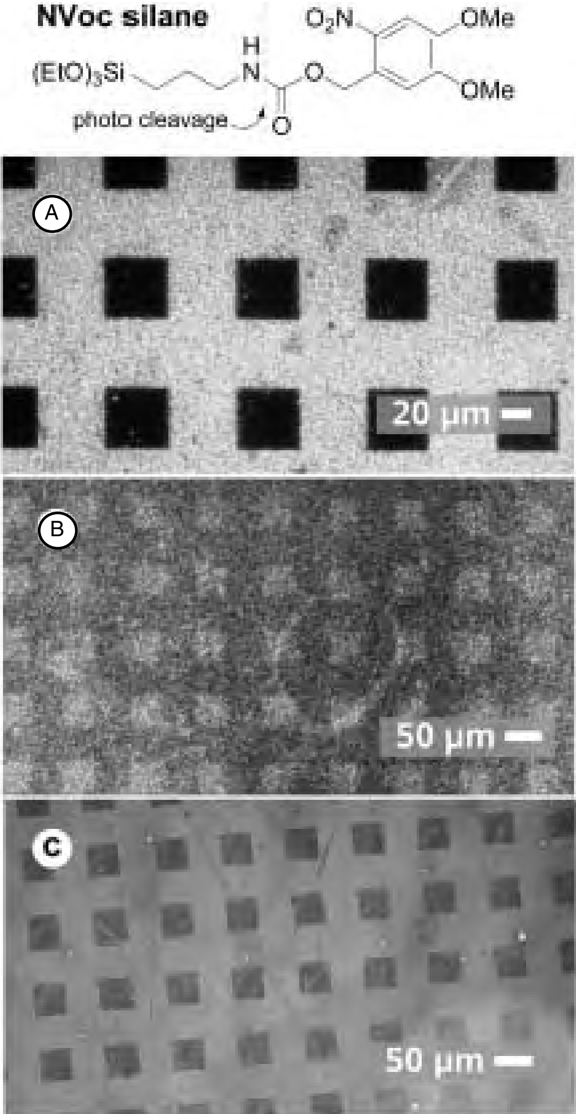


Figure 22. (A) Optical micrograph of latex particles assembled onto an NVoc silane monolayer after photo-deprotection with an argon laser (364 nm) through a gold mask on quartz (dark squares were protected from light). (B) Particle assembly (pH 4.5) on a mixed monolayer of the NVoc and ammonium silane after irradiation (similar to A). Bright squares were protected from light and show a lower particle density in the SEM. (C) Fluorescence micrograph of a mixed monolayer (NVoc and ammonium silane, similar to B) after irradiation and staining with a fluorescent probe (Alexa Fluor 488). Reprinted with permission from [237], U. Jonas et al., PNAS 99, 5034 (2002). © 2002, National Academy of Science.

relevant substrates such as silicon with an oxide layer, glass, and quartz. The particle adsorption process driven essentially by electrostatic and polar interactions in combination with adhesion inhibition at hydrophobic surface layers is a self-controllable and highly parallel process which should allow the simultaneous fabrication of many devices in one assembly step with essentially no size limitations from nanometer to the millimeter length scales.

3.5. Optically Directed Self-Organization

At the current stage of development, all of the above-mentioned methods are only capable of generating colloidal arrays built up by small domains, and the largest single domain usually contains fewer than 10,000 colloidal spheres [240]. Most of these methods can only form 2D hexagonal arrays in which the colloidal spheres are in physical contact. As a result, it is very hard to independently vary the lattice constant and the particle size. The approach based on optical forces seems to have the potential to overcome these difficulties [241–242]. In this method, colloidal spheres are organized into a highly ordered 2D structure in a liquid by creating an optical standing wave pattern having a regular array of intensity antinodes. The colloidal spheres are then driven to the antinodes maxima by the optical forces. Depending on how many laser beams are used to create the standing wave, patterns as complex as a quasicrystal have also been produced. As demonstrated by Misawa et al. and by Mio and Marr, the method based on optical forces was also capable of generating an arbitrary 2D or 3D pattern by adding individual colloidal spheres to an array one at a time [243–244].

Several years ago, Ashkin et al. experimentally demonstrated that small particles (25 nm to 10μm) in a suspension could be trapped by a single-focused laser beam [245]. Following their work, many researchers have used optical trapping techniques (also known as optical tweezers) to manipulate micrometer and submicrometer-sized objects [246–247]. Because of their nondestructive, sterile nature, optical tweezers have found great success in manipulating biological systems, including bacteria, viruses, chromosomes, and DNA [248–251]. More recently, Hayward et al. has shown the usefulness of light-induced orientation in what they called "optically tunable micropatterns" using ultraviolet light illumination [252]. Though the mechanism is largely unexplained, it could result from the light-induced changes on the wetting behavior of the substrate—a mixture of optical illumination for chemical patterning [253].

3.6. Electric or Magnetic Field-Assisted Self-Organization

Application of an electric field results in structural transitions in the colloidal suspension because the interparticle electrostatic interactions due to polarization are stronger than Brownian forces. The tendency of particles in suspension to form structures such as chains upon application of an electric field was reported centuries ago by scientists such as Priestly and Pohl [254–255]. Quantitative experiments on the electrorheological effect were first performed by Winslow in 1949, when he reported that suspensions of silica gel particles in low-viscosity oils tend to fibrillate upon application of electric fields, with fibers forming parallel to the field [256]. It was reported that at fields larger than ~3 kV/mm, the suspensions behaved like a solid, which flowed like a viscous fluid above a yield stress that was proportional to the square of the applied electric field. A recent and comprehensive survey of electrorheological (ER) fluids, where the issue of particle aggregation is addressed, is provided by Parthasarathy and Klingenberg [257]. At low fields, below ~100 V/mm, where the interparticle electrostatic interaction energies are low compared to thermal energies, no significant particle association are generally observed and the nanoparticles tend to sediment to the bottom electrode, if their density is larger than the solvent density. However, at field strengths ~1000 V/mm, field-induced structures, such as chains of touching particles, are formed that do not break up due to thermal fluctuations, implying that the electrostatic energy at contact was many times kT [258].

As an example of the ordering of nanoparticles onto structured surface, Kumacheva et al. [259] has reported an experiment with poly(methylmethacrylate) (PMMA) spheres electrodeposited onto Indium tin oxide (ITO)-covered substrates. The speed of colloid crystal growth can be controlled via electrodeposition parameters such as voltage and time, and electrophoretic mobility of the particles, thus control over layer-by-layer deposition is possible. They further reported that the particles get into ordered phase after reaching the electrode, which they attributed to the reorganization of the microspheres in the grooves by squeezing the newly arriving spheres between the already deposited particles, and followed by synergistic particle rearrangement. This resulted in large-scale particle ordering, as is shown in Figure 23(D) for assembly of the colloidal spheres in 4.2 μm-thick grooves.

The use of nonuniform electric fields for the manipulation of μm-sized particles are reasonably well-documented [260–268]. A variety of biological applications including collection, fusion, and separation of biological cells [269], immobilization of DNA [270], and a collection of viral particles [271] have been demonstrated. Dielectrophoresis (DEP) has been used in the assembly of a biosensor by concentrating α-protein-coated latex microspheres in the area between two oppositely charged electrodes [272]. Alignment of nanospheres [273] and dielectrophoresis for the trapping of submicron latex spheres inside a 3-D, multielectrode array [263], and alignment of metallic needle-shaped particles with diameters between 70–350 nm and a few μm in length using interdigitated electrodes has also been reported; an example is shown in Figure 24 [274]. The possibility of, contact-free handling for assembly and mechanical processing of microbeads has been demonstrated [275]. Today DEP has not yet been sufficiently exploited for the self-organization of nanoparticles, though the fundamental principles of DEP are reasonably well-understood. This is an attractive method for the self-organization of nanoparticles and will almost certainly gain more importance in the future works for the self-organization of nanoparticles.

The electric field forces on particles can be controlled by monitoring the charge on the colloidal particles (through the control of the double layer). Recent work within this area, has shown that the electric field is more than just a substitute for gravity, and the formation of highly ordered colloidal layers has been reported for electrophoretic deposition of micrometer-sized silica and polystyrene particles [277–279], and nanometer-sized gold particles [277, 280–281]. The self-organization phenomenon was explained by electro-osmotic or electro-hydrodynamic microflows around the particles, which are induced by the electric field and lead to attractive forces between the particles at the surface [277, 282]. While the exact mechanism of ordering is still open to debate, it is clear that improved ordering results from the fact that the colloidal particles are only weakly bound ("physisorbed") to the substrate and remain fairly mobile. Only after an additional fixing step (e.g., application of a high field), the

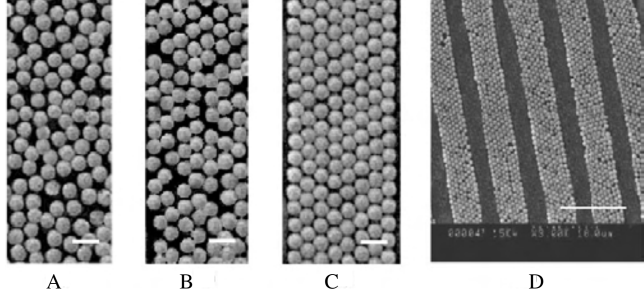


Figure 23. SEM images of the colloidal arrays of latex spheres of PMMA electrodeposited on nonpatterned (A) and patterned (B–D) ITO-coated surfaces. The widths of the grooves are 5.5 μm (B) and 4.2 μm (C, D). The SEM images of the colloidal arrays with the width 4.2 μm were cropped in (A) and (B). Scale bars are 1 μm (A–C) and 10 μm (D). Reprinted with permission from [259], E. Kumacheva et al., *Adv. Mater.* 14, 221 (2002). © 2002, Wiley-VCH.

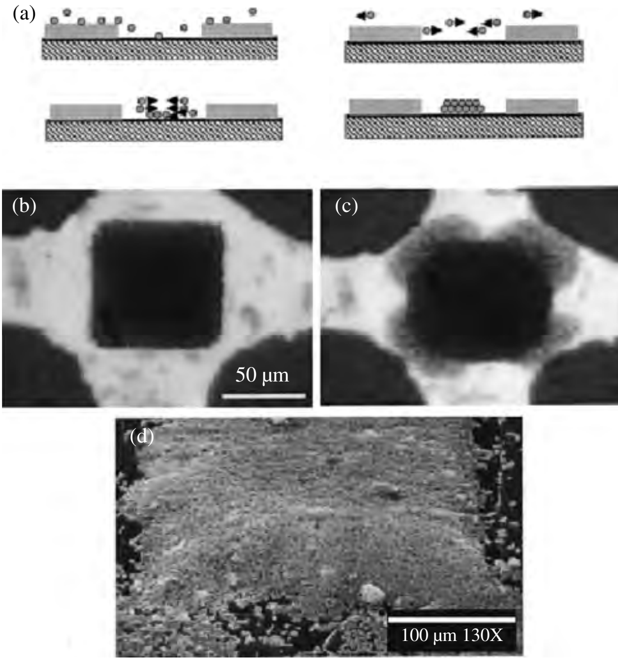


Figure 24. (a) Schematic presentation of the 3D particle organization achieved *via* DEP. The first particles on the plane are pushed to the center of the electrode setup (negative DEP) and form an ordered structure. Particles that later arrive in the vicinity of the electrodes deposit on the top of the 3D structure, creating additional layers. (b) A freestanding, pyramid-shaped structure of silica particles is shown as a result of dielectrophoretic organization. (c) The structure collapses immediately after the electric field is switched off. (d) SEM image of a 3D, freestanding formation of silica particles in the area between two long and parallel electrodes. Reprinted with permission from [276], A. Docoslis and P. Alexandris, *Electrophoresis.* 23, 2174 (2002). © 2002, Wiley-VCH.

colloidal film is anchored to the substrate. Based on these ideas, it has been proposed that highly ordered multilayers of colloidal particles could be grown by using dc or low frequency ac fields (one such example is shown in Figure 25 [277]).

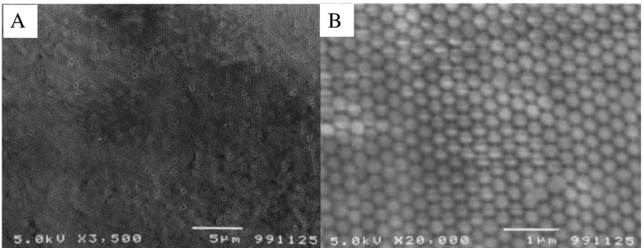


Figure 25. (A) Large single domain of the opal made of 300 nm silica spheres. The long-range ordering can be observed over 100 μm. The sample for the SEM observation was derived by carefully removing one substrate of the sandwich-like cell. (B) Hexagonal arrangement of the silica spheres with excellent arrangement was observed. It was found that the domain size exceeds 100 mm. The hexagonal arrangement was assigned to the close-packed surface of a face-centered cubic or hexagonal close-packed structure. It is interesting to note that opaline materials synthesized by conventional sedimentation methods show both the hexagonal and tetragonal arrangements. Reprinted with permission from [230], H. Miguez et al., *Langmuir* 13, 6009 (1997). © 1997, American Institute of Physics.

Magnetic fields can, in principle, be used to direct particles onto a substrate, making use of the motion of magnetic particles in magnetic field gradients [283], and there have been reports where magnetic field forces have been used to direct the colloidal crystal formation [284]. In a recent publication, Bizdoaca and Lin et al. have reported on the fabrication of grids of micrometer-sized core-shell particles, utilizing magnetophoretic deposition techniques in a water-based colloidal suspension with core-shell type spherical polystyrene particles of 640 nm diameter core, covered with a shell of five layers of 12 nm diameter Fe_3O_4 nanocrystals. The separation and length of the individual chains were tuned by the magnetic field [285–286]. Magnetic field forces may play an important role in the organization of colloidal nanoparticles into ordered arrays most likely in conjunction with electric field forces.

3.7. Biologically Assisted

Nature allows structural building blocks to hierarchically organize into an ensemble with an atomic order resolution, which are formed spontaneously with high accuracy and minimum energy similar to the bottom-up self-organization process [287]. On the contrary, it is quite difficult for chemists to achieve the total synthesis and structural control of macromolecules by covalent chemical synthesis methods only [288]. Nobel Laureate Jean-Marie Lehn used the words of Leonardo da Vinci 'Where nature finishes producing its own species, man begins, using natural things and with the help of this nature, to create an infinity of species,' while, writing about the perspectives on the future and perspectives of supramolecular chemistry [289]. Natural evolution has led to highly functional assemblies of proteins, nucleic acids, and other (macro)molecules which perform complicated tasks that are still difficult to emulate. For example, the 20-nm ribosome particle is an effective supramolecular machine which spontaneously self-assembles from more than 50 individual protein and nucleic acid building blocks, thereby impressively demonstrating the power of biologically programmed molecular recognition [290–292]. The bacterial flagellar motor, which is 30 nm in diameter, resembles the electric rotary motor [293–295], while muscle proteins resemble a linear motor, but the "engine" size is approximately 30 nm [296]. Assemblies of these proteins are self-assembled molecular machines that are small in size with highly efficient energy transduction properties. These characteristics make proteins good solutions for constructing sophisticated devices on the nanometric scale.

Protein molecules have the ability to build nanometer-sized supramolecules by self-assembly, which can themselves create higher order structures like cell components. Some of the schemes that attach protein molecules are illustrated in Table 2. Mirkin et al. [75] and Alivisatos et al. [297] have showed that complementary DNA oligonucleotides offer possibilities for self-assembled nanoparticles. Protein-based conjugation offers various advantages like a large number of complementary systems with a wide range of free-association energies are synthetically available, and the systems are well known in biochemistry and molecular immunology which can be handled by standard procedures [298–299]. From a materials science point-of-view,

Table 2. Various Methods to Attach Protein to Self-Assembled Surfaces

Functional side group on peptide	Available surface derivatization	Type of binding
	Natural	
-COOH (carboxylic acid) Asp	Amino	Electrostatic covalent amide (after carboxy activation)
NH_2 (amino) Lys, Gln, Arg	Carboxylic acid, active ester, epoxy	Electrostatic covalent amide
-SH (thiol) Cys	Maleimide	Covalent thio-ether
-OH Ser, Thr	Epoxy	Covalent ether
	Synthetic	
His-Tag	Ni-NTA Complex	Coordination complex
Strep-Tag	Strep-Tactin, Streptavidin	Supramolecular complex
Biotin	Streptavidin	Supramolecular complex

polypeptides offer many advantages over conventional synthetic polymers due to their inherent ability to adopt stable conformations and self-assemble into precisely defined structures that allow unprecedented control over materials morphology and properties. Synthetic strategies for the preparation of polypeptides can be divided essentially into three classes: ring-opening polymerization, solid-phase synthesis, and protein engineering. Recent advances in these methods have allowed the preparation of (poly)peptides and (poly)peptide hybrids, which can be assembled in a controlled fashion into supramolecular architectures and materials that mimic the structure and function of proteins [300].

Right from the discovery of the double-helix structure of DNA, biology has evolved from a purely descriptive and phenomenological discipline to exact science in the molecular regime, and recombinant DNA technology has brought insights into the basic principles of many biochemical processes and is also shaping developments in modern biotechnology. Today, we can genetically engineer relatively simple bacterial cells, and hopefully, in the future, we will be able to tailor complex organisms. In view of such revolutionary developments, scientists and researchers have reported wide-ranging developments in the fusion of biotechnology with nanoparticles, since biomolecular components have typical size dimensions in the range of about 5 to 200 nm [301]. The concept of DNA hybridization-based self-organization of molecular compounds has been applied to the assembly of DNA-derivatized gold colloids [75, 297, 302]. Such defined arrangements of nanocrystal metal clusters (quantum dots) have been investigated for their novel physical properties and possible applications in the field of opto-electronic technology [303–306]. For this, superlattices or quantum dot molecules are necessary, in which crystallites from different materials are spatially assembled to tailor the properties. Recently, Takahagi et al. [307] described the fabrication of 3D particle structures based on DNA hybridization. The results indicate that DNA hybridization is a very useful method for assembling nanoparticles into a 3D nano-structure. To fabricate this structure, two colloidal gold suspensions containing particles of 9 nm and 20 nm in diameter, and oligonucleotides that are functionalized by heptanethiol at their 5′ termini were used by the authors. The thiol group at the end of the oligonucleotides was adsorbed to the particle surfaces by a chemical bond upon mixing with the colloidal gold suspension, resulting in oligonucleotide monolayers on the particles. The 9-nm particle suspensions were mixed with thiol-5′-GGTCGGCACA-3′, and the 20 nm particle suspension was mixed with thiol-5′-TGTGCCGACC-3′. After keeping for 48 h at room temperature, the two mixture suspensions were blended and diluted at pH 7. A 3D structure has formed with alternating superposition of the big and small particles. Though the internal structure cannot be observed with SEM, the 20 nm gold particles can be seen to exclusively connect with the 9 nm particles on their surfaces. These results suggest that the internal structure of DNA-linked particles might consist of alternating small and large particles, and 3D particle networks had formed.

The highly specific recognition properties of antibodies and antigens can be used to assemble a wide range of nanoparticle-based structures with specific cross-linking, compositions, and macroscopic architectures; for example, IgE and IgG antibodies have been used to attach on surfaces followed by interparticle conjunction in the presence of bivalent antigens with appropriate double-headed functionalities [297]. Similarly, streptavidin/biotin cross-linking has been used to bind nanocrystals onto a substrate [299]. Some protein-supramolecules have the ability to accommodate inorganic materials and this process is called biomineralization. Their surfaces are thus designed to sequester inorganic materials into composites like seashells or bone [308]. This biomineralization process has been used to make proteins that accommodate nanometer-size inorganic quantum dots [300]. As proteins are more easily damaged or destroyed compared to inorganic materials, there are several potential means to eliminate the protein moiety from an assembled array by heat treatment and UV irradiation, or through other chemical means.

Combining the abilities of proteins to self-assemble, organize materials, and then be selectively eliminated, it is expected that nanometer-size inorganic materials can be positioned by protein supramolecules into nanostructures. The protein, therefore, acts as a temporary scaffold, which determines the nature of the final structure without being part of it. As a feasibility study of this method, the ferritin molecule assembly was employed to make an array of inorganic nano-dots suitable for a quantum electronic device. Several methods have been reported to obtain arrays of ferritin molecules and it has become possible to make large 2D crystals or arrays of ferritin at an air-water interface [309–313].

Another process has been proposed which exploits protein molecules as scaffolds for producing inorganic, functional nanostructures on a flat surface. An array of conductive FeO dots with a diameter of approximately 6 nm was fabricated and a hexagonally packed array of ferritin molecules was formed at an air-water interface by self-assembly. Following this, the ordered ferritin was transferred onto a modified Si surface and then the protein shell of the ferritin molecule was eliminated by 450 °C heat treatment

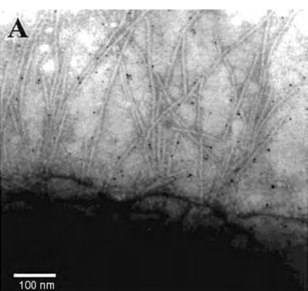
under nitrogen, which resulted in an ordered array of conducting nano-dots bound to the Si surface [314]. Combining this result with the fact that ferritin molecules have the ability to accommodate various metals and metal complexes, it is possible to organize nano-dot arrays suitable for quantum electronic devices by the so-called bio-mimetic “Bio Nano Process” method [315–321]. As the ferritin molecules are produced from the same gene, the cavity size of each apoferritin molecule is identical and hence, the nanoparticles so produced are expected to be identical in size—a key requirement in producing quantum opto-electronic devices. Combined with the self-assembly of the protein, it may be possible to build more complex functional inorganic nanostructures on a flat surface than simple arrays.

Nanoscale-ordered arrays of metal and semiconductor quantum dots have been fabricated by binding presynthesized nanoparticles onto crystalline protein templates made from genetically engineered hollow double-ring structures called chaperonins [322]. Using structural information as a guide, a thermostable recombinant chaperonin subunit was modified to assemble into chaperonins with either 3 nm or 9 nm apical pores surrounded by reactive thiols. These were crystallized into 2D templates up to 20 μm in diameter. The periodic solvent-exposed thiols within these crystalline templates were used to size-selectively bind and organize either gold or CdSe-ZnS quantum dots into arrays.

The cell membrane is another example of a self-assembled system. Simplified models of cell membranes consider solid-supported membranes as model systems for fundamental biophysical research, biosensors, and the design of phantom cells exhibiting well-defined adhesive properties and receptor densities [323]. Several strategies for the assembly of biomembranes onto various solid supports have been reviewed in great detail [324–328]. The major techniques are direct vesicle fusion [329], the LB technique [330], and molecular self-assembly from dilute organic solutions [331]. Three types of supported membranes can be assembled [332]:

(a) supported, lipid bilayer membranes prepared by vesicle spreading or LB deposition at hydrophilic surfaces,
(b) covalently tethered SAMs of functionalized lipid derivatives, prepared by coadsorption of membrane-forming components,
(c) asymmetric or hybrid bilayer membranes, composed of an outer lipid layer and an inner SAM. The most natural membrane bilayer environment is the first approach.

Another innovative approach has been reported by Walboomers et al. [333] and Britland et al. [334], where they have demonstrated that cells can be “contact-guided” along deep grooves on a patterned surface without any chemical treatments to the surface itself. Britland et al. [334] modified glass substrates with both chemical and topographical cues. The chemical adhesive tracks were patterned both parallel and perpendicular to the topographical grooves with laminin, an ECM protein carrying the RGD peptide sequence. The cells were found to orient along chemical adhesive track when the grooves on the substrate are less than 500 nm in depth. For deeper grooves, the contact guidance by the topography dominates the cellular behavior (alignment). If nanoparticles are attached to the cells, then by virtue of the cell alignment, the particles will get aligned into the structure that has been engraved, analogous to reports on decorating such structures by capillary forces that are quite difficult to control.



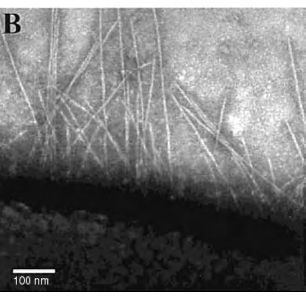


Figure 26. Typical TEM images of sectioned areas of (A) pili of the *E. coli* ORN178 strain bound with m-AuNP; (B) the *E. coli* ORN208 strain deficient of the *fimH* gene without m-AuNP binding. The experiments were performed in LB at room temperature. Reprinted with permission from [337], C. C. Lin et al., *J. Amer. Chem. Soc.* 127, 3508 (2002). © 2002, American Chemical Society.

In another recent article, biodirected epitaxial nanodeposition of polymers was achieved on a template with an oriented molecular surface. *Acetobacter xylinum* synthesized a ribbon of cellulose I microfibrils onto a fixed, nematic, ordered substrate of glucan chains with unique surface characteristics. The substrate directed the orientation of the motion due to the inverse force of the secretion during biosynthesis, and the microfibrils were aligned along the orientation of the molecular template. This approach also promises some outstanding possibility of self-organization of nanoparticles [335]. Bacteriophages offer another way to form large-dimension, patterned assemblies by engineering the phage to specifically bind functional components, as has been done for a variety of inorganic materials [336]. Because the phage is strongly bound to the inorganic components, the phages are participants rather than bystanders in the assembly process. Relatively small (~20 nm in diameter) particles are easily organized into the layered structure formed by the viral rods, leading to well-populated lines only a few tens of nanometers across, spaced by the length of the bacteriophages. In a recent article recently published, Lin et al. has mannose-encapsulated gold nanoparticles to type 1 pili in *Escherichia coli* [337]. Type 1 pili are filamentous, proteinaceous appendages that extend from the surface of many gram-negative organisms and are composed of FimA, FimF, FimG, and FimH proteins. FimA accounts for more than 98% of the pilus protein, and FimH is uniquely responsible for the binding to D-mannose (Fig. 26).

4. CONCLUSIONS

The synthesis of nanostructured materials with tunable properties is central to the development of varying applications in nanoscale science and technology. The convergence of physics, chemistry, and biology will actually lead to an explosive increase in possibilities of R&D directions. Choices will be difficult and sometimes limited by the breadth of expertise available. Encouragement of

multidisciplinary and multilaboratory collaborations would accelerate progress to the benefit of all. Through devices such as quantum well lasers and their enabling roles in applications, ranging from telecommunications to home entertainment, it is evident that nanotechnology has and is having a major impact on our daily lives which will only multiply over the decades ahead. Self-assembly of nanometric particles will definitely play an important role in this. Non-lithographic approaches based on thermodynamically driven self-organization processes are especially appealing, because of their potential for large-scale production with very small infrastructure investments. The spontaneous organization of monolayer-protected metal nanoparticles into periodic 2D arrays is archetypal of this approach, with many of these arrays demonstrating novel optical or electronic properties as a function of particle size or interparticle spacing. Stabilized particles beyond a nanometric-size threshold become increasingly prone to multilayer or 3D aggregate formation, which can be attributed to the rapid increase in van der Waals attraction and the loss of surfactant chain mobility on the planar facets of the nanoparticles as a function of size.

The other approach of positioning molecules ranging from small alkanethiols to larger biosystems, such as proteins or DNA on surfaces that can be used as a generic approach to position other molecules such as enzymes, collagens, fibronectin, or synthetic moieties, is another interesting possibility to assemble ordered structures. As the interdisciplinarity of this field develops further, we will see the birth of innovative techniques aimed at mimicking nature. An interesting article on the biomimetic approach to produce materials has been published recently [338]. The interface between biological systems and functional semiconductors remains a relatively unexplored area that requires research in the near future. The ability of electronic/optical devices to sense minute samples of biological/medical materials *in-situ*, and to relay this information to a central location opens up possibilities for quick diagnostics and the remote monitoring of patients.

A research field, which is not yet treated in detail, is the influence of size and ordered arrangement of particles in layers on the heat conductivity. The phonon heat conduction mechanism in nanostructures differs significantly from the behavior in bulk material. Phonon size effects include increased phonon scattering at grain boundaries, modification of the phonon dispersion relation, and phonon rarefaction surrounding small structures have been predicted and also observed experimentally [339–340]. It also seems that change of the phonon spectrum, due to confinement effects, could lead to phonon-phonon part of the thermal conductivity, like it occurs in superlattices. The fact that thermal transport properties in nanostructured materials are affected by the size opens up new possibilities for material scientists to engineer structures adapted to the desired properties depending on its applications. Provided we understand the phenomena in detail, one can engineer the structure of the materials by assembly of particles in such a way that the phonon conduction (phononics or phonon engineering) in different directions could be possible. For microelectronics, phonon engineering can provide improved device design to minimize the impact of adverse size effects on the device temperature [341].

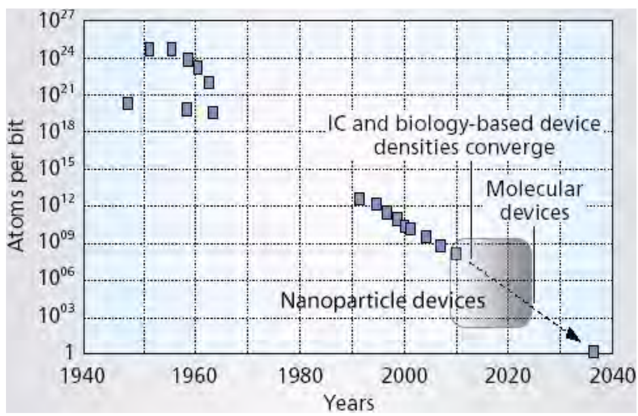


Figure 27. Moore's law predictions. Adapted with permission from [342], V. V. Khirnov and D. J. C. Herr, *Computer*, January 2001, 34. © 2001, IEEE.

The extension of Moore's law is expected to continue until the next decade (as shown in Fig. 27) with devices in their present form, which will ultimately give way to disruptive technologies involving nanoparticles, single molecule electronics, and single electron transistors [342]. Again, there will be a need for the design of complex novel architectures to accommodate these new devices with novel materials and gate designs and interconnects. Today, the major shortfall of nanoscience and nanotechnology is the lack of understanding of the physics of the interaction of objects (surfaces, particles, individual molecules) at the nanoscale. Questions need to be answered regarding how nanoparticles can be stabilized, and in what media, how nanoparticles interact and influence each other, what proportions in a hybrid system make a critical difference, and how the ordered structures can be created and retained. How can these characteristics be predicted? This area of "Extreme Nanotechnology" requires vast amounts of imagination and intuition if nanoscience research is to have any practical impact on human lives. Its potential is enormous, and the need to understand underpinning nanoscience should be properly supported so that the bottom-up nanoscience and nanotechnology is not overshadowed by the reluctance to invest in revolutionary technologies, since the benefits from this area will be considerable in the long run.

GLOSSARY

2D arrays Two dimensional arrangement of nanoparticles.

3D structures Three dimensional arrangement of nanoparticles.

Acetobacter xylinum It is a gram negative bacterium and is unique in its prolific synthesis of cellulose.

Atomic force microscopy (AFM) An instrument able to image surfaces to molecular accuracy by mechanically probing their surface contours. A kind of proximal probe.... A device in which the deflection of a sharp stylus mounted on a soft spring is monitored as the stylus is moved across a surface. If the deflection is kept constant by moving the surface up and down by measured increments, the result

(under favorable conditions) is an atomic-resolution topographic map of the surface. Also termed a scanning force microscope.

Biomimetic Imitating, copying, or learning from nature. Nanotechnology already exists in nature; thus, nanoscientists have a wide variety of components and tricks already available.

"Bottom Up" method Building larger objects from smaller building blocks. Nanotechnology seeks to use atoms and molecules as those building blocks. The advantage of bottom-up design is that the covalent bonds holding together a single molecule are far stronger than the weak.

Chemical vapor deposition (CVD) A chemical reaction that transforms gaseous molecules, called precursor, into a solid material, in the form of thin film of powder, on the surface of a substrate.

Coulombic repulsion If bonds in the molecule are polar, partial electrostatic charges will reside on the atoms. This electrostatic interaction is represented with a Coulombic potential function according to Coulomb's law. The repulsion due to this electrostatic interaction is termed as Coulombic repulsion.

Deoxyribonucleic acid (DNA) A macromolecule consisting of one or two strands of linked deoxyribolucleotides.

Dielectrophoresis (DEP) The lateral motion imparted on uncharged particles as a result of polarization induced by non-uniform electric fields.

Electrostatic stabilization Electrostatic stabilization is used to prevent agglomeration of nanoparticles by absorbed molecules. It involves the creation of an electrical double layer arising from ions absorbed on the surface and associated counter ions that surround the particle. Thus if the electric potential associated with the double layer is sufficiently high, the Coulombic repulsion between the particles will prevent their agglomeration.

Escherichia coli It is one of hundreds of strains of the bacterium. Although most strains are harmless and live in the intestines of healthy humans and animals, this strain produces a powerful toxin and can cause severe illness.

Extracellular matrix (ECM) protein Example: procollagen peptide 1 (P1CP), procollagen peptide 3 (P3P), hyaluronic acid (HA).

IgG immunoglobulins Immunoglobulins are thiol-containing plasma proteins, which include all antibody molecules. Members of the immunoglobulin super family share structural and genetic features with immunoglobulin molecules and contain at least one immunoglobulin domain. An immunoglobulin domain is made of two □-pleated sheets held together by a disulfide bond.

Langmuir–Blodgett The name of a nanofabrication technique used to create ultrathin films (monolayers and isolated molecular layers), the end result of which is called a "Langmuir-Blodgett film".

Moore's Law Coined in 1965 by Gordon Moore, future chairman and chief executive of Intel, it stated at the time that the of number transistors packed into an integrated circuit had doubled every year since the technology's inception four years earlier. In 1975 he revised this to every two years, and most people quote 18 months. The trend cannot continue indefinitely with current lithographic techniques, and a limit is seen in ten to fifteen years. However, the baton could be passed to nanoelectronics, to continue the trend (though the smoothness of the curve will very likely be disrupted if a completely new technology is introduced).

Nanocrystals Also known as nanoscale semiconductor crystals. "Nanocrystals are aggregates of anywhere from a few hundred to tens of thousands of atoms that combine into a crystalline form of matter known as a "cluster." Typically around ten nanometers in diameter, nanocrystals are larger than molecules but smaller than bulk solids and therefore frequently exhibit physical and chemical properties somewhere in between. Given that a nanocrystal is virtually all surface and no interior, its properties can vary considerably as the crystal grows in size."

Nanomaterials Can be subdivided into nanoparticles, nanofilms and nanocomposites. The focus of nanomaterials is a bottom up approach to structures and functional effects whereby the building blocks of materials are designed and assembled in controlled ways.

Physisorption (or physical adsorption) Is adsorption in which the forces involved are intermolecular forces (van der Waals forces) of the same kind as those responsible for the imperfection of real gases and the condensation of vapors, and which do not involve a significant change in the electronic orbital patterns of the species involved.

Polymethylmethacrylate spherical particles (PMMA spheres) These particles are inert and nonbiodegradable.

Quantum Describes a system of particles in terms of a wave function defined over the configuration of particles having distinct locations is implicit in the potential energy function that determines the wave function, the observable dynamics of the motion of such particles from point to point. In describing the energies, distributions and behaviours of electrons in nanometer-scale structures, quantum mechanical methods are necessary. Electron wave functions help determine the potential energy surface of a molecular system, which in turn is the basis for classical descriptions of molecular motion. Nanomechanical systems can almost always be described in terms of classical mechanics, with occasional quantum mechanical corrections applied within the framework of a classical model.

Quantum dots (QD) Nanometer-scale "particles" for selectively holding or releasing electrons. It is a particle of matter so small that the addition or removal of an electron changes its properties in some useful way.

RGD peptide Peptides containing the Arg-Gly-Asp (RGD) sequence. They can inhibit platelet aggregation.

Scanning electron micrograph (SEM) The picture formed by the scanning beam of electrons in a scanning electron microscope.

Scanning force microscopy (SFM) The SFM works by detecting the vertical position of a probe while horizontally scanning the probe or the sample relative to the other. The probe is in physical contact with the sample and its vertical position is detected by detecting the position of a reflected laser beam with a photo diode that consists of two or four segments.

Scanning probe microscopy (SPM) SPM technologies share the concept of scanning an extremely sharp tip (3–50 nm radius of curvature) across the object surface. The tip is mounted on a flexible cantilever, allowing the tip to follow the surface profile.

Scanning tunneling microscopy (STM) It is a device that obtains images of the atoms on the surfaces of materials. The governing principle of STM is the quantum tunneling of electrons through a thin potential barrier separating two electrodes. The instrument basically consists of a very sharp tip which position is controlled by piezoelectric elements (converting voltage in mechanical deformation). By applying a voltage between the tip and a metallic or semiconducting sample, a current can flow between these electrodes when their distance is reduced to a few atomic diameters. The amplitude of the current strongly depends on the distance between the tip and the sample, and of course also on the potential difference.

Self-assembly In chemical solutions, self-assembly (also called Brownian assembly) results from the random motion of molecules and the affinity of their binding sites for one another. Also refers to the joining of complementary surfaces in nanomolecular interaction.

Serum albumins Thiol-containing protein. It is a lipid carrier in blood serum.

Surface plasmon resonance (SPR) It is a phenomenon that occurs when light is reflected off thin metal films. A fraction of the light energy incident at a sharply defined angle can interact with the delocalised electrons in the metal film (plasmon) thus reducing the reflected light intensity.

Thiol This is a subclass of chemical compound that contains sulphur, esp. denoting that a compound is derived from a specified compound by the replacement of an oxygen atom with a sulphur atom. More specifically a compound that contain

$$-SH \text{ or } -\overset{|}{\underset{|}{C}}-S-\overset{|}{\underset{|}{C}}-$$

are called thiol.

"Top down" method One approach for the building up of nanoscopic features. In this method, the features are written directly onto a substrate for example by electron beam and then by applying appropriate etching and deposition processes, the nanoscopic features are engraved.

Transmission electron micrograph (TEM) In TEM the magnification of the objective remains fixed while the focal length of the projector is changed to vary magnification.

ACKNOWLEDGMENTS

The authors would like to thank the Swiss National program in Nanotechnology (TOP NANO 21; 5971.2 TNS) for financial support. They would also like to thank colleagues and students for assisting in the preparation of this article. Special thanks to Dr. P. Bowen, Dr. R. Houriet, Dr. Y. Houst, and Mr. F. Juillerat for their support and encouragement.

REFERENCES

1. V. Paillard, P. Melinon, J. P. Perez, V. Dupuis, A. Perez, and B. Champagnon, *Phys. Rev. Lett.* 71, 4170 (1993).
2. P. Melinon, V. Paillard, V. Dupuis, A. Perez, P. Jensen, A. Hoareau, M. Broyer, J.-L. Vialle, M. Pellarin, B. Baguenard, and J. Lerme, *Internat. J. Mod. Phys. B* 9, 339 (1995).
3. H. Haberland M. Moseler, Y. Qiang, O. Rattunde, T. Reiners, and Y. Thurner, *Surf. Rev. Lett.* 3, 887 (1993).
4. J. Dutta, H. Hofmann, Ch. Hollenstein, and H. Hofmeister, "Nanoparticles in Solids and Solutions: Preparation, Characterization and Utilization", J. H. Fendler, Ed. Wiley-VCH, Weinheim, Germany, 1998, Chapter 8 p. 173.
5. H. Hofmann, J. Dutta, S. Scholz-Odermatt, and J.-Ch. Valmalette, *Ceramic Engineering and Science Proceedings*, 18(4B) (1997) 687.
6. S. Fafard, Z. R. Wasilewski, and M. Spanner, *Appl. Phys. Lett.* 75, 1866 (1999).
7. N. Carlsson, T. Junno, L. Montelius, M.-E. Pistol, L. Samuelson, and W. Seifert, *J. Cryst. Growth* 191, 347 (1998).
8. V. I. Klimov, D. W. McBranch, C. A. Leatherdale, and M. G. Bawendi, *Phys. Rev. B* 60, 13740 (1999).
9. F.-Y. Tsai and C. P. Lee, *J. Appl. Phys.* 84, 2624 (1998).
10. D. Gammon, E. S. Snow, B. V. Shanobrook, D. S. Katzer, and D. Park, *Science* 273, 87 (1996).
11. Z. Y. Zhong, B. Gates, Y. N. Xia, and D. Qin, *Langmuir* 16, 10369 (2000).
12. M. Zahn, *J. Nanoparticle Res.* 3, 73 (2001).
13. J. Shi, S. Gider, K. Babcock, and D. D. Awschalom, *Science* 271, 937 (1996).
14. C. Lebreton, C. Vieu, A. Pépin, M. Mejias, F. Carcenac, Y. Jin, and H. Launois, *Microelectr. Eng.* 41/42, 507 (1998).
15. A. D. Yoffe, *Adv. Phys.* 50, 1 (2001).
16. L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, *Rep. Prog. Phys.* 64, 701 (2001).
17. P. Ball, "The Self-Made Tapestry: Pattern Formation in Nature" Oxford University Press, New York, 1999.
18. A. Miller, "The Developmental Biology of Plants and Animals" (C. F. Graham and P. F. Wareing, Eds.), pp. 249–269, Oxford: Blackwell Scientific Publications, 1976.
19. S. Camazine and J.-L. Deneubourg, in "Les Insectes Sociaux" 12th Congress of the I.U.S.S.I. Paris, Sorbonne, (A. Lenoir, G. Arnold, and M. Lepage), p. 228, Université Paris Nord, Paris 1994.
20. Y. Schiffmann, *Prog. Bio. Mol. Biol.*, 68, 145 (1997).
21. G. Nicolis, G. Deweland, and J. Turner (Eds.), "Order and Fluctuations in Equilibrium and Nonequilibrium Statistical Mechanics," Wiley, New York, 1981.
22. M. C. Cross and P. C. Hohenberg, *Rev. Mod. Phys.* 65, 851 (1993).
23. P. Borckmans, G. Dewel, De A. Wit, and D. Walgraef, in "Chemical Waves and Patterns," (R. Kapral and K. Showalter, Eds.), Kluwer, Dordrecht 1994, pp. 323–363.
24. R. Vacassy, S. M. Scholz, J. Dutta, C. J. G. Plummer, R. Houriet, and H. Hofmann, *J. Amer. Cer. Soc.* 81, 2699 (1998).
25. E. Bodenschatz, W. Zimmermann, and L. Kramer, *J. Phys.*, 49 (1988) 875.
26. I. Rehberg, B. L. Winkler, M. de la Torre Juarez, S. Rasenatand, and W. Schopf, *Festkorperprobleme/Adv. Solid State Phys.* 29, 35 (1989).
27. F. H. Busse, *Rep. Prog. Phys.* 41, 1929 (1978).
28. G. I. Sivashinsky, *Physica D* 8, 243 (1983).
29. M. Q. López-Salvans, F. Saguésand, and J. Claret, "Proc. 1st Latin American School on Materials Instabilities," Valparaiso, Chile, Kluwer Publication, Dordrecht 2000.
30. H. Neuhauser, "Plastic Instabilities and the Deformation of Metals, in: Patterns, Defects and Materials Instabilities" (D. Walgraef, and N. M. Ghoniem, Eds.). Kluwer Academic Publishers, Dordrecht, 1990.
31. J. Kratochvil, *Rev. Phys. Appl.* 23, 419 (1988).
32. R. Kossowsky and S. C. Singhal, "Surface Engineering, Surface Modification of Materials," Martinus Nijhoff, Dordrecht, 1984.

33. D. Bauerle, "Laser Processing and Chemistry," Springer-Verlag, New York, 1996.
34. J. H. Evans, *Nature* 29, 403 (1971).
35. M. Seul and R. Wolfe, *Phys. Rev. A* 46, 7534 (1992).
36. M. C. Roco, S. Williams, and P. Alivisatos (Eds.), "Nanotechnology Research Directions: IWGN Workshop Report Vision for Nanotechnology Research and Development in the Next Decade," WTEC, Loyola College, Maryland, September, 1999.
37. I. Borgia, B. Brunetti, I. Mariani, A. Sgamellotti, F. Cariati, P. Fermo, M. Mellini, C. Viti, and G. Padeletti, *Appl. Surf. Sci.* 185, 206 (2002).
38. G. Decher, *Science* 277, 1232 (1997).
39. J.-M. Lehn, PNAS 4763 (2002).
40. M. Jose-Yacaman, L. Rendon, J. Arenas, and M. C. S. Puche, *Science* 273, 223 (1996).
41. http://www.liv.ac.uk/Chemistry/Links/selfassembly.html
42. E. Hu and D. Shaw, http://www.wtec.org/loyola/nano/final/ch2.pdf, p. 25
43. F. Schreiber, *Prog. Surf. Sci.* 65, 151 (2000).
44. J. A. Larson, M. Nolan, and J. C. Greer, *J. Phys. Chem.* B 106, 593 (2002).
45. J. Flink, C. J. Kiely, D. Bethell, and D. Schiffrin, *Chem Mater.* 10, 922 (1998).
46. G. Carrot, J. C. Valmalette, C. J. G. Plummer, S. M. Scholz, J. Dutta, H. Hofmann, and H. Hilborn, *Coll. Poly. Sci.* 276, 853 (1998).
47. M. Brust and C. J. Kiely, *Coll Surf.* A202, 175 (2002).
48. D. H. Napper, "Polymeric Stabilization of Colloidal Dispersions," Academic Press, London 1983.
49. J. T. G. Overbeek, "Colloidal Dispersions," Royal Society of Chemistry, London, 1981.
50. C. Bossel, J. Dutta, R. Houriet, J. Hilborn, and H. Hofmann, *Mater. Sci. Eng. A* 204, 107 (1995).
51. P. C. Hiemenz, "Principles of Colloid and Surface Chemistry," Marcel Dekker, New York, 1986.
52. D. Myers, "Surfaces, Interfaces and Colloids," VCH, Weinheim, 1991.
53. F. T. Hesselink, A. Vrij, and J. T. G. Overbeek, *J. Phys. Chem.* 75, 2094 (1971).
54. D. J. Meier, *J. Phys. Chem.* 71, 1861 (1967).
55. P. J. Flory and W. R. Krigbaum, *J. Chem. Phys.* 18, 1086 (1950).
56. F. R. Eirich, *J. Coll. Internat. Sci.* 58, 423 (1977).
57. A. J. Silberberg, *J. Chem. Phys.* 46, 1105 (1967).
58. G. Steinberg, *J. Phys. Chem.* 71, 292 (1967).
59. T. F. Tadros, *J. Polym.* 23, 683 (1991).
60. J. M. Stouffer and T. J. McCarthy, *Macromolecules* 21, 1204 (1988).
61. E. Killman, J. Eisenlauer, and M. Korn, *J. Polym. Sci. Polym. Symp.* 61, 413 (1977).
62. P.-G. de Gennes, *Macromolecules* 13, 1069 (1980).
63. R. Zerushalmi-Royen, J. Klein, and L. Fetters, *Science* 263, 793 (1994).
64. G. Reiter, *Europhys. Lett.* 33, 29 (1996).
65. R. G. Nuzzo and D. L. Allara, *J. Am. Chem. Soc.* 105, 4481 (1983).
66. T. L. Freeman, S. D. Evans, and S. D. Ulman, *Langmuir* 11, 4411 (1995).
67. R. Vacassy, L. Lemaire, J.-C Valmalette, J. Dutta, and H. Hofmann, *J. Mater. Sci. Lett.* 17, 1665 (1998).
68. M. Brust, M. Walker, D. Bethell, D. J. Schiffrin, and R. Whyman, *J. Chem. Soc., Chem. Commun.* 7, 801 (1994).
69. C. P. Collier, T. Vossmeyer, and J. R. Heath, *Ann. Rev. Phys. Chem.* 49, 371 (1998).
70. D. J. Lavrich, S. M. Wetterer, S. L. Bernasek, and G. Scoles, *J. Phys. Chem. B* 102, 3456 (1998).
71. S. S. Ghosh, P. M. Kao, A. W. McCue, and H. L. Chappelle, *Bioconjugate Chem.* 1, 71 (1990).
72. X. Yang, L. A. Wenzler, J. Qi, X. Li, and N. C. Seeman, *J. Am. Chem. Soc.* 120, 9779 (1998).
73. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature* 394, 539 (1998).
74. W.-L. Shaiu, D. D. Larson, J. Vesenka, and E. Henderson, *Nucleic Acids Res.* 21, 99 (1993).
75. C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, *Nature* 382, 607 (1996).
76. S.-J. Park, A. A. Lazarides, C. A. Mirkin, P. W. Brazis, C. R. Kannewurf, and R. L. Letsinger, *Angew. Chem.* 112, 4003 (2000).
77. R. L. Letsinger, R. Elghanian, G. Viswanadham, and C. A. Mirkin, *Bioconjugate Chem.* 11, 289 (2000).
78. H. Mattoussi, J. M. Mauro, E. R. Goldman, G. P. Anderson, V. C. Sundar, F. V. Mikulec, and M. G. Bawendi, *J. Am. Chem. Soc.* 122, 12142 (2000).
79. R. S. Ingram, M. J. Hostetler, and R. W. Murray, *J. Am. Chem. Soc.* 119, 9175 (1997).
80. G. Schmid, "Clusters and Colloids," VCH, Weinheim, 1994.
81. M. Kerker, "The Scattering of Light and Other Electromagnetic Radiation," Academic Press, New York, 1969.
82. U. Kreibig and P. Z. Zacharias, *Physics* 231, 128 (1970).
83. R. H. Doremus and P. J. Rao, *Mater. Res.* 11, 2834 (1996).
84. G. L. Hornyak, C. J. Patrissi, C. R. Martin, J.-C Valmalette, J. Dutta, and H. Hofmann, *Nano. Mater.* 9, 575 (1997).
85. T. Schalkhammer, *Monatshefte für Chemie* 129, 1067 (1998).
86. T. Okamoto and I. Yamaguchi, *Opt. Lett.* 25, 372 (2000).
87. J. J. Storhoff, R. Elghanian, R. C. Mucic, C. A. Mirkin, and R. L. Letsinger, *J. Am. Chem. Soc.* 120, 1959 (1998).
88. Y. Dirix, C. Bastiaansen, W. Caseri, and P. Smith, *Adv. Mater.* 11, 223 (1999).
89. A. H. Lu, G. H. Lu, A. M. Kessinger, and C. A. Foss Jr., *J. Phys. Chem. B* 101, 9139 (1997).
90. L. G. Olson, Y. Lo, T. P. Beebe, Jr., and J. M. Harris, *Anal. Chem.* 73, 4268 (2001).
91. A. N. Shipway, E. Katz, and I. Willner, *Chemphyschem.* 1, 18 (2000).
92. A. Ulman, *Chem. Rev.* 96, 1533 (1996).
93. X. Michalet, F. Pinaud, T. D. Lacoste, M. Dahan, M. P. Bruchez, A. P. Alivisatos, and S. Weiss, *Single Mol.* 2, 261 (2001).
94. A. L. Elfros and M. Rosen, *Ann. Rev. Mater. Sci.* 30, 475 (2000).
95. R. Vacassy, S. M. Scholz, J. Dutta, C. J. G. Plummer, R. Houriet, and H. Hofmann, *J. Am. Cer. Soc.* 81, 2699 (1998).
96. M. L. Steigerwald, A. P. Alivisatos, J. M. Gibson, T. D. Harris, R. Kortan, A. Muller, A. M. Thayer, T. M. Duncan, D. C. Douglass, and L. E. Brus, *J. Am. Chem. Soc.* 110, 3046 (1988).
97. N. Herron, Y. Wangand, and H. Eckert, *J. Am. Chem. Soc.* 112, 1322 (1990).
98. S. M. Scholz, R. Vacassy, L. Lemaire, J. Dutta, and H. Hofmann, App. Organo. Chem. 12, 327 (1998).
99. K. M. Choi and K. J. Shea, *J. Phys. Chem.* 98, 3207 (1994).
100. M. P. Pileni, L. Motte, and C. Petit, *Chem. Mater.* 4, 338 (1992).
101. X. K. Zhao, S. Baral, R. Rolandi, and J. H. Fendler, *J. Am. Chem. Soc.* 110, 1012 (1988).
102. R. S. Urquhart, D. N. Furlong, T. Gengenbach, N. J. Geddes, and F. Grieser, *Langmuir* 11, 1127 (1995).
103. G. L. Hornyak, C. J. Patrissi, C. R. Martin, J. C. Valmalette, L. Lemaire, J. Dutta, and H. Hofmann, *Nano. Mater.* 9, 571 (1997).
104. K. K. W. Wong and S. Mann, *Adv. Mater.* 8, 928 (1996).
105. C. T. Dameron, R. N. Reese, R. K. Mehra, A. R. Kortan, P. J. Carroll, M. L. Steigerwald, L. E. Brus, and D. R. Winge, *Nature* 338, 596 (1989).
106. R. N. Reese, C. A. White, and D. R. Winge, *Plant Physiol.* 98, 225 (1992).
107. C. T. Dameron and D. R. Winge, *Inorg. Chem.* 29, 1343 (1990).
108. I. Soten and G. A. Ozin, *Curr. Opin. Coll. Interf. Sci.* 4, 325 (1999).
109. N. Malikova, I. Pastoriza-Santos, M. Schierhorn, N. A. Kotov, and L. M. Liz-Marzan, *Langmuir* 18, 3694 (2002).
110. P. C. Andersen and K. L. Rowlen, *Appl. Spectro.* 56, 124A (2002).

111. X. M. Lin, H. M. Jaeger, C. M. Sorensen, and K. J. Klabunde, *J. Phys. Chem. B* 105, 3353 (2001).
112. C. P. Collier, T. Vossmeyer, and J. R. Heath, *Ann. Rev. Phys. Chem.* 49, 371 (1998).
113. C. B. Murray, C. R. Kagan, and M. G. Bawendi, *Science* 270, 1335 (1995).
114. R. P. Andres, J. D. Bielefeld, J. I. Henderson, D. B. Janes, V. R. Kolagunta, C. P. Kubiak, W. J. Mahoney, and R. G. Osifchin, *Science* 273, 1690 (1996).
115. R. L. Whetten, J. T. Khoury, M. M. Alvarez, S. Murthy, I. Vezmar, Z. L. Wang, P. W. Stephens, C. L. Cleveland, W. D. Luedtke, and U. Landman, *Adv. Mater.* 8, 428 (1996).
116. V. M. Shalaev and M. Moskovits, "Nanostructured Materials: Clusters, Composites and Thin Films," ACS Symposium Series,Vol. 679, Washington DC, American Chemical Society, 1997.
117. C. S. Weisbecker, M. V. Merritt, and G. M. Whitesides, *Langmuir* 12, 3763 (1996).
118. S. Biggs and P. Mulvaney, *J. Chem. Phys.* 100, 8501 (1994).
119. V. Kane and P. Mulvaney, *Langmuir* 14, 3303 (1998).
120. J. Israelachvili, "Intermolecular and Surface Forces," 2nd ed., Academic Press, New York 1992, Chapter 10.
121. T. Fujimura, T. Itoh, A. Imada, R. Shimada, T. Koda, N. Chiba, H. Muramatsu, H. Miyazaki, and K. Ohtaka, *J. Lumin.* 87–89, 954 (2000).
122. A. Rogach, A. Susha, F. Caruso, G. Sukhorukov, A. Kornowski, S. Kershaw, H. Möhwald, A. Eychmuller, and H. Weller, *Adv. Mater.* 12, 333 (2000).
123. R. Shimada, Y. Komori, T. Koda, T. Fujimura, T. Itoh, and K. Ohtaka, *Mol. Cryst. Liq. Cryst. Sci. Technol. Sect. A* 349, 5 (2000).
124. G. Subramania, K. Constant, R. Biswas, M. M. Sigalas, and K. M. Ho, *J. Lightwave Technol.* 17, 1970 (1999).
125. Y. Xia, B. Gates, and S. H. Park, *J. Lightwave Technol.* 17, 1956 (1999).
126. A. K. Boal, F. Ilhan, J. E. DeRouchey, T. Thurn-Albrecht, T. P. Russell, and V. M. Rotello, *Nature* 404, 746 (2000).
127. F. Auer, M. Scotti, A. Ulman, R. Jordan, B. Sellergren, J. Garno, and G.-Y. Liu, *Langmuir*, 16, 7554 (2000).
128. R. Meallet-Renault, P. Denjean, and R. B. Pansu, *Sens. Actuators B* 59, 108 (1999).
129. F. Burmeister, W. Badowsky, T. Braun, S. Wieprich, J. Boneberg, and P. Leiderer, *Appl. Surf. Sci.* 144–145, 461 (1999).
130. C. P. Collier, T. Vossmeyer, and J. R. Heath, *Ann. Rev. Phys. Chem.* 49, 371 (1998).
131. J. Boneberg, F. Burmeister, C. Schae.e, P. Leiderer, D. Reim, A. Fery, and S. Herminghaus, *Langmuir* 13, 7080 (1997).
132. J. C. Halteen and R. P. Van Duyne, *J. Vac. Sci. Technol. A* 13, 1553 (1995).
133. M. H. Charles, M. T. Charreyre, T. Delair, A. Elaissari, and C. Pichot, *S. T. P. Pharm. Sci.*, 11, 251 (2001).
134. O. D. Velev and E. W. Kaler, *Langmuir* 15, 3693 (1999).
135. G. Bauer, F. Pittner, and T. Schalkhammer, *Mikrochim. Acta* 131, 107 (1999).
136. H. Takei, in "Microfluidic Devices and Systems" (A. B. Frazier and C. H. Ahn, Eds.) *SPIE-Internat. Soc. Opt. Eng.*, Santa Clara, CA, 3515, p. 278 (1998).
137. C. G. J. Koopal and R. J. M. Nolte, *Enzyme Microb. Technol.* 16, 402 (1994).
138. F. Ganachaud, A. Elaiessari, C. Pichot, A. Laayoun, and P. Cros, *Langmuir* 13, 701 (1997).
139. S.-C. Huang, H. Swerdlow, and K. D. Caldwell, *Anal. Biochem.* 222, 441 (1994).
140. W. Härtl, C. Beck, and R. Hempelmann, *J. Chem. Phys.* 110, 7070 (1998).
141. Y. Solomentsev, M. Böhmer, and J. L. Anderson, *Langmuir* 13, 6058 (1997).
142. D. Rudhardt, C. Bechinger, and P. Leiderer, *Polymers* 112, 163 (1999).
143. C. Bechinger, M. Brunner, and P. Leiderer, *Phys. Rev. Lett.* 86, 930 (2001).
144. E. Kim, Y. Xia, and G. M. Whitesides, *Nature* 376, 581 (1995).
145. K. M. Chen, X. Jiang, L. C. Kimerling, and P. T. Hammond, *Langmuir* 16, 7825 (2000).
146. J. Aizenberg, P. V. Braun, and P. Wiltzius, *Phys. Rev. Lett.* 84, 2997 (2000).
147. P. A. Kralchevsky and N. D. Denkov, *Curr. Opin. Coll. Interf. Sci.* 6, 383 (2001).
148. C. D. Dushkin, G. S. Lazarov, S. N. Kotsev, H. Yoshimura, and K. Nagayama, *Coll. Polym. Sci.* 227, 914 (1999).
149. K.-H. Lin, J. C. Crocker, V. Prasad, A. Schofield, D. A. Weitz, T. C. Lubensky, and A. G. Yodh, *Phys. Rev. Lett.* 85, 1770 (2000).
150. M. Heniand and H. Löwen, *Phys. Rev. Lett.* 85, 3668 (2000).
151. P. V. Braun, R. W. Zehner, C. A. White, M. K. Weldon, C. Kloc, S. S. Patel, and P. Wiltzius, *Adv. Mater.* 13, 721 (2001).
152. J. W. G. Tyrrell and P. Attard, *Phys. Rev. Lett.* 87, 176104 (2001).
153. Y. Lu, Y. Yin, and Y. Xia, *Adv. Mater.* 13, 34 (2001).
154. Y. Yin, Y. Lu, and Y. Xia, *J. Am. Chem. Soc.* 123, 771 (2000).
155. O. D. Velev, K. Furusawa, and K. Nagayama, *Langmuir* 12, 2374 (1996).
156. W. T. S. Huck, J. Tien, and G. M. Whitesides, *J. Am. Chem. Soc.* 120, 8267 (1998).
157. O. D. Velev, A. M. Lenhoff, and E. Kaler, *Science* 87, 2240 (2000).
158. G. A. Ozin and M. Y. Yang, *Adv. Func. Mater.* 11, 95 (2001).
159. E. Kim, Y. Xia, and G. M. Whitesides, *Adv. Mater.* 8, 245 (1996).
160. A. Stein and R. C. Schroeden, *Curr. Opin. Solid State Mater. Sci.* 5, 553 (2001).
161. O. D. Velev, T. A. Jede, R. F. Lobo, and A. M. Lenhoff, *Nature* 389, 447 (1997).
162. O. D. Velev, T. A. Jede, R. F. Lobo, and A. M. Lenhoff, *Chem. Mater.* 10, 3597 (1998).
163. S. H. Park and Y. Xia, *Adv. Mater.* 10, 1045 (1998).
164. B. Gates, Y. Yin, and Y. Xia, *Chem. Mater.* 11, 2827 (1999).
165. B. T. Holland, C. F. Blanford, and A. Stein, *Science* 281, 538 (1998).
166. B. T. Holland, C. F. Blanford, T. Do, and A. Stein, *Chem. Mater.* 11, 795 (1999).
167. S. A. Johnson, P. J. Ollivier, and T. E. Mallouk, *Science* 283, 963 (1999).
168. J. S. Yin and Z. L. Wang, *Adv. Mater.* 11, 469 (1999).
169. H. Yan, C. F. Blanford, B. T. Holland, M. Parent, W. H. Smyrl, and A. Stein, *Adv. Mater.* 11, 1003 (1999).
170. Y. A. Vlasov, N. Yao, and D. J. Norris, *Adv. Mater.* 11, 165 (1999).
171. G. Subramanian, V. N. Manoharan, J. D. Thorne, and D. J. Pine, *Adv. Mater.* 11, 1261 (1999).
172. O. D. Velev, P. M. Tessier, A. M. Lenhoff, and E. W. Kaler, *Nature* 401, 548 (1999).
173. P. Jiang, J. Cizeron, J. F. Bertone, and V. L. Colvin, *J. Am. Chem. Soc.* 121, 7957 (1999).
174. P. V. Braun and P. Wiltzius, *Nature* 402, 603 (1999).
175. A. A. Zakhidov, R. H. Baughman, Z. Iqbal, C. Cui, I. Khayrullin, S. O. Dantas, J. Marti, and V. G. Ralchenko, *Science* 282, 897 (1998).
176. P. Yang, T. Deng, D. Zhao, P. Feng, D. Pine, B. F. Chmelka, G. M. Whitesides, and G. D. Stucky, *Science* 282, 2244 (1998).
177. F. Meseguer, A. Blanco, H. Miguez, F. Garcia-Santamaria, M. Ibisate, and C. Lopez, *Coll. Surf. A*: 202 (2002) 281.
178. Y. Yin, Y. Lu, B. Gates, and Y. Xia, *J. Am. Chem. Soc.* 123, 8718 (2001).
179. A. Badia, S. Singh, L. Demers, L. Cuccia, G. R. Brown, and R. B. Lennox, *Eur. J. Chem.* 2, 359 (1996).
180. J. Schmitt, G. Decher,W. J. Dressick, S. L. Brandow, R. E. Geer, R. Shashidhar, and J. M. Calvert, *Adv. Mater.* 9, 61 (1997).
181. T. Hanaoka, H. P. Kormann, M. Kröll, T. Sawitowski, and G. Schmid, *Eur. J. Inorg. Chem.* 807 (1998).
182. M. E. Davis, *Nature* 417, 813 (2002).

183. A. Ulman, “An Introduction to Ultrathin Organic Films: From Langmuir–Blodgett to Self-Assembly,” Academic Press, San Diego, CA, 1991.
184. P. Ganguly, D. V. Paranjape, K. R. Patil, M. Sastry, and F. Rondelez, *Langmuir* 13, 5440 (1997).
185. M. Clemente-Leon, C. Mingotaud, B. Agricole, C. J. Gomez-Garcia, E. Coronado, and P. Delhaes, *Angew. Chem. Internat. Ed. Engl.* 36, 1114 (1997).
186. A. Riccio, M. Lanzi, F. Antolini, C. De Nitti, C. Tavani, and C. Nicolini, *Langmuir* 12, 1545 (1996).
187. U. Raedler, C. Heiz, P. Luigi, and R. Tampe, *Langmuir* 14, 6620 (1998).
188. P. Pieranski, *Phys. Rev. Lett.* 45, 569 (1980).
189. A. J. Hurd and D. W. Schaefer, *Phys. Rev. Lett.* 54, 1043 (1985).
190. A. J. Armstrong, R. C. Mockler, and W. J. O’Sullivan, *J. Phys. Condens. Matt.* 1, 1707 (1989).
191. J. C. Earnshaw, *J. Phys. D* 19, 1863 (1986).
192. A. Ulman, *Adv. Mater.* 3, 298 (1991).
193. J. H. Fendler, *Chem. Mater.* 8, 1616 (1996).
194. K. S. Mayya, V. Patil, M. Kumar, and M. Shastry, *Thin Solid Films* 312, 308 (1998).
195. H. Schuller, *Kolloid-Z*. 216–217, 380 (1967).
196. E. Sheppard and N. Tcheurekdjian, *J. Coll. Interf. Sci.* 28, 48 (1968).
197. J. Kumaki, *Macromolecules*, 19, 2258 (1996).
198. K.-U. Fulda, D. Pieche, B. Tieke, and H. Yarmohammadipour, *Prog. Coll. Polym. Sci.* 101, 178 (1996).
199. M. Bardosova, P. Hodge, L. Pach, V. Smatko, and R. H Tredgold, “8th European Conference on Organised Films, Book of Abstracts,” 2001.
200. C. A. Murray and D. H. V. Winkle, *Phys. Rev. Lett.* 58, 1200 (1987).
201. A. T. Skjeltorp and P. Meakin, *Nature* 335, 424 (1988).
202. N. D. Denkov, O. D. Velev, P. A. Kralchevsky, I. B. Ivanov, H. Yoshimura, and K. Nagayama, *Nature* 361, 26 (1993).
203. G. Picard, *Langmuir* 14, 3710 (1998).
204. M. Kondo, K. Shinozaki, L. Bergstrom, and N. Mizutani, *Langmuir* 11, 394 (1995).
205. H. W. Deckman, J. H. Dunsmuir, and S. M. Gruner, *J. Vac. Sci. Technol.* B7, 1832 (1989).
206. F. Lenzmann, K. Li, A. H. Kitai, and H. D. H. Stover, *Chem. Mater.* 6, 156 (1994).
207. K. U. Fulda and B. Tieke, *Adv. Mater.* 6, 288 (1994).
208. F. Burmeister, C. Schafle, T. Matthes, M. Bohmisch, J. Boneberg, and P. Leiderer, *Langmuir* 13, 2983 (1997).
209. N. D. Denkov, O. D. Velev, P. A. Kralchevsky, I. B. Ivanov, H. Yoshimura, and K. Nagayama, *Langmuir* 8, 3183 (1992).
210. A. S. Dimitrov and K. Nagayama, *Langmuir* 12, 1303 (1996).
211. O. D. Velev, N. D. Denkov, V. N. Paunov, P. A. Kralchevsky, and K. Nagayama, *Langmuir* 9, 3702 (1993).
212. S. Rakers, L. F. Chi, and H. Fuchs, *Langmuir* 13, 7121 (1997).
213. A. S. Dimitrov, T. Miwa, and K. Nagayama, *Langmuir* 15, 5257 (1999).
214. A. S. Dimitrov, C. D. Dushkin, H. Yoshimura, and K. Nagayama, *Langmuir* 10, 432 (1994).
215. G. S. Lazarov, N. D. Denkov, O. D. Velev, P. A. Kralchevsky, and K. Nagayama, *J. Chem. Soc. Faraday Trans.* 90, 2077 (1994).
216. B. Kim, S. L. Tripp, and A. Wei, Materials Research Society Symposium Proceedings, Purdue University, 2002, 676 (Synthesis, Functional Properties and Applications of Nanostructures), Y6.1.1–Y6.1.7.
217. L. M. Goldenberg, J. Wagner, J. Stumpe, B.-R. Paulke, and E. Görnitz, *Langmuir* 18, 5627 (2002).
218. R. Aveyard, J. H. Clint, D. Nees, and V. N. Paunov, *Langmuir* 16, 8820 (2000).
219. D. H. Everett, “Basic Principles of Colloid Science,” Royal Society of Chemistry, London 1988.
220. W. B. Russel, D. A. Saville, and W. R. Schowalter, “Colloidal Dispersions,” Cambridge University Press, Cambridge, 1989.
221. R. J. Hunter, “Introduction to Modern Colloid Science,” Oxford University Press, Oxford, 1993.
222. A. K. Arora and B. V. R. Tata (Eds.), “Ordering and Phase Transitions in Colloidal Systems,” VCH, Weinheim, 1996.
223. K. E. Davis, W. B. Russel, and W. J. Glantschnig, *Science* 245, 507 (1989).
224. P. N. Pusey and W. van Megen, *Nature* 320, 340 (1986).
225. P. J. Darragh, A. J. Gaskin, and J. V. Sanders, *Aust. Gemmol.*, November, 109 (1977).
226. T. C. Simonton, R. Roy, S. Komarneni, and E. Breval, *J. Mater. Res.* 1, 667 (1986).
227. A. P. Philipse, *J. Mater. Sci. Lett.* 8, 1371 (1989).
228. V. N. Bogomolov, S. V. Gaponenko, I. N. Germanenko, A. M. Kapitonov, E. P. Petrov, N. V. Gaponenko, A. V. Prokofiev, A. N. Ponyavina, N. I. Silvanovich, and S. M. Samoilovich, *Phys. Rev.* E55, 7619 (1997).
229. J. V. Sanders, *Nature* 204, 1151 (1964).
230. H. Miguez, F. Meseguer, C. Lopez, A. Mifsud, J. S. Moya, and L. Vazquez, *Langmuir* 13, 6009 (1997).
231. L. V. Woodcock, *Nature* 388, 235 (1997).
232. A. van Blaaderen, R. Ruel, and P. Wiltzius, *Nature* 385, 321 (1997).
233. P. Jiang, J. F. Bertone, K. S. Hwang, and V. L. Colvin, *Chem. Mater.* 11, 2132 (1999).
234. H. Kuhn, D. Möbius, and H. Bücher, in “Physical Methods of Chemistry” (A. Weisenberger and B. Rossiter Eds.) Vol. 1 (3b), 577. Wiley, New York, 1972.
235. M. Gleiche, L. F. Chi, and H. Fuchs, *Nature* 403, 173 (2000).
236. M. Himmelhaus and H. Takei, *Phys. Chem. Chem. Phys.* 4, 496 (2002).
237. U. Jonas, A. del Campo, C. Krüger, G. Glasser, and D. Boos, *PNAS* 99, 5034 (2002).
238. S. Friebel, J. Aizenberg, S. Abad, and P. Wiltzius, *Appl. Phys. Lett.* 77, 2406 (2000).
239. M. Nakagawa, S.-K. Oh, and K. Ichimura, *Adv. Mater.* 12, 403 (2000).
240. G. S. Lazarov, N. D. Denkov, O. D. Velev, P. A. Kralchevsky, and K. Nagayama, *J. Chem. Soc. Faraday Trans.* 90, 2077 (1994).
241. M. M. Burns, J. M. Fournier, and J. A. Golovchenko, *Science* 249, 749 (1990).
242. W. Hu, H. Li, B. Chang, J. Yang, Z. Li, J. Xu, and D. Zhang, *Opt. Lett.* 20, 964 (1995).
243. H. Misawa, K. Sasaki, M. Koshioka, N. Kitamura, and H. Masuhara, *Appl. Phys. Lett.* 60, 310 (1992).
244. C. Mio and M. D. W. Marr, *Langmuir* **15**, 8565 (1999).
245. A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu, *Opt. Lett.* 11, 288 (1986).
246. M. M. Burns, J. M. Fournier, and J. A. Golovchenko, *Science* 249, 749 (1990).
247. S. P. Smith, S. R. Bhalotra, A. L. Brody, B. L. Brown, K. Boyda, and M. Prentiss, *Am. J. Phys.* 67, 26 (1999).
248. A. Ashkin, J. M. Dziedzic, and T. Yamane, *Nature* 330, 769 (1987).
249. A. Ashkin and J. M. Dziedzic, *Science* 235, 1517 (1987).
250. M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, *Biophys. J.* 72, 1335 (1997).
251. S. Sato and H. Inaba, *Opt. Quantum Electron.* 28, 1 (1996).
252. R. C. Hayward, D. A. Saville, and I. A. Aksay, *Nature* 404, 56 (2000).
253. N. Sakai, A. Fujishima, T. Watanabe, and K. Hashimoto, *J. Phys. Chem.* B105, 3023 (2001).
254. J. Priestley, “The History and Present State of Electricity with Original Experiments,” 2nd ed, J. Dodsley, London 1769.
255. H. A. Pohl, “Dielectrophoresis,” Cambridge University Press, Cambridge, UK, 1978.
256. W. M. Winslow, *J. Appl. Phys.* 20, 1138 (1949).
257. M. Parthasarathy and D. Klingenberg, *Mater. Sci. Eng.* R.17, 57 (1996).
258. U. Dassanayake, S. Fraden, and A. van Blaaderen, *J. Chem. Phys.* 112, 3851 (2000).

259. E. Kumacheva, R. K. Golding, M. Allard, and E. H. Sargent, *Adv. Mater.* 14, 221 (2002).
260. J. S. Crane and H. A. Pohl, *Electrostatics* 5, 11 (1978).
261. Y. Huang and R. Pethig, *Meas. Sci. Technol.* 2, 1142 (1991).
262. T. L. Mahaworashilpa, H. G. L. Coster, and E. P. George, *Biophys. Acta* 1193, 118 (1994).
263. T. Schnelle, T. Müller, G. Gradl, S. G. Shirley, and G. Fuhr, *Electrophoresis* 21, 66 (2000).
264. T. B. Jones, "Electromechanics of Particles," Cambridge University Press, New York, 1995.
265. A. Docoslis, N. Kalogerakis, L. A. Behie, and K. V. I. S. Kaler, *Biotechnol. Bioeng.* 54, 239 (1997).
266. M. P. Hughes, M. F. Flynn, and H. Morgan, *Electrostatics* 163, 81 (1999).
267. N. G. Green and H. Morgan, *J. Phys. Chem.* B103, 41 (1999).
268. A. Docoslis and P. Alexandridis, "Proc. 2001 Annual Meeting American Electrophoresis Society," AIChE Pub. No. 148, 91 (2001).
269. C. Reichle, T. Schnelle, T. Müller, T. Leya, and G. Fuhr, Biochim. *Biophys. Acta* 1459, 218 (2000).
270. T. Yamamoto, O. Kurosawa, H. Kabata, N. Shimamoto, and M. Washizu, *IEEE Trans. Ind. Appl.* 36, 1010 (2000).
271. H. Morgan, M. P. Hughes, and N. G. Green, *Biophys. J.* 77, 516 (1999).
272. O. D. Velev and E. W. Kaler, *Langmuir* 15, 3693 (1999).
273. K. D. Hermanson, S. O. Lumsdon, J. P. Williams, E. W. Kaler, and O. D. Velev, *Science* 294, 1082 (2001).
274. P. A. Smith, C. D. Nordquist, T. N. Jackson, T. S. Mayer, B. R. Martin, J. Mbindyoand, and T. E. Mallouk, *Appl. Phys. Lett.* 77, 1399 (2000).
275. G. Fuhr, C. Rechle, T. Müller, K. Kahlke, K. Schutze, and M. Stuke, *Appl. Phys. A Mater. Sci. Process.* 69, 611 (1999).
276. A. Docoslis and P. Alexandris, *Electrophoresis* 23, 2174 (2002).
277. M. Trau, D. A. Saville, and I. A. Aksay, *Langmuir* 13, 6375 (1997).
278. M. Trau, D. A. Saville, and I. A. Aksay, *Science* 272, 706 (1996).
279. M. Bohmer, *Langmuir* 12, 5747 (1996).
280. M. Giersig and P. Mulvaney, *J. Phys. Chem.* 97, 6334 (1993).
281. M. Giersig and P. Mulvaney, *Langmuir* 9, 3408 (1993).
282. Y. Solomentsev, M. Böhmer, and J. L. Anderson, *Langmuir* 13, 6058 (1997).
283. R. E. Rosensweig, "Ferrohydrodynamics," Dover Publications, Mineola, 1997.
284. M. Golosovsky, Y. Saado, and D. Davidov, *Appl. Phys. Lett.* 75, 4168 (1999).
285. E. L. Bizdoaca, M. Spasova, M. Farle, M. Hilgendor, and F. Caruso, *J. Magnetism Magnetic Mater.* 240, 44 (2002).
286. J. Lin, *J. Solid State Chem.* 159, 26 (2001).
287. E. Baer, A. Hiltner, and H. D. Keith, *Science* 235, 1015 (1987).
288. T. Shimizu, *Macromol. Rapid Comm.* 23, 311 (2002).
289. J.-M. Lehn, "Supramolecular Chemistry Concepts and Perspectives," VCH, Weinheim, 1995.
290. T. Moore and A. Steitz, *Science* 289, 905 (2000).
291. E. Westhof and N. Leontis, *Angew. Chem.* 112, 1651 (2000).
292. D. M. J. Lilley, *Chem. Bio. Chem.* 2, 31 (2001).
293. M. L. DePamphilis and J. Adler, *J. Bacteriol.* 105, 376 (1971).
294. R. M. Macnab, *Ann. Rev. Genet.* 26, 131 (1992).
295. N. R. Francis, G. E. Sosinsky, D. Thomas, and D. J. DeRosier, *J. Mol. Biol.* 235, 1261 (1994).
296. A. Houdusse, V. N. Kalabokis, D. Himmel, A. G. Szent-Gyorgyi, and C. Cohen, *Cell* 97, 459 (1999).
297. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. Bruchez, and P. G. Schultz, *Nature* 382, 609 (1996).
298. W. Shenton, S. A. Davis, and S. Mann, *Adv. Mater.* 11, 449 (1999).
299. S. Connolly and D. Fitzmaurice, *Adv. Mater.* 11, 1202 (1999).
300. H. A. Klok, *Angew. Chem. Int. Ed.* 41, 1509 (2002).
301. D. Philip and J. F. Stoddart, *Angew. Chem. Internat. Ed. Engl.* 35, 1154 (1996).
302. R. Elghanian, J. J. Storhoff, R. C. Mucic, R. L. Letsinger, and C. A. Mirkin, *Science* 277, 1078 (1997).
303. H. Weller, *Angew. Chem.* 108, 1159 (1996).
304. D. Bethell and D. J. Schiffrin, *Nature* 382, 581 (1996).
305. A. P. Alivisatos, *Science* 271, 933 (1996).
306. R. F. Service, *Science* 271, 920 (1996).
307. T. Takahagi, S. Huang, G. Tsutsui, H. Sakaueand, and S. Shingubara, *Mater. Res. Soc. Symp. Proc.* 707, 87 (2002).
308. K. K. W. Wong and S. Mann, *Curr. Opin. Coll. Inter. Sci.* 3, 63 (1998).
309. T. Furuno, H. Sasabe, and K. M. Ulmer, *Thin Solid Films* 180, 23 (1989).
310. H. Yoshimura, M. Matsumoto, S. Endo, and K. Nagayama, *Ultra-Microscopy* 32, 265 (1990).
311. H. Yoshimura, T. Scheybani, W. Baumeister, and K. Nagayama, *Langmuir* 10, 3290 (1994).
312. S. Takeda, H. Yoshimura, S. Endo, T. Takahashi, and K. Nagayama, *Proteins* 23, 548 (1995).
313. T. Scheybani, H. Yoshimura, W. Baumeister, and K. Nagayama, *Langmuir* 12, 431 (1996).
314. I. Yamashita, *Thin Solid Films* 393, 12 (2001).
315. K. M. Towe, in "Origin, Evolution and Modern Aspects of Biomineralization in Plants and Animals," (R. E. Crick, Ed.), Plenum Press, New York, p. 265, 1990.
316. F. C. Meldrum, V. J. Wade, D. L. Nimmo, B. R. Heywood, and S. Mann, *Nature* 349, 684 (1991).
317. F. C. Meldrum, B. R. Heywood and S. Mann, *Science* 257, 522 (1 1992).
318. F. C. Meldrum, T. Douglas, S. Lei, P. Arosio, and S. Mann, *J. Inorg. Biochem.* 58, 59 (1995).
319. T. Douglas, D. P. E. Dickson, S. Betteridge, J. Charnock, C. D. Garner, and S. Mann, *Science* 269, 54 (1995).
320. S. R. Sczekan and J. G. Joshi, *Biochim. Biophys. Acta* 990, 8 (1998).
321. K. K. W. Wong and S. Mann, *Adv. Mater.* 8, 928 (1996).
322. A. McMillian, C. D. PMVOLA, J. Howard, S. L. Chaw, N. J. Zaluzec, and J. D. Trent, *Nature Mater.* 1, 247 (2002).
323. E. Sackmann, *Science* 271, 43 (1996).
324. C. A. Steinem, A. Janshoff, W. P. Ulrich, M. Sieber, and H. J. Galla, *Biochim. Biophys. Acta-Biomembr.* 1279, 169 (1996).
325. G. Puu and I. Gustafson, *Biochim. Biophys. Acta* 1327, 149 (1997).
326. A. L. Plant, *Langmuir* 15, 5128 (1999).
327. E. Sackmann and M. Tanaka, *Trends Biotechnol.* 18, 58 (2000).
328. S. G. Boxer, *Curr. Opin. Chem. Biol.* 4, 704 (2000).
329. R. G. Horn, *Biochim. Biophys. Acta* 778, 224 (1984).
330. L. K. Tamm and H. M. McConnell, *Biophys. J.* 47, 105 (1986).
331. C. Duschl, M. Liley, H. Lang, A. Ghandi, S. M. Zakeeruddin, H. Stahlberg, A. Nemetz, W. Knoll, and H. Vogel, *Mater. Sci. Eng.* C4, 7 (1996).
332. A. N. Parikh, J. D. Beers, A. P. Shreve, and B. I. Swanson, *Langmuir* 15, 5369 (1999).
333. X. F. Walboomers, W. Monaghan, A. S. G. Curtis, and J. A. Jansen, *J. Biomed. Mater. Res.* 46, 212 (1999).
334. S. Britland, C. Perridge, M. Denyer, H. Morgan, A. Curtis, and C. Wilkinson, *Exp. Biol. Online* 1, 2 (1996).
335. T. Kondo, M. Nojiri, Y. Hishikawa, E. Togawa, D. Romanovicz, and R. M. Brown, Jr., *PNAS*, 99, 14008 (2002).
336. S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, A. M. Belcher, *Nature*, 405, 665 (2000).
337. C. C. Lin, Y. C. Yeh, C. Y. Yang, C. L. Chen, G. F. Chen, C. C. Chen, and Y. C. Wu, *J. Am. Chem. Soc.* 124, 3508 (2002).
338. J. F. V. Vincent, *Materials Today*, Dec. 2002, p. 28.
339. G. Chen, *Internat. J. Therm. Sci.* 39, 471 (2000).
340. L. Braginsky, N. Lukzen, V. Shklover, and H. Hofmann, *Phys. Rev. B* 66, 134203 (2002).
341. L. B. Kish, *Phys. Lett.* A 305, 144 (2002).
342. V. V. Zhirnov and D. J. C. Herr, *Computer*, January 2001, p. 34.

Self-Organized Nanostructure Formation on Surfaces

Andrew T. S. Wee
National University of Singapore, Singapore

CONTENTS

1. Introduction
2. Self-Organized Nanostructures
3. Step and Facet-Templated Nanostructures
4. Molecular Nanostructures
5. Ion-Beam Induced Nanostructures
6. Photon-Induced Nanostructures
7. Summary
Glossary
References

1. INTRODUCTION

There is much interest in recent years in the methods of creating nanometer-scale structures on surfaces to form one- (1D) and two-dimensional (2D) ordered structures. When adatoms are nucleated into ordered arrays of sufficiently small dimensions, quantum mechanical effects determine their properties, for example, in quantum dots or quantum wires. Such formation of ordered nanostructures on surfaces is indeed a challenge in both fundamental and applied science research. Fabrication of uniform-size clusters in the 1 to 2-nm size regime is challenging because fluctuation in size of a few atoms could substantially alter their electronic properties. Such a length scale is important for quantum applications since the Fermi wavelength for most metals is around 1 nm.

The importance of the fundamental properties of nanostructures on surfaces has been highlighted in the National Science and Technology Council report on nanotechnology [1]. Many key questions relate to the structure or arrangement of atoms in a nanostructure. The relative stability of different structural phases is altered in the nanometer regime, affected by both kinetic and thermodynamic factors. Variations may arise due to factors such as surface energies, defects, and electronic quantum-size effects. There is, therefore, a need to map out the kinetics and thermodynamics of phase transformations in nanostructures. For a complete understanding of the physical properties of nanostructures, the structure of the surface must also be determined. Surfaces of nanostructures are likely to vary significantly from the well-characterized bulk structures. Hence, the ability to measure and systematically control the surface will continue to be an important field of research over the next decade.

The direct study of surface nanostructures at the atomic scale became possible with the invention of the scanning tunneling microscope (STM) by Binnig and Rohrer [2]. Recent advances in scanning tunneling microscopy and spectroscopy of nanostructures have been reviewed by Schneider [3], and these techniques provide new opportunities for investigating locally the geometric and electronic properties of nanostructures on surfaces. Ushioda [4] presents an overview of the experimental method and physical principles of STM light emission spectroscopy. By using this new spectroscopic technique, one can obtain the optical emission spectra of individual surface nanostructures that are imaged by the STM. An overview of the status and prospects of STM is presented by Binnig and Rohrer [5]. They suggest the potential applications of the scanning probe microscopies to a relatively broad range of biological, chemical, and technological areas.

The production of semiconductor nanostructures has been under intensive investigation due to its potential for applications in optoelectronic and quantum devices. Techniques such as electron-beam lithography [6] are widely used, but due to the low throughput of such techniques, alternative methods based on self-organization to form large arrays of nanostructures have recently attracted much attention. Most of such work is based on the Stranski-Krastanov (SK) growth of semiconductor thin film on lattice-mismatched semiconductor substrates [7]. The influence of growth temperatures and growth rates, with varying composition and coverage on the size and number of self-assembled islands has been discussed in detail in previous publications [8, 9]. However, there is a significant drawback in the use of self-assembled dots since their sizes are often heterogeneous, resulting in a wide spread of electronic states.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

This review will focus on nonlithographic methods of nanostructure formation on single-crystal metal and semiconductor surfaces, with less emphasis on the well-studied III–V alloy semiconductors. Since adsorbate-induced surface reconstructions have been well studied by surface scientists over the past few decades, this review will also exclude such studies where the reconstructed unit cell is generally below 1 nm. The mechanisms for nanostructure formation on surfaces are not fully understood, and this review attempts to classify various adatom/molecule-substrate systems where nanostructure formation is observed in terms of the proposed mechanisms.

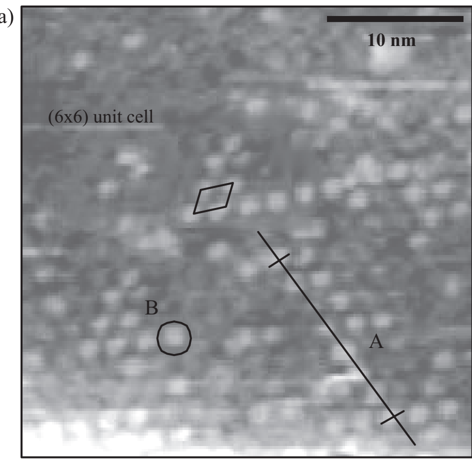
2. SELF-ORGANIZED NANOSTRUCTURES

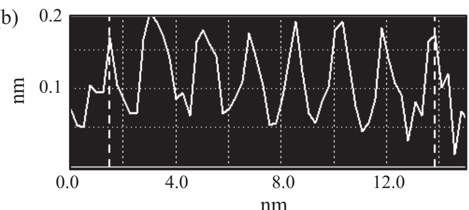
A. Spontaneous Assembly of Nanoclusters

The most well-studied surface reconstruction is perhaps the Si(111)–(7 × 7) surface, which has a unit cell divided into faulted and unfaulted halves [10]. Vitali et al. have demonstrated the formation of a superlattice of metallic nanoclusters, when thallium is evaporated onto clean Si(111)–(7 × 7) surfaces [11]. The metallic nanoclusters comprise nine adatoms each and are trapped within the attractive potential wells in the faulted half of the (7 × 7) unit cell, thus forming an ordered array of two-dimensional nanodot structures.

Li et al. reported the spontaneous assembly of identical-size In nanocluster arrays on Si(111)–(7 × 7) [12]. They show that substrate-induced spontaneous clustering is locally stable but globally metastable. Homogeneous cluster formation can be realized by delicate control of growth kinetics, i.e., within a suitable temperature range (100–200 °C) and deposition rate. If the substrate temperature is too low or the deposition rate too high, the In atoms do not have enough time to arrive at the expected location and form immobile nuclei of smaller sizes. On the other hand, if the substrate temperature is too high, the more energetic In atoms will agglomerate into larger clusters that coalesce into islands.

Si nanoclusters have also been observed to self-assemble on the 6H–SiC(0001)–(3 × 3) Si–rich surface [13]. Regular-sized Si "magic clusters" are ejected when this surface is heated to 950 °C and quenched to room temperature (Fig. 1). The Si clusters spontaneously form a metastable ordered (6 × 6) superstructure and appear to be driven by the need to reduce the dangling bond density per unit cell. These Si magic clusters have been previously reported on the Si(111)–(7 × 7) when it is quenched, although no self-assembly was observed on that surface [14].





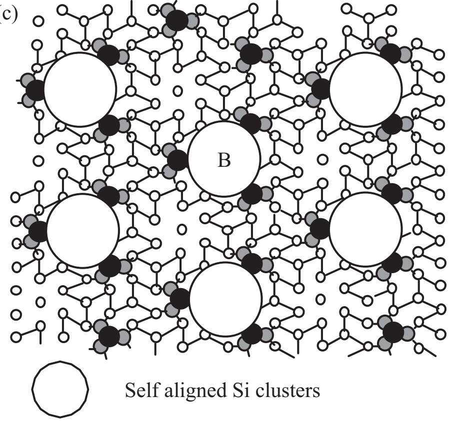


Figure 1. (a) Filled-state 30 nm×30 nm STM image (0.2 nA,–2.8 V) showing initial formation of cluster periodicity after heating the (3 × 3) reconstruction to 95 °C and subsequent cooling for 60 minutes. (b) STM cross-sectional line profile. A showing that the clusters posses a periodicity 6.1 ± 0.1 times that of the bulk. (c) Proposed model of 6H–SiC(001) (6 × 6) reconstruction due to cluster self-assembly. The atoms shown in the third Si adlayer layer represent unrelaxed surface after Si tetramer ejection. Reprinted with permission from [13], W. J. Ong et al., *Appl. Phys. Lett.* 80, 3406 (2002). © 2002, American Institute of Physics.

B. Surface Lattice Dislocations

Ordered nanocluster formation also has been reported on the Au(111) surface. For instance, Chambliss et al. reported that Ni islands grow with a spacing of 7.3 nm in rows 14.0 nm apart [15]. This regular nucleation is caused by the long-range "herringbone" reconstruction of Au(111), which induces periodic surface-lattice dislocations. Statistics of island size and number show that atoms diffusing on the surface bind at these dislocations with a low sticking probability and act as nuclei for further aggregation. It also has been reported that for Fe [16] and Co [17] deposition on Au(111), islands are nucleated at the elbows of the herringbone reconstruction. However, such preferential nucleation on Au(111) was not observed for other metals such as Al, Cu, Ag [18], and Au [19]. Meyer et al. [20] used their STM images for the Ni/Au(111) system to explain this disparity by proposing a two-step process: first, the place exchange of Ni adatoms with Au surface atoms occurs

at the elbow sites, followed by nucleation of Ni islands atop substitutional Ni atoms. The three metals that form islands at the herringbone elbows all have a higher surface free energy and heat of sublimation than Au, suggesting that place-exchange sites are favored. These place-exchange atoms then act as preferential nucleation sites for subsequent island formation.

Fruchart et al. [21] have harnessed the self-organized arrays of nanosized Co dots on Au(111) to fabricate Co pillars for high-density magnetic storage applications. This is achieved by sequentially depositing a fraction x of an atomic layer of Co and $1-x$ atomic layer of Au. At each step, Co adatoms aggregate on top of the previous dots, gradually forming columns. Helveg et al. [22] used STM to study the atomic scale structure of a single layer of MoS_2 nanoparticles on Au(111) as a model system in hydrodesulfurization catalysis.

C. Adsorbate-Induced Superstructures

The structures of crystalline surfaces are modified by the adsorbates, often by minor rearrangements to accommodate the adsorbate atoms, but sometimes by major restructuring of the surface [23]. Restructuring occurs in order to maximize the bonding and stability of the adsorbate-substrate complex. Thermodynamic processes drive this restructuring wherein stronger adsorbate-substrate bonds form to compensate for the weakening of bonds between the substrate atoms, an inevitable consequence of the chemisorption-induced restructuring process.

By using the Cu surface as an example, the structure of O chemisorption reconstructions on low-index Cu surfaces is well known [24]. O chemisorption on both Cu(110) and (100) surfaces leads to substantial reconstruction of the topmost Cu atom layer. Both reconstructions appear to receive stability from being able to accommodate O–Cu–O building blocks, of the sort needed to construct bulk Cu_2O [25]. The O atoms are fourfold coordinated, but are not quite centered in the Cu tetrahedra. Each O atom achieves a four-coordinate status with reasonable O–Cu bond lengths of 1.85–1.90 Å. On Cu(100), oxygen induces a $(\sqrt{2} \times 2\sqrt{2})$R45°–O reconstruction involving the removal of every fourth Cu [001] atom row. For Cu(110), the (2×1)–O phase is an "added row" structure with every other [001] Cu atom row absent, and O atoms occupy near collinear long-bridge sites in the added rows forming O–Cu–O chains.

One approach to increase the periodicity of the adsorbate-induced superstructures formed into the nanometer scale is the use of high index surfaces. High-Miller-index surfaces appear to exhibit high adatom mobility, and coordinatively unsaturated surface atoms move easily toward new and more bulk-like equilibrium positions. Wee et al. [26] have demonstrated the formation of a series of Cu(210)–$(n \times 1)$O $(n = 2, 3, 4)$ superstructures by using STM and low energy electron diffraction (LEED). Furthermore, they observed that adsorption of O at room temperature, followed by high-temperature annealing, leads to a series of mesoscopic superstructures ordered on length scales ranging between 1.5 to 15 nm [27]. A periodic supergrating for the Cu(110)–(2×1)O system also has been reported with a periodicity of between 6 to 14 nm, depending on oxygen coverage and substrate temperature [28].

Adsorbate-induced 2D nanostructures also have been reported for the Cu(210)–Br and Cu(100)–Br systems. When Cu(210) was dosed with Br at room temperature and subsequently annealed, a mesoscopic triangular checkerboard was observed [29]. The basic triangular unit comprises 4 or 5 atom rows (made up of 10 or 15 atoms, respectively) in the {210} plane, with edges of the triangle along the [001], $[1\bar{2}1]$ and $[12\bar{1}]$ directions. Br is believed to have etched the Cu(210) surface, leaving the mesoscopic checkerboard structure that reflects {210} symmetry.

For the Cu(100)–Br system, a corresponding mesoscopic chessboard structure has been observed [30]. Under these conditions, no desorption of Br was observed. The step edges are oriented along the {100} directions, reflecting the Cu(100) substrate symmetry, and the "white" Br squares have a local $c(2 \times 2)$ reconstruction. It appears that in this system, the adsorption and etching due to bromine provides a mechanism for the development of strain within the mesoscopic islands, thus limiting their island size.

Leibsle and coworkers have templated low-index Cu surfaces with self-assembled arrays of atomic nitrogen islands and then used these islands as masks to fabricate nanoscale structures [31]. Annealing of the Cu(100) surface with 0.5 ML (monolayer) atomic nitrogen coverage results in self-assembled regular island arrays (island size of 5×5 nm^2) of the $c(2 \times 2)$N reconstruction [32]. Nanometer scale lines and 2D arrays of islands of various metals such as Fe, Cu [33], Ag [34, 35], Co [36], have been demonstrated. Similarly, Ag nanostructures on N-modified Cu(111) [37], and Co nanowires on N-modified Cu(110) [38] have been reported. Ohno et al. [39] studied the growth mechanism of Fe nanoisland arrays on Cu(0 0 1)–$c(2 \times 2)$N surfaces where 5×5 nm^2 patches of N-adsorbed $c(2 \times 2)$ structure are squarely arranged and separated by a few nm-wide lines of clean Cu surface. Preferential island growth at the intersections of the clean Cu lines is ascribed to the large sticking probability. Difference of surface strain on the clean Cu lines is the origin of the site selections both for the Fe inclusion in the surface and for the sticking probability of mobile Fe adatoms to the edge of the islands.

The Si(100)–$(2 \times n)$ surface also has been used for the formation of metal nanowire structures. This $(2 \times n)$ reconstruction can be obtained by various methods, such as quenching the sample after flashing and contaminating the surface with small amounts of Ni [40]. Li et al. [41] demonstrated the spontaneous formation of ordered In nanowire arrays on Si(100)–$(2 \times n)$. The deposited in atoms predominantly occupy the normal (2×1) dimer row structure and develop into a uniform array of in nanowires at a coverage of about 0.2 ML. Kida et al. [42] have formed Fe nanowire arrays on a water-terminated Si(100)–$(2 \times n)$ surface. The average wire width was 2 nm and the period was 3 nm, reflecting the $(2 \times n)$ surface reconstruction.

D. Self-Organized Surface Phases

Self-organized surface nanostructures such as atomic wires also may be formed by one element of a binary substrate, without the need of an adsorbate. Soukiassian [43]

has observed the formation of highly stable self-organized atomic lines on cubic β–SiC(100) and has highlighted their potential use in SiC-based nanoelectronic devices. Highly stable and long Si atomic lines are composed of Si dimers and form at the phase transition between the (3×2) and $c(4 \times 2)$ reconstructions by selective Si atom organization [44]. These Si atomic lines coincide systematically with the lateral mismatch between the $c(4 \times 2)$ Si-dimer rows. Their number and spacing can be controlled by annealing time and temperature, resulting in single isolated atomic lines or large superlattices. Carbon atomic chains also have been observed on C-terminated β–SiC(100) [45]. These carbon atomic lines are extremely stable up to 1200 °C, and comprise buckled single bond C–C dimers (sp^3), having a direction perpendicular to the C≡C triple bond dimers (sp) forming a $c(2 \times 2)$ surface reconstruction. The authors suggest that anisotropic (tensile and compressive) surface stress is the main driving force in this carbon $sp \rightarrow sp^3$ diamond-like temperature-controlled transformation.

E. Strain-Relief

Brune et al. [46] have demonstrated the formation of ordered 2D metal nanostructure arrays on substrates with periodic patterns defined by dislocations that form to relieve strain. The strain-relief patterns are created spontaneously when 1 to 2 ML of one material is deposited on a substrate with a different lattice constant, for example, Fe arrays on a 2 ML Cu on Pt(111) (Fig. 2), and Ag arrays on 2 ML Ag on Pt(111). Nedelmann et al. [47] studied the evolution of the strain-induced stripe pattern on Cu/Ni(100). Compressive strain resulting from the 2.6% lattice mismatch between Cu and Ni is relieved by the formation of V-shaped defects in the Cu film. This strain relief mechanism gives rise to a characteristic pattern of orthogonal stripes at the film surface.

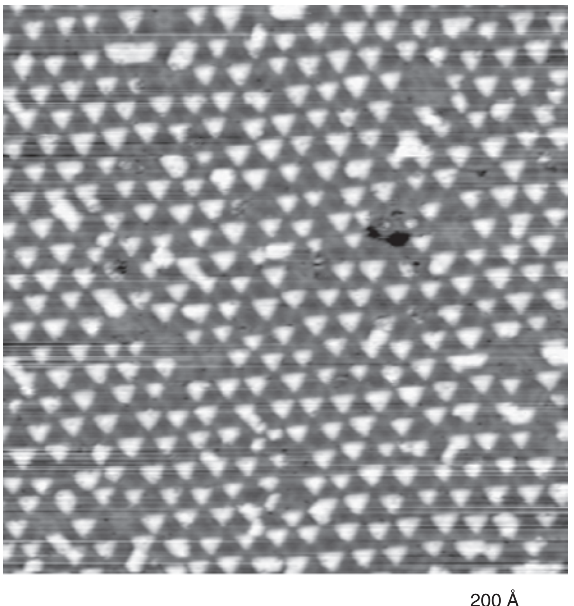


Figure 2. STM image of a periodic array of Fe islands nucleated on the dislocation network of a Cu bilayer on Pt(111) at 250 K. Reprinted with permission from [46], H. Brune et al., *Nature*, 394, 451 (1998). © 1998, Macmillan Magazines Ltd.

Mass selected Ag_7 clusters have been deposited at two different impact energies (20 and 95 eV) on a similar dislocation network formed by evaporating 2 ML of Ag on Pt(111) [48]. The surface has been investigated by low temperature STM as a function of annealing temperature. The clusters adsorption site, stability, and ordering depend on impact energies and annealing temperatures. It was observed that the hcp small triangle and dislocation region of the reconstructed unit cell form preferred pinning centers at low temperature; at higher temperature, most clusters are situated in the fcc hexagon of the unit cell. It is shown that the surface morphologies obtained by energetic cluster deposition are different from those obtained by thermal atom deposition; they depend on impact energy, deposition temperature, and annealing temperature. Although island ordering is not as good as that obtained by atom deposition, the results indicate that the large parameter space in cluster surface collisions (cluster size and impact energy) could be used to tailor metal nanostructures on structured surfaces.

The deposition of Ge on Si(100) leads to strain-induced formation of Ge hut clusters when the Ge film exceeds a critical thickness of 3 to 4 ML [49]. This layer-plus-island growth or SK growth occurs due to the lattice mismatch between Ge and Si, but, because of the initial Ge layer growth, the Ge quantum dots (QD) grown are not isolated. The presence of an adsorbate can alter the nucleation behavior of the Ge dots. Leifeld et al. [50] have shown that with a 0.11 ML precoverage of carbon, Ge dots form below the critical thickness for Ge islanding on bare Si(100). Das et al. [51] showed that by coating the Si surface with a polymer, isolated Ge islands and nanowires can form. This is explained by the fact that with a polymer overlayer, there is no strain-induced lattice mismatch and SK growth is not expected. Either layer-by-layer or Frank van der Merwe growth occurs under wetting conditions, or (3D) island growth, or Volmer-Weber growth occurs under nonwetting conditions. In this case, the authors have used the nonwetting criterion to grow isolated Ge nanoclusters on polymer surfaces.

Strain-induced growth of nanostructures also has been observed in several binary and ternary semiconductor systems. Leonard et al. [52] observed the 2D–3D growth mode transition during the initial stages of growth of highly strained InGaAs on GaAs, and this is used to obtain quantum-sized dot structures. Nözel et al. [53] described the self-organized formation of box-like structures during the interrupted epitaxial growth of strained InGaAs/AlGaAs multiplayer structures on (311)B GaAs substrates. Moison et al. [54] showed that the deposition of InAs on GaAs proceeds first by 2D growth, and above 1.75-ML regular nanometer-scale InAs dots form. Epitaxially self-assembled $In_xGa_{1-x}As$ QD within a GaAs substrate have atom-like properties and potential applications in optical and optoelectronic devices, quantum computing, and information storage [55].

Schmidt et al. [56] present different growth techniques and strategies for achieving different levels of order in the field of self-assembled semiconductor nanostructures. In a first step, they report a significant improvement of the

nanostructure size homogeneity by using either a repetitive desorption and regrowth procedure or by applying extremely low growth rates at high-growth temperatures. With this approach, an InAs/GaAs QD ensemble with a height distribution of ±5%, and a final photoluminescence (PL) peak line width of 19 meV at room temperature was fabricated. Regular square arrays of self-assembled SiGe islands with a strong size and shape homogeneity is achieved by deposition near their thermodynamic equilibrium using liquid phase epitaxy. The combination of self-assembly with conventional prepatterning leads to a long-range lateral order of In(Ga)As QDs on GaAs(001). A 3D crystal is fabricated by stacking multiple layers and vertically aligned In(Ga)As QDs onto a prepatterned GaAs(001) substrate, and good PL properties at room temperature are observed.

3. STEP AND FACET-TEMPLATED NANOSTRUCTURES

A. Stepped Surfaces

Stepped surfaces can serve as templates for producing 1D wires or stripes by decorating the step edges in a step-flow growth mode [57, 58]. Silicon is among the preferred substrates for step-flow growth of quantum wires because of the maturity of silicon-processing techniques in microelectronics. By appropriately controlling the annealing conditions, Viernow et al. [59] have produced highly regular array of steps on vicinal Si(111)–(7 × 7) with steps edges that are atomically straight for up to 2×10^4 lattice sites. The terraces are single domain, which produces a minimum kink width of 2.3 nm (half a 7 × 7 unit cell) and is, therefore, a high barrier for kink creation. Takagi et al. [60] used vicinal Si(111) to fabricate arrays of iron silicide dots of nanometer size on Si(111)– $\sqrt{3} \times \sqrt{3}$Ag substrates. Formation and magnetic properties of the dots have been investigated with STM, LEED, AES, and magnetic *ex-situ* measurements. The dots were found to nucleate preferentially at step edges than on terraces. They were ferromagnetic, and their composition was deduced to be $\sim Fe_3Si$.

Chain-like reconstructions of metals on semiconductors provide a natural vehicle to explore 1D electrons, provided that the electronic states are sufficiently decoupled from each other and from the bulk states of the substrate. Shibata et al. [61] studied the adsorption of Au on vicinal Si(111) by using STM and found ordering of Au in the steps that corresponded to (557) orientation. STM of Au on Si(557) showed double rows of protrusions [62], and 3D x-ray diffraction measurements were used to elucidate the detailed structure of the Au chain [63]. The structure is consistent with the 1D metallic behavior seen by photoemission. A 1D 5 × 2 reconstruction of Gd also has been reported on vicinal Si(111), which had stabilized in single domain form [64]. It contained zig-zag rows of occupied orbitals alternating with rows of unoccupied orbitals, as inferred from spectroscopic STM. Chain-like metal structures also has been observed for the Si(111)–(4 × 1)In system [65].

Nanowires grown by step decoration on metal surfaces also have been demonstrated. Depending on the overlayer–substrate combination, all classical thin film growth modes have been observed in one less dimension. The Cu/Mo(110) system exhibits row-by-row growth along the steps [66]. On the other hand, the Cu/W(110) system displays SK growth [67] and the Co/Cu(111) system island growth [68]. The growth modes are a result of competing effects, such as surface energy, interface energy, misfit strain, and step energy. The quantized electronic states in nanowires fabricated by step decoration have been observed [69].

B. Faceted Surfaces

It can be seen from the above discussion that vicinal surfaces with regular distributions of straight steps can act as natural templates for the growth of 1D nanostructures on surfaces due to their strong uniaxial anisotropy. This is possible if the substrate is stable during adsorption. In many cases, adsorbates induce significant surface faceting to minimize surface energy, and this also can be used for surface nanostructure formation. For example, regular arrays of 1D Ag stripes can be grown by adsorption and annealing of Ag on vicinal Cu(111) [70]. The Ag adlayer induces a periodic hill-and-valley nanostructure with alternating Ag-covered (112)-oriented facets and clean Cu stepped areas. The structure of the Ag-covered stripes is consistent with a rotated, uniaxially compressed, close-packed layer that wets the Cu(112) plane.

Nanoscale surface patterning by adsorbate-induced faceting and selective growth has been demonstrated on the insulator/metal growth system NaCl/Cu(211), which is inherently unstable against lateral pattern formation on the nanometer scale [71, 72]. Cu(211) is a vicinal surface with (111) terraces and intrinsic (100) steps, and restructures upon NaCl deposition when the growth temperature exceeds 270 K. The initially flat surface is transformed into a periodic 1D hill-and-valley structure consisting of (311) and (111) facets. NaCl grows selectively on (311) facets only, thereby creating a regular surface pattern with alternating stripes of bare Cu and chemically inert NaCl-covered areas. A highly regular assembly of three-sided pyramids can be fabricated by growing the ionic insulator NaCl on the kinked metal surface Cu(532) [73]. Only two pyramid faces are covered by NaCl, resulting in an overall surface structure that is modulated in surface chemical behavior. STM shows that electrostatic interactions plays an important role in the binding between an ionic insulator and a stepped or kinked metal template. This criterion provides a means to create nano- to mesoscopic surface structures combined with spatial modulation in surface chemical behavior. Mauch et al. [74] investigated the submonolayer growth of NaCl on Cu(100) by STM and LEED. The initial growth is characterized by decoration of step edges and rather strong restructuring of terraces of the Cu(100) surface. Further deposition of NaCl leads to the formation of narrow NaCl stripes oriented along the $[0\ \bar{1}\ 1]$ and $[0\ 1\ \bar{1}]$ directions. For higher coverages, these stripes form regular patterns with a fivefold superstructure.

Self-organized nanoscale pattern formation has been demonstrated on vicinal Si(111) surfaces via a two-stage faceting transition [75]. All the patterns, consisting of a periodic array of alternating (7 × 7) reconstructed terraces and step-bunched facets, have the same periodicity and facet structure, independent of the miscut angle, while the width of the facets increases linearly with miscut angle. The

pattern formation is attributed to a surface faceting transition that involves two transition stages: the first stage forms a stress-domain structure defining the universal periodicity; the second stage forms the low-energy facets controlling the facet width.

One-dimensional nanostructures of noble metals (Ag, Au) also have been grown on the high index Si(5 5 12) surface [76]. When low coverages (<0.25 ML) of noble metals (Ag, Au) are deposited onto Si(5 5 12) and annealed to 450 °C, they form overlayer nanowires with the periodicity of the Si surface (5.4 nm). At higher coverages and temperatures, significant restructuring occurs leading to the creation of facet planes.

4. MOLECULAR NANOSTRUCTURES

Self-assembly of regular arrays of thermodynamically stable nanostructures from appropriately functionalized molecules is a promising approach to future mass production of nanoscale devices [77]. Molecular self-assembly is the spontaneous association of molecules under equilibrium conditions into stable, structurally well-defined aggregates joined by noncovalent bonds. Molecular self-assembly is ubiquitous in biological systems and underlies the formation of a wide variety of complex biological structures. It has been shown that a self-assembled monolayer of thiolates deposited on suitable substrates could be made into resonant-tunneling transistors.

Molecules adsorbed on surfaces are particularly appealing because their arrangement is directly observable with the STM. For example, 2D supramolecular clusters and chains are observed upon submonolayer deposition of 1–nitronaphthalene (NN) onto reconstructed Au(111) [78, 79]. The molecules become pseudo-chiral upon adsorption. At higher NN coverage (0.3–0.75 ML), well separated molecular double chains are formed, either perpendicular or parallel to the Au(111) reconstruction domains. At low coverages (0.05–0.2 ML), NN forms 2D supermolecular clusters consisting predominantly of 10 molecules (decamers). Their handedness is determined from high-resolution STM images and local-density calculations. Modeling shows that hydrogen bonds cause the observed self-assembly. In contrast to atoms, the intermolecular electrostatic interaction of polar molecules leads to thermodynamically stable structures. Clusters and chains mutually interact via electrostatic repulsion. Comparison of images with submolecular resolution to local density calculations elucidates the thermodynamical stability, as well as the internal structure of the decamers.

Contini et al. [80] studied the adsorption of 2–mercaptobenzoxazole (MBO) molecules (C_7H_5NOS) on Cu(100) at room temperature by STM and found that after 0.8 L exposure, it is possible to locate ordered domains with well-defined parallel rows. After 38 L of deposition the surface appears completely covered by rounded protrusions randomly distributed. These results suggest a SK growth mode: initially an ordered monolayer forms and above this layer other molecules are adsorbed with a random distribution. The aromatic thione MBO molecule presents a S-containing head, which reacts with the Cu surface, and an aromatic ring originating lateral van der Waals interactions and molecular self-organization. A long-range-ordered $p(2\times2)$ structure of MBO on Cu(100) is obtained by MBO sublimation in highly controlled ultrahigh-vacuum conditions. The MBO molecular orbitals and interface electronic levels have been identified by angular-resolved high-resolution Ultra Violet (uv) photoelectron spectroscopy [81]. The role of S as a chemical hook for the molecule to the Cu surface is evidenced, and a comparison with similar compounds suggests that the adsorption mechanism and the related electronic structure are rather general results for π-conjugated molecules with a S-containing head.

STM observations of pentacene molecules on Au(111) at more than a monolayer coverage revealed periodic rows of molecules spaced 61 ± 5 Å apart [82]. These periodic rows consist of flat and edge-on molecules aligned with the initial layer of pentacene. Comparison of STM images and calculations show that the formation of widely spaced periodic rows takes place because of the total free energy gain from commensurate growth on the substrate and bulk pentacene-pentacene interactions. Lopinski et al. [83] demonstrated an approach for fabricating nanoscale organic structures on silicon surfaces, using minimal intervention by the tip of a STM and a spontaneous self-directed chemical growth process. Growth of straight molecular styrene lines, each composed of many organic molecules, and the crystalline silicon substrate determines both the orientation of the lines and the molecular spacing within these lines. This process should, in principle, allow parallel fabrication of identical complex functional structures.

The demand for enantiomerically pure compounds by the pharmaceutical and chemical industries has spurred the development of chiral technologies, which aim to exert control over a chemical reaction by directing its enantioselectivity. Heterogeneous enantioselective catalysis allows the production and ready separation of large quantities of chiral product, while using only small quantities of catalyst. Heterogeneous enantioselectivity usually is induced by adsorbing chiral molecules onto catalytically active surfaces. A mimic of one such catalyst is formed by adsorbing (R, R)–tartaric acid molecules on Cu(110) surfaces, which generates a variety of surface phases, only one of which is potentially catalytically active [84]. Ortega Lorenzo et al. [85] show that the active phase consists of extended supramolecular assemblies of adsorbed (R, R)–tartaric acid, which destroy existing symmetry elements of the underlying metal and directly bestow chirality to the modified surface. The adsorbed assemblies create chiral channels exposing bare metal atoms, and it is these chiral spaces are believed to be responsible for imparting enantioselectivity, by forcing the orientation of reactant molecules docking onto catalytically active metal sites.

Experiments on individual molecules by using scanning probe microscopies have demonstrated a diversity of physical, chemical, mechanical, and electronic phenomena [86]. They have permitted deeper insight into the quantum electronics of molecular systems and have provided unique information on their conformational and mechanical properties. Concomitant developments in experimentation and theory have allowed a diverse range of molecules to be studied, varying in complexity from simple diatomics to biomolecular systems. At the level of an individual molecule,

the interplays of mechanical and electronic behavior and chemical properties manifest themselves in an unusually clear manner.

Reed et al. [87] showed that molecules of benzene–1, 4–dithiol self-assembled onto the two facing gold electrodes of a mechanically controllable break junction form a statically stable gold-sulfur-aryl-sulfur-gold system, allowing for direct observation of charge transport through the molecules. This study provides a quantitative measure of the conductance of a junction containing a single molecule, which is a fundamental step in the emerging area of molecular-scale electronics. De Ventra et al. [88] have performed first-principles calculations of transport through a benzene–1,4–dithiolate molecule with a third capacitive terminal (gate). They found that the resistance of the molecule rises from its zero-gate-bias value to a value roughly equal to the quantum of resistance (12.9 kΩ) when resonant tunneling through the π^* antibonding orbitals occurs. Despite the growing interest in such systems, many fundamental properties of the molecule–substrate interface still are unanswered.

The construction of nanoscale electronic devices involves the fabrication of narrow gap electrodes and subsequent entrapment of a few or a single molecule between the metallic electrodes. The demonstrated techniques for fabricating electrodes include mechanical break junctions, nanopore methods [89], electrochemical plating [90], shadow evaporation [91], electromigration [92], and electron beam lithography [93]. Except for electron beam lithography, all the other techniques have a low yield and involve a series of steps that are difficult to control. Commercial electron beam lithography systems have the ability to focus electrons to diameters less than 10 nm such that metal-insulator-metal junctions for single-molecule electronics could be fabricated. So far, simple and reliable schemes for sub-10 nm and sub-5 nm electrode gaps by using electron beam lithography have been reported [94, 95].

One nm electrode gaps have been successfully prepared by using the combined electron beam lithography and electromigration technique [96, 97]. In this approach, a narrow gold bridge first was fabricated by electron beam lithography. The closely packed gold electrodes (1 nm apart from one another) then were created by the electromigration-induced break-junction technique. Trapping a molecule between two metal electrodes to make nanoscale transistor is a difficult technological challenge. The molecule needs suitable terminations that reliably bind it chemically to the two electrodes, bridging the gap between them. Moreover, there is presently no viable imaging technique for direct confirmation of successful trapping without perturbing the system, and the embedding of molecules could only be indirectly established from its conduction properties.

Lin and Kagan [98] used layer-by-layer self-assembly of metal–metal bonded supramolecules to fabricate lateral nanoelectronic devices. Mercaptoethylpyridine is used to bind to Au substrates and to template the sequential assembly of alternating layers of redox active dirhodium complexes $[Rh_2(DAniF)_2]_2$ $(O_2CCH_2CO_2)_2$ (DAniF = *N*, *N′*–di–*p*–anisylformamidinate) and conjugated organic ligands *trans*-1,2–bis(4–pyridyl)ethylene. Optical spectroscopy and atomic force microscopy show that the structure and composition of these thin films are similar to those found in tightly packed single crystals. Electrochemical studies of these films, grown on Au substrates, reveal a reversible oxidation wave at ~406 mV, corresponding to the one electron oxidation of the Rh_2^{4+} center. This directed assembly technique was used to fabricate lateral nanoelectronic devices in which the supramolecules span the channels. Tailoring the chemistry of the templating ligand enables assembly on desired surfaces, and engineering the chemistry of the supramolecules' dimetal units and coordinating ligands may tune the device characteristics.

Information storage technology relies on multistable systems that can be controllably switched between different configurations of comparable free energy. Multistability is present in many molecular and supramolecular systems through a variety of properties (conformations, redox and spin states, shape and dimensionality) that can be influenced by external stimuli. Cavallini et al. [99] have proposed a novel strategy by using thin films (3- to 35-nm thickness) of each of three amide-based rotaxanes. Rotaxanes are molecules in which a macrocycle is locked mechanically onto a "thread" by two bulky "stoppers." Their architecture, analogous to that of an abacus, suggests that they could be used as switchable components for artificial machines that function through mechanical motion at the molecular level [100]. Controlling such motion in the solid state could be used to store information. The films can be imaged by atomic force microscopy (AFM) in contact mode with a set point force below a threshold value of 2 nN. An increase in the load force to just above the 2-nN threshold results in a mechanical perturbation whose effect is localized at the contact area of the tip. When the tip is continuously scanned along a line, a string of regularly spaced dots appears. The scanning AFM tip gives the energy to the molecules along the line to reorganize into nuclei. As nuclei coarsen by ripening, a characteristic distance emerges. Finally, stable nuclei grow by incorporation of the mobile molecules to form the dots. Writing of multiple dots simultaneously could be achieved by using lithography based on multiple sources of perturbation, for instance by using a stamp. With such an approach, information storage on a thin film could reach densities of 1 to 10 Gbit/in^2.

5. ION-BEAM INDUCED NANOSTRUCTURES

Ion sputtering has been demonstrated as an alternative method for nanopatterning without the need for standard lithographic techniques. Recent studies using the high-spatial-resolution STM and AFM reveal that ion bombardment produces repetitive structures at the nanometre scale, creating peculiar surface morphologies ranging from self-affine patterns to "fingerprint-like" and even regular structures, for instance waves (ripples), checkboards, or pyramids. The phenomenon is related to the interplay between ion erosion and diffusion of adatoms (vacancies), which induces surface reorganization. Valbusa et al. [101] reviews the use of sputter etching to modify *in-situ* surfaces and thin films, producing substrates with well defined vertical roughness, lateral periodicity, and controlled step size and orientation.

In another review, Murty [102] summarizes the ion-solid collision process leading to the ejection of atoms and discusses different aspects of sputtering relevant to applications.

Mayer et al. [103] characterized the development of nanometer scale topography on SiO surfaces as a result of low energy, off-normal ion bombardment, using *in-situ* energy dispersive X-ray reflectivity and AFM. Surfaces roughen during sputtering by heavy ions, with roughness increasing approximately linearly with ion fluence up to 10^{17} cm^{-2}. A highly coherent ripple structure with wavelength of 30 nm and oriented with the wave vector parallel to the direction of incidence is observed after Xe sputtering at 1 keV. A linear model is presented to account for the experimental observations that include roughening both by random stochastic processes and by development of a periodic surface instability due to sputter yield variations with surface curvature that leads to ripple development.

Gago *et al.* [104] reported the production of ordered assemblies of silicon nanostructures by means of irradiation of a Si (100) substrate with 1.2 keV Ar^+ ions at normal incidence. Atomic force microscopy and high-resolution transmission electron microscopy show that the silicon structures are crystalline, display homogeneous height, and spontaneously arrange into short-range hexagonal ordering. Under prolonged irradiation, all dot characteristics remain largely unchanged and a small corrugation develops at long wavelengths. The formation of the dots is interpreted as a result of instability due to the sputtering yield dependence on the local surface curvature.

Grazing incidence small-angle X-ray scattering and AFM have been used to investigate the evolution of Si(100) surface morphology during normal-incidence Ar^+ sputtering as a function of ion energy in the range of 100–500 eV [105]. In the energy range of 100–300 eV, two structures with distinct individual length scales and behaviors form on the surface. The smaller scale (lateral size of 20–50 nm) morphology grows in scattering intensity and coarsens with time, while the larger scale (lateral size of approximately 100 nm) morphology grows in scattering intensity but does not coarsen significantly with time. At higher energies (400–500 eV), sputtering causes the Si(100) surface to become smoother on length scales smaller than 200 nm.

Spectroscopic light scattering was used to monitor periodic ripple evolution on Si(001) *in-situ* during oblique incidence (67.5°) Ar^+ sputtering with an energy of 750 eV [106]. Analysis indicates that, under high flux, the concentration of mobile species on the surface is temperature and ion-flux independent. This is due to an effect of ion collision cascades on the concentration of mobile species. The migration energy on the surface is measured to be 1.2 ± 0.1 eV.

Shang et al. [107] fabricated high-density Si cone arrays with a defined orientation over a large area by ion beam sputtering. The density of the Si cones is as high as 6×10^8 cm^{-2}, and a silicide tip was found on top of each Si cone. The Si cone arrays have a relatively low turn-on field and have potential applications in electronic vacuum devices.

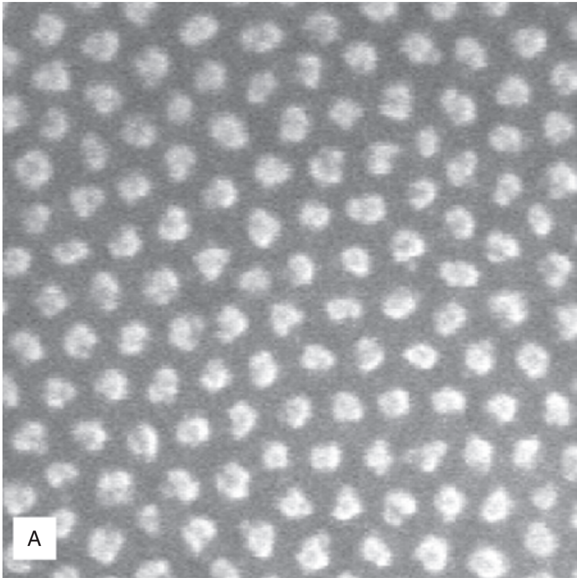
Chason et al. [108] investigated the temperature-dependent roughening kinetics of Ge surfaces during low energy ion sputtering by using energy dispersive X-ray reflectivity. At 150 °C and below, the surface is amorphized by ion impact and roughens to a steady state small value. At 250 °C, the surface remains crystalline, roughens exponentially, with time, and develops a pronounced ripple topography. At higher temperatures, this exponential roughening is slower, with an initial sublinear time dependence. A model that contains a balance between smoothing by surface diffusion and viscous flow, and roughening by atom removal explains the kinetics. Ripple formation is a result of a curvature-dependent sputter yield.

Facsko et al. [109] observed that well-ordered nanoscale islands appear on GaSb (001) surfaces during low-energy normal incidence Ar^+ sputtering. Crystalline dots 35 nm in diameter arranged in a regular hexagonal lattice were produced on the GaSb surface, the formation mechanism replying on the interplay between ion-induced surface roughening and smoothing due to surface diffusion (Fig. 3). The characteristic length of the generated pattern was observed to scale with the square root of ion energy over the energy range 75–1800 eV [110]. Similar nanoscale islands formation on lnP was achieved by Frost et al. when the lnP surfaces were bombarded with oblique incident Ar^+ ions at 500 eV with simultaneous sample rotation [111].

Periodic Co/Pd magnetic nanostructures have been fabricated by depositing perpendicular magnetic films on nanopatterned GaSb(100) substrates formed under normal incidence 420 eV Ar^+ bombardment [112]. Similar stripe domain patterns and reduced coercivity are observed for the as-deposited sample as compared with those for a reference unsputtered substrate. Further processing steps such as etching and polishing will be needed to isolate the magnetic elements from each other in order to make ultrahigh density magnetic storage media.

Ni films grown at room temperature on flat and prenanostructured Cu(001) surfaces have been studied *in-situ* by STM and the magneto-optic Kerr effect [113]. Ar^+ bombardment of 30–45 monolayer thick Ni/Cu(001) films results in the formation of self-organized pseudoperiodic nanostructures. On flat Cu(001), the nanostructuring process rotates the easy-magnetization axis of the 45 monolayer Ni/Cu(001). During the growth of Ni on prenanostructured Cu(001), neither preferential decoration of step-edges by material nor filling of the Cu-craters occurs.

Ion sputtering-induced nanopatterns also have been observed on metal surfaces. The morphology of the Ag(110) surface after Ar^+ ion sputtering has been studied by a variable temperature STM [114]. For ion energies greater than 800 eV and in the temperature range $270 < T_S < 320$ K (where T_S is the sputtering temperature), a peculiar ripple structure well aligned along the $\langle 1\bar{1}0 \rangle$ direction and independent of the ion beam angle has been observed. A simple model of this effect, based on the anisotropy and on the temperature hierarchy of the Ag adatom and vacancy surface diffusion, explains the observed features. The evolution of the Cu(110) surface morphology during low temperature (180 K) ion sputtering also was studied as a function of the incident ion beam angle by means of STM [115]. The morphology was dominated by a ripple structure with the wave vector parallel or perpendicular to the direction of the incident beam. The time evolution of the interface shows that the ripple wavelength increases in time following a scaling law $\lambda \propto t^z$, with $z = 0.26 \pm 0.02$. These results are ascribed to the effect of a Schwoebel barrier on



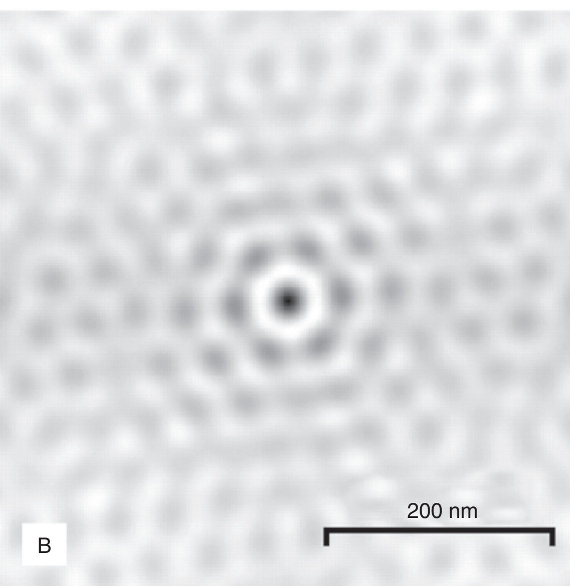


Figure 3. (A) SEM image showing highly ordered cones on a GaSb(100) surface after ion sputtering; (B) the corresponding 2D autocorrelation reveal the regularity and hexagonal ordering of the dots, which extends over more than six periods. Reprinted with permission from [109], S. Facsko et al. *Science* 285, 1551 (1999). © 1999, American Association for the Advancement of Science.

the interlayer diffusion of the recoiling atoms produced during ion sputtering.

Nanopatterns on metal surfaces can be created easily by ion sputtering. However, due to the fast diffusion processes characterizing these materials, the nanostructures are often unstable at room temperature and above. This effect prevents the use of such patterned substrates in nanotechnology applications. Sekiba et al. have proposed a simple oxidation process to stabilize these features durably [116]. The method has been tested on Cu, but its generality suggests that it can be applied to many other metals.

Glancing-incidence ion-beam irradiation has been used both to ease kinetic constraints that otherwise restrict the establishment of long-range order and to impose external control on the orientation of nanowire arrays formed during stress-field-induced self-ordering of calcium atoms on a CaF_2(111) surface [117]. The arrays exhibit exceptional long-range order, with the long axis of the wires oriented along the azimuthal direction of ion-beam incidence. Transport measurements reveal a highly anisotropic electrical conductivity, whose maximum lies in the direction of the long axis of the 10.1-nm-period calcium wires.

Kahng et al. [118] show that sputtering-induced nanostructures have an inherently nonlinear origin, first appearing when the nonlinear terms start to dominate the surface dynamics. Depending on the sign of the nonlinear terms, determined by the shape of the collision cascade, the surface can develop regular islands or holes with identical dynamical features. While the size of these nanostructures is independent of flux and temperature, it can be modified by tuning the ion energy. By using numerical simulations with the KS equation, Kim et al. [119] demonstrate that rotated ripple structures (RRS) on sputter-eroded surfaces are potential candidates for nanoscale wire fabrication. The RRS can form when the width of the collision cascade in the longitudinal direction is larger than that in the transverse direction and the incident angle of ion beam is chosen in a specific window. By calculating the structure factor for the RRS, it is found that the RRS are more regular and their amplitude is more enhanced compared with the much studied ripple structure forming in the linear regime of sputter erosion.

Lau et al. report another way of creating nanostructures on surfaces by ion sputtering of strained superlattices that possess periodic interfacial elemental segregation [120]. Arrays of nanoscale depressions were observed during sputtering of Si/SiGe heterostructures by using 500 eV O_2^+ ions at angles of incidence between 46° and 56°, with simultaneous sample rotation in a secondary ion mass spectrometry system. Both AFM and TEM results suggest that these depressions were produced as a result of preferential sputtering of Ge and only occur when there are undulations present at the Si/SiGe interface. The formation of the interface undulation appears to occur in tandem with the increase in the Ge surface segregation at the Si/SiGe interface for layers grown at elevated temperatures.

Manipulating matter at the nanometre scale is important for many electronic, chemical, and biological advances, but present solid-state fabrication methods do not reproducibly achieve dimensional control at the nanometre scale. Li et al. have reported a means of fashioning matter at these dimensions that uses low-energy ion beams and reveals surprising atomic transport phenomena that occur in a variety of materials and geometries [121]. The method is implemented in a feedback-controlled sputtering system that provides fine control over ion beam exposure and sample temperature. They name the method "ion-beam sculpting," and apply it to the fabrication of a molecular-scale hole, or nanopore, in a thin insulating solid-state membrane. Such pores can serve to localize molecular-scale electrical junctions and switches, and function as masks to create other small-scale structures.

6. PHOTON-INDUCED NANOSTRUCTURES

Pulsed-laser irradiation is another method for growing nanoparticles on surfaces [122]. The high photon flux impinging on a target produces the ejection of a high density of neutral atoms, ions, and electrons. The background gas pressure confines the ejected material to a certain distance from the target that can be collected on a substrate or redeposited on the target. Fowlkes et al. [123] used pulsed-laser irradiation to induce the formation of linear arrays of nanoparticles that can extend over millimeter distances. On flat surfaces, the irradiation induces the fragmentation and clustering of a thin silicon film pulsed-laser deposited on silicon into nanoparticles that grow to 30 to 40 nm in diameter. The nanoparticles aggregate into clusters that migrate, forming short curvilinear groups that exhibit a short-range ordering. If a region containing a microscopic roughness is introduced, the nanoparticles are forced to align into long and remarkably straight lines with line spacing very close to the laser wavelength. A close connection is established between the nanoparticle alignment and the evolution of laser-induced periodic surface structures, and the microscopic roughness serves as a trigger to produce the alignment.

Kawamura et al. [124] used femtosecond laser pulses (wavelength ~800 nm) to encode two cross-superposed holographic gratings on silica glass. A variety of periodic nanostructures from a 1D wire array to 2D arrays of holes or islands were formed by changing the energy density and the incidence angle of the irradiation laser beams. The smallest dimensions were a width of ~15 nm for wires and a diameter of ~20 nm for holes. Laser-driven microexplosions occurring within the microcylindrical-lens array created by the first laser pulse are suggested as a mechanism for the formation of these structures. Only two pulses are required to encode these periodic structures, and this method could be applied to the fabrication of photonic crystals and QD or wire arrays.

7. SUMMARY

While lithographic techniques (electron, optical, UV, X-ray, ion projection, etc.) still dominate today in micro- and nanoscale fabrication, this review has focused on methods of self-organized nanostructure formation on surfaces, which could provide an alternative approach to nanodevice fabrication with potentially smaller feature sizes, even approaching the atomic scale. In particular, emphasis is placed on how nanostructures can form spontaneously due to surface dislocations, adsorbates, strain-relief, and the presence of steps and facets. Molecular nanostructures are described because of their potential applications in the biomedical sciences. It also is shown that broad beams of ions and photons can induce nanostructure formation under appropriate conditions.

GLOSSARY

Adsorbate An atom or molecule that adheres to a surface through chemical or physical bonding.

Atomic force microscope (AFM) An instrument that is able to image surfaces to molecular accuracy by mechanically probing their surface contours. A device in which the deflection of a sharp stylus mounted on a soft spring is monitored as the stylus is moved across a surface. If the deflection is kept constant by moving the surface up and down by measured increments, the result (under favorable conditions) is an atomic-resolution topographic map of the surface.

Chemisorption Adsorption in which the forces involved are binding forces of the same kind as those operating in the formation of chemical compounds. Chemisorbed atoms or molecules are therefore more strongly bound to the surface than physisorbed atoms or molecules.

Clusters Aggregates of atoms (or molecules) that contain between three and a few thousand atoms which have properties intermediate between those of the isolated monomer (atom or molecule) and the bulk or solid state material.

Dislocations Linear defects inside a crystal lattice which govern the plastic behaviour of a material. Since dislocations cannot end somewhere inside a crystal, they must either form dislocation loops or reach the surface.

Fermi wavelength At low temperatures the current is mainly driven by electrons with energy close to the Fermi energy; hence the relevant length scale is the Fermi wavelength, typically around 35 nm. In particular, when the dimension of metallic nanowires is comparable to the electron Fermi wavelength, the conductance is quantized and no longer obeys classical Ohm's Law.

Ions An ion is any atom or group of atoms that bears one or more positive or negative electrical charges. Ions are typically formed by the addition of electrons to, or the removal of electrons from, neutral atoms or molecules or other ions.

Mesoscopic The Greek word *meso* means "in between". Mesoscopic physics refers to the physics of structures of intermediate sizes, ranging from a few atomic radii to a few microns. A mesoscopic sample is too big to study its properties by standard methods in the physics of individual atoms, and is too small for the application of the familiar physical laws of the macro world. A macroscopic device, when scaled down to a meso-size, starts revealing the quantum signatures of conventional characteristics. For example, in the macro world, the conductance of a wire increases continuously with its diameter, but in the meso world the wire's conductance is quantized, i.e., the increases occur in steps.

Nanostructures Regular structures formed, for example on a surface, at the nanometer scale.

Photons Energy packets of electromagnetic radiation; light consists of photons.

Scanning tunneling microscope (STM) An instrument that is able to image conducting surfaces to atomic accuracy and has been used to manipulate atoms/molecules on a surface. The normal operation of the device is to maintain a constant tunneling current between the probe tip and the surface kept at constant bias difference, so the vertical movement of the tip maps the surface electronic topography.

Self-organized The spontaneous regular arrangement of atoms, molecules or clusters, for example on a surface, due to interactions between the entities and the surface.

Stranski–Krastanov (SK) growth The growth of 3D islands on a surface after the strain energy exceeds some critical limit of a 2D structure (wetting layer).

Surface facets More stable low index surface planes (with higher atom density) which a high index surface (with lower atom density) often prefers to transform itself into.

Surface steps Discontinuities on a surface typically due to an additional atomic layer. These steps are usually favorable sites for atoms, molecules or clusters to bind at.

REFERENCES

1. M. C. Roco, S. Williams, and P. Alivisatos (Eds.), "*Nanotechnology Research Directions: IWGN Workshop Report*." National Science and Technology Council Committee on Technology; The Interagency Working Group on Nanoscience, Engineering and Technology; WTEC, Loyola College in Maryland, p. 4. http://itri.loyola.edu/nano/IWGN.Research.Directions/
2. G. Binnig and H. Rohrer, *Hev. Phys. Acta* 55, 726 (1982).
3. W.-D. Schneider, *Surf. Sci.* 514, 74 (2002).
4. S. Ushioda, *J. Electron Spectrosc. Relat. Phenom.* 109, 169 (2000).
5. G. Binnig and H. Rohrer, *IBM J. Res. Dev.* 44, 279 (2000).
6. T. I. Kamins and R. S. Williams, *Appl. Phys. Lett.* 71, 1201 (1997).
7. V. A. Shcukin and D. Bimberg, *Rev. Mod. Phys.* 71, 1125 (1999).
8. P. Schittenhelm, G. Abstreiter, A. Darhuber, G. Bauer, P. Werner, and A. Kosogov, *Thin Solid Films* 294, 291 (1997).
9. P. Schittenhelm, M. Gail, and G. Abstreiter, *J. Cryst. Growth* 157, 260 (1995).
10. K. Takayanagi et al., *J. Vac. Sci. Technol. A* 3, 1502 (1981).
11. L. Vitali, M. G. Ramsey, and F. P. Netzer, *Phys. Rev. Lett.* 83, 316 (1999).
12. J.-L. Li, J.-F. Jia, X.-J. Liang, X. Liu, J.-Z. Wang, Q.-K. Xue, Z.-Q. Li, J. S. Tse, Z. Zhang, and S. B. Zhang, *Phys. Rev. Lett.* 88, 066101 (2002).
13. W. J. Ong, E. S. Tok, H. Xu, and A. T. S. Wee, *Appl. Phys. Lett.* 80, 3406 (2002).
14. I.-S. Hwang, M.-S. Ho, and T. T. Tsong, *Phys. Rev. Lett.* 83, 120 (1999).
15. D. D. Chambliss, R. J. Wilson, and S. Chiang, *Phys. Rev. Lett.* 66, 1721 (1991).
16. J. A. Stroscio, D. T. Pierce, R. A. Dragoset, and P. N. First, *J. Vac. Sci. Technol., A* 10, 1981 (1992).
17. B. Voigtländer, G. Meyer, and N. M. Amer, *Phys. Rev. B* 44, 10354 (1991).
18. M. M. Dovek, C. A. Lang, J. Nogami, and C. F. Quate, *Phys. Rev. B* 40, 11973 (1989).
19. C. A. Lang, M. M. Dovek, J. Nogami, and C. F. Quate, *Surf. Sci.* 224, L947 (1989).
20. J. A. Meyer, I. D. Baikie, E. Kopatzki, R. J. Behm, *Surf. Sci.* 365, L647 (1996).
21. O. Fruchart, M. Klaua, J. Barthel, and J. Kirschner, *Phys. Rev. Lett.* 83, 2769 (1999).
22. S. Helveg, J. V. Lauritsen, E. Lægsgaard, I. Stensgaard, J. K. Nørskov, B. S. Clausen, H. Topsøe, and F. Besenbacher, *Phys. Rev. Lett.* 84, 951 (2000).
23. G. A. Somorjai, in "*Introduction to Surface Chemistry and Catalysis*," Wiley, New York, p. 412 ff. 1994.
24. F. Besenbacher and J. K. Nørskov, *Prog. Surf. Sci.* 44, 5 (1993), and references therein.
25. W. Liu, K. C. Wong, H. C. Zeng, and K. A. R. Mitchell, *Prog. Surf. Sci.* 50, 247 (1995).
26. A. T. S. Wee, Y. P. Guo, K. C. Tan, H. Q. Wang, T. K. Leong, and C. H. A. Huan, *Surf. Interface Anal.* 32, 228 (2001).
27. A. T. S. Wee, J. S. Foord, R. G. Egdell, and J. B. Pethica, *Phys. Rev. B* 58, R7548 (1998).
28. K. Kern, H. Niehus, A. Schatz, P. Zeppenfeld, J. Goerge, and G. Comsa, *Phys. Rev. Lett.* 67, 855 (1991).
29. A. T. S. Wee, T. W. Fishlock, R. A. Dixon, J. S. Foord, R. G. Egdell, and J. B. Pethica, *Chem. Phys. Lett.* 298, 146 (1998).
30. T. W. Fishlock, J. B. Pethica, and R. G. Egdell, *Surf. Sci.* 445, L47 (2000).
31. F. M. Leibsle, *Surf. Sci.* 514, 33 (2002).
32. F. M. Leibsle, S. S. Dhesi, S. D. Barrett, and A. W. Robinson, *Surf. Sci.* 317, 309 (1994).
33. T. M. Parker, L. K. Wilson, N. G. Condon, and F. M. Leibsle, *Phys. Rev. B* 56, 6458 (1997).
34. S. L. Silva and F. M. Leibsle, *Surf. Sci.* 440, L835 (1999).
35. S. M. York and F. M. Leibsle, *Appl. Phys. Lett.* 78, 2763 (2001).
36. S. L. Silva, C. R. Jenkins, S. M. York, and F. M. Leibsle, *Appl. Phys. Lett.* 76, 1128 (2000).
37. S. M. York, C. R. Jenkins, S. L. Silva, and F. M. Leibsle, *Surf. Sci.* 464, L752 (2000).
38. S. M. York and F. M. Leibsle, *Phys. Rev. B* 64, 033411 (2001).
39. S. Ohno, K. Nakatsuji, and F. Komori, *Surf. Sci.* 523, 189 (2003).
40. K. Muller, E. Lang, L. Hammer, W. Grimm, P. Heilman, and K. Heinz, in "*Determination of Surface Structure by LEED*," p. 483. (P. M. Marcus and F. Jona, Plenum, New York, 1984.
41. J.-L. Li, X.-J. Liang, J.-F. Jia, X. Liu, J.-Z. Wang, E.-G. Wang, and Q.-K. Xue, *Appl. Phys. Lett.* 79, 2826 (2001).
42. A. Kida, H. Kajiyama, S. Heike, T. Hashizume, and K. Koike, *Appl. Phys. Lett.* 75, 540 (1999).
43. P. Soukiassian, *Mater Sci. Eng. B* 61–62, 506 (1999).
44. P. Soukiassian, F. Semond, A. Mayne, and G. Dujardin, *Phys. Rev. Lett.* 79, 2498 (1997).
45. V. Derycke, P. Soukiassian, A. Mayne, G. Dujardin, and J. Gautier, *Phys. Rev. Lett.* 81, 5868 (1998).
46. H. Brune, M. Giovannini, K. Bromann, and K. Kern, *Nature (London)* 394, 451 (1998).
47. L. Nedelmann, B. Müller, B. Fischer, K. Kern, D. Erdös, J. Wollschläger, and M. Henzler, *Surf. Sci.* 376, 113 (1997).
48. H. Jödicke, R. Schaub, R. Monot, J. Buttet, and W. Harbich, *Surf. Sci.* 475, 109 (2001).
49. Y.-W. Mo, D. E. Savage, B. S. Swartzentruber, and M. G. Lagally, *Phys. Rev. Lett.* 65, 1020 (1990).
50. O. Leifeld, E. Müller, D. Grützmacher, B. Müller, and K. Kern, *Appl. Phys. Lett.* 74, 994 (1999).
51. A. K. Das, J. Kamila, B. N. Dev, B. Sundaravel, and G. Kuri, *Appl. Phys. Lett.* 77, 951 (2000).
52. D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, and P. M. Petroff, *Appl. Phys. Lett.* 63, 3203 (1993).
53. R. Nötzel, J. Temmyo, and T. Tamamura, *Nature (London)* 369, 131 (1994).
54. J. M. Moison, F. Houzay, F. Barthe, L. Leprince, E. André, and O. Vatel, *Appl. Phys. Lett.* 64, 196 (1994).
55. P. M. Petroff, A. Lorke, and A. Imamoglu, *Phys. Today* 54, 46 (2001).
56. O. G. Schmidt, S. Kiravittayaa, Y. Nakamuraa, H. Heidemeyera, R. Songmuanga, C. Müllera, N. Y. Jin-Phillippa, K. Eberla, H. Wawrab, S. Christiansenc, H. Gräbeldingerd, and H. Schweizer, *Surf. Sci.* 514, 10 (2002).
57. P. M. Petroff, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.* 45, 620 (1984).
58. M. Tsuchiya, P. M. Petroff, and L. A. Coldren, *Appl. Phys. Lett.* 54, 1690 (1989).
59. J. Viernow, J.-L. Lin, D. Y. Petrovykh, F. M. Leibsle, F. K. Men, and F. J. Himpsel, *Appl. Phys. Lett.* 72, 948 (1998).
60. Y. Takagi, A. Nishimura, A. Nagashima, and J. Yoshino, *Surf. Sci.* 514, 167 (2002).
61. M. Shibata, I. Sumita, and M. Nakajima, *Phys. Rev. B* 57, 1626 (1998).
62. R. Losio, K. N. Altmann, A. Kirakosian, J.-L. Lin, D. Y. Petrovykh, and F. J. Himpsel, *Phys. Rev. Lett.* 86, 4632 (2001).

63. I. K. Robinson, P. A. Bennett, and F. J. Himpsel, *Phys. Rev. Lett.* 88, 096104 (2002).
64. A. Kirakosian, J. L. McChesney, R. Bennewitz, J. N. Crain, J. L. Lin, and F. J. Himpsel, *Surf. Sci.* 498, L109 (2002).
65. H. W. Yeom, S. Takeda, E. Rotenberg, I. Matsuda, K. Horikoshi, J. Schaefer, C. M. Lee, S. D. Kevan, T. Ohta, T. Nagao, and S. Hasegawa, *Phys. Rev. Lett.* 82, 4898 (1999).
66. D. Y. Petrovykh, F. J. Himpsel, and T. Jung, *Surf. Sci.* 407, 189 (1998).
67. Y. M. Mo, and F. J. Himpsel, *Phys. Rev. B* 50, 7868 (1994).
68. J. de la Figuera, M. A. Huerta-Garnica, J. E. Prieto, C. Ocal, and R. Miranda, *Appl. Phys. Lett.* 66, 1006 (1995).
69. F. J. Himpsel, *J. Electron Spectrosc. Relating Phenom.* 75, 187 (1995).
70. A. R. Bachmann, A. Mugarza, J. E. Ortega, and S. Speller, 64, 153409 (2001).
71. S. Fölsch, A. Helms, S. Zöphel, J. Repp, G. Meyer, and K. H. Rieder, *Phys. Rev. Lett.* 84, 123 (2000).
72. S. Fölsch, A. Helms, A. Riemann, J. Repp, G. Meyer, and K. H. Rieder, *Surf. Sci.* 497, 113 (2002).
73. S. Fölsch, A. Riemann, J. Repp, G. Meyer, and K. H. Rieder, *Phys. Rev. B* 66, 161409 (2002).
74. I. Mauch, G. Kaindl, and A. Bauer, *Surf. Sci.* 522, 27 (2003).
75. F. K. Men, F. Liu, P. J. Wang, C. H. Chen, D. L. Cheng, J. L. Lin, and F. J. Himpsel, *Phys. Rev. Lett.* 88, 096105 (2002).
76. A. A. Baski, K. M. Saoud, and K. M. Jones, *Appl. Surf. Sci.* 182, 216 (2001).
77. G. M. Whitesides, J. P. Mathias, and C. T. Seto, *Science* 254, 1312 (1991).
78. M. Böhringer, K. Morgenstern, W.-D. Schneider, R. Berndt, F. Mauri, A. De Vita, and Roberto Car, *Phys. Rev. Lett.* 83, 324 (1999).
79. K. Morgenstern, E. Laegsgaard, I. Stensgaard, F. Besenbacher, M. Bohringer, W. D. Schneider, R. Berndt, F. Mauri, A. De Vita, and R. Car, *Appl. Phys. A* 69, 559 (1999).
80. G. Contini, V. Di Castro, A. Angelaccio, N. Motta, and A. Sgarlata, *Surf. Sci.* 470, L7 (2000).
81. C. Mariani, F. Allegretti, V. Corradini, G. Contini, V. Di Castro, C. Baldacchini, and M. G. Betti, *Phys. Rev. B* 66, 115407 (2002).
82. C. B. France, P. G. Schroeder, and B. A. Parkinson, *Nano Letters* 2, 693 (2002).
83. G. P. Lopinski, D. D. M. Wayner, and R. A. Wolkow, *Nature (London)* 406, 48 (2000).
84. M. Ortega Lorenzo, S. Haq, P. Murray, R. Raval, and C. J. Baddeley, *J. Phys. Chem.* (*London*) 103, 10661 (1999).
85. M. Ortega Lorenzo, C. J. Baddeley, C. Muryn, and R. Raval, *Nature (London)* 404, 376 (2000).
86. J. K. Gimzewski and C. Joachim, *Science* 283, 1683 (1999).
87. M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, and J. M. Tour, *Science* 278, 252 (1997).
88. M. Di Ventra, S. T. Pantelides, and N. D. Lang, *Appl. Phys. Lett.* 76, 3448 (2000).
89. M. A. Reed, *N. Y. Acad. Sci.* 852, 133 (1998).
90. A. F. Morpurgo and C. M. Marcus, *Appl. Phys. Lett.* 74, 2084 (1999).
91. G. Phillip, T. Weimann, P. Hinze, M. Burghard, and J. Weis, *Microelectron. Eng.* 46, 157 (1999).
92. H. Park, A. K. L. Lim, A. P. Alivisatos, J. Park, and P. L. McEuen, *Appl. Phys. Lett.* 75, 301 (1999).
93. P. B. Fischer and S. Y. Chou, *Appl. Phys. Lett.* 62, 2989 (1993).
94. K. Liu, Ph. Bucchignano, R. Martel, and S. Sun, *Appl. Phys. Lett.* 80, 866 (2002).
95. M. S. M. Saifullah, T. Ondarcuhu, D. K. Koltsov, C. Joachim, and M. E. Welland, *Nanotechnology* 13, 659 (2002).
96. J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruña, P. L. McEuen, and D. C. Ralph, *Nature (London)* 417, 722 (2002).
97. W. Liang, M. P. Shores, M. Bockrath, J. R. Long, and H. Park, *Nature (London)* 417, 725 (2002).
98. C. Lin and C. R. Kagan, *J. Am. Chem. Soc.* 125, 336 (2003).
99. M. Cavallini, F. Biscarini, S. Léon, F. Zerbetto, G. Bottari, and D. A. Leigh, *Science* 299, 531 (2003).
100. C. P. Collier, E. W. Wong, M. Belohradský, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, and J. R. Heath, *Science* 285, 391 (1999).
101. U .Valbusa, C. Boragno, and F. B. de Mongeot, *J. Phys. Condens Matter* 14, 8153 (2002).
102. M. V. R. Murty, *Surf. Sci.* 500, 523 (2002).
103. T. M. Mayer, E. Chason, and A. J. Howard, *J. Appl. Phys.* 76, 1633 (1994).
104. R. Gago, L. Vázquez, R. Cuerno, M. Varela, C. Ballesteros, and J. M. Albella, *Appl. Phys. Lett.* 78, 3316 (2001).
105. F. Ludwig Jr., C. R. Eddy Jr., O. Malis, and R. L. Headrick, *Appl. Phys. Lett.* 81, 2770 (2002).
106. J. Erlebacher, M. J. Aziz, E. Chason, M. B. Sinclair, and J. A. Floro, *Phys. Rev. Lett.* 82, 2330 (1999).
107. N. G. Shang, F. Y. Meng, F. C. K. Au, Q. Li, C. S. Lee, I. Bello, and S. T. Lee, Adv. Mat. 14, 1308 (2002).
108. E. Chason, T. M. Mayer, B. K. Kellerman, D. T. McIlroy, and A. J. Howard, *Phys. Rev. Lett.* 72, 3040 (1994).
109. S. Facsko, T. Dekorsy, C. Koerdt, C. Trappe, H. Kurz, A. Vogt, and H. L. Hartnagel, *Science* 285, 1551 (1999).
110. S. Facsko, H. Kurz, and T. Dekorsy, *Phys. Rev. B* 63, 165329 (2001).
111. F. Frost, A. Schindler, and F. Bigl, *Phys. Rev. Lett.* 85, 4116 (2000).
112. Y. J. Chen, J. P. Wang, E. W. Soo, L. Wu, and T. C. Chong, *J. Appl. Phys.* 91, 7323 (2002).
113. J. Lindner, P. Poulopoulos, M. Farle, and K. Baberschke, *J. Magn. Magn. Mater* 218, 10 (2000).
114. S. Rusponi, C. Boragno, and U. Valbusa, *Phys. Rev. Lett.* 78, 2795 (1997).
115. S. Rusponi, G. Costantini, C. Boragno, and U. Valbusa, *Phys. Rev. Lett.* 81, 4184 (1998).
116. D. Sekiba, S. Bertero, R. Buzio, F. B. de Mongeot, C. Boragno, and U. Valbusa, *Appl. Phys. Lett.* 81, 2632 (2002).
117. M. Batzill, F. Bardou, and K. J. Snowdon, *Phys. Rev. B* 63, 233408 (2001).
118. B. Kahng, H. Jeong, and A. L. Barabasi, *Appl. Phys. Lett.* 78, 805 (2001).
119. J. Kim, B. Kahng, and A. L. Barabasi, *Appl. Phys. Lett.* 81, 3654 (2002).
120. G. S. Lau, E. S. Tok, R. Liu, A. T. S. Wee, W. C. Tjiu, and J Zhang, *Nanotechnology* 14, 1187 (2003).
121. J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, and J. A. Golovchenko, *Nature (London)* 412, 166 (2001).
122. T. Yoshida, S. Takeyama, Y. Yamada, and K. Mutoh, *Appl. Phys. Lett.* 68, 1772 (1996).
123. J. D. Fowlkes, A. J. Pedraza, D. A. Blom, and H. M. Meyer III, *Appl. Phys. Lett.* 80, 3799 (2002).
124. K. Kawamura, N. Sarukura, M. Hirano, N. Ito, and H. Hosono, *Appl. Phys. Lett.* 79, 1228 (2001).

www.aspbs.com/enn

Semiconductor Nanodevice Modeling

Eric A. B. Cole

University of Leeds, Leeds, United Kingdom

CONTENTS

1. Nanodevices and Modeling
2. Equations and Their Derivation
3. Numerical Methods of Solution
4. Finite Difference Methods
Glossary
References

1. NANODEVICES AND MODELING

1.1. Why Semiconductor Nanodevices?

We are used to the presence of computer silicon chips in everyday objects—automobile engines, washing machines, and relatively inexpensive children's toys. They are even being inserted into furniture flat-packs to ensure that the components are assembled in the correct order [1]. A common feature of these everyday chips is that they do not need to work at very high speed (relatively speaking in chip terms) and they do not need to be too small in size. These are devices whose size is measured in micrometers (one micrometer equals one-millionth of a meter, or 1000 nanometers), and the prefix "nano" is never applied to them because the prefix "micro" is sufficient to describe their size. However, the rate at which new electronic devices is being developed grows ever larger. There is a need for new scientific measuring instruments and detectors to become more sensitive, and success in the consumer market means that devices must be smaller and more efficient. Their development depends crucially on the development of the individual components which make up the devices, and there is an overall trend for these components also to be smaller and work more rapidly. In particular, developments in the manufacture of thin films have made it possible to fabricate more complex layers of material, down to the limits of a single atomic layer. As described below, these pave the way for the preparation of very small and fast transistors onto chips, leading to devices which operate at very high frequencies.

So what types of new devices are we talking about? The development of mobile telephones affords a very good example of the way in which devices are being developed for minutarization and speed of operation. In any communications channel, a higher bandwidth means that information can be sent faster and in greater amounts, and a higher bandwidth requires that communication channels must be able to handle signals of higher frequency. Frequency is measured in *hertz* (Hz), or cycles per second. Processors in today's personal computers operate at frequencies typically between 500 MHz and 2.5 GHz (1 MHz, or one megahertz, is one million hertz; 1 GHz, or one gigahertz, is 1000 megahertz). The first generation of mobile telephones operated at frequencies between 900 MHz and 1.9 GHz. The new generation of mobile telephones allows much more information to be sent, including photographs, and so much more bandwidth is needed for these devices; a group at the University of Leeds is now working on designing and producing devices which will operate up to 2000 GHz. More bandwidth also means more efficient microwave devices for use in, for example, receivers for satellite TV and radio transmitters [2]. A recent development has taken place in the use of high frequency devices for use in hidden weapons detection; metal detectors cannot locate plastic or ceramic weapons, and X-ray imaging is potentially harmful. To overcome this problem, a group at the Pacific Northwest National Laboratory has developed a method of producing holographic screened images using detectors working at frequencies of typically 100 GHz. An increase in bandwidth will also mean that one day we could have nano cameras circulating through the human body sending information to outside receivers. Another major area of development is in optoelectronic devices, which directly interface optics and electronics. An example of this technology lies in the amplification of optical signals passing through fibers. In long stretches of fiber, for example in undersea communications, optical signals degrade and must be boosted before passing down the next stretch of fiber. Previous methods have relied on the conversion of the weakened optical signal to electrical ones, but a much more efficient system is to allow the optical signal to stimulate semiconducting material which itself will directly generate light for onward transmission. Another application of this optoelectronic technology lies in environmental control through the detection of light which has been degraded by passing through atmospheric polution [3]. However, our old friend silicon (Si) is not efficient at generating light,

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (653–667)

and much more efficient semiconductor generators such as gallium arsenide (GaAs) are being developed.

What are the physical processes in our devices which determine how fast they work? The answer contains many strands, but the two main ones are size and material composition. To show how size influences the situation, consider a very basic device consisting of a lump of semiconducting material with two metal contacts at some distance apart on its surface—we will describe more specific and complex devices in some detail later. The device is switched on by applying a voltage across the contacts, and a resulting current is caused by the transit of *electrons* through the material, from one contact to the other. (current is also carried by the *holes* left behind in the lattice structure which electrons have vacated, but more of this later). If we want to use this device for rapid switching, then the speed of this switching (the frequency of which is given by the number of switches per second) depends on how rapidly an electron can pass from one contact to the other. Clearly, by making the device smaller we are able to cut down on the transit time and hence increase the number of switches per second, or the frequency range, of the device, and this is partly why smaller devices are being developed. Higher frequencies mean larger bandwidth, resulting in the ability to send more traffic down a given channel. It is therefore truly remarkable that if we want to increase this traffic then we have to use smaller devices! It is not only the smallness of size that is important in increasing this capacity, but the composition of the material used is also important in reducing transit times. If electrons were swimmers, they would prefer not to swim through a treacle-like substance such as silicon, but they would find the journey much less arduous to swim through a less viscous substance such as gallium arsenide. In material terms, we say that gallium arsenide has greater *mobility* than silicon; we will show later how the value of the mobility is used. This swimming analogy, although extremely simplistic, is quite a reasonable one; as we shall see later, it turns out that in some theoretical process we can treat the electron transport in materials as a fluid flow.

However, it is plain that a further reduction in size must affect the way in which we think of electron transport in our transistors. Let us do a simple calculation involving some typical physical values to illustrate this point. Suppose that in the simple device described above, the distance between the contacts is one micrometer (a micrometer, denoted by the symbol μm, is one-millionth of a meter), and it is one-tenth of a micrometer thick and one centimeter wide (or 10,000 micrometers). Note that one of these lengths is very much larger than the others; we will use this fact when we come to mathematically model such a device. Then the volume of the device is 1000 cubic micrometers (10^3 μm^3). If we take a typical electron density of 100 electrons per cubic micrometer, then at any one time there will be $10^3 \times 10^2 = 10^5$ electrons in the device. We can easily reduce this number by a factor of 10, to a value of 10,000, by just over halving each dimension. Thus we soon get to the stage where, with only 10,000 particles in the transistor at any one time, we can no longer treat the electron flow as we would a continuous fluid. This picture of electron flow as a continuous fluid, together with the use of the Boltzmann transport equation which we will describe later, has been the basis of our efforts to describe the physical nature of electron transport in semiconductors. Any further reduction in size of our devices will mean that we will have to employ drastically new theoretical methods to describe this transport. Extra problems arise if our well-known devices are reduced further in size. The current that such a device is able to pass will be limited if the device is too thin. Further, hotspots with extremely high temperatures can occur in nanoscale devices, and then lattice vibrations will scatter the conduction electrons in an uncontrollable way. It must therefore be clear that there is a limit to which we are able to reduce the size of our well tried and tested devices.

We have described a device containing 10,000 electrons at one time, but amazingly we can create devices which contain only 1 electron at a time, leading to the exciting prospect of single-electron transistors [4–6]. Such devices exploit the essentially quantum nature of electron behavior, because they cannot be treated as classical particles which obey Newton's laws of motion. To illustrate this point, imagine what would happen if you were to suddenly stand up and run into the nearest wall—please, *do not* try this. Experience, and the predictions of Newtonian mechanics, suggests that you will bounce back off the wall. However, quantum theory predicts that electrons can perform the trick of passing through an equivalent wall and travel on the other side. This process is called *tunneling*—they pass through potential barriers that would be impenetrable to them classically. This behavior is exploited in the single-electron transistor, in which an electron can tunnel into a "box" made up of suitably positioned potential barriers approximately 1 nanometer apart. By altering the applied voltage around the barriers, electrons can be made to tunnel in and out of the box one at a time so that the charge inside the box can be counted in discrete units. These electrons can be confined in all three space directions, and for this reason such a device is known as a *quantum dot*. Since electrons traverse the dot one at a time, such an arrangement can be used for extremely sensitive measurements of electric charge. Another exciting possibility is that these dots could be used for manipulating quantum bits that will be needed for the implementation of *quantum computing* [7]. Another exciting prospect arises from the fact that a number of electrons confined within a quantum dot have interactions unlike those seen between electrons in an atom. Effectively, artificial atoms can be created, and an exciting challenge is to make new material composed of these new artificial "atoms." Many problems lie ahead in exploiting these new ideas—for example, how to create these dots to a consistent pattern rather than randomly, how to link these "atoms" together to form new material, and how to do all this at room temperature rather than at the low temperatures required at present. These are physical constructs for which the prefix "nano" really is more appropriate than the prefix "micro."

1.2. Process of Nanodevice Modeling

Novel manufacturing techniques must be developed in order to produce these new nanotransistors, but first new prototypes and experimental devices must be made before they are ready for the mass market. It is extremely expensive

to set up the fabrication tools for these prototype devices, especially if they are found not to work as required. Thus it is no longer economically feasible to fabricate experimental designs without at first producing mathematical models which will predict how these devices will perform. In the same way that new materials and fabrication processes are being developed, so too are new mathematical techniques being developed in order to predict the behavior of these new devices.

There are three main strands to the modeler's task. First, it must be decided what physical processes are to be included, and how to incorporate them into equations. Too many physical effects will make the task cumbersome and impossible to complete, while too few will make the model physically unrealistic. The equations which arise at this stage of the modeling process are almost always nonlinear partial differential equations. The second strand involves the solution of these equations, always assuming of course that they have been set up in a way that they *do* possess a solution. Since the equations are highly nonlinear [for example, the lifetime τ_e appearing in Eq. (4) is itself a function of the solution, whose form must be inserted numerically], then it is possible to obtain exact analytical solutions to the equations only if the physical content is idealized to such an extent that the device is physically unrealistic. Hence, the solution of the equations for realistic devices invariably involves numerical computing techniques. The third strand involves the interpretation of the numerically computed results. Having obtained a numerical solution to the equations which model a particular physical configuration of a device, physical characteristics may then be computed and compared with experimental results, if there are any. If a prototype device has already been manufactured, and the computed results *do not* compare favorably with the measured ones, then the mathematical model must be changed in some way—for example, by modeling extra physics. If the computed results *do* compare favorably, then the modeler will have the confidence to suggest improvements to the physical configuration by suggesting, for example, that doping profiles or device dimensions be changed at manufacture.

We will illustrate the role of device modeling by concentrating on one particular nanodevice—the high electron mobility transistor (HEMT). This is a device which has improved transport qualities compared with the more basic metal Semiconductor field effect transistor (the MESFET) in the way described below. Although these devices are only two of many devices in existence, a description of the process of modeling them illustrates common features applicable to the modeling of many nanodevices. A typical HEMT or MESFET structure is illustrated schematically in Figure 1 and consists of layers of different semiconducting material, together with at least three contacts—the *source*, *gate*, and *drain*. The source and drain are ohmic contacts (that is, they are perfect sources and sinks providing no injection of charge carriers), while the gate is a Schottky contact (a reverse contact bias at such a metal—semiconductor interface blocks the flow of electrons and creates a region that is depleted of electrons; this is modeled using the built-in potential ψ_b introduced in Section 2) [8]. The gate is often recessed into the device as shown. A typical gate length

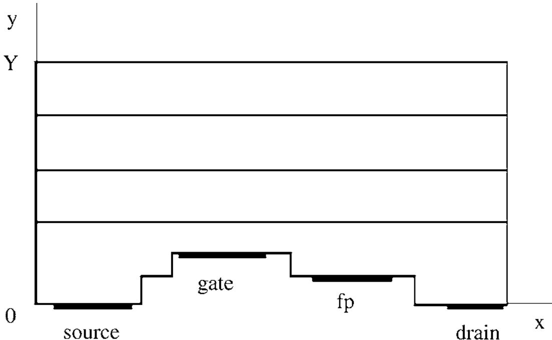


Figure 1. A typical HEMT or MESFET structure showing the source, gate, and drain contacts. There are four layers and two recesses in this figure. A fourth field plate contact (fp) is often added in order to spread the electric field in the device.

usually lies in the range of 100 to 500 nanometers. Voltages V_g and V_d relative to the source are applied to the gate and drain respectively, and a current flows through the device. The current–voltage (I_d–V_d) characteristic curve has a device-dependent profile for a fixed value of V_g. By varying V_g, this characteristic profile is changed because the effect of changing V_g is to cause a larger or smaller electron depletion region around the gate. Generally in a semiconductor, current is carried by both electrons and holes (they are *bipolar* devices), but in the MESFET or HEMT the current is dominated by the electron flow, and hole current is neglected (the *unipolar* model).

The layers of the MESFET consist mainly of *n*- or *p*-type semiconducting material, most often of GaAs which has a higher electron mobility than that of silicon (that is, for given electron density and electric field values, the current is higher). The disadvantage of this device is that the conduction electrons suffer many collisions with the structure of the crystal lattice. This limits the speed at which the device works. A much more efficient arrangement, as far as shorter transit time is concerned, is the HEMT whose structure consists of, for example, GaAs and aluminum gallium arsenide (AlGaAs). This device works because conduction electrons fall into potential wells near the AlGaAs interfaces where they suffer far fewer collisions. These wells are typically only several nanometers wide, and the electrons are able to transit the device virtually unhindered in thin sheets parallel to the interfaces. Again, an analogy of a fluid flow springs to mind here, where rainwater falling onto a roof rolls down into a gutter which carries it quickly away.

A typical HEMT device will consist of many different layers of GaAs and $Al_uGa_{1-u}As$ with differing mole fractions u of Al. The depth of the device in the direction of the axis Oy is typically of an order of magnitude smaller than its length in the direction of the axis Ox. Its length along the third axis Oz (not shown in Fig. 1) is typically many orders of magnitude larger, and it is common to consider only a two-dimensional model for a slice perpendicular to Oz. The modeler may seek to predict the effects of altering many physical parameters of the system; for example, device characteristics may be varied by changing the various physical dimensions of the device, or by changing the number of layers, or by changing the Al mole fractions u of the layers, or by changing the electron doping densities. A fairly recent

development has been to add an extra contact—the *field plate* (denoted by *fp* in Fig. 1) [9, 10]. This has the effect of raising the operating voltages of HEMTs by spreading out hotspots in the electric field.

There are several different levels at which the modeling of devices may be achieved, with the simpler processes not achieving the degree of flexibility required in dealing with multilayer structures. Computer aided design (CAD) packages exist which exploit most of these processes. The first main method to be used was the *equivalent circuit method* [11–14] in which equivalent electric circuits are constructed to mimic device characteristics. Although still used in many CAD packages, it is a very blunt approach and is not capable of taking the subtleties of solid state processes into account. At the other end of the scale is the *Monte Carlo* technique [15–17] in which individual electron paths are computed once all scattering data are given. This approach can produce excellent results but, given the very large number of particles needing to be tracked, it is extremely expensive in computational cost. A third approach, which is the one we shall concentrate on in this work, is that of *semi-classical modeling*. In this method, the Poisson equation is solved together with equations which arise by taking the first few moments of the Boltzmann transport equation. Data for lifetimes, mobilities, etc. are input in numerical form having first been calculated (in previous simulations) using the Monte Carlo technique. This method allows fairly rapid implementation while at the same time allowing subtle solid state processes to be incorporated. One of the simplest methods in this semiclassical approach is the *drift diffusion approximation* [18, 19] in which the Poisson equation is solved along with the current continuity equation. This method produces reasonable results for devices with relatively large physical dimensions. However, for nanodevices which are considerably smaller, in which high electric fields are generated, the *hydrodynamic* model [20, 21] must be used to account for the effect of electron heating. In this approach, the Poisson equation is solved with both the current continuity and energy transport equations. In the case of the HEMT in which quantum wells are formed at the interfaces, we must also solve the Schrödinger eigenvalue equation of quantum mechanics. A number of eigensolutions of this equation are then used to compute the electron number and energy densities at each point in the device.

In Section 2 we introduce the differential equations which describe the hydrodynamic model for the HEMT. In Section 3 we discuss several main methods which are used to obtain numerical solutions of the equations. In Section 4 we will concentrate on the most widely used of these methods, namely, the *finite difference* method. We will discuss the main techniques which are employed in implementing this method and give an example of a solution for a specific multilayer HEMT.

2. EQUATIONS AND THEIR DERIVATION

2.1. Main Equations

When attempting to model a particular device, we must first decide what variables will best describe its physical nature. In semiconductor device modeling, the most natural variables will be the electrostatic potential ψ, electron temperature T_e, electron density n, hole density p, quasi Fermi level $E_F = -q\phi$, total energy density W, electron current density $\mathbf{J}$, electric field $\mathbf{E}$, and the average carrier energy ξ. Not all of these quantities are independent, and the relations between them will be outlined below.

The transport of charge carriers in semiconductors must obey *Maxwell's electromagnetic equations*

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$
$$\nabla \cdot \mathbf{D} = \rho, \qquad \nabla \cdot \mathbf{B} = 0$$

where $\mathbf{B}$ is the magnetic induction, $\mathbf{D}$ is the electric displacement, $\mathbf{H}$ is the magnetic field, and ρ is the electric charge density. This last equation implies that $\mathbf{B} = \nabla \times \mathbf{A}$ for some vector potential $\mathbf{A}$. The first equation then implies that $\nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = \mathbf{0}$, or equivalently that $\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla\psi$ for some scalar potential ψ. In the modeling of many devices, and especially those being considered here, the term $\frac{\partial \mathbf{A}}{\partial t}$ may be neglected due to its relatively small size, and so the electric field will be given by $\mathbf{E} = -\nabla\psi$. We further assume that the permittivity is independent of time and that polarization due to mechanical forces is negligible, so that we may write $\mathbf{D} = \epsilon_0\epsilon_r\mathbf{E}$ where $\epsilon_0 = 8.85 \times 10^{-12}$ F m^{-1} is the permittivity of free space, and ϵ_r is the relative permittivity of the material (13.18 in the case of GaAs). The third of the four Maxwell equations then gives $\nabla \cdot (\epsilon_0\epsilon_r\nabla\psi) = -\rho$. If N_D is the doping density imposed at manufacture, then we may write $\rho = q(N_D - n)$ where $q = 1.6 \times 10^{-19}$ C is the magnitude of the electron charge. We finally have Poisson's equation

$$\nabla \cdot (\epsilon_0\epsilon_r\nabla\psi) = -q(N_D - n) \tag{1}$$

The starting point for the modeling of the transport dynamics is the Boltzmann transport equation

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{k}} \cdot \frac{\partial \mathbf{k}}{\partial t} + \frac{\partial f}{\partial \mathbf{r}} \cdot \frac{\partial \mathbf{r}}{\partial t} = 0 \tag{2}$$

for the distribution function $f(\mathbf{k}, \mathbf{r}, t)$ which is given in terms of the particle momentum $\hbar\mathbf{k}$, position $\mathbf{r}$, and time t. The interpretation of this function is that $f(\mathbf{k}, \mathbf{r}, t)\mathbf{d}^3\mathbf{r}\mathbf{d}^3\mathbf{k}$ is the number of particles in the space region $\mathbf{d}^3\mathbf{r}$ at position $\mathbf{r}$ and momentum region $\mathbf{d}^3\mathbf{k}$ at momentum $\hbar\mathbf{k}$ at time t, where $2\pi\hbar$ is Planck's constant $h = 6.624 \times 10^{-34}$ J s. We then take a *moment* of this equation by multiplying through by some function g of the variables and integrating over all $\mathbf{k}$-space [22]. We take the first three moments of this equation by taking $g = 1$, $m\mathbf{v}$, and $\frac{1}{2}mv^2$ in turn, where m and $\mathbf{v}$ are the particle mass and velocity respectively. Then using various simplifying assumptions [23, 24] we obtain the following set of equations:

(i) *the continuity equation for electrons*

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J} \tag{3}$$

Normally, terms involving generation and recombination rates for charge carriers are present on the right hand side of this equation. However, they are neglected here for the unipolar MESFET model.

(ii) *the energy transport equation*

$$\frac{\partial W}{\partial t} = \mathbf{J} \cdot \mathbf{E} - \frac{(W - W_0)}{\tau_e} - \nabla \cdot \mathbf{s} \tag{4}$$

Here, the term $\mathbf{J} \cdot \mathbf{E}$ represents the Joule heating (in any conductor carrying a current, heat called the Joule heat is produced). The second term on the right hand side of this equation represents the relaxation of the total energy density W back to the equilibrium energy density W_0. If we write $W = n\xi$, then the energy relaxation lifetime τ_e is generally found to be a function of ξ. The quantity $\mathbf{s}$ is the energy flux described below.

2.2. Subsidiary Equations

We first consider the set of equations in relation to MESFET modeling, in which there are no quantum wells formed at the layer interfaces. In this case there is no need to solve the Schrödinger equation explicitly. For the *nondegenerate approximation*, which is valid when the pure semiconductor is not too highly doped with impurities, the electron density is given by the simple exponential form

$$n = N_c \exp\left(\frac{q}{k_B T_e}(\psi - \phi)\right) \tag{5}$$

Here, $k_B = 1.38 \times 10^{-23}$ J K^{-1} is the Boltzmann constant, and $N_c \equiv 2(2\pi m^* k_B T_e h^{-2})^{1.5}$ where m^* is the effective mass of the electron. In this nondegenerate case, the expression for the current density $\mathbf{J}$ in Eq. (3) and the energy flux $\mathbf{s}$ in Eq. (4) can be written

$$\mathbf{J} = -qn\mu\nabla\psi + k_B\mu T_e \nabla n + ak_B\mu n\nabla T_E + bk_B T_e n\nabla\mu \tag{6}$$

$$\mathbf{s} = -\mu W\mathbf{E} - \frac{k_B}{q}\nabla(\mu W T_e) \tag{7}$$

where μ is the electron mobility. The various terms on the right hand side of Eq. (6) are explained as follows. Remembering that the electric field is $\mathbf{E} = -\nabla\psi$, the first term is $qn\mu\mathbf{E}$ and represents the drift of the charged electrons in this field. The second term represents the diffusion of the particles, irrespective of their charge. The third term is the contribution from the diffusion of the energy, and the fourth term should really be included as part of the term involving ∇n. The expression for $\mathbf{s}$ is similarly interpreted; the first term is the energy drift and the second is the energy diffusion. The values of the constants a and b are taken differently by several authors. Tang [25] takes $a = 1$ and $b = 1$, McAndrew et al. [26] and Selberherr [27] take $a = \frac{1}{2}$ and $b = 0$, and Snowden and Loret [28] and Feng and Hintz [29] take $a = 1$ and $b = 0$. The position-dependent average electron energy ξ is normally given by

$$\xi = \frac{1}{2}m^* v^2 + \frac{3}{2}k_B T_e$$

and it is usual to omit the kinetic energy term $\frac{1}{2}m^* v^2$ which is typically an order of magnitude lower than the thermal energy. Since ξ is the average electron energy, we further have $W = n\xi$ for the total electron energy.

2.3. Physical Data

The above expressions require that the electron mobility μ, effective mass m^*, and lifetime τ must be specified in terms of ξ. These specifications are obtained using Monte Carlo simulations and experimental data on the steady state transport charateristics of the bulk material. This provides data for ξ, τ, m^*, and other quantities in terms of the static electric field E_{ss} which then acts as an intermediate variable. The electron mobility is given by

$$\mu = \frac{300\mu_0}{T_0}\left(\frac{1 + \frac{v_s E_{ss}^3}{\mu_0 E_0^4 (1 - 5.3 \times 10^{-4} T_0)}}{1 + (E_{ss}/E_0)^4}\right) \tag{8}$$

where the parameters take the values shown in Table 1 for several materials. For $Al_uGa_{1-u}As$, we take [30] $m^* = (0.067 + 0.083u)m_e$ where $m_e = 9.11 \times 10^{-31}$ kg is the electron rest mass, and $\epsilon_r = 13.18 - 3.12u$. Similar expressions apply for InGaAs.

In the case of the HEMT, we have quantum wells formed at the interfaces of the different material layers, and so the Schrödinger equation must be solved for the discrete eigenstates which result from these wells. We may explain the physical nature of these wells by studying the nature of the energy levels allowed to electrons in a semiconductor. When the Schrödinger equation is solved for a potential caused by a periodic crystal lattice, the energies allowed to the electrons fall into bands, separated by forbidden bands. Electrons which are free to move through the crystal have energies in the *conduction band*. These electrons can lose energy, for example through emitting photons or by colliding with the crystal lattice, and fall into the next lowest band—the *valence band*—where they are tied to the lattice sites. They may also be knocked out of these sites to become conduction electrons again. If E_c is the lowest allowed energy in the conduction band, and E_v is the highest in the valence band, the above generation and recombination processes will depend on the value $E_g = E_c - E_v$ of the energy gap between the two bands. When two different materials with different energy gaps are brought together there will be a mismatch in their values of E_c. The discontinuity step E_h in these values gives rise to the quantum wells, and its value again depends on the mole fraction u of Al or In in the $Al_uGa_{1-u}As$ or $In_uGa_{1-u}As$. We take [30] $E_h = 0.65(1.155u - 0.37u^2)$ for an $Al_uGa_{1-u}As$–GaAs interface.

2.4. Schrödinger Equation for the HEMT

The modeling and numerical simulation of the HEMT involve an extra degree of difficulty, because quantum wells are formed at the interfaces of the different materials,

Table 1. Values of the parameters used in the expression for the electron mobility μ for several materials.

Material	E_0 (V m^{-1})	v_s (m s^{-1})	μ_0 (m^2 V^{-1} s^{-1})
GaAs	4.0×10^5	8.5×10^4	0.8
AlGaAs	7.0×10^5	5.0×10^4	0.2
InGaAs	3.0×10^5	1.0×10^5	1.0

and so the energies allowed to the electrons are partially quantized there. The quantized energy levels $\lambda_{(\nu)}$ with corresponding eigenfunctions $\xi_{(\nu)}$ satisfy the Schrödinger equation

$$-\frac{\hbar^2}{2}\nabla\cdot\left(\frac{1}{m^*}\nabla\xi_{(\nu)}\right)+q(V_{xc}+E_c)\xi_{(\nu)} = q\lambda_{(\nu)}\xi_{(\nu)} \qquad (\nu=0,1,2,\dots) \tag{9}$$

where $E_c = E_h - \psi$ is the conduction band minimum and V_{xc} is the exchange correlation energy. This last term arises purely as a quantum effect and comes when we change from a single-particle description to a many-particle description. Its form may be approximated by [31]

$$V_{xc} = -\left(1+0.03683 r_s \ln\left(1+\frac{21}{r_s}\right)\right)\left(\frac{2}{\pi\alpha r_s}\right)R_y^* \tag{10}$$

where

$$\alpha=\left(\frac{4}{9\pi}\right)^{1/3} \qquad a^*=\frac{4\pi\epsilon_0\epsilon_r h^2}{q^2 m^*} \qquad r_s=\left(\frac{4}{3}\pi a^{*3}n\right)^{-1/3}$$

and $R_y^* = q^2/(8\pi\epsilon_0\epsilon_r a^*)$ is the effective Rydberg constant. Note the careful positioning of the quantity m^* in Eq. (9). This positioning is needed to ensure that the kinetic energy operator is Hermitian (Hermitian operators are necessary because they possess real eigenvalues).

When obtaining a numerical solution of the equations, the most time-consuming part is the solution of the two-dimensional Schrödinger equation. In order to speed up the process, it is convenient to solve the eigenvalue problem in one-dimensional slices perpendicular to the layer structure, that is, along Oy, since the potential varies much more slowly along Ox. Consequently we solve the one-dimensional equation

$$-\frac{\hbar^2}{2}\frac{d}{dy}\left(\frac{1}{m^*}\frac{d\xi_{(\nu)}}{dy}\right)+q(V_{xc}+E_c)\xi_{(\nu)} = q\lambda_{(\nu)}\xi_{(\nu)} \qquad (\nu=0,1,2,\dots) \tag{11}$$

The expressions for the total electron density n and energy density W are more complicated for the HEMT structure, since we must take the contributions from the quantized structure into account. In particular, the electron density n does not obey the nondegenerate expression of Eq. (5). We take the first L eigensolutions (a discussion of how this value is taken is given in the next section) ordered in the sense that $\lambda_{(0)} < \lambda_{(1)} < \cdots < \lambda_{(L-1)}$ and then take integrals involving the density of states. The total electron density is given by $n = n_2 + n_3$ where, outside the well defined by $\lambda_{(L-1)} < E_c$, we take

$$n_2 = 0 \qquad n_3 = N_{c3}F_{\frac{1}{2}}\left(\frac{1}{k_B T_e}(E_F - E_c)\right) \tag{12}$$

where the *Fermi integral* F is defined by

$$F_r(a) \equiv \frac{1}{\Gamma(r+1)}\int_0^\infty \frac{x^r}{1+e^{x-a}}\,dx \tag{13}$$

where $\Gamma(r+1) \equiv \int_0^\infty x^r e^{-x}\,dx$. This replaces the simple exponential form for the nondegenerate case given in Eq. (5). Inside the well we take

$$\begin{aligned} n_2 &= N_{c2}\sum_{\nu=0}^{L-1}|\xi_i(y)|^2 \ln\left(1+e^{\frac{1}{k_BT_e}(E_F-\lambda_{(\nu)})}\right) \\ n_3 &= \frac{1}{2}\pi\left(\frac{8m^*}{h^2}\right)^{\frac{3}{2}}\int_{\lambda_{(L-1)}}^\infty \frac{(E-E_c)^{\frac{1}{2}}}{1+\exp(\frac{1}{k_BT_e}(E-E_F))}\,dE \end{aligned} \tag{14}$$

where the coefficients are defined by $N_{c2} \equiv 4\pi m^* k_B T_e/h^2$ and $N_{c3} \equiv 2(2\pi m^* k_B T_e/h^2)^{1.5}$. Similar expressions hold for the total energy density $W = n\xi = W_2 + W_3$; outside the well we take

$$W_2 = 0 \qquad W_3 = W_{c3}F_{\frac{3}{2}}\left(\frac{1}{k_BT_e}(E_F-E_c)\right) \tag{15}$$

while inside the wells,

$$\begin{aligned} W_2 &= W_{c2}\sum_{\nu=0}^{L-1}|\xi_i(y)|^2\left(F_1\left(\frac{1}{k_BT_e}(E_F-\lambda_{(\nu)})\right)-\frac{1}{k_BT_e}\right. \\ &\qquad \left.\times(E_c-\lambda_{(\nu)})\ln\left(1+e^{\frac{1}{k_BT_e}(E_F-\lambda_{(\nu)})}\right)\right) \end{aligned} \tag{16}$$

$$W_3 = \frac{1}{2}\pi\left(\frac{8m^*}{h^2}\right)^{\frac{3}{2}}\int_{\lambda_{(L-1)}}^\infty \frac{(E-E_c)^{\frac{3}{2}}}{1+\exp(\frac{1}{k_BT_e}(E-E_F))}\,dE$$

where

$$W_{c2} \equiv \frac{4\pi m^* k_B^2 {T_e}^2}{h^2} \qquad \text{and} \qquad W_{c3} \equiv 3\left(\frac{2\pi m^* k_B T_e}{h^2}\right)^{1.5} k_B T_e$$

The complete solution of the equations requires that the boundary conditions be specified. On all of the free surfaces we require that the normal derivatives of ψ, ϕ, and T_e are zero, so that no current or energy flows through these surfaces. On the ohmic source and drain we set ψ equal to the applied voltages V_s and V_d, respectively, while n is set to the doping density. On the Schottky gate we set $\psi = V_g + \psi_b$ where ψ_b is the built-in potential described in Section 1 whose value is taken as −0.8 V, while n is taken as a very small nominal value to account for the depletion region around the gate. The electron temperature T_e is set to the constant lattice temperature T_0 on all three contacts. The eigenfunctions $\xi_{(\nu)}$ are taken to vanish at the extremities of y. If a field plate is incorporated into the structure then we may treat it, in a crude approximation, as a second Schottky gate but with a different value of ψ_b.

The extreme nonlinearity of the problem is now apparent. A study of the close relation between the Poisson equation (1) and the Schrödinger equation (11) illustrates this point very clearly. To solve the Poisson equation for the electrostatic potential ψ we need to know the electron density n on the right hand side of the equation. We calculate n from the solution of the Schrödinger equation using Eqs. (12)–(14), but to solve the Schrödinger equation we need to know ψ because it appears in the potential of that equation (remember that $E_c = E_h - \psi$). Add in the fact that that the equations of current continuity and energy transport also contain these variables, both implicitly and explicitly, and we realize that the only way of solving the equations simultaneously is by numerical means.

3. NUMERICAL METHODS OF SOLUTION

3.1. Overview

The equations described in Section 2 are highly nonlinear: for example, the lifetime τ_e appearing in Eq. (4) is itself a function of the solution, and numerical values of this function must be fed in from experimental data. For any realistic device, therefore, no analytic solutions are possible, and numerical methods must be used. In this section we will outline a number of main numerical methods which have often been used in simulations. One of these—the finite difference method—will be picked out as being the most convenient method to use for many devices, and we will give details of the main techniques for implementing this method in Section 4.

The main solution methods will be illustrated for a situation in which a two-dimensional cross section is appropriate. This situation will apply when, for example, one length (taken as the length in the z direction) is very much larger than the other two—this case will certainly apply, for example, in HEMT or MESFET simulations. This condition may not apply in some situations, and the full three dimensional simulation is appropriate there. On the other hand, some simulations of the *p–n* junction use only a one-dimensional model. In all cases, the methods described here for the two-dimensional situation can be extended for a three-dimensional simulation or restricted for a one-dimensional simulation.

3.2. Finite Difference Method

In this method, the space of the device is overlaid with a rectangular mesh whose lines are parallel to the axes Ox and Oy. Meshpoints $x_0, x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{NI}$ are taken along the x axis and meshpoints $y_0, y_1, \ldots, y_{NJ}$ are taken along the y axis [32, 33]. Along Ox therefore, there are a total of $(NI+1)$ meshpoints, with $(NI-1)$ being internal, and there are NI intervals. Any function $f(x, y)$ which appears in the defining equations is then evaluated at any of the $(NI+1) \times (NJ+1)$ meshpoints, with the notation that $f_{i,j} \equiv f(x_i, y_j)$. In this way, the governing equations are written down at each meshpoint (i, j), and a set of simultaneous equations is obtained for the values of the various physical quantities at each meshpoint: for example, the total electron density n will take the $(NI+1) \times (NJ+1)$ values $n_{i,j}$. The exact ways in which the partial differential equations are written in terms of the discretized values, and the ways in which the resulting nonlinear finite difference equations are solved, are the subjects of a vast body of ongoing research.

The mesh need not be uniform; that is, the mesh spacings need not be all of the same width. We will use the notation that the mesh spacings along Ox and Oy are $h_i \equiv x_{i+1} - x_i$ and $k_j \equiv y_{j+1} - y_j$, respectively, so that, for example, h_i and h_{i-1} are not necessarily equal. It is convenient to use such a nonuniform mesh in devices in which a variable is rapidly changing in some regions and is slowly changing in others. The competing needs for numerical accuracy and speed of solution require that as many meshpoints as possible are used in these regions of rapid change while relatively few are used in the other regions. As the numerical solution progresses, certain mechanisms are employed to check whether or not extra meshpoints should be inserted at various positions. Since the overall size $(NI+1) \times (NJ+1)$ of the mesh determines the speed of solution of the equations, extra meshpoints should not be inserted in regions where they are not needed.

It is often convenient to specify the value of a physical quantity, not at the meshpoint itself, but at the halfpoint which lies equidistant between two meshpoints (we will see this in the specification of the current density **J** later). In this case, the obvious notation to use is $f_{i+(1/2),j} \equiv f(x_i + \frac{1}{2}h_i, y_j)$ and $f_{i,j+(1/2)} \equiv f(x_i, y_j + \frac{1}{2}k_j)$. This notation allows various derivatives to be discretized in a simple manner (for simplicity we will temporarily use one-dimensional notation to describe this process). To obtain the first derivative of the function f at meshpoint i we expand both $f(x_{i+(1/2)})$ and $f(x_{i-(1/2)})$ about $x = x_i$. On subtracting the resulting equations and neglecting the term $-(h_i - h_{i-1})f_i''/4$ and terms involving higher orders of h_i and h_{i-1}, we obtain

$$f_i' = \frac{f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}}{\frac{1}{2}(h_i + h_{i-1})} \tag{17}$$

A similar expression for the second derivative of f at x_i may be derived in the form

$$f_i'' = \frac{\frac{1}{h_i}f_{i+1} - (\frac{1}{h_i} + \frac{1}{h_{i-1}})f_i + \frac{1}{h_{i-1}}f_{i-1}}{\frac{1}{2}(h_i + h_{i-1})} \tag{18}$$

These two expressions are greatly simplified if a uniform mesh is used: with $h_i = h_{i-1} \equiv h$, we have

$$f_i' = \frac{f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}}{h} \qquad f_i'' = \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2}$$

However, if a uniform mesh is used with a sufficient number of meshpoints in regions of rapidly changing variables, then an unnecessarily large number will be inserted in regions where they are not needed. The modeler must decide whether to use too many meshpoints using the simplification of code involving a uniform mesh, or a smaller number of meshpoints using a nonuniform mesh with a resulting complication of the numerical code.

Discretization of the boundary conditions is necessary, and the exact way in which this is done can have an important bearing on the accuracy of the solution. The simplest boundary condition occurs when the value of a quantity is given at a boundary point; for example, the value of the electron temperature T_e may be specified at the contacts. In other cases, the value of the *outward* derivative of the quantity may be specified. The simplest and least accurate method is to use at the end $x = x_0$, for example, the two-point expression $f_0' = (f_0 - f_1)/h_0$. A more accurate expression is the Newton formula

$$f_0' = -\frac{h_1(h_1 + 2h_0)f_0 - (h_0 + h_1)^2 f_1 + {h_0}^2 f_2}{h_0 h_1 (h_0 + h_1)}$$

which involves the value of the function at three points.

The finite difference method can be applied most usefully in situations where the boundaries of various regions

of the device are all parallel to the rectangular axes Ox and Oy, and this is particularly true in the case of MESFET and HEMT modeling. In this situation the boundary conditions are relatively simple to implement. In the case of the HEMT, the interfaces between the various layers of different material are all parallel to one axis (the x-axis Ox). This greatly simplifies the solution of the Schrödinger eigenvalue equation because although a full two-dimensional solution may be obtained, considerable saving may be made in the simulation time if the equation is solved in one-dimensional slices perpendicular to the layer structure. The use of the finite difference approach is very well suited to this construction. This is the most widely used method, and we will describe methods of implementing this method more fully in Section 4.

However, when the device under consideration is not regularly shaped, implementation of boundary conditions along irregularly shaped boundaries becomes difficult when we try to impose a rigid rectangular mesh which is parallel to only two fixed directions. Other methods are then needed, and we will briefly describe two of them—the finite element method and the boundary element method.

3.3. Finite Element Method

This method is particularly suited in application to devices which have irregular geometries, that is, those in which boundaries and interfaces are not parallel to an obvious set of fixed rectangular axes [27, 34–39]. The region of the device is partioned into a finite set of nonoverlapping subregions, or *elements*. In order to apply this method to finding the solution $f(x, y)$ of a certain differential equation it is necessary to approximate f by a simple function defined in each element. Suppose that we are using triangular elements, and a typical triangular element (e) has vertices A, B, and C. The simplest such approximate function defined on triangle ABC is a linear function

$$f_{(e)} = a_0 + a_1 x + a_2 y \tag{19}$$

where a_0, a_1, and a_2 are constants. Such a function is defined on each element (e). Another way of writing this expression is in the form

$$f_{(e)} = N_A(x, y) f_A + N_B(x, y) f_B + N_C(x, y) f_C \tag{20}$$

where f_A, f_B, and f_C are the values of f at A, B, and C respectively, and the functions $N_A(x, y)$, etc. are called the *shape functions*. These shape functions are chosen in such a way that $N_A(x, y) = 1$ when the values of x and y are the coordinates of the vertex A, and $N_A(x, y) = 0$ at all the other vertices of the element. The chosen shape functions must be linear if the expressions in Eqs. (19) and (20) are to be consistent. The next step is to substitute the expression for $f_{(e)}$ into the differential equation satisfied by the function f. An appropriate use of the divergence theorem then produces three simultaneous equations containing f_A, f_B, and f_C on the element (e). This process is done on each element, and the resulting sets of equations are then *assembled* to produce a set of M simultaneous equations for the values of f at each of the total of M vertices. This assembly process is really a matter of bookkeeping, since a vertex will be a member of several elements, and a record has to be kept of all the elements which contain a given vertex.

Standard procedures are then used to solve the set of simultaneous equations for the values of f at each vertex. Clearly, it is advantageous to keep the value M as small as possible, consistent with accuracy, and an important part of the finite element method lies in choosing the initial triangulation [34]. A further important aspect of this method is that the mesh must be refined in certain regions where, for example, the electric field is high. This refinement leads to the insertion of extra vertices resulting in a greater number of triangular elements. Triangular elements are the simplest shapes, and it is possible to extend this triangulation method by inserting extra vertices at the midpoints of the edges. However, triangular elements are not the only ones which may be used. For example, in the case of the MESFET or HEMT in which all boundaries are parallel to either Ox or Oy, rectangular elements may be used. If a typical rectangular element (e) has vertices A, B, C, and D at its corners, then a bilinear function

$$f_{(e)} = a_0 + a_1 x + a_2 y + a_3 xy$$

can be used. The use of rectangular elements rather than triangular elements in appropriate circumstances will reduce the number of vertices needed and hence reduce the computational cost of the simulation.

The finite difference method is easier to code than the finite element method. However, once coded, the finite element method is very adaptable in that it is easy to incorporate changes in geometry and the shape functions. Finite difference methods require much less storage space for bookkeeping arrays, but in these days of inexpensive computer memory the storage requirement of the finite element method presents no problem.

3.4. Boundary Element Method

The basic idea behind this method is that we obtain a set of equations for the discretized variables only at points around the boundary of the device. Clearly, this reduces the number of discretized equations which must be solved. Once the values on the boundary have been determined, it is then possible to determine the values on the internal meshpoints by performing integrals. We will outline this method in the case of obtaining a numerical solution of the Poisson equation

$$\nabla \cdot (\epsilon \nabla \psi) = 0 \tag{21}$$

in which the permittivity ϵ may be a given function of ψ: $\epsilon \equiv \epsilon(\psi)$.

First consider the solution of the Laplace equation

$$\nabla^2 u = 0 \tag{22}$$

for the function u in some internal two-dimensional region Ω. The value of the function is given on some parts of the boundary $\partial\Omega$ and its derivative is given on other parts of $\partial\Omega$. We start with Green's theorem in the form

$$\int_{\partial\Omega} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) dl = \iint_{\Omega} (u \nabla^2 v - v \nabla^2 u)\, ds \tag{23}$$

where we choose the second function v to be such that $\nabla^2 v = \delta(\mathbf{r}_0 - \mathbf{r})$. Here, dl is the infinitesimal increment of length along the boundary, ds is the infinitesimal element of surface area, δ is the Dirac delta function, and $\mathbf{n}$ is the outward normal to $\partial\Omega$. Then Eq. (23) becomes

$$\int_{\partial\Omega}\left(u\frac{\partial v}{\partial n}(\mathbf{r}_0,\mathbf{r}) - v(\mathbf{r}_0,\mathbf{r})\frac{\partial u}{\partial n}\right) dl = \psi(\mathbf{r}_0)\iint_{\Omega}\delta(\mathbf{r}_0,\mathbf{r})\, ds = \eta(\mathbf{r}_0)\psi(\mathbf{r}_0) \qquad (24)$$

where $\mathbf{r}_0$ is any point inside or on the boundary of the region. If we choose v such that $v(\mathbf{r}_0,\mathbf{r}) \equiv v(|\mathbf{r}-\mathbf{r}_0|)$, then it may be shown that the function η is given by

$$\eta(\mathbf{r}_0) = \begin{cases} 2\pi & \text{for } \mathbf{r}_0 \epsilon \Omega \\ \theta & \text{for } \mathbf{r}_0 \epsilon\, \partial\Omega \\ 0 & \text{otherwise} \end{cases}$$

In this expression for η, θ is the angle between the tangents to the boundary curves on opposite sides of $\mathbf{r}_0$ on the boundary, so that boundaries with corners of any angle may be considered. In particular, if the boundary is smooth at $\mathbf{r}_0$ then $\theta = 2\pi$. It further follows from the defining conditions on v that $v(\mathbf{r}_0,\mathbf{r}) = \ln|\mathbf{r}-\mathbf{r}_0|$. We may now apply this to the solution of the Poisson equation (21). If we take $u \equiv \int^{\psi}\epsilon(\phi)\, d\phi$ and define the function h by $h(\psi) \equiv u/\psi$, then Eq. (21) implies Eq. (22), and Eq. (24) may be written

$$\int_{\partial\Omega}[\psi(\mathbf{r})h(\psi(\mathbf{r}))\ln'|\mathbf{r}_0-\mathbf{r}| - \psi'(\mathbf{r})\epsilon(\psi(\mathbf{r}))\ln|\mathbf{r}-\mathbf{r_0}|]\,dl = \eta(\mathbf{r}_0)\psi(\mathbf{r}_0)h(\psi(\mathbf{r}_0)) \qquad (25)$$

The boundary is now discretized into N elements, and Eq. (25) is then written down at the midpoint of each element. This provides N simultaneous nonlinear equations for the $2N$ values of ψ and ψ' at these boundary nodes. However, the number of unknowns is actually N because either ψ or ψ' is given at each node. Once the values on the boundary have been found, the value at any interior point $\mathbf{r}_0$ may be found using Eq. (25) with $\eta(\mathbf{r}_0) = 2\pi$. Further details on the boundary element method may be found in the works of Ingham [40] and Kelmanson [41].

3.5. Time Discretization

When modelers turn their attention to a particular device, they are interested in obtaining numerical solutions for two main modes of operation. The first mode is the *steady state* mode in which, after a certain time has elapsed since switching on the device, it has reached a steady state in which none of the internal variables change with time (this obviously does not apply to devices which operate in oscillator mode). The second mode is the *transient* mode in which numerical results are obtained for the period prior to the steady state mode. This clearly introduces an extra degree of complexity into the simulation, since the number of variables which describe the internal state of the device must be supplemented by the time variable t.

An appropriate time discretization of the equations must be made when we are looking for the transient solution. Take a time step Δt. Then we will use the notation that a function f evaluated at the pth time point given by $t = p\Delta t$ will be denoted by f^p. Suppose, for the sake of illustration, that the current density $\mathbf{J}$ is a function of the electrostatic potential ψ, quasi Fermi level E_F, electron temperature T_e, and electron density n: $\mathbf{J} \equiv \mathbf{J}(n, \psi, E_F, T_e)$. Its value at meshpoint (i, j) at this time is $\mathbf{J}^p_{i,j} \equiv \mathbf{J}(n^p_{i,j}, \psi^p_{i,j}, E^p_{Fi,j}, T^p_{ei,j})$. Then if we discretize the current continuity equation (3) in the form

$$\frac{n^{p+1}_{i,j} - n^p_{i,j}}{\Delta t} = \frac{1}{q}\nabla\cdot\mathbf{J}^p_{i,j}$$

we have an equation which is simple to solve for $n^{p+1}_{i,j}$ in terms of quantities evaluated at the earlier time point:

$$n^{p+1}_{i,j} = n^p_{i,j} + \frac{\Delta t}{q}\nabla\cdot\mathbf{J}^p_{i,j}$$

This explicit equation is trivial to iterate since the right hand side is a function of the variables which are calculated at the earlier time point. Unfortunately, the method will give a solution which diverges from the true solution unless the time step Δt is excessively small [42]. The fully implicit Crank–Nicholson scheme [32]

$$n^{p+1}_{i,j} = n^p_{i,j} + \frac{\Delta t}{2q}(\nabla\cdot\mathbf{J}^{p+1}_{i,j} + \nabla\cdot\mathbf{J}^p_{i,j})$$

will provide greater accuracy and better convergence, but it is difficult to implement since the right hand side now contains all of the variables for which we are solving at the new time point $(p+1)$. A scheme which is easier to implement is the *semi-implicit* scheme in which the only variable involving the new time point on the right hand side is n itself [29, 43]:

$$n^{p+1}_{i,j} = n^p_{i,j} + \frac{\Delta t}{2q}(\nabla\cdot\mathbf{J}(n^{p+1}_{i,j}, \psi^p_{i,j}, E^p_{Fi,j}, T^p_{ei,j}) + \nabla\cdot\mathbf{J}(n^p_{i,j}, \psi^p_{i,j}, E^p_{Fi,j}, T^p_{ei,j}))$$

This is an equation which is implicit in the variable n but explicit in all of the other variables. It does not have the full accuracy and convergence of the fully implicit scheme, but this limitation can be avoided when the space discretization of $\mathbf{J}$ involves a basic upwinding scheme using the Bernoulli function. This scheme will be detailed in Section 4.

4. FINITE DIFFERENCE METHODS

4.1. Overview

Because the geometries of many devices have a rectangular structure, the use of finite differences seems to be the most appropriate approach to use [44–48] and is certainly the most widely used. In this section we will describe some of the more important aspects of the finite difference approach in their applications to device modeling. The initial problem in implementing this approach is how to actually construct the initial mesh, and how it may subsequently be modified as the numerical solution progresses. However the mesh is constructed, the net result is that sets of nonlinear equations are formed which must be solved for the finite number of discretized variables. The main method used is the *Newton*

method, and this method will be outlined with particular reference to device simulation. We will not describe the details of the solution of the equations which result from the application of the Newton method, since details are described in many standard texts [49].

The regular nature of the device geometry also lends itself to the method of *multigrids*, in which an exact solution on a relatively coarse mesh is transferred onto successively finer meshes; this often produces a rapid and accurate solution to the numerical problem.

Any simulation involving the flow of some quantity, whether it be flow of fluid or flow of electrons, benefits from using a discretization involving the technique of *upwinding*. This produces more satisfactory convergence and accuracy properties and will be described in this section.

We discuss the discretization of the Schrödinger equation with particular reference to ensuring that its Hermitian properties are preserved. In all cases, our methods will be illustrated for cases in which two-dimensional approaches are appropriate. Straightforward extensions and restrictions will apply in three-dimensional or one-dimensional cases. Finally, we show how these techniques may be implemented by presenting simulation results for a typical HEMT structure.

4.2. Mesh Construction

We first describe the construction of the mesh used to overlay the device. We need a rectangular mesh which fits with the various physical edges, corners, and interfaces of the device, and one which is capable of being upgraded (downgraded) when it is found that more (fewer) meshpoints are needed in certain regions. Hence, a routine involving dynamic allocation of memory is needed, and the specification of a meshpoint must include links to neighboring meshpoints together with the values of the physical variables there. For example, using the C programming language, we may specify a meshpoint in a two-dimensional mesh as indicated in Table 2. Other quantities may be included in the definition; for example, the components E_x and E_y of the electric field, and the *status* of the meshpoint (interior, exterior, or edge) used in the bookkeeping of the mesh. If a three-dimensional mesh is needed, the typedef definitions will include extensions such as *struct something *up; struct something *down; double z;* etc. Meshpoints are initially assigned to the various corners of the device shown in Figure 1, to the edges of the contacts, and also to the intersections of the layer interfaces with the device edges. Meshpoints are created on opposite edges corresponding to each created meshpoint, and the connections are made. We must then decide on the maximum values of the intervals h and k of the initial mesh, and extra meshpoints are inserted into the basic mesh to ensure that the edge lengths of each new interval are below these maximum values. As the iterations proceed, checks may be made at intervals to determine whether new meshpoints need to be inserted; for example, if the difference in the value of ψ between consecutive meshpoints is greater than some predetermined value then an extra mespoint is inserted midway between the two, and a new line of meshpoints is then drawn through the inserted point. In the same way, unwanted meshpoints may be removed if they are not needed. In both cases, links must be re-established between new adjacent meshpoints.

Table 2. Specification of a meshpoint using the C programming language, indicating typical values defined at a point together with links to four neighboring meshpoints.

```
typedef struct something {
        double x, y;
        double h, k;
        struct something *north;
        struct something *south;
        struct something *east;
        struct something *west;
        double psi, EF, Te;
        etc....................
} meshpoint;
```

In this method, if a new meshpoint is needed, for example, near a gate edge, then a whole line of new meshpoints is inserted. This procedure will put extra meshpoints in regions where perhaps they are not needed, for example, near the far edges of the device. One way of eliminating this problem is to use the *method of boxes* [50] in which meshlines are terminated so that they do not continue into these regions; such a method eliminates unwanted meshpoints but requires extra bookkeeping resources. The method we have described in detail above lies midway in complexity between the method of boxes and that of using a fixed uniform rectangular mesh with no refinement.

4.3. Newton Method

However the mesh is constructed, the values of all the variables which describe the system are given at each meshpoint. Since the partial differential equations which describe the system are second order in the space derivatives and are written down at each internal meshpoint, the result is a set of equations at each point which includes contributions of quantities at the four nearest neighbors of the point. For example, by a simple extension of Eq. (18), a quantity $\nabla^2 f$ will be discretized at the internal meshpoint (i, j) by the five-point formula

$$\nabla^2 f_{i,j} = \frac{\frac{1}{h_i} f_{i+1,j} - (\frac{1}{h_i} + \frac{1}{h_{i-1}}) f_{i,j} + \frac{1}{h_{i-1}} f_{i-1,j}}{\frac{1}{2}(h_i + h_{i-1})} + \frac{\frac{1}{k_j} f_{i,j+1} - (\frac{1}{k_j} + \frac{1}{k_{j-1}}) f_{i,j} + \frac{1}{k_{j-1}} f_{i,j-1}}{\frac{1}{2}(k_j + k_{j-1})} \tag{26}$$

Hence the value of $\nabla^2 f_{i,j}$ involves the values of f evaluated at the five meshpoints (i, j), $(i-1, j)$, $(i+1, j)$, $(i, j-1)$, and $(i, j+1)$. In the approach we are discussing here, the governing equations at the internal points are the Poisson equation, the current continuity equation, and the energy transport equation described in Section 2. These will be supplemented by the equations for the boundary conditions on the edges. Hence there will be three equations written down at each meshpoint, and so there will be three dependent variables for which we must solve at each meshpoint. We will take these as the electrostatic potential ψ, quasi Fermi level

E_F, and the electron temperature T_e. It is true that quantities such as the total electron density n, electron energy density W, and many other quantities enter the calculation and need to be found. But all of these extra quantities are effectively intermediate variables in the calculation. Writing $N \equiv (NI+1)(NJ+1)$ for the total number of meshpoints in the mesh, which has $(NI+1)$ and $(NJ+1)$ points along Ox and Oy respectively, we must solve for a total of $3N$ quantities $\psi_{i,j}$, $E_{Fi,j}$, and $T_{ei,j}$, $(i=0,1,\ldots,NI; j=0,1,\ldots,NJ)$. These equations are highly nonlinear, and the method which is used most widely for solving them is the Newton method [49], or some variant of it. This method will be outlined now. Suppose we have $3N$ nonlinear equations for $3N$ unknowns x_i,

$$F_1(x_1,x_2,\ldots,x_{3N})=0 \quad F_2(x_1,x_2,\ldots,x_{3N})=0 \quad \cdots F_{3N}(x_1,x_2,\ldots,x_{3N})=0$$

which can be written more compactly in vector notation as

$$\mathbf{F}(\mathbf{x})=\mathbf{0} \tag{27}$$

where both $\mathbf{F}$ and $\mathbf{x}$ are $3N$-component vectors. These equations will be iterated to a solution. Let $\mathbf{x}^r$ be the solution vector obtained after r iterations, and let us look for a better solution $\mathbf{x}^{r+1}=\mathbf{x}^r+\delta\mathbf{x}^r$. Then expanding in a Taylor series up to the first order,

$$\mathbf{0}=\mathbf{F}(\mathbf{x}^r+\delta\mathbf{x}^r)\approx\mathbf{F}(\mathbf{x}^r)+\mathbf{J}^r\delta\mathbf{x}^r$$

where $\mathbf{J}^r$ is the $3N\times 3N$ Jacobian matrix whose elements are $(\mathbf{J})_{ij}\equiv\partial F_i/\partial x_j$ evaluated at the rth iteration. Hence the correction $\delta\mathbf{x}^r$ to the rth step satisfies the equation

$$\mathbf{J}^r\delta\mathbf{x}^r=-\mathbf{F}(\mathbf{x}^r) \tag{28}$$

so that a better estimate of the solution is

$$\mathbf{x}^{r+1}=\mathbf{x}^r-(\mathbf{J}^r)^{-1}\mathbf{F}(\mathbf{x}^r) \tag{29}$$

Hence starting with an initial guess $\mathbf{x}^1$, Eq. (29) is iterated until some suitable stopping criterion is met, for example, until all of the residuals $|F_i(\mathbf{x}^r)|$ become smaller than some prescribed values. At each iteration number r, we must solve the set of $3N$ linear equations (28) for the $3N$ values δx_1^r, $\delta x_2^r,\ldots,\delta x_{3N}^r$. On applying this process to our device problem, we see that this amounts to solving a set of $3(NI+1)\times(NJ+1)$ equations for the same number of unknowns. Even taking a rather coarse mesh with $NI=NJ=99$, for example, we see that this amounts to solving 30,000 equations for 30,000 unknowns, or equivalently inverting the $30{,}000\times 30{,}000$ matrix $\mathbf{J}^r$ at each iteration. However, we have seen that each equation written down at internal meshpoint (i,j) involves the values only at that point and its four nearest neighbors, so that most of the elements of $\mathbf{J}^r$ are zero. It is then appropriate and common to use *sparse matrix theory* [51] in the solutions of Eq. (28).

A simplification may be made in the case of our device modeling. We are solving for a set of three quantities ψ, E_F, and T_e at each meshpoint, and we may form a three-component vector $\mathbf{y}_{i,j}$ at each meshpoint: $\mathbf{y}_{i,j}=(\psi_{i,j},E_{Fi,j},T_{ei,j})$. We also have three equations $\mathbf{F}_{i,j}=(F_{1i,j},F_{2i,j},F_{3i,j})=\mathbf{0}$ at each meshpoint (i,j). We therefore have N vector equations, each with three components,

$$\mathbf{F}_{i,j}(\{\mathbf{y}_{r,s}: r=0,\ldots,NI;\quad s=0,\ldots,NJ\})=\mathbf{0}$$
$$(i=0,\ldots,NI;\quad j=0,\ldots,NJ).$$

Although this may appear to be merely a change of notation, this form of the problem does lead to a useful modification of the solution in that, in further neglecting the four contributions $\mathbf{y}_{i-1,j}$, $\mathbf{y}_{i+1,j}$, $\mathbf{y}_{i,j-1}$, and $\mathbf{y}_{i,j+1}$ to the resulting equations, we need only invert a 3×3 matrix at each meshpoint [52]. It can be shown that this procedure leads to the same solution, but at a slightly slower convergence rate.

Mention has been made of the initial guess $\mathbf{x}^1$, and the successful convergence of this method depends crucially on this initial guess [53]. Even for a "good" initial guess, the iteration process is fragile in the early stages, and all too often the process fails. To overcome this, the update process is replaced by

$$\mathbf{x}^{r+1}=\mathbf{x}^r+\alpha^r\delta\mathbf{x}^r$$

where $0\le\alpha^r\le1$. In the early stages of the iteration, the value of α^r is taken to be very small, typically 0.001. The value is allowed to increase as the iterations progress. Either this can be done manually or its value can be determined using the values of the residuals; if a residual suddenly increases then the value of α is reset to a smaller value. In any event, a value of typically 0.4 is usually the maximum sustainable value for α^r in order that the iterations continue without failing [54].

4.4. Multigrid Method

The multigrid method applied to device simulation can considerably speed the process of solving the resulting set of nonlinear simultaneous equations in the $3(NI+1)\times(NJ+1)$ variables $\psi_{i,j}$, $E_{Fi,j}$, and $T_{ei,j}$ $(i=0,\ldots,NI;\ j=0,\ldots,NJ)$. For simplicity of notation, we will describe the method applied to solving for a single function ψ on a one-dimensional grid; its extension to the problem of solving for more functions on two or three problems is straightforward.

Consider a progression of G meshes, or *grids*, $g=1,2,\ldots,G$. Grid (g) will have an even number N_g of intervals; that is, it will have (N_g+1) meshpoints. Grid $(g+1)$ is obtained from grid (g) by halving each interval of grid (g) so that grid $(g+1)$ will have $N_{g+1}=2N_g$ intervals. Thus if grid (1) has 25 intervals, grid (4) will have 200 intervals. Hence grid (1) is the coarsest grid, and we wish to eventually have a device solution on the finest grid (G).

In the *full approximation storage method* of Brandt [55], an exact solution is obtained on the coarsest grid (1): this should not prove too onerous if the value of N_1 is small enough. The solution is then imposed on grid (2) and a small number of iterations (perhaps 20) is performed on this grid. The new approximate solution is then imposed back onto grid (1) and an exact solution is refound. Haven't we found this already? No, because as we shall see, we will obtain the exact solution to a slightly different equation. The process is repeated up to grid (3) with a small number of iterations performed on each grid which is visited on the upward and downward paths. The process is repeated in stages until

the final fine grid (G) is reached. Thus for four grids, the sequence of visits will be $1 \to 2 \to 1 \to 2 \to 3 \to 2 \to 1 \to 2 \to 3 \to 4 \to 3 \to 2 \to 1 \to 2 \to 3 \to 4$ with exact solutions being found on grid (1) and a finite number of iterations performed on all the other grids.

A sequence of equations is iterated on each grid. Suppose that the equation we wish to solve is written formally as $L\psi = 0$ where L is the appropriate differential operator. For example, if we are solving the Laplace equation $\nabla^2\psi = 0$ then L is the Laplacian ∇^2. This equation is discretized on grid (g) as

$$L_{(g)}\psi_{(g)} = 0 \tag{30}$$

where $L_{(g)}$ is the discretization of L on that grid. For example, if L is the Laplacian, then $L_{(g)}$ will be the difference operator given in Eq. (18). Our goal will therefore be to solve Eq. (30) on the fine grid (G). We must first decide how to take our solutions between successive grids. The *restriction operator* R takes a value of ψ from a grid (g) to the next coarse one ($g - 1$). An example of such an operator is the *half-weighting* process

$$\begin{aligned}\psi_{(g-1)j} &= \frac{1}{2}\psi_{(g)2j} + \frac{1}{4}\psi_{(g)2j+1} + \frac{1}{4}\psi_{(g)2j-1} \qquad \text{internally}\\ &= \psi_{(g)2j} \qquad \text{on boundary}\end{aligned}$$

(remember, we are using a one-dimensional illustration in which j labels the meshpoint). The reverse *prolongation operator* P takes a solution from grid ($g - 1$) onto the next fine grid (g). *Half-weighting* prolongation takes the form

$$\begin{aligned}\psi_{(g)2j} &= \psi_{(g-1)j}\\ \psi_{(g)2j+1} &= \frac{1}{2}(\psi_{(g-1)j+1} + \psi_{(g-1)j})\end{aligned}$$

We iterate a succession of equations on each grid. The original equation (30) is generalized to

$$L_{(g)}\psi_{(g)} = f_{(g)} \qquad (g = 1, \ldots, G) \tag{31}$$

where the functions $f_{(g)}$ are calculated on the downward paths as

$$f_{(g-1)} = L_{(g-1)}(R\psi'_{(g)}) + R(f_{(g)} - L_{(g)}\psi'_{(g)})$$

where $\psi'_{(g)}$ is the *approximate* solution to Eq. (31) and R is the restriction operator. The functions $f_{(g)}$ are initialized to zero on the first upward path and are generated on the downward paths. An exact solution is found only on the coarsest grid, and a small number of iterations are performed on each visit to finer grids.

The effort in programming this method is rewarded by achieving fast solution times. It was found [56] that this method produced a solution with a speedup of 4.9 times on a corresponding simple nonmultigrid method. Care must be taken in choosing the coarsest grid when applying this method to device modeling. Clearly, this grid should be as coarse as possible in order that the task of finding an exact solution on this grid is not too onerous. However, the grid must be fine enough to take in all of the physical features of the device. For example, it would not be acceptable to take a grid so coarse that the boundaries between three consecutive layers fell into one mesh interval.

4.5. Upwinding and the Bernoulli Function

The use of the method of upwinding is very common in numerical simulations of fluid flows [57–60] and is applicable to the study of electron flow in semiconductor devices [61]. There is a very simple physical idea behind this technique. Suppose that a particular quantity V is to be evaluated at the midpoint of an interval, that is, halfway between two meshpoints x_i and x_{i+1}. A crude estimate will give equal relevant contributions from both meshpoints. However, if an electric field exists (as it always does in our devices), then the value of V at the midpoint will be influenced more strongly by conditions at one meshpoint than the other—in effect, this “influence” will be swept to the midpoint by the electric field. A simple calculation will illustrate this point. Consider a simplification of the expression in Eq. (6) for the current density, for a one-dimensional case in which T_e is constant:

$$J = -qn\mu\frac{d\psi}{dx} + k_B\mu T_e\frac{dn}{dx}$$

Assume that $E = -\frac{d\psi}{dx}$, and J and μ are constants in the interval. Then we have a simple linear first order differential equation

$$\frac{dn}{dx} + \left(\frac{qE}{k_BT_e}\right)n = \frac{J}{\mu k_BT_e}$$

which has the solution

$$n = Ae^{-\frac{qE}{k_BT_e}x} + \frac{J}{q\mu E} \qquad (E \neq 0) \tag{32}$$

where A is an arbitrary constant. If we write down this solution at both x_i and $x_{i+1} = x_i + h_i$, we can solve the resulting linear simultaneous equations for A and J. On substituting these values back into Eq. (32) and defining $P \equiv qEh_i/k_BT_e$, we finally have the result

$$\frac{n - n_i}{n_{i+1} - n_i} = \frac{1 - e^{-P\frac{(x-x_i)}{h_i}}}{1 - e^{-P}} \tag{33}$$

The quantity P is called the *Peclet number*, and it has the same sign as the electric field E. The interpretation of this result is as follows: if P is large and positive (that is, the electric field is large and in the positive Ox direction), then the exponential terms in Eq. (33) are very small, giving $n \simeq n_{i+1}$. Thus in this case the value of n at any point in the interval is more strongly influenced by its value at x_{i+1}. On the other hand, if the electric field is reversed so that P is large and negative, then $n \simeq n_i$, and the value of n at any point in the interval is more strongly influenced by its value at x_i. If the electric field is zero, then Eq. (33) gives the linear relation

$$n = n_i + \left(\frac{x - x_i}{h_i}\right)(n_{i+1} - n_i)$$

[we must let $P \to 0$ in Eq. (33) to obtain this result]. For x midway between x_i and x_{i+1}, this result for $E = 0$ gives $n = \frac{1}{2}(n_i + n_{i+1})$ as expected.

This is the basic idea behind the Scharfetter–Gummel approach [62], in which errors and instabilities can arise in iterative procedures unless an exponential variation is taken

between meshpoints. On making the same assumptions as above, namely, that $E = -\frac{d\psi}{dx}$ is a constant in the interval (x_i, x_{i+1}) (this is a weaker condition than assuming that ψ itself is a constant in the interval), we find that the current density along Ox at the half-point $(i+\frac{1}{2}, j)$ is

$$J_{xi+\frac{1}{2},j} = \frac{\mu_{i+\frac{1}{2},j} k_B T_e}{q h_i} \left(B(-P_{i+\frac{1}{2},j}) n_{i+1,j} - B(P_{i+\frac{1}{2},j}) n_{i,j} \right) \tag{34}$$

where

$$P_{i+\frac{1}{2},j} \equiv \frac{q}{k_B T_e} (\psi_{i,j} - \psi_{i+1,j})$$

and B is the *Bernoulli function* defined by

$$B(t) \equiv \frac{t}{e^t - 1} \tag{35}$$

Again, the sign of $P_{i+(1/2),j}$ is determined by the sign of $E_{xi+(1/2),j} = (1/h_i)(\psi_{i,j} - \psi_{i+1,j})$, and so result (34) may be interpreted in light of the brief description given above of the upwinding process. We may form similar expressions for $J_{xi-(1/2),j}$, $J_{yi,j+(1/2)}$, and $J_{yi,j-(1/2)}$. The expression for $\nabla \cdot \mathbf{J}$ in the current continuity equation (3) is then discretized using

$$\nabla \cdot \mathbf{J}_{i,j} = \frac{1}{h_i}(J_{xi+\frac{1}{2},j} - J_{xi-\frac{1}{2},j}) + \frac{1}{k_j}(J_{yi,j+\frac{1}{2}} - J_{yi,j-\frac{1}{2}})$$

The Bernoulli function has the properties that $B(t) > 0$, $B(0) = 1$, $B(y) \to 0$ as $t \to \infty$, and $B(t) \to -t$ as $t \to -\infty$. In any numerical implementation using the function $B(t)$, care must be taken to avoid overflows and underflows when evaluating the function. Selberherr [20] approximates this function by

$$B(t) = \begin{cases} -t & \text{for } t \le t_1 \\ \frac{t}{e^t - 1} & \text{for } t_1 < t \le t_2 \\ 1 - \frac{1}{2}t & \text{for } t_2 < t \le t_3 \\ \frac{te^{-t}}{1 - e^{-t}} & \text{for } t_3 < t \le t_4 \\ te^{-t} & \text{for } t_4 < t \le t_5 \\ 0 & \text{for } t_5 < t \end{cases}$$

where the values of $t_1, \ldots, t_5$ delimit certain regions of the t-axis. A clever aspect of this method is that $t_1, \ldots, t_5$ are determined by the particular machine on which the numerics are performed; for example, the value of t_1 may be found by asking the machine to solve the equation $-t = t/(e^t - 1)$ to its own accuracy. The author found the values $t_1 = -t_4 = -22.874$, $t_2 = -t_3 = -2.527 \times 10^{-3}$, and $t_5 = 89.416$ when implementing the routine on a PC.

The above discussion concerned the case in which the electron temperature T_e was constant. This situation does not apply in realistic devices, and so the result in Eq. (34) must be modified. In this more general case, influences will be propagated not only by the electric field but also by the temperature gradient. It can be shown [63] that the generalization is

$$\begin{aligned} J_{xi+\frac{1}{2},j} = {} & \frac{k_B}{h_i} \mu_{i+\frac{1}{2},j} \\ & \times \left[C_a\left(\frac{q}{k_B}(\psi_{i+1,j} - \psi_{i,j}), T_{ei+1,j}, T_{ei,j} \right) n_{i+1,j} \right. \\ & \left. - C_a\left(-\frac{q}{k_B}(\psi_{i+1,j} - \psi_{i,j}), T_{ei,j}, T_{ei+1,j} \right) n_{i,j} \right] \end{aligned} \tag{36}$$

where the function C is defined as

$$C_a(x, T_1, T_2) \equiv t(T_1, T_2) B\left(\frac{x - a(T_1 - T_2)}{t(T_1, T_2)} \right)$$

and the function t is defined as

$$t(T_1, T_2) \equiv \frac{T_2 - T_1}{\ln T_2 - \ln T_1} \qquad \text{for } T_1 \ne T_2$$

The quantity a is that which appears in the coefficient of ∇T_e in Eq. (6); the value of b in that equation has been taken as zero. It can be shown that the function $t(T_1, T_2)$ is actually a *mean value* of T_1 and T_2 which lies between their arithmetic and geometric means. Similar expressions can be written for $J_{xi-(1/2),j}$, $J_{yi,j+(1/2)}$, and $J_{yi,j-(1/2)}$ so that $\nabla \cdot \mathbf{J}$ may be evaluated. The result in Eq. (36) has been derived for the electrons in HEMTs and MESFETs. However, it is equally valid for electron and hole currents in bipolar devices when the appropriate electron and hole densities are used.

4.6. Solution of the Schrödinger Equation

The numerical solution of the Schrödinger equation presents one of the largest costs in computing time when performing the complete numerical simulation. Before each of the Newton iterations described previously, we must first solve the Schrödinger equation in order that its solutions may be used to update the values of $n_{i,j}$ and $W_{i,j}$ of the electron density and electron energy density at each meshpoint using Eqs. (12)–(16). Trellakis and Ravaioli [64] have developed a fast way of producing a three-dimensional spectral solution of the Schrödinger equation, and Grinstein et al. [65] have solved the equation using a multigrid method. Whatever method is used, the cost of the solution may be reduced by calculating as few eigensolutions as possible consistent with accuracy of solution [66].

The cost may also be reduced by solving the equation in one-dimensional slices perpendicular to the layer structure, and this is the method which will be described here. In this method it is assumed that changes in potential occur much more rapidly along Oy (across the layer structure) than along Ox where device lengths are much larger. We proceed by discretizing Eq. (11). The first term in the equation is the kinetic energy contribution, and it has been written in such a way as to preserve its Hermitian character since the effective mass m^* is a function of position (Hermitian operators are necessary because they possess real eigenvalues). Discretization of this operator must be done so that its

Hermitian character is preserved [67]. Halfway discretization gives

$$-\frac{\hbar^2}{2}\left(\frac{\left(\frac{1}{m^*}\frac{d\xi_{(\nu)}}{dy}\right)_{j+\frac{1}{2}} - \left(\frac{1}{m^*}\frac{d\xi_{(\nu)}}{dy}\right)_{j-\frac{1}{2}}}{\frac{1}{2}(k_j + k_{j-1})}\right)$$

and, on completing the process, Eq. (11) becomes

$$a_j\xi_{(\nu)j-1} + b_j\xi_{(\nu)j} + c_j\xi_{(\nu)j+1} = \lambda_{(\nu)}\xi_{(\nu)j} \tag{37}$$

where the coefficients a_i, b_i, and c_i are defined as

$$c_j \equiv -\frac{\hbar^2}{qk_j(k_j + k_{j-1})m^*_{j+\frac{1}{2}}}$$

$$a_j \equiv -\frac{\hbar^2}{qk_{j-1}(k_j + k_{j-1})m^*_{j-\frac{1}{2}}}$$

$$b_j \equiv -(a_j + c_j) + V_{xci,j} - E_{hi,j} - \psi_{i,j}$$

Equation (37) is a set of linear equations for the $\xi_{(\nu)j}$ and can be written in the matrix form $\mathbf{A}\xi_{(\nu)} = \lambda_{(\nu)}\xi_{(\nu)}$ where the matrix $\mathbf{A}$ is defined by

$$\mathbf{A} = \begin{pmatrix} b_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & \cdots & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & \cdots & 0 \\ 0 & 0 & a_4 & \ddots & c_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

This matrix is tridiagonal, but it is only symmetric if the mesh is uniform. Its eigenvalues can be found using a QL algorithm with implicit shifts [49]. This algorithm exploits the fact that we can write A = QL where Q is an orthogonal matrix, and L is a left triangular matrix with zeros above the diagonal. If these eigenvalues are then put into ascending order, we only need to solve for the first L of the corresponding eigenfunctions for use in evaluating n and W. This is done by solving the equations using the Newton method with an initial guess [56]

$$\xi_{(\nu)j} = \sin\left(\frac{\nu\pi j}{NJ}\right) \qquad (\nu = 0, \ldots, L-1; \quad j = 0, \ldots, NJ)$$

4.7. Example Solution

We end by giving an illustration of the type of output we may expect from the numerical modeling of a typical HEMT structure; we present a solution for a simple five-layer HEMT with one recess and no field plate. The structure will consist of layers of—starting with the layer closest to the edge containing the contacts—GaAs, AlGaAs, InGaAs, AlGaAs, and GaAs. In each AlGaAs layer the mole fraction u is taken as 0.23, while the In mole fraction of the InGaAs is taken as 0.2. The overall length along the contact edge is 1400 nm, the overall thickness of the device is 250 nm, and the depth of the recess is 25 nm. Since the results presented here are for illustrative purposes only, we will not state the remaining physical data for the device. The results of any simulation will present the device modeler and designer with a vast amount of data to analyze, and one of the most useful ways of describing these data is by graphical plotting. Figure 2 illustrates just one set of data—that of the conduction band profile E_c throughout the device—when the source and gate voltages are set to 0.0 V and the drain voltage is set to 2.0 V. A total of 10 eigensolutions ($L = 9$) was used in this solution. This particular plot illustrates in a very clear way the quantum well structure of the device. Such plots of the quantity E_c and all of the other parameters of the system—for example, the electron density n and the electric field $\mathbf{E}$—must be carefully examined by the designer in order to achieve a full picture of the operation of the particular device being modeled.

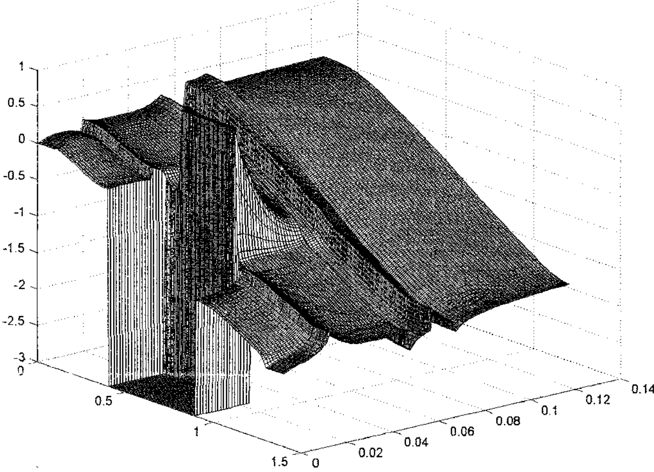


Figure 2. A plot of the conduction band profile E_c for a five-layer HEMT with one recess and no field plate. The horizontal axes are in units of micrometers, and the vertical axis is in units of eV. The plot along Oy has been truncated to 0.14 micrometers in order to show the quantum well structure more clearly.

GLOSSARY

Bandwidth The range of frequencies in a spectral line for which its Fourier transform has significant values.

Bipolar devices Devices in which both electron and hole transport are important.

Hermitian operator An operator A is hermitian if the inner product of any vector **v** with A**v** is real. Such an operator possesses real eigenvalues.

High electron mobility transistor (HEMT) This is a MESFET containing layers of material which give rise to quantum wells at their interfaces.

Metal semiconductor field effect transistor (MESFET) Uses a metal to form a Schottky gate.

Mobility The ratio of the average particle drift velocity to the applied electric field.

Nondegenerate approximation Applies when Fermi integrals can be approximated by exponential functions, mainly in cases where doping densities are not too high.

Schottky effect The lowering of the potential barrier when negative charges are brought near a metal interface and an electric field is applied.

Tunneling In quantum mechanics, the ability of an electron to cross a potential barrier which would be impenetrable using classical theory.

REFERENCES

1. E. Samuel, *New Scientist* p. 20, (7 September 2002).
2. I. Malsch, *Vacuum Solutions* 17, 17 (2000).
3. P. Lugli, F. Compagnone, A. DiCarlo, and A. Reale, *VLSI Design* 13, 23 (2001).
4. M. Devoret and C. Glatti, *Physics World* (Sept. 1998),
5. D. K. Ferry and S. M. Goodnick, "Transport in Nanostructures." Cambridge Univ. Press, Cambridge, UK, 1997.
6. L. P. Kouwenhoven and P. L. McEuen, Single electron transport through a quantum dot, in "Nano-Science and Technology" (G. Timp, Ed.). Am. Inst. of Phys., New York, 1996.
7. M. A. Nielsen and I. L. Chuang, "Quantum Computation and Quantum Information." Cambridge Univ. Press, Cambridge, UK, 2000.
8. B. G. Streetman, "Solid State Electronic Devices." Prentice Hall, New York, 1991.
9. S. Karmalkar and U. K. Mishra, *IEEE Trans. Electron Devices* 48, 1515 (2001).
10. A. Wakejima, K. Ota, K. Matsunaga, W. Contrata, and M. Kuzuhara, in "IEEE Radio Frequency Integrated Circuit Symposium," NEC Corporation, Japan, 2001.
11. F. Filicori, G. Ghione, and C. U. Naldi, *IEEE Trans. Microwave Theory Tech.* MTT-40, 1333 (1992).
12. L. O. Chua, *IEEE Trans. Circuits Syst.* CAS-27, 1014 (1980).
13. J. M. Golio, "Microwave MESFETs and HEMTs." Artech House, Norwood MA, 1990.
14. R. E. Anholt and S. E. Swirhun, *IEEE Trans. Electron Devices* ED-39, 2029 (1992).
15. T. Mietzner, J. Jakumeit, and U. Ravaili, *VLSI Design* 13, 175 (2001).
16. R. W. Kelsall, *VLSI Design* 8, 367 (1998).
17. P. Lugli, Monte Carlo models and simulations, in "Compound Semiconductor Device Modeling" (C. M. Snowden and R. E. Miles, Eds.), pp. 210–231. Springer-Verlag, Berlin, 1993.
18. M. Reiser, *IEEE Trans. Electron Devices* ED-20, 35 (1973).
19. C. M. Snowden, M. J. Howes, and D. V. Morgan, *IEEE Trans. Electron Devices* ED-30, 1817 (1983).
20. Y.-K. Feng and A. Hintz, *IEEE Trans. Electron Devices* ED-35, 1419 (1988).
21. T. Shawki, G. Salmer, and O. El-Sayed, *IEEE Trans. Computer Aided Design* 9, 1150 (1990).
22. J. J. Duderstadt and W. R. Martin, "Transport Theory." Wiley, New York, 1979.
23. C. M. Snowden, Classical and semi-classical models, in "Semiconductor Device Modelling" (C. M. Snowden, Ed.), pp. 16–33. Springer-Verlag, Berlin, 1989.
24. H. C. de Graaff and F. M. Klaassen, "Compact Transistor Modelling for Circuit Design." Springer-Verlag, New York, 1990.
25. T.-W. Tang, *IEEE Trans. Electron Devices* ED-31, 1912 (1984).
26. C. C. McAndrew, K. Singhal, and E. L. Heasell, *IEEE Electron Devices Lett.* EDL-6, 446 (1985).
27. S. Selberherr, "Analysis and Simulation of Semiconductor Devices." Springer-Verlag, New York, 1984.
28. C. M. Snowden and D. Loret, *IEEE Trans. Electron Devices* ED-34, 212 (1987).
29. Y.-K. Feng and A. Hintz, *IEEE Trans. Electron Devices* ED-35, 1419 (1988).
30. S. Adachi, *J. Appl. Phys.* 58, R1 (1985).
31. F. Stern, *Phys. Rev. B* 30, 840 (1984).
32. G. D. Smith, "Numerical Solution of Partial Differential Equations: Finite Difference Methods." Clarendon Press, Oxford, 1985.
33. E. A. B. Cole, Numerical methods and their application to device modelling, in "Compound Semiconductor Device Modelling" (C. M. Snowden and R. E. Miles, Eds.), pp. 1–25. Springer-Verlag, Berlin, 1993.
34. S. D. Mobbs, Numerical techniques—The finite element method, in "Semiconductor Device Modelling" (C. M. Snowden, Ed.), pp. 49–59. Springer-Verlag, Berlin, 1989.
35. M. Shen, M.-C. Cheng, and J. J. Liou, *VLSI Design* 13, 79 (2001).
36. T. A. Manteuffel and K. J. Ressel, *SIAM J. Numer. Anal.* 35, 806 (1998).
37. M. Fortin and G. Yang, *Compel.* 15, 4 (1996).
38. Z. Chen, *Comp. Math. Appl.* 31, 17 (1996).
39. N. R. Aluru, A. Raefsky, and P. M. Pinsky, *Compu. Methods Appl. Mech. Eng.* 107, 269 (1993).
40. D. B. Ingham, Numerical techniques: Finite difference and boundary element methods, in "Semiconductor Device Modelling" (C. M. Snowden, Ed.), pp. 34–48. Springer-Verlag, Berlin, 1989.
41. M. A. Kelmanson, *Math. Comput. Modelling* 31, 45 (2000).
42. M. Reiser, *IBM Res. J.* RZ499 (1972).
43. C. M. Snowden, "Semiconductor Device Modelling." Peregrinus, London, 1988.
44. A. M. Anile, S. F. Liotta, G. Mascali, and S. Rinaudo, *VLSI Design* 13, 355 (2001).
45. W.-S. Choi, J.-G. Ahn, Y.-J. Park, H.-S. Min, and C.-G. Hwang, *IEEE Trans. Comput. Aided Design* 13, 899 (1994).
46. S.-H. Lo and C.-P. Lee, *IEEE Trans. Electron Devices* 39, 1564 (1992).
47. H. Xue, M. J. Howes, and C. M. Snowden, *Int. J. Num. Modelling*, 4, 107 (1991).
48. R. R. Pantoja, M. J. Howes, J. R. Richardson, and C. M. Snowden, *IEEE Trans. Microwave Theory* 37, 2039 (1989).
49. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C." Cambridge Univ. Press, Cambridge, UK, 1992.
50. A. F. Franz, G. A. Franz, S. Selberherr, C. Ringhofer, and P. Markowich, *IEEE Trans. Electron Devices* ED-30, 1070 (1983).
51. R. Drury, The Physical Modelling of Heterojunction Field Effect Transistors, Ph.D. Thesis, University of Leeds, 1993.
52. E. A. B. Cole, in "Proc. of the International Workshop on Computational Electronics," University of Leeds, 1993, pp. 11–24.
53. R.-K. Perng and C.-Y. Wu, *Solid State Electron.* 33, 287 (1990).
54. S. Hussain, *Numerical Modelling of Microwave Semiconductor Devices*, Ph.D. Thesis, University of Leeds, 2001.
55. A. Brandt, S. McCormick, and J. Ruge, *SIAM J. Sci. Stat. Comput.* 4, 244 (1983).
56. E. A. B. Cole, C. M. Snowden, and T. Boettcher, *Int. J. Numer. Modelling* 10, 121 (1997).
57. J. L. Stegger and R. F. Warming, *J. Comp. Phys.* 40, 263 (1981).
58. B. van Leer, in "Lecture Notes in Physics," Vol. 170, p. 507. Springer-Verlag, New York/Berlin, 1982.
59. P. L. Roe, *J. Comp. Phys.* 43, 357 (1981).
60. S. Osjer, *Math. Stud.* 47, 179 (1981).
61. M. Shen, W.-K. Yip, M.-C. Cheng, and J. J. Liou, *VLSI Design* 13, 329 (2001).
62. D. L. Scharfetter and H. K. Gummel, *IEEE Trans. Electron Devices* ED-16, 64 (1969).
63. E. A. B. Cole and C. M. Snowden, *Int. J. Num. Modelling* 8, 13 (1995).
64. A. Trellakis and U. Ravaioli, *VLSI Design* 13, 341 (2001).
65. F. F. Grinstein, H. Rabitz, and A. Askar, *J. Comp. Phys.* 51, 423 (1983).
66. E. A. B. Cole, T. Boettcher, and C. M. Snowden, *Semicond. Sci. Thechnol.* 12, 100 (1997).
67. R. A. Morrow and K. R. Brownstein, *Phys. Rev. B* 30, 678 (1984).

www.aspbs.com/enn

Semiconductor Nanoparticles for Photocatalysis

W. Li, S. Ismat Shah

University of Delaware, Newark, Delaware, USA

CONTENTS

1. Introduction
2. Synthesis of Semiconductor Nanoparticles
3. Mechanism of Nanoparticle Photocatalysis
4. Enhancement of Photocatalytic Activity
5. Prospects for TiO_2 Nanoparticles
6. Summary

   Glossary

   References

1. INTRODUCTION

With the development of human society, both the amount and the type of pollutants are increasing in the ground, air, and water systems. Pollutants usually are toxic organic chemicals emanating from various sources and include industrial products and by-products. Halogenated organic chemicals and commercial colorants are two main types of pollutants threatening the well-being of human society. However, hazardous organic molecules can be rendered nontoxic and broken down to carbon dioxide, water, or simple mineral acids, etc., by a degradation treatment [1]. Increased public concern for environmental cleanup has promoted the development of highly efficient photocatalysts that can participate in detoxification reactions. Environmental remediation by photocatalysts comes with several advantages: direct conversion of pollutants to nontoxic by-products without the necessity for any other associated disposal steps; use of oxygen as an oxidant and elimination of expensive oxidizing chemicals; the potential of using free and abundant solar energy; self-regeneration and recycling of the photocatalyst, etc. It is, therefore, no surprise that the research and development activities in this field have been very vigorous in recent years.

Semiconductor nanoparticles are successfully utilized in photocatalytic applications [2–10]. Colloidal semiconductors were first introduced in 1976 [11]. Since then, the promising application of nanoparticles in photocatalysis has achieved great attention and been extensively investigated. This is in addition to their application in photovoltaics and electronic devices, etc. The semiconductor catalysts form a heterogeneous photocatalysis system with pollutant molecules and exhibit a high degradation effectiveness. The selection of semiconductor nanoparticles for photocatalysis is based on several aspects that will be described later.

Semiconductors make generation and transfer of charge carriers possible with the assistance of extra energy. For photocatalysis, electrons and holes are crucial for reduction and oxidation reactions, respectively. Thus, the capability to supply charge carriers is a basic consideration when a catalyst is being selected. Electronically, the three common types of material are metal, insulator, and semiconductor. They are categorized by the electric conduction that is related to the electron energy. The conduction and valence bands are the two main energy bands. The lowest point of the conduction band and the highest point of the valence band are the conduction band edge and the valence band edge, respectively. The separation between the conduction band edge and valence band edge is called the band gap or energy gap, within which no electronic state exists. For metals, the electronic states form a continuum, and no energy gap exists between the two band edges. The electrons in a partly filled conduction band and holes in a valence band contribute to the electric current because they have sufficient mobility. Metals are a good conductor because their conduction bands are partly filled with electrons. It is extremely hard to form mobile electrons and holes in insulators because either their conduction band is empty, with a band gap too large to excite electrons, or both conduction and valence

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (669–695)

bands are fully occupied by electrons with no electronic state left in the conduction band. Although semiconductors also have empty conduction bands at absolute zero, the magnitude of the band gap is smaller than that of insulators, so electrons can be excited into the empty conduction band, leaving holes in the valence band for electronic balance if extra energy equal to or greater than its band-gap energy is provided. Therefore, semiconductors can provide both electrons and holes. Furthermore, semiconductors with band gaps large enough for the redox potential of the $H_2O/^{\bullet}OH$ couple (~ -2.8 eV) to lie within them can be viable for photocatalysis processes [12] because, in such cases, excited electrons and holes can be transferred to the pollutant to participate in the reduction and oxidation reactions, respectively, that occur on the catalyst surface. A detailed mechanism of photocatalysis is discussed in Section 3.

Particles with sizes less than several tens of nanometers (1 nm $= 10^{-9}$ m) are known as nanoparticles. They are in powder form or are nanocrystalline thin films on certain supports. Nanoparticles are distinguished from bulk by their high surface-to-volume ratio of nanoparticles that causes the structural and electronic changes that, in turn, induce other properties to become different from those of the bulk. Unique photophysical, photochemical, photoelectronic, and photocatalytic properties can occur in semiconductor nanoparticle systems [8, 13–17]. The differences can be found at both equilibrium and nonequilibrium states, including thermodynamics and kinetics [18, 19]. They critically depend on particle size, shape, and surface characteristics. Lower melting temperatures [20], altered phase transformation behaviors [19, 21], and nonlinear optical phenomena [22, 23] have been observed in nanoparticles. More pertinent to our discussion, with the decrease of particle size, an extremely high surface area-to-volume ratio is obtained, leading to an increase in surface-specific active sites for chemical reactions and photon absorption to enhance the reaction and absorption efficiency. The enhanced surface area also increases surface states, which changes the activity of electrons and holes and affects the chemical reaction dynamics. For instance, the constant of the charge carrier transfer rate is relatively higher in nanoparticles than in bulk. When particle size decreases below the Bohr radius of the first excitation state, the quantum size effect could occur because of the confinement of charge carriers [12, 24, 25]. Such particles are also referred to as quantized or Q-particles. The quantum size effect splits both conduction and valence bands into discrete electronic states [26], and the size dependence of optical and electronic properties becomes much more pronounced. For instance, size quantization can increase the band gap of photocatalysts to enhance the redox potential of conduction band electrons and valence band holes [27]. Finally, nanoparticles can lead to the possibility of indirect electron transitions at the boundary of the crystals and realize the essential enhancement of light absorption. More details on the significance of nanoparticles are presented in Section 4.2.

In this review we will cover the most frequently used nanoparticles. They include oxides (TiO_2, ZnO, Fe_2O_3, WO_3, SnO_2, Ag_2O, V_2O_5, $SrTiO_3$), sulfides (ZnS, CdS, MoS_2, Cu_xS, Ag_2S, PbS), selenides (CdSe, PbSe, HgSe), iodides (AgI), and modified systems such as coupled semiconductor systems (CdS/TiO_2, $CdSe/TiO_2$, SnO_2/TiO_2, ZnO/TiO_2, ZnO/CdS), etc. Among them, TiO_2 nanoparticles and modified TiO_2 nanoparticles are presented in much more detail because they are the most extensively studied and most efficient photocatalysts [28–30]. Other semiconductor nanoparticles generally have less photocatalytic activity than TiO_2, and some have problems associated with stability, reactivity, etc. [10]. Fe_2O_3 easily undergoes photocathodic corrosion [12], and its active form α-Fe_2O_3 also has high selectivity for the reactant [31]. ZnO is not stable in water, and $Zn(OH)_2$ can be formed on its surface, which is detrimental to its photoreactivity [30]. CdS, PbS, and CdSe are not only toxic but also unstable and undergo photoanodic corrosion in aqueous media [32].

TiO_2 is an important semiconductor oxide photocatalyst because of its chemical stability, nontoxicity, low cost, and high catalytic performance. TiO_2 occurs in three main crystal structures at normal pressure: anatase, rutile, and brookite. Although commercial standard Degussa P25 is composed of 80% anatase and 20% rutile, anatase is believed to have the highest photoreactivity [16, 33–35]. Anatase belongs to the $I4_1/amd$ space group and has a tetragonal structure. Figure 1 [36, 37] is a schematic of the anatase structure, and it gives the calculation for the a and c values. Ti is in an octahedron of two apical and four equatorial oxygen atoms. The neighboring octahedra share adjacent edges. TiO_2 became the focus of attention in 1972 with the discovery of photoelectrochemical splitting of water by TiO_2 [38]. Early studies were mainly related to photovoltaics for energy conversion [39, 40]. It has only been in the past decade that a large number of research papers on TiO_2 photocatalysis have been published. This activity has continued, and various scientists are engaged in research to improve the photocatalytic performance of TiO_2.

2. SYNTHESIS OF SEMICONDUCTOR NANOPARTICLES

2.1. Requirements

The earliest synthesis of nanoparticles can be tracked back to the formation of relatively monodispersed colloids of gold nanoparticles reported by Faraday in the nineteenth century. With the appearance of new semiconductor photocatalysts and miniaturization of electronic devices, smaller nanoparticles are preferred. More interestingly, the methods that can produce very fine particles (1–25 nm) are intensively pursued because of the unique properties of these nanoparticles. Currently, a variety of synthesis methods are being used for the preparation of semiconductor nanoparticles. They include both the chemical methods and the physical approaches. Ideally, the methods employed for synthesis are expected to form nanoparticles with monodispersion or narrow size distribution and to easily modify particle properties, such as size, surface, doping, etc., during the synthesis. Not only are these requirements useful for the

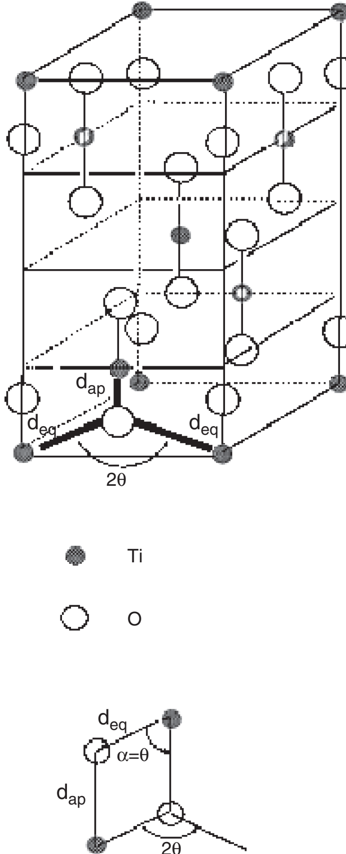


Figure 1. TiO_2 anatase structure. The lattice constants *a* and *c* are expressed as $a = 2d_{eq} \sin\theta$ and $c = 4(d_{ap} + d_{eq} \cos\theta)$, respectively, where d_{eq} and d_{ap} are equatorial and apical Ti-O bond lengths, respectively. d_{eq} is 1.934 Å and d_{ap} is 1.980 Å for anatase [37]. θ is the O-Ti-O angle, which is formed by one apical and one equatorial Ti-O bond. Reprinted with permission from [36], A. Fahmy et al., *Phys. Rev. B* 47, 11717 (1993). © 1993, American Physical Society.

improvement of semiconductor photocatalysis; they are also needed in the fabrication of semiconductor devices. However, most of the methods can only produce nanoparticles with near monodispersion or relatively narrow size distribution due to the fluctuations in growth conditions such as gas flow, growth temperature, or dispersion of precursors. Therefore, a well-controlled synthesis process is still a challenge.

2.2. Chemical Methods

These methods utilize chemical precursors and chemical reactions to produce nanoparticles. They include various modifications of chemical vapor deposition [41–46], sol–gel [47–52], chemical impregnation [53], coprecipitation [54, 55], hydrothermal methods [56–59], flame synthesis [60–63], and traditional or novel chemical reactions [64, 65]. There are several good review papers describing some of these techniques [64, 66, 67]. In the following section two of the most used techniques for nanoparticle synthesis, the metallorganic chemical vapor deposition (MOCVD) and sol–gel techniques, are described.

2.2.1. MOCVD

The basic idea in utilizing MOCVD is that in this process the chemical reactions among precursors occur at a certain temperature in either a vertically or horizontally aligned hot wall chamber. Therefore, no postsynthesis processes like calcination, centrifugation, or hydrothermal processing are required for crystallization or particle refinement. The control of size distribution can be achieved by controlling the temperature of the reactor and/or the flow rate of the precursors. Dopants can also be easily introduced into the reactor through a solid source, separated from the other liquid-precursor, or mixed in with the precursor. Figure 2 is a schematic of a MOCVD system for the preparation of TiO_2 nanoparticles. The most frequently selected Ti precursors are titanium tetraisopropoxide (TTIP) and $TiCl_4$. TTIP is preferred because $TiCl_4$ produces HCl as a toxic by-product. The carrier gas for the precursor can be an inert gas, such as Ar, or other nonreactive gases. In some cases the carrier gas also participates in the chemical reaction. The formation of TiO_2 nanoparticles is basically divided into four stages [41–43, 46]: (I) Introduction of mechanically mixed Ti precursor vapor (with carrier gas) and O_2 from the low-temperature region, (II) production of TiO_2 monomers based on the reaction of TTIP or $TiCl_4$ and O_2 in the higher-temperature region, (III) condensation of TiO_2 monomers to form larger clusters by homogeneous nucleation in the gas phase based on a collision mechanism, (IV) formation of TiO_2 particles in the gas phase and on the substrates after attachment of large clusters to the substrate. Therefore, the reaction and condensation are two crucial stages in the formation of semiconductor nanoparticles by MOCVD. Either thin films [68–70] or free-standing nanoparticles [41–46] can be deposited on certain substrates, depending upon the experimental conditions, such as deposition rate and temperature.

2.2.2. Sol–Gel

The main advantage of this process is that it makes it easy to prepare nanocomposite materials such as organic-coated nanoparticles and nanoparticles in an organic network. The process starts with a precursor solution, and the solvent can

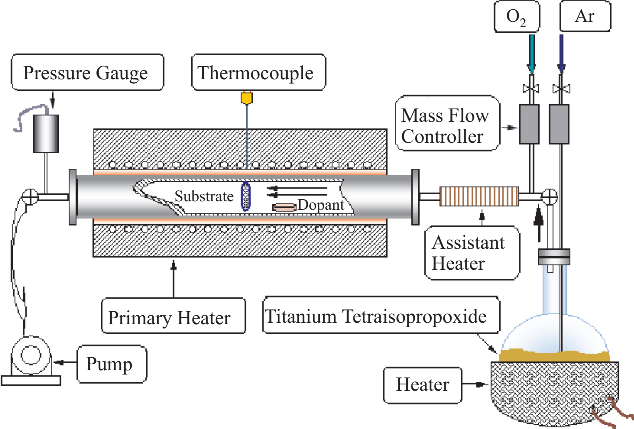


Figure 2. MOCVD system schematic for the synthesis of TiO_2 nanoparticles.

be ethanol or alcohol, etc., which has a low evaporation point. For the purpose of surface modification or doping, organic material or dopant precursor solutions can be mixed with the initial solution. The chemical reactions are carried out by dip-coating precursor solutions on desired substrates in a hood (usually at room temperature). In most cases, the samples are required to rest in the hood for enough time to dissipate gas-phase by-products and to remove excess solvent. The as-grown nanoparticles are amorphous, and a calcination process is needed to form crystals. Thus, the sol–gel process is normally followed by heat treatment for crystallization. Nanoparticles or nanophase films are achieved after calcination. Furthermore, to obtain fully crystallized ultrafine nanoparticles (1.5–5 nm), the particle growth of metal oxides in sol–gel can be inhibited by replacing the surface hydroxyl group before annealing with another functional group. This can produce small secondary-phase particles to restrict the advancement of grain boundaries at high temperatures [50].

2.3. Physical Methods

In general, thermal evaporation and sputtering are two primary physical approaches for the synthesis of semiconductor nanoparticles. For these methods, a relatively strict vacuum environment is required compared with the chemical methods to prevent contamination of the sample, the evaporation source, or the sputtering source. Most nanoparticles synthesized by these methods are in film form on certain specific substrates.

Vacuum thermal evaporation is a process in which evaporated flux from a desired material source is allowed to condense on a substrate. According to the different energy sources for the evaporation, this method is usually categorized as resistive evaporation, electron beam evaporation, ion vapor evaporation, cathodic arc deposition, or laser ablation. Detailed descriptions of these methods can be found in [71] and [72]. Typically, these evaporative deposition methods are performed in high vacuum without a gas atmosphere because the evaporated flux has low energy and a short mean free path. However, if an inert gas is used during the evaporation, the evaporated flux loses its energy quickly by collision with gas molecules. This results in the formation of powder by condensation. This process is also known as inert gas condensation (IGC) [72].

In sputtering, plasma is generated between an anode and a cathode. A large positive bias on the cathode, which is also the target to be sputtered, attracts the ions to the target. The ions accelerate across a dark space and impinge on the target, causing the target atoms to sputter. The sputtered flux can be deposited as a thin film and, depending upon the temperature and pressure, can form a nanostructured thin film. Moreover, the sputtered flux can be thermalized by collision with the sputtering gas such that nanoparticles are obtained. Specifically, TiO_2 thin films have been deposited by reactive magnetron sputtering [73], and CdTe quantum dots were synthesized by RF magnetron sputtering [74–76]. Ordered semiconductor dots were also formatted by ion-beam sputtering [77, 78].

3. MECHANISM OF NANOPARTICLE PHOTOCATALYSIS

3.1. General Description

Photocatalysis can be a homogeneous or a heterogeneous process. Homogeneous photocatalysis is a process that utilizes oxidizers such as O_3 and H_2O_2, while simultaneously coupling with light for the mineralization of organic molecules. However, photoreaction with the irradiation of semiconductor nanoparticles is heterogeneous photocatalysis and does not require oxidizers. Instead, it takes advantage of excited electrons and/or holes from nanoparticles to alter the kinetics of chemical reactions. At the end of each catalytic cycle, nanoparticle photocatalysts can usually be regenerated. It is normally supposed that a good photocatalyst always accelerates the reaction rate if there is no special restriction. A comprehensive description of homogeneous and heterogeneous photocatalysis systems can be found in [3]. Heterogeneous photocatalysis is categorized as a sensitized reaction or a catalyzed reaction, based on the origin of the electrons and holes. If the generation of charge carriers is initiated from absorbate species and they are transferred to the catalyst at ground state, which then takes part in the reaction, then the process is a catalyzed photoreaction. On the other hand, if initial photoexcitation occurs from a nanoparticle catalyst and charge carriers are transferred to ground-state molecules, the process is called a sensitized photoreaction [29].

The whole process of nanoparticle photocatalysis is simplified to electronic excitation, trapping, recombination, interfacial charge transfer, and chemical reactions. The slowest kinetic process becomes the rate-limiting step and has the most influence on the resulting photocatalytic activity. The final efficiency of the chemical reaction is the overall interaction of these theoretically divided parts. They are dynamically connected to each other with a certain sequence to complete the photocatalysis process. For simplicity, a heterogeneous photocatalysis process is expressed by Eqs. (1) to (5):

Electronic excitation:

$$\left|\mathrm{NP} + h\nu \rightarrow e^-_{\mathrm{cb}} + h^+_{\mathrm{vb}}\right| \tag{1}$$

Charge trapping:

$$\left|\begin{array}{l} e^-_{\mathrm{cb}} \rightarrow e^-_{\mathrm{tr}} \\ h^+_{\mathrm{vb}} \rightarrow h^+_{\mathrm{tr}} \end{array}\right| \tag{2}$$

Charge recombination:

$$\left|\begin{array}{l} e^-_{\mathrm{cb}} + h^+_{\mathrm{vb}}(h^+_{\mathrm{tr}}) \rightarrow \text{photons/phonons} \\ e^-_{\mathrm{tr}} + h^+_{\mathrm{vb}}(h^+_{\mathrm{tr}}) \rightarrow \text{photons/phonons} \end{array}\right| \tag{3}$$

Interfacial charge transfer:

$$\left|\begin{array}{l} e^-_{\mathrm{cb}}(e^-_{\mathrm{tr}}) + \mathrm{A} \rightarrow \mathrm{A}^- \\ h^+_{\mathrm{vb}}(h^+_{\mathrm{tr}}) + \mathrm{D} \rightarrow \mathrm{D}^+ \end{array}\right| \tag{4}$$

Chemical reactions:

$$\left|\mathrm{D(A)} \xrightarrow{\text{oxidation (reduction)}} \mathrm{Min}\right| \tag{5}$$

where NP and Min stand for nanoparticles and mineralized products (including H_2O and CO_2), respectively. A and D are acceptor- and donor-type chemical substrates on the surface of nanoparticles, respectively. A pictorial diagram of these processes is shown in Figure 3 [79]. The details for photocatalysis will be discussed step by step in the following sections. The mechanism to be discussed is general and fundamental. Specific photocatalysis cases for modified nanoparticles are included in Section 4.

3.2. Electronic Excitation

In photocatalysis processes, electronic excitation is the initial step for producing highly reactive and electronically excited states by photon absorption. Photoexcitation can occur from either nanoparticle catalysts or adsorbate molecules. Details of molecular electronic excitation are presented in Section 4.4.4. The excitation in semiconductor nanoparticles is addressed in the following section.

Semiconductors have an energy band gap (E_g) that ranges from the top of a filled valence band to the bottom of a vacant conduction band. Figure 4 shows the potential positions of valence and conduction bands for several semiconductors [29]. These data are for bulk materials. Nanoparticles have a small blue shift of the band gap, and, especially for Q-particles, the band gap increases with decreasing particle size [80, 81]. Electrons in the valence band require a certain energy to overcome the band gap and arrive at the conduction band. Photons that have energy ($h\nu$) equivalent to the band-gap energy can excite the electrons. With the interaction between incident light and nanoparticles, electron and hole pairs can be simultaneously generated because the same number of positive holes are left in the valence band to keep the nanoparticles electrically neutral upon the excitation of negative electrons to the conduction band. According to the semiconductor band-gap theory, direct and indirect absorption processes can occur in direct and indirect band-gap semiconductors, respectively. There is no phonon, which is the energy of lattice harmonic oscillation, involved in direct photon absorption to create free electrons and holes. Extra phonon energy is absorbed or emitted to excite electrons for momentum and energy conservation in an indirect absorption process. Indirect interface absorption can be induced in both direct and indirect band nanoparticles because large amounts of surfaces and grain boundaries help the electron transition and thus enhance the light absorption.

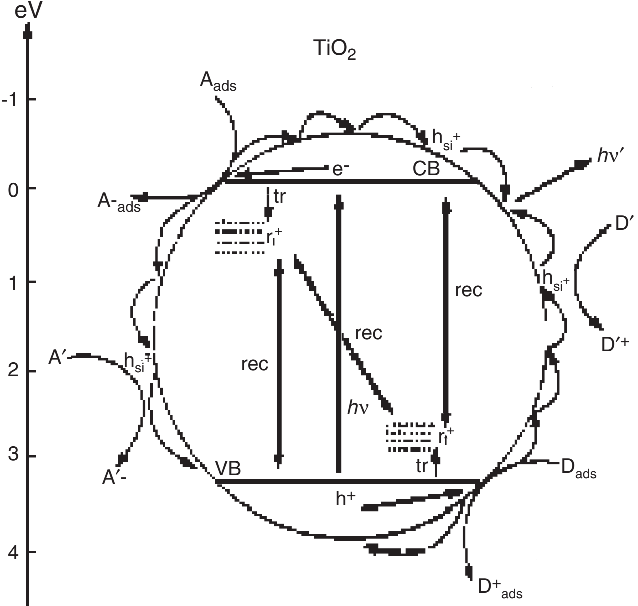


Figure 3. Pictorial diagram of heterogeneous photocatalysis in TiO_2. CB, conduction band; VB, valence band; ST, shallow trapped; DT, deep trapped; SR, surface recombination; VR, volume recombination; A, acceptor; D, donor; $h\nu$, energy of incident light. Reprinted with permission from [79], N. Serpone et al., *J. Phys. Chem.* 99, 16655 (1995). © 1995, American Chemical Society.

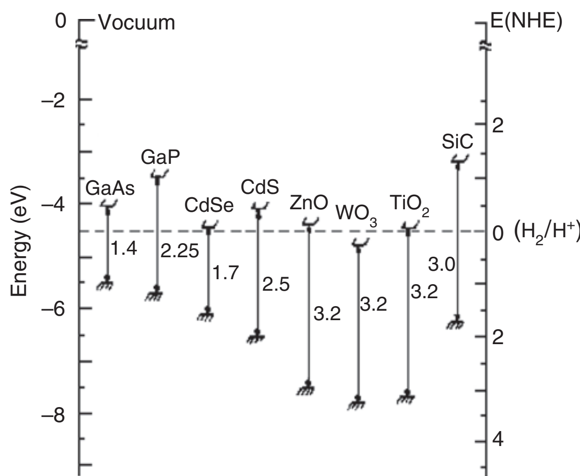


Figure 4. Conduction-band edge, valence-band edge, and band gap of some semiconductors in the aqueous electrolyte at pH 1. Reprinted with permission from [29], A. L. Linsebigler et al., *Chem. Rev.* 95, 735 (1995). © 1995, American Chemical Society.

The rate of charge carrier generation is closely related to the efficiency of photon absorption. The photon absorption is a function of active absorption area and the band gap of nanoparticles. Smaller particles have a larger surface area and can provide more absorption sites to increase the absorbed photon numbers in a given time. Also, as mentioned above, nanoparticles can increase the possibility of indirect interface electron transition to increase the light absorption. The photoresponse can be extended to longer wavelengths by dye sensitization, attachments of foreign species on the surface, and band-gap narrowing by doping, etc., so that charge carriers are more easily produced with lower-energy irradiation.

3.3. Charge Carrier Transfer

3.3.1. Trapping

The photoexcited free electrons and holes experience charge trapping and recombination during their bulk migration and interfacial transfer for chemical reactions on the surface of nanoparticles. The separated charge carriers have a lifetime on the order of nanoseconds to undergo transfer [82].

Therefore, the charge trapping and recombination processes in this period are critical. They greatly affect the final total number, lifetime, and dynamics of free charge carriers on the surface of nanoparticles to change the photoreactivity because trapping is helpful for charge carrier separation and negates the carrier losses due to recombination. The dynamics of the charge carrier trapping process has been studied with various methods, especially time-resolved techniques. A stochastic kinetic model was used to interpret the charge-trapping process in colloidal TiO_2 particles [83]. Picosecond and femtosecond transient absorption spectroscopy has also been utilized to investigate the trapping processes in ZnO and TiO_2 nanoparticles [9, 79, 84, 85]. Laser flash photolysis has been performed to study the electron trapping in WO_3 [86, 87], SnO_2 [88], and ZnO [89–91]. Time-resolved photon echo experiments were done to deduce the trapping time of CdSe nanoparticles [92]. The trapped charge carriers can be characterized by electron paramagnetic resonance (EPR), absorption spectroscopy, and emission methods because their existence is sufficiently long and can be from microseconds to seconds [9].

The trapping sites can be surface defects, lattice sites, and dopants in nanoparticles. The surface in nanoparticles is large and is usually not ideal or regular. The nature of surface defect sites is related to the preparation method and can be changed by various surface modification methods (Section 4.4). The surface irregularities, such as vacancies, and coverage of foreign species, form surface defect sites which, in turn, introduce surface electron states into the band gap of semiconductor nanoparticles. They become important localized trapping centers for charge carriers by the energy difference between the trap and the bottom of the conduction band or the top of the valence band. Charges trapped at the surface of CdS single-crystal electrodes were shown by photoreflectance spectroscopy [93]. Both low-temperature EPR [94] and ambient-temperature spin-trap EPR [95] showed that holes can be trapped as surface-bound $O^{\bullet}$ or $^{\bullet}OH$ radicals in the presence of O_2 and H_2O on the surface of TiO_2. Moreover, the absorbed O_2 molecules on the surface are very important electron traps for forming O_2^- in most photocatalytic reactions [96, 97]. The lattice sites, such as host atoms and oxygen vacancies, can also become trapping centers. Some trapping in the lattice sites may cause chemical changes. Time-resolved analysis showed that CdS nanoparticles have chemical changes caused by reactions between trapped electrons and the Cd^{2+} site [98–101]. The chemical changes can be found elsewhere for CdSe [101] and PbS [102]. EPR studies showed that Ti^{4+} can trap electrons to form Ti^{3+} within anatase TiO_2 particles [103–106]. Oxygen vacancy is also implicated in electron trapping in TiO_2 [107]. The most interesting and significant trapping is from dopants. The impurities can be added as dopants during or after synthesis of nanoparticles. Dopants as trapping sites are limited to a certain concentration beyond which recombination becomes dominant because a high density of trapped charge carriers also increases the possibility of recombination with following opposite charge carriers generated by the photon.

Trapping can form shallow traps and deep traps. Trapping on surfaces is determined by the surface properties. The basic hydroxide groups on TiO_2 surfaces can cause deep traps for valence band holes [9]. Trapping by dopants depends on the potential of dopants in the band gap of semiconductors. Trapping from dopants that have a potential energy just below the conduction band or just above the valence band is considered shallow trapping; otherwise it is deep trapping. Shallow trapping is preferred for photoreactivity enhancement. Although deep trapping can also extend charge carrier lifetime, it may cause lowering of the redox potential of nanoparticles and cause a decrease in photoreactivity [12].

The trapping times for electrons and holes are different. Generally, electrons are easier and quicker to trap than holes. The hole and electron trapping time in CdS nanoparticles was found to be 1 ps [108] and 100 fs or less [109, 110], respectively. The trapping of holes generally occurs within 250 ns [10, 29], but electron trapping is usually completed within 180 fs for TiO_2 and most other semiconductor nanoparticles [104]. This is compatible with the results of a femtosecond time-resolved diffuse reflectance spectroscopy study for wet and dry Q-TiO_2 and dry P25 particles which showed that electron trapping takes place in less than 200 fs in all three samples [111]. However, a discrepancy in trapping time still exists. Longer electron trapping times of 0.5 to 8 ps for CdSe [92] and 30 ps for CdS [108, 112] were reported. The difference may be caused by the quality of samples, the nature of the nanoparticles, and trapping states (shallow traps or deep traps) [18].

Trapped electrons and holes are likely to be released. Free electrons or holes with extended lifetimes will participate in chemical reactions. If trapped electrons and holes are also to contribute to the reactions, first of all they have to be detrapped from trapping sites and obtain sufficient mobility for transfer without being recombined before reaching the reactants. This condition requires that trapped charge carriers have survived for a long time.

3.3.2. Recombination

The excited electrons and holes are thermodynamically unstable and have a tendency to recombine to lower system energy. This process is called charge carrier annihilation or recombination. Recombination can be radiative or radiationless. Radiative recombination emits light (photons), and radiationless recombination emits heat (phonons). Time-resolved techniques are used to study the recombination processes.

Surface recombination and volume recombination are two events for electronic decay in nanoparticles. The recombination strongly depends on particle size. For particles with sizes larger than several tens of nanometers, including bulk, volume recombination is dominant. Volume recombination is the direct or indirect recombination between conduction-band electrons and valence-band holes. After excitation, electrons and holes have electronic relaxation to the bottom of conduction and the top of valence bands, respectively. The relaxation process occurs on the order of 100 fs by interaction between electrons and phonons. Direct recombination of charges in the bands could occur, and it is often radiative with intense band edge luminescence if there are only a few or no electronic states within the band gap [18]. This would last for nanoseconds. With the increase

in band-gap states, volume recombination is dominated by direct exciton-exciton annihilation rather than recombination of trapped charge carriers. When there are electronic states induced by surface defects and dopants in the band gap, part of the conduction band electrons and/or valence band holes can be trapped during their movement back to a lower energy state. The trapped charge carriers then undergo radiative or irradiative recombination. This recombination can happen between conduction-band electrons and trapped holes, valence holes and trapped electrons, and trapped electrons and trapped holes. The recombination rate is slower than that of direct annihilation [18] and is related to the trapping depth. However, when particle size decreases below several nanometers and causes a high density of charge carriers on the surface, surface recombination starts to take place. Surface recombination refers to the decay between surface trapped electrons and holes, surface nontrapped electrons and holes, or surface trapped and nontrapped charge carriers. Therefore, a size window exists for nanoparticles to have a low rate of charge carrier decay beyond which volume recombination and surface recombination will become too large for large and ultrafine particles, respectively.

First-order and second-order recombinations are two primary kinetic processes. Second-order recombination is faster than first-order recombination. Serpone et al. proposed a model of recombination kinetics for trapped electrons and trapped holes based on the results of time-resolved emission decays. They demonstrated that the decay is first order in 2.1-nm TiO_2, whereas second-order decay occurs for 13.3 and 26.7 nm [79]. The equations are

First-order recombination:

$$I(t) = A' \exp(-k_r t)$$

Second-order recombination:

$$I(t) = \frac{A}{(1 + ([e^-/h^+]/VN)k_t t)^2}$$

where $I(t)$ is the ratio of the average concentration of electron/hole pairs at times t and 0 ps, A and A' are constants, k_r is a rate constant, $[e^-/h^+]$ is the average concentration of charge pairs at time t, V is the particle volume, and N is Avogadro's number. Earlier, picosecond and nanosecond transient absorption methods showed second-order decay of trapped electrons in TiO_2 [83], which was confirmed by femtosecond time-resolved diffuse reflectance spectroscopy by Colombo and Bowman [111] and other researchers [113].

The decay time of the charge carrier is quantitatively studied. EPR measurements showed that the recombination of shallow trapped and deep trapped charge carriers for Cu_2S nanoparticles are 1.1 ps and 80 ps, respectively [18]. Femtosecond transient absorption/bleach spectroscopy indicated that deep trapped electrons recombine with associated recombination times greater than 1 ns in Ag_2S nanoparticles [18]. Decay times for 2.1-, 13.3-, and 26.7-nm TiO_2 were determined to be 67, 405, and 66 ps, respectively, by transient emission decays [79]. The dependence of the recombination rate on the particle size is consistent with the size window for volume recombination and surface recombination, as discussed above.

3.3.3. Interfacial Transfer

Interfacial transfer is not only the continuation of bulk migration; it is also a process connected to the chemical reaction. To participate in the chemical reaction, electrons or holes must undergo interfacial transfer to the acceptors or donors attached on the nanoparticle surface, in addition to their bulk migration. The driving force for the electron transfer is the energy difference between the conduction band (or energy of trapped electrons) and the reduction potential of the acceptor redox couple. Similarly, the driving force for the hole transfer is the energy difference between the valence band (or energy of trapped holes) and the oxidation potential of the donor redox couple. As mentioned above, before the charge carrier transfer on the particle surface, the photogenerated electrons and holes have to move from the bulk to the surface by diffusion. The time of bulk migration is largely dependent on the mobility and lifetime of charge carriers and particle size. The average times of electron and hole migration from bulk part to the surface for 2.1-, 13.3-, and 26.7-nm TiO_2 are 0.05, 1.8, and 9.9 ps, respectively [103, 104]. The overall efficiency of interfacial transfer is determined by the competition between recombination and trapping followed by the competition between recombination of trapped carriers and interfacial charge transfer [12, 114]. The interfacial electron transfer directly competing with the recombination process was also verified by Kamat et al. [115] and other scientists [111].

Size and surface characteristics of nanoparticles have an extremely special role in interfacial charge transfer. The main reason for this is that the trapping and recombination processes are critically affected by surface-related factors, as mentioned earlier. For Q-particles, one additional reason for this is the increase in charge carrier potential, which can increase the driving force of transfer due to the band-gap enlargement. The electron transfer process from CdS nanoparticles to surface-absorbed viologens was investigated, and the ratio of recombination rate constant to electron transfer rate constant was found to be dependent on the particle size. It increases by a factor of 20 with a doubling of particle radius [116]. The interfacial hole transfer dynamics of P-25 TiO_2/SCN^- was probed as a function of thiocyanate ion concentration on the surface by femtosecond time-resolved diffuse reflectance spectroscopy because of the hole-scavenging capability of SCN^- [111]. The repression of e^-/h^+ recombination is also contributed by valence-band h^+ consumption in the oxidation of H_2O_2 by terephthalic acid fluorescence probing [117]. The enhancement of electron transfer on the surface of semiconductor Q-particles is attributed to both an increase in electron energy in the conduction band and changes in the physical and chemical nature of the surface, such as change in surface charge and increase in chemical nonstoichiometry [118]. Generally, the rate constant of interfacial electron transfer is determined to be greater than 5×10^{10} s^{-1} [119–121].

3.4. Chemical Reactions

3.4.1. Pollutants

The surviving electrons and holes may accelerate or initiate the photoreactions by interfacial charge transfer to chemical substrates. The type of chemical depends on the

selectivity of semiconductor nanoparticles, most of which have low selectivity. Nanoparticle photocatalysts can join in many chemical reactions, among which pollutant degradation is of the most concern for environmental protection.

Organic and inorganic compounds are two types of common pollutants. Organic compounds include halogenated organics {chlorophenols (CP) [2-CP, 4-CP, pentachlorophenol], chlorinated compounds [polyvinyl chloride, chlorinated dioxins, chlorinated dibenzofurans, trichloroethylene (TCE), and DDT], fluorophenol [*m*-fluorophenol], and PCBs, etc.} and commercial colorants {dyes [naphthol blue black, acid orange 7, disperse blue] and pigments}. Inorganic compounds include heavy metal species (Pb^{2+}, Sn^{4+}, As^{5+}, Hg^{+}) and ammonia, etc. Table 1 contains a list of typical organic pollutants ([12] and references therein).

3.4.2. Advanced Oxidation Processes

The function of photocatalysis in chemical reactions is versatile. First of all, there are two types of photocatalysis, homogeneous and heterogeneous. In homogeneous photocatalysis, reactions are oxidative because of the direct oxidizing agents (O_3, H_2O_2, O_3/H_2O_2, etc.). For heterogeneous photocatalysis, which involves semiconductor nanoparticles, the reaction type largely depends on the redox potential of chemical molecules. If the redox level is below the conduction band edge of nanoparticles, reduction reaction can take place. On the other hand, oxidation reaction will occur if the redox level lies above the valence band edge. In the case of redox levels nearly centered in the band gap, both oxidation and reduction reactions could occur. As a result, nanoparticles are capable of photodecomposition of small inorganic molecules, photooxidation of organic and inorganic species, and photoreduction of organic and inorganic species. Extensive examples of reaction types over TiO_2 were given by Howe [30] and Linsebigler et al. [29]. From the viewpoint of pollutant remediation, most reactions are oxidation reactions.

Photocatalysis is an advanced oxidation process (AOP), as are radiolysis [122–126], sonolysis [127–131], and other methods [9]. These oxidation processes are distinguished from each other by the way hydroxyl radicals ($^\bullet OH$) are generated because they have common $^\bullet OH$ as a primary oxidation agent to make the oxidation reaction possible in an aqueous system. Radiolytical processes produce $^\bullet OH$ by the irradiation of water with γ-ray or high-energy electron beams from ionizing radiation sources. Sonochemical methods use ultrasound to sonicate water in the presence of dissolved gas to form bubbles first and then dissociate H_2O to form $^\bullet OH$. A photocatalysis reaction forms $^\bullet OH$ by the incorporation of water with holes that form because of the irradiation of direct oxidizing agents in homogeneous systems or nanoparticles in heterogeneous systems. In addition to $^\bullet OH$ radicals, direct h^+ from catalyst and hydrogen peroxide radicals ($HO_2^\bullet$, $^\bullet O_2^-$) can also be possible oxidizing agents for heterogeneous photocatalysis. $^\bullet OH$ is considered to be a very active and effective chief radical for the oxidation of organic molecules in aqueous slurry systems. The existence of $^\bullet OH$ can be detected by radical detection methods with electron spin resonance spectroscopy [132–135] and other analysis techniques [136–138]. Extensive studies in the field of AOP photocatalysis have been conducted [3, 9, 139, 140]. By photocatalytic reaction, hydrogen, oxygen, and carbon are transformed into H_2O and CO_2. In addition to this, other typical elements in organic contaminants, such as nitrogen, sulfur, phosphorus, and halogen, can form NO_3^-, SO_4^{2-}, PO_4^{3-}, and other acid minerals by oxidation processes.

Liquid-solid and gas-solid are two representative heterogeneous photocatalysis reactions for pollutant remediation in water and air, respectively. Vast amounts of photocatalysis research have dealt with the liquid phase [2, 5, 7], with relatively less work carried out on gas-phase systems [7, 140–145]. For gas-phase reactions, O_2, which is adsorbed with H_2O molecules on the catalyst surface, can largely improve the oxidation reaction. Although O_2 molecules are not only electron scavengers but also the main radicals with oxygen-derived formations ($^\bullet O_2^-$ and $HO_2^\bullet$, etc.) that oxidize organic molecules, the existence of O_2 alone is found to be inefficient for oxidation. H_2O is necessary to improve photooxidation reactions in gas-phase systems. The role of H_2O in the gas-phase reaction is complicated and still not very clear. Fourier transform infrared studies revealed that the photodegradation rate of 4-CP was dramatically enhanced in the presence of water, but the action of water was not known [146]. Water was determined to form $^\bullet OH$ with holes, but it was not confirmed to be responsible for the oxidation of TCE as studied by the effect of different water vapor pressures on the photoreactivity of TiO_2 [147, 148]. The initial oxidation of TCE was due to oxygen-derived radicals but not $^\bullet OH$ radicals. This result was obtained by the investigation of the photoreactivity of hydroxylated TiO_2, with the use of isotopically labeled oxygen and water [29]. The presence of water was found to be helpful for the removal of reaction intermediates or products from the TiO_2 surface to prevent poisoning in both gas and liquid phases [149]. However, for the liquid phase reaction, both O_2 and H_2O were required for the photooxidation process. O_2 is mainly responsible for scavenging of electrons, and water is used for $^\bullet OH$ formation, so that oxidation reactions can be successfully completed. As mentioned above, $^\bullet OH$ radicals are the primary oxidation agents in this case. The aqueous reaction is commonly used for water purification. The reaction pathways in heterogeneous photocatalysis are complicated. For different pollutants and catalysts, the reaction pathways may be different.

4. ENHANCEMENT OF PHOTOCATALYTIC ACTIVITY

4.1. Quantum Yield

Photodegradation of pollutants is a direct way to examine the performance of semiconductor nanoparticles for photocatalysis. Quantum yield (quantum efficiency) is a criterion for measuring the efficiency of a photocatalytic reaction. It is defined as the number of consumed or destroyed molecules per absorbed photon [150]. In the case of multiple reactants, quantum yield is sometimes based on the consumption of a particular reactant. Nevertheless, precise determination of actual absorbed photons is much more difficult in heterogeneous than in homogeneous media because of the light

Table 1. List of typical organic compounds.

Chlorinated aromatics	Nitrogenous compounds	Hydrocarbons, carboxylic acids, alcohols, halocarbons, and heteroatom compounds
2-Chlorophenol	2-, 3-, and 4-Nitrophenol	1,3-Butadiene
3-Chlorophenol	2,5-Dinitrophenol	1,2-Dibromoethane
4-Chlorophenol	Trinitrophenol	2,2,5-Trimethylpentane
2,4-Dichlorophenol	Atrazine	2-Ethoxyethanol
3,4-Dichlorophenol	Dimethylformamide	2-Methoxyethanol
2,6-Dichlorophenol	Nitrobenzene	Acenaphthene
2,4,5-Trichlorophenol	4-Nitrophenyl ethylphenylphosphonate	Acetone
Pentachlorophenol	4-Nitrophenyl diethyl phosphate	Benzene
Chlorobenzene	4-Nitrophenyl isopropylphyl-phosphonate	Ethylbenzene
1,2,4-Trichlorobenzene	Azobenzenes	Benzoic acid
1,3-Dichlorobenzene	Cyclophosphamide	Bromoform
1,2-Dichlorobenzene	EDTA	Catechol
1,4-Dichlorobenzene	Methyl orange	Cresols
2,3,4-Trichlorobiphenyl	Methylene blue	Cyclohexane
2,7-Dichlorodibenzo-*p*-dioxin	Methyl viologen	Diethyl phthalate
2-Chlorodibenzo-*p*-dioxin	Monuron	Di-*n*-butyl phthalate
2,4,5-Trichlorophenoxyacetic acid	Nitrotoluene	Ethylene
2,4-Dichlorophenoxyacetic acid	Picoline	Formaldehyde
Hexachlorobenzene	Piperidene	Hexane
PCBs	Proline	2-Propanol
DDT	Pyridine	Malathion
Chlorinated surfactants	Simazine	Methanol
	Theophylline	Methyl vinyl ketone
Chlorinated aliphatic	Thymine	Naphthalene
and olefinic compounds	Trietazine	Phenol
1,1,1-Trichloroethane		Polynuclear aromatics
1,1,2,2-Tetrachloroethane		Propene
1,1,1,2-Tetrachloroethane		*tert*-Butyl alcohol
1,1,2-Trichloroethane		Toluene
1,1,2-Trichloro-1,2,2-trifluoroethane		Xylene
1,1,1-Trifluoro-2,2,2-trichloroethane		1,3-Diphenylisobenzofuran
1,1-Difluoro-1,2,2-trichloroethane		Bromodecane
1,1-Difluoro-1,2-dichloroethane		Bromododecane
1,1-Dichloroethane		Dodecanol
1,2-Dichloroethane		1-Propanol
1,2-Dichloroethylene		Fluorophenols
1,2-Dichloropropane		(4-Thiophenyl)-1-butanol
bis(2-Chloroethyl) ether		4-Hydroxybenzyl alcohol
Carbon tetrachloride		Acetic acid
Chloroacetic acid		Acetophenone
Chloroethane		Adipic acid
Chloroform		Alkylphenols
Methylene chloride		1-Butanol
Tetrachloroethylene		Butadiene
Trichloroethylene		Butyric acid
Chloral hydrate		Cyclohexene
Chloranil		Cyclohexanedicarboxylic acid
Chloroethylammonium chloride		Dibromomethane
Dichloroacetic acid		Diphenyl sulfide
Trichloroacetic acid		Dodecane
		Dodecyl sulfate
		Dodecylbenzene sulfonate
		Ethane
		Ethanol
		Ethyl acetate
		Formic acid

continued

Table 1. Continued

Chlorinated aromatics	Nitrogenous compounds	Hydrocarbons, carboxylic acids, alcohols, halocarbons, and heteroatom compounds
		Isobutane
		Isobutene
		Lactic acid
		Oxalic acid
		Propionic acid
		Pyridine
		Salicylic acid
		Sucrose
		Tetrafluoroethylene

Source: From [12] and references therein.

reflection and scattering on the surface of nanoparticles, absorption by the supports, and/or transmission in transparent colloidal sols. It is, therefore, ideally assumed that all of the photons are absorbed. The yield, therefore, is the apparent quantum yield. It is the most frequently used and can be calculated from experimental degradation rate curves. Another widely used criterion is the apparent reaction rate constant, because most photodegradations of dilute organic reactions are pseudo-first-order. However, significant variations in apparent quantum yield of the same catalyst for the same type of reactant in different photosystems were observed [151–153]. A perfect measurement should give consistent photocatalytic activity in different test conditions. It turns out that neither the apparent reaction rate constant nor the apparent quantum yield can fully reflect the intrinsic efficiency of nanoparticles. It was found that apparent quantum yield changes with the geometric size of the reaction zone [154], whereas the apparent reaction rate constant varies with the concentration of the catalyst and the power of irradiation [155]. They only provide apparent photocatalytic activity and can be used to compare relative photoreactivities in the same laboratory.

The factors affecting the apparent photocatalytic activity can be either operational variables or the quality of nanoparticles. The operational variables are external factors, such as irradiation conditions, photoreactor features, and other experiment parameters (amount of catalyst, pH value, temperature, and volume of the solution, etc.). For instance, the photodecomposition rate of water by Pt-RuO_2/TiO_2 was shown to be a linear function of the irradiation intensity, an exponential function of reaction temperature, and an exponential function of steady-state pH value [156]. The oxidation rate of phenol in aqueous solution with anatase TiO_2 was also affected by the initial pH of the suspension and the presence of Cl^- and H_2O_2 [157]. To find intrinsic photoreactivity, the same quality of photocatalyst is used while experimental conditions are varied.

Radiation conditions and photoreactor features are unstable aspects varying the apparent photoreactivity. For solution degradation experiments, three typical photoreactors are used: annular photoreactors, flat-plate reactors, and optical fiber reactors [158]. These reactors were described in detail by Howe [30]. The actual light absorption does vary with irradiation conditions, concentration of particles in solution, and photoreactor features. The irradiation conditions include uniformity of photon absorption in solution and other properties (wavelength, intensity, etc., of the incident photons). Light with a wavelength longer than the absorption edge of nanoparticles will be less energetic. An increase in light intensity can speed up the absorbed number of photons in unit time and vary the steady-state concentration of active radicals to change the photoreactivity. The rate of photochemical reaction on the surface was approximated as either a linear [144] or a square root function of irradiation intensity [159]. Radiant flux profile is not a constant, especially in large annular reactors and thick slurries, because of the resistance of light transfer. Therefore, particles in different positions of the reactor may have different absorption rates. The nonuniformity of radiation distribution affects reaction kinetics, and the overall quantum yield is decreased by increased reaction zone thickness [160]. The effect of catalyst concentration on the light absorption is not linear. With the increase in particle concentration, light absorption could be correspondingly enhanced because more absorption sites are active. But the resistance of light transfer becomes high when particles become thick. In this case, the farther the particles are from the illumination source, the lower the absorption coefficient. This situation is more noticeable in annular photoreactors than in others in which particles are coated with thin films.

The apparent kinetics of the degradation reaction is also a function of the amount of catalyst, pH value, temperature, volume of the solution, and foreign species in the solution. Increases in particle amounts to enhance the light absorption, as mentioned above, can supply more charge carriers. The same amount of particles and a different volume of solution give different particle concentrations and affect the degradation rate. The pH value can vary the surface charge and possibly change the band energies of metal oxide semiconductors with 0.059 mV/pH unit [9]. Higher pH values can provide more OH^- and form more $^{\bullet}OH$, which increase the oxidation reaction rate. Higher temperatures are favorable for endothermic reactions. The temperature factor can be avoided by cooling degradation cells with cold water or air to keep their temperature constant. Finally, there are other factors that affect the degradation results, including supplied oxygen and the presence of foreign ions in the solution. O_2 is an electron scavenger and is helpful for charge carrier separation. No slurry photooxidation reactions can be completed without O_2. The oxidation reaction rate on TiO_2 is related to the reduction rate of O_2 by electrons. Furthermore, the rate of radiationless

recombination was enhanced if the reduction rate was not sufficiently high [161–163]. Actually, photooxidation experiments are always supplied with saturated O_2, so the effect of O_2 can be minimized.

Precise quantum yield measurements are carried out to obtain intrinsic photocatalytic activity. The measurements are based on the approach of intrinsic reaction kinetics or the redefinition of quantum yield for the correction of apparent photocatalytic activity. In fact, it is very difficult to correct all of the operational variables to achieve intrinsic photoreactivity. Most efforts focus on the effects of irradiation conditions and photoreactor features because conditions such as amount of catalyst, pH value, temperature, and volume of solution are relatively easy to control and repeat precisely in different systems. As a result, irradiation and photoreactor factors are theoretically minimized to make the derived intrinsic photodegradation rates by a particular photocatalyst independent. Some attempts have been made to obtain a universal mean of photocatalytic activity. The ratio of hydrogen production rate per unit mass of catalyst to the transmittance of incident light through slurry solution was defined as the intrinsic production rate of hydrogen in the photocatalytic cleavage of water. This production rate was not dependent on reactor geometry and catalyst loading [156]. A kinetic model, comprising a single equation for low and high irradiation conditions, was successfully applied to the photodegradation of trichloroethylene in water [164]. Relative photonic efficiency, as a correction factor, was suggested for the comparison of photocatalytic activities of different photocatalysts by Serpone [165]. Recently, a correction factor representing the ratio of the radical generation rate in the laboratory to that in reference-ideal photocatalytic reactors has been introduced to obtain the intrinsic reaction rate constant from the measurement of the apparent reaction rate constant. The intrinsic constant was found to be independent of the reactor size, light power, and catalyst concentration [160].

The overall intrinsic value of photocatalytic activity clearly relies on the quality of the nanoparticles. The quality is defined by multiple factors, such as the efficiency of the photocatalytic reaction process, the wavelength range of response, the stability of particles under irradiation, and the selectivity of the reactants. For example, most semiconductor nanoparticles have large band gaps, which limit their use with short-wave irradiations such as ultraviolet light. Shifting the absorption edge to longer wavelengths can increase the photoabsorption efficiency, and visible light absorption in particular can greatly extend the application of catalysts to sunlight. Improving photoreactivity is always a challenge in the area of photocatalysis. The quality of nanoparticles varies with the structure, size, shape, doping, and various surface modifications of the particles. The enhancement of quantum yield is based on these interior factors. These issues are discussed in Sections 4.2–4.4.

4.2. Structure and Size Effects

4.2.1. Purpose

The significance of nanoparticles for photocatalysis was mentioned in the Introduction. In the following, the structural, electronic, and optical properties of nanoparticles, especially quantum-sized particles, Q-particles, will be discussed further. Because of the specialty of Q-particles, their properties have more influence on quantum yield than do the properties of common nanoparticles. Thus, understanding quantization effects is the prerequisite to studying the size effects on the photoreactivity of nanoparticles. Enhancement of photoreactivity can be expected by structure and phase selection, as well as size modification. Four aspects are considered to be helpful for photoreactivity improvement with a decrease in particle size: (1) increasing redox potential by band-gap enlargement, (2) shortening the pathway of charge transfer, (3) decreasing volume recombination, and (4) enhancing light absorption and the effective reaction area with increased surface area.

4.2.2. Structural Change

Nanoparticles can be amorphous or crystalline, depending upon the synthesis methods and conditions. Atoms in an amorphous structure have only short-range order because the atoms do not have enough activation energy and/or time for diffusion during synthesis. Amorphous particles can be crystallized if proper treatment is used to meet the energy requirement. Crystals are either polycrystalline or single crystalline, which have Bravis lattice structures with atoms in ordered sites. Polycrystals are composed of single crystals with different orientations separated by grain boundaries. For crystals, different structures or phase transitions could occur with changing temperature, pressure, and composition. For instance, TiO_2 crystals can be in at least eight formations: anatase, rutile, brookite, α-PbO_2, β-VO_2, hollandite, fluorite, and baddelleyite (ZrO_2) [19, 166, 167]. Anatase, rutile, and brookite are known to be relatively simple structures at normal pressure.

The surface structure of nanoparticles is different from that of bulk. In nanoparticle systems, atoms always terminate with a surface, and the surface area thus becomes extremely large. Moreover, the periodic arrangement may become different at the surface to meet the requirement of minimization of the overall system energy. It is called structural relaxation and can be achieved by surface reconstruction and lattice distortion. Once a surface is created, the surface accumulates unbalanced electric charges and has a composition different from that of the bulk counterpart. For instance, oxide surfaces nonstoichiometrically have excess oxygen with dangling bonds. Surface free energy and surface stress are two crucial factors for surface relaxation. Surface excess free energy is generated because of the deviation of surface electronic structure from that of bulk. It is quantitatively equal to the total energy difference between nanoparticles and bulk of the same amount, divided by the surface area. Surface stress is produced because of the force exerted among excess dangling bonds on the surface, and it is given as the force per unit length. The surface atoms can reconstruct or rearrange themselves when the difference between surface stress and surface free energy is large [168, 169]. The lattice distortion or reconstruction occurs to reduce dangling bonds on the surface and to relax the lattice of nanocrystals. With the decrease in particle size, more distorted lattice cells appear because of the surface stability. The crystal structure thus tends to be hard to identify

because the lattice is so distorted and so many atoms are displaced. Banfield and Zhang proposed a molecular dynamics simulation of a structural model for nanosized TiO_2 [19]. Dangling oxygen ions that prefer to form bridged oxygen pairs on the surface are the key to the structural change. By structural relaxation, the interatomic distance between two titanium atoms on the surface is either shortened when the surface tensile stress exists or enlarged when compressive stress forms to lower the system energy. The surface structure was clearly identified as being composed of three layers: surface, near surface, and interior. The lattice is most distorted on the surface layer and then gradually reaches the bulk structure at the interior layer. The interior layer decreases with the decrease in particle size and finally disappears when TiO_2 nanoparticles are very fine.

Phase structure and composition affect the photoreactivity. Iron oxide has Fe_2O_3, α-Fe_2O_3, β-Fe_2O_3, and γ-Fe_2O_3 structure formations, of which α-Fe_2O_3 is the most stable and effective catalyst [170]. However, α-Fe_2O_3 was found to be active only for the oxidation of sulfite, oxalate [170], and iodide [171] but not for the oxidation of cyanide [172] and the formation of H_2O_2 [31]. A significant increase in activity of TiO_2 nanoparticles for the degradation of phenol was observed with a decrease in amorphous content, although the size of particles with amorphous content was smaller than that of the completely crystallized sample [173]. One possible reason for this was that the amorphous phase acted as a recombination center to decrease the number of photogenerated electrons and holes. Anatase is believed to have a higher photoreactivity than rutile. The effect of anatase mass fraction on the photodegradation of methylene blue and photodecomposition of ammonia gas was investigated, and it was found that the particles with a larger percentage of anatase had higher reactivity [174]. However, commercial standard Degussa P25 is composed of 20% rutile and 80% anatase, and the small content of rutile is demonstrated to have improved photoreactivity.

4.2.3. Quantization Effects

Decreasing particle size leads to a larger surface area and introduces more electronic states and unbalanced charges onto the surface. The structural changes are the basis of unique properties of semiconductor nanoparticles because surface structure and particle size greatly affect electronic structures, which in turn have effects on physical and chemical properties. Consequently, with the reduction of particle size, many novel properties can be observed. Quite drastic property changes, known as size quantization or quantum-size (Q-size) effects, occur because of the spatial confinement of charge carriers when the particle size decreases to less than the Bohr radius of the first excitation state [175].

There are discrepancies for the critical particle size or the threshold of Q-size effects. First of all, the critical size varies with the type of the particle because of the difference in effective charge. Q-size effects were observed by Henglein when the size of CdS particles was below 10 nm [175]. Quantization of MoS_2 nanoparticles occurred within 2 nm [176]. The critical size for TiO_2 was considered to be 10 nm [30]. However, even for the same particles, differences in critical size still exist. The first excitation value of TiO_2 was determined to be 0.75–1.9 nm [177], but various values for the size quantization regime were reported, including $\leq$53 nm (especially $\leq$10 nm) [105], 2.5–10 nm [178], and 2–50 nm [26], etc. More detailed descriptions of the size discrepancies can be found in [10]. The different results for size threshold determination can be caused by variations in surface characteristics, the shape of particles, the precision of measurement, or other undetermined factors.

Electronic properties of nanoparticles are strongly affected by size quantization. Q-particles are actually molecular clusters without complete electron delocalization. They have discrete or quantized energy levels for both conduction band electrons and valence band holes because the charge carriers are confined in potential walls with very small dimensions less than their de Broglie wavelength. This situation is like simple particles in a box. Serpone et al. demonstrated that semiconductor nanoparticles with a few molecular units can maintain discrete lowest unoccupied molecular orbitals (LUMOs) and highest occupied molecular orbitals (HOMOs) [25]. Thus, the electronic structure of nanoparticles can be studied with the molecular orbital (MO) method. The MO model for particles grown with N monomeric units was schematically illustrated by Hoffmann et al. [12], as shown in Figure 5. By the semiempirical MO approach, the influence of surface structures on the electronic state of CdS was directly calculated, and low-coordinated Cd atoms were found to produce surface states [179].

Optical absorption and photoluminescence are dependent on particle size. With a decrease in nanoparticle size, the effective band gap increases because photoexcited electron and hole pairs have insufficient space to coexist within the ultrafine particles. The change in band-gap energy is a strict function of particle size and the effective mass of charge carriers. By using the exciton model, Brus derived the first excitonic state of a semiconductor cluster [80, 81]

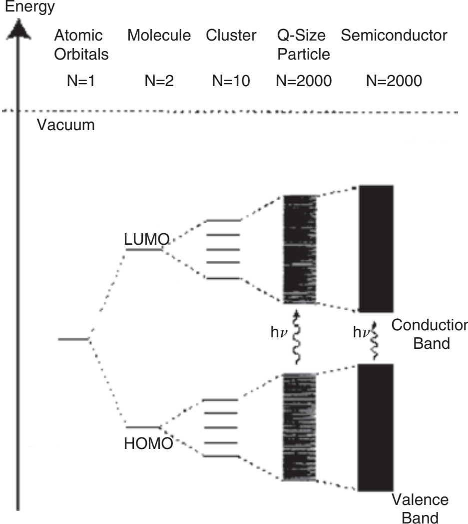


Figure 5. Molecular orbital model for electronic structure change with the growth of particle size. Reprinted with permission from [12], M. R. Hoffmann et al., *Chem. Rev.* 95, 69 (1995). © 1995, American Chemical Society.

as expressed by

$$\Delta E_g = E_g(R) - E_g(R = \infty) = \frac{\hbar^2 \pi^2}{2R^2 \mu} - \frac{1.8e^2}{\varepsilon R}$$

where $E_g(R)$ and $E_g(R = \infty)$ are band gaps of nanoparticles and the bulk state, respectively, R is the nanoparticle size, ε is the relative dielectric constant of the semiconductor, and μ is the exciton reduced mass, with a value of $(1/m_e + 1/m_h)^{-1}$ (m_e and m_h are the effective masses of electron and hole, respectively). The equation shows that nonnegligible band-gap energy change takes place and pronounced blue shift of the absorption threshold occurs in very fine particles, typically Q-particles. For comparison, the size dependency of the band gap is illustrated in Figure 5. Experimentally, the absorption as a function of particle size in CdS nanoparticles was observed and explained by Brus [80, 81], Henglein et al. [180, 181], and other scientists [179, 182, 183]. The comparison of absorption of ZnO with different particle sizes was made by Kamat et al. [89, 184]. Also, similar effects can be seen from other nanoparticles, such as MoS_2 [185], PbS [182], HgSe [186], PbSe [186], and CdSe [187]. For TiO_2, a 0.15-eV band-gap blue shift of 2–4-nm anatase and a 0.1-eV shift of 2.5-nm rutile were reported [177]. TiO_2 nanoparticles showed a blue shift in the absorption spectra with decreasing size, and the band gap increased from 3.027 to 3.167 eV with the particle size ranging from 39 to 3 nm [188]. However, no blue shift of 1.5-nm TiO_2 was reported, and the disability of the effective mass approximation in the band-gap energy change is a possible reason for this [189]. Different photoluminescence properties are observed in fine nanoparticles, and a large blue shift is indicated by different colors of Q-particles. For instance, CdS became colorless from a normally yellow color when the particle size was less than 2.2 nm [190].

4.2.4. Size Optimization

Size effects on the photocatalytic activity originate primarily from size quantization and size-related surface characteristics such as surface area and defects. With a decrease in particle size, charge carrier transfer can be accelerated because of either the ultrafine diameter of the crystals or the elevation of the redox potential of charge carriers by band-gap enlargement. At the same time, small particles offer increased light absorption area and active reaction sites due to larger surface area. For example, the adsorption rate and adsorbability of methylene blue (MB) were found to increase with the size reduction of TiO_2 nanoparticles [191]. However, decreasing particle size does not necessarily enhance the activity of nanoparticles. When particle size is reduced below a certain value, and most photogenerated electrons and holes are close to the surface, surface recombination will dominate, which shortens the lifetime of charge carriers. Moreover, blue shift of the band gap hampers the extension of the light absorption range and decreases the light absorption efficiency under the same irradiation energy range. These causes are unfavorable for the improvement of photoreactivity, so Q-particles do not necessarily show increased photoreactivity. Thus, the size dependency of photoreactivity is due to several competing factors, and an optimal size can be selected for the highest photocatalytic performance. Furthermore, size is one of the most important factors affecting the rate-limiting step. For instance, ZnO Q-particles may show enhanced photoefficiency for systems that have a rate-limiting step of interfacial charge transfer [192]. One more point should be mentioned here: that the photoreactivity may vary even for the same type of particles with the same size, because of the different preparation methods, which may cause the surface characteristics to be different.

Size effects on the photocatalytic activity of nanoparticles are investigated in various reaction systems. Anpo et al. [105] observed blue shift of the absorption edge and increased photocatalytic efficiency of TiO_2 for the photohydrogenolysis reaction of CH_3CCH in water with a decrease in anatase particle diameter from 53 to 3.8 nm, especially below 10 nm. This result was attributed to the enhancement of activity of charge carriers and/or suppression of radiationless transfer of the absorbed photon energy because of size quantization-related electronic structure modification. Improved catalytic efficiency with finer TiO_2 particles was also reported for the degradation of phenol [173] and MB [174, 191]. However, a linear increase in the degradation rate as a function of the anatase crystalline size, which was changed by varying the calcination temperature, was also reported [193]. Wang et al. demonstrated that photocatalytic activity could not monotonically increase with a decrease in particle size, and 11 nm was the optimal size of TiO_2 nanoparticles synthesized by hydrothermal methods for the decomposition of chloroform [48]. Particles with a larger or smaller size could have decreased photoreactivity due to the deficient surface area and high charge carrier recombination rate, respectively. These discrepancies can be understood based on the chemicals to be degraded, preadsorption of chemicals, actual activation surface area, number of surface states, and other factors.

4.3. Dopant Effects

4.3.1. Purpose

Dopants as foreign elements in nanoparticles are commonly used to enhance the photocatalytic activity. Cation, anion, and co-doping are common. Dopants can be introduced into nanoparticles either during the synthesis or after synthesis by processes such as ion implantation [194], thermal diffusion [195], or conventional solid-state reactions [196]. The dopant sources (precursors) are either ion compounds or neutral atoms, which can interact with the host lattice to build chemical bonds. Thus, the type of doping in this section refers to the final dopant chemical state(s) in the semiconductors. Dopants change the lattice thermal dynamics [197–200], electronic structures [201–203], and photocatalytic properties, etc., of semiconductor nanoparticles. Enhanced photocatalytic activity is expected by doping. In this situation, dopants might (1) increase charge separation efficiency by charge carrier trapping; (2) mediate interfacial charge transfer; (3) improve light absorption and change photoluminescence properties by introducing electronic states in the band gap or extend the light absorption range by narrowing the band gap; (4) reduce the particle size or change the phase composition. For doping, not only the enhancement of photoreactivity but also the effect on

visible light absorption is of interest. The actual role of dopants during the heterogeneous photocatalytic reaction is determined by their nature, position, and concentration in nanoparticles.

The natural properties refer to potential energy levels and oxygen affinity (for cations), etc. Dopants act as shallow traps of electrons if their energy level is near the conduction band of nanoparticles. They act as deep traps of electrons if their energy level is far below the conduction band of nanoparticles. Similarly, shallow trappings become possible if the dopant level is just above the valence band, and deep hole trappings are possible if the dopant level is far above the valence band. Shallow trapping increases the lifetime of charge carriers by successful separation of electron-hole pairs. However, deep traps reduce the energy difference between redox potential and trapped electrons or holes to slow down the charge carrier transfer, which is a crucial step for the overall reaction rate. As discussed in Section 3.3.1, the potential energy level of dopant in the semiconductor band gap should be considered. Accordingly, the dopants that introduce deep trapping are not suitable for the improvement of photoreactivity.

The dopants can be on the surface or in the lattice. The cases of dopant with high concentration on the surface are exclusively analyzed in Section 4.4 as surface modifications. Inside the lattice the dopant can be substitutional, interstitial, or on both sites. The effect of dopants on photoreactivity critically depends on their location and coordination in the particles. For example, substitutional dopant ions contribute mostly to the change in electronic structure and an increase in light absorption efficiency. Thus, determination of dopant position is necessary for investigations of the role of the dopant. However, the locations of dopant ions are affected not only by the nature (such as ionic radii) and concentration of ions, but also by the synthesis method. For dopant ions with sizes comparable to those of host ions, it is easier to occupy the host sites, as opposed to those that have radii much larger or much smaller than that of the host. The amount of dopant in the lattice is limited by the maximum equilibrium solid solubility limit of the dopants in the semiconductor nanoparticles. For different synthesis processes, dopant position also changes. Dopant ions may only be absorbed on the surface of particles during the initial hydrolysis step in the sol–gel method, and some of these ions are then incorporated into substitutional or interstitial sites of TiO_2 or form separate dopant-related phase(s) during calcination [30]. Recently, Li et al. [45] indicated that metal ions were in the lattice of TiO_2, and no dopant accumulation was detected on the surface for the low-level doping of metal ions by MOCVD. The method dependency of dopant position is one of the reasons why contradicting results for the same type of dopant ion, even with the same concentration levels, were reported by different researchers. For example, there is disagreement about whether Cr^{3+} dopant enhances or inhibits the photoreactivity of TiO_2 [204–207]. It is important to locate the dopant position. The location of 1% Fe in TiO_2 prepared by sol–gel dip coating was studied by Fourier transforms of extended X-ray absorption fine structure (EXAFS) analysis. EXAFS showed that Fe was incorporated completely into substitutional sites in the anatase lattice when high annealing temperatures were used [30].

Dopants can enhance the photoreactivity only within a certain range of concentration. As mentioned in Section 3.3, both too low and too high dopant concentrations are not desired, because of either insufficient amounts for achieving dopant effects or deterioration of photoactivity caused by fast recombination of charge carriers. The increased recombination may come from the recombination between trapped charges and subsequent photon-generated counterpart charges and/or recombination between trapped electrons and trapped holes through quantum tunneling [208] if the dopant traps both electron and hole. Thus, there exists an optimal dopant concentration that realizes the highest photoreactivity. The optimal dopant is size dependent. By studying the degradation efficiency of $CHCl_3$ with Fe^{3+}-doped TiO_2 nanoparticles, Zhang et al. [49] found that the optimal Fe^{3+} doping level decreased with increasing particle size, and it decreased from 0.2 at% for 6 nm to 0.05 at% for 11-nm particles. They reasoned that for the same dopant concentration, the larger particles have a higher possibility of multiple trappings of charge carriers because the average path length becomes longer and multiple trappings will lead to a high volume recombination rate. The limit of dopant concentration is not the same for different aspects of dopant effects. The optimal concentration for trapping efficiency might not be the one for the highest light absorption. Thus, an increase in light absorption does not necessarily mean the improvement of overall quantum yield, which is mainly dependent on the limiting steps of photocatalysis.

Doping-induced visible light absorption is of extreme interest. Although the band-gap narrowing can cause lower photoredox potential energy, which reduces the transfer rate of charge carriers, the increased photon absorption efficiency or generation rate of charge carriers usually can compensate for the negative effect and largely enhance the overall photoreactivity. As Asahi et al. [209] analyzed for doped TiO_2, to realize visible light reaction, the requirements for dopants in general semiconductor systems are: (1) doping must produce electronic states in the band gap of nanoparticles; (2) the redox potential of $H_2O/{}^{\bullet}OH$ should be within the band gap of doped nanoparticles to make photoreaction possible; (3) the doping-induced states should be very shallow, either below the conduction band edge or above the valence band edge. Additionally, only substitutional dopants can have a strong enough interaction with conduction-band electrons or valence-band holes to substantially modify the electronic structure of the host. The detailed mechanism of dopant-induced visible light absorption will be individually analyzed for cation doping and anion doping in the following sections.

4.3.2. Cation Doping

Metal ions have been extensively used for cation doping. The dopants cover primary metal ions, transition metal ions, and lanthanide ions (e.g., Nd^{3+}, Os^{3+}, Re^{5+}). Potential energy levels of some frequently used metal ions in the band gap of TiO_2 nanoparticles are schematically shown in Figure 6 [47]. Valence values of dopants less than and larger than that of Ti^{4+} are known as n- and p-type doping, respectively. An early comprehensive metal ion doping

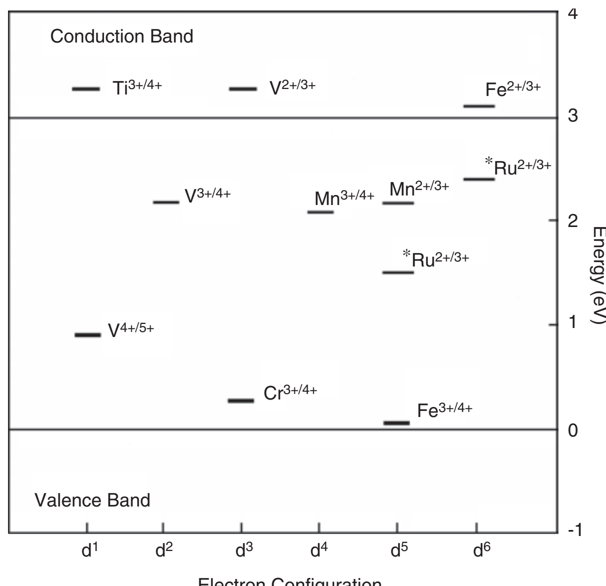


Figure 6. Energy levels of some metal ion dopants in the band gap of TiO_2 nanoparticles. Reprinted with permission from [47], W. Choi et al., *J. Phys. Chem.* 98, 13669 (1994). © 1994, American Chemical Society.

study was carried out by Choi et al. for 21 different ion-doped Q-TiO_2 [47]. Among metal ion dopants, the influence of transition metal ions on the photoreactivity of semiconductor nanoparticles is frequently studied because transition metals are known for their d-electronic configurations, which are helpful for the electronic structure modification of semiconductor nanoparticles.

Fe^{3+}, Cr^{3+}, and V^{4+} are the metal ion dopants that are used most often. There are no general agreements about whether they can enhance the photoreactivity of nanoparticles or not.

(1) Fe: Fe^{3+}-doped TiO_2 showed increased photoreduction of N_2 [210, 211] and methyl viologen [212]. Fe^{3+}-doped (0.11 to 1.76 wt%) TiO_2 prepared by coprecipitation exhibited high photoreduction of sodium nitrite because of the extended lifetime of electron-hole pairs and an increased diffusion length of electrons [213]. The same range of Fe was also indicated as the optimum iron content for TiO_2 prepared by the sol–gel method [213]. Also, negligible activity was found for very high Fe concentrations, and this was attributed to the photon absorption of aggregated iron oxide on the surface of TiO_2 due to too high dopant concentration. Doping Fe^{3+} at 0.5 at% in Q-TiO_2 was good for both oxidation of chloroform and reduction of tetrachloride [47]. Laser flash photolysis and time-resolved microwave conductivity measurements confirmed that the lifetime of electrons in doped particles increased from 200 μs to 50 ms. Low doping concentration (0.15 and 0.5 wt%) Fe doping showed considerably better performance for the degradation of maleic acid compared with pure TiO_2 [214, 215], and 0.75 at% Fe was also found to be an optimal concentration for increasing photoxidation of certain pyrimidine bases [216]. However, Fe-doped TiO_2 has only little influence [54, 217] or detrimental effect on the photodegradation of phenol [45, 218].

(2) Cr: TiO_2 doped with Cr^{3+} was found to induce visible light cleavage of water [204] and an increased reduction rate of N_2 [54]. In contrast, detrimental effects of Cr^{3+} doping were found by Herrmann et al. for the oxidation of liquid cyclohexane [206] and hydrocarbons [219]. Cr^{3+} was incorporated in substitutional sites of Ti^{4+} by ion implantation, which made remarkable decomposition of NO in visible light range possible [207, 220, 221]. But the photoreactivity in visible light was 25–1000 times lower than that under UV irradiation, although Cr^{3+}-doped TiO_2 could absorb visible light [205]. EPR study indicated a detrimental effect of the Cr^{3+} for the degradation of dimethylsulfoxide (DMSO) [222]. Similarly, Cr^{3+}-doped TiO_2 was also found to show no improvement for rhodamine B (RB) degradation because of the reduced lifetime of charge carriers [223].

(3) V: Vanadium commonly has V^{3+}, V^{4+}, and V^{5+} valence states. The photoreactivity of V^{3+}- and V^{4+}-doped Q-TiO_2 underwent very similar enhancement because V^{3+} was easily oxidized to V^{4+}, which can trap both electron and hole. V^{5+} doping also increased the photoactivity, but it was not as efficient as V^{3+} and V^{4+} doping because V^{5+} trapped only electrons [47]. Moreover, V ion-implanted TiO_2 also showed high photoactivity under visible light irradiation [220, 221]. However, Hoffmann et al. [224] performed a detailed investigation of vanadium-doped TiO_2. The coprecipitation methods followed by various temperature heat treatments showed different vanadium existence states, including surface-bound VO^{2+}, surface V_2O_5 islands, interstitial V^{4+}, and substitutional V^{4+}. All of these vanadium forms promoted charge carrier recombination and reduced the photoreactivity of TiO_2. Statistically, from the above examples, these three transition ions have either detrimental effects or positive functions with a competitive quantity of cases. It is hard to predict their effects in different reaction systems.

Many other cases of transition metal ion doping has been investigated. For example, Cu^{2+}-doped ZnS was found to have a visible light response and high photoreactivity [225]. Ni^{2+}-doped ZnS showed a visible light response and a narrowed band gap (2.3 eV) [226]. It had high efficiency for the photocatalytic H_2 evolution under visible light irradiation. Low levels of Ag^{+}-doped CdS showed enhanced efficiency for the oxidation of aniline, and its luminescence was quenched by absorbed aniline because efficient charge transfer occurred at the Ag_2S/CdS interface [227]. But, at a high concentration, the surface of CdS had Ag_2S, which prevents holes from interacting with reactant and reduced photoreactivity. Mn^{2+}-doped TiO_2 was tested to have low efficiency on DMSO degradation [222]. Mo^{5+} (>1 at%) was declared to increase RB adsorption on the surface of TiO_2 and enhance the photoreactivity [223].

There has been relatively less literature about primary group metal ion and lanthanide ion doping. Wang et al. [228] found that the photoresponses of Eu^{3+}-, La^{3+}-, Nd^{3+}-, and Pr^{3+}-doped TiO_2 nanoparticle electrodes for the degradation of RB were much stronger than that of undoped TiO_2. Sm^{3+} showed little increase because of the efficient separation of charge carries. Li et al. [45] confirmed the significance of Nd^{3+}-doped TiO_2 for the degradation of 2-chlorophenol. High-level doping (20 mol%) of alkaline earth ions (Ca^{2+}, Sr^{2+}, Ba^{2+}) doubly increased the photoreactivity of anatase, whereas 2-mol% of doping exhibited only half the enhancement for the destruction of oxalic acid [229].

Dramatic reduction of particle size by a high content of dopants and its related high adsorption rate could be the reason for this, although the doped TiO_2 had substantial amorphous parts.

Low concentrations of metal oxides, sulfides, or selenides can also be doped as ions into the nanoparticles. The optimal concentration, 1.5–3 mol% Nb_2O_5, was determined to be helpful in TiO_2 for the degradation of 1,4-dichlorobenzene (DCB) [230]. But Nb_2O_5-doped nanocrystalline WO_3 thin films had decreased photoreactivity with increases in Nb_2O_5 concentration due to an increased amorphous phase [231]. Studies of Nb_2O_5-doped TiO_2 can also be found elsewhere [232–234]. WO_3-doped TiO_2 exhibited high photoreactivity for the degradation of DCB, and the optimal content of dopant was 3 mol% [235]. The enhancement was attributed to the increased electron transfer rate from TiO_2 to trapped in W^{5+}. Al_2O_3 doping was selected to suppress the photocatalytic activity of rutile pigment for the oxidation of alkyd resins because Al_2O_3 could provide recombination sites [236]. Moreover, Al_2O_3 in the bulk of TiO_2 was considered a much more efficient recombination center than that on the surface. Doping of NiO in $SrTiO_3$ powder was found to have a photoactivity 100 times higher than that of $SrTiO_3$ alone for decomposition of both vapor and liquid water [237, 238].

Appropriate metal ion doping can make visible light absorption of nanoparticles possible. Doping Mn^{2+} ($<1\%$) is known to show orange luminescence or red shift of absorption for large band-gap semiconductors such as CdS and ZnS [239–242]. Cu^{2+}- and Ag^{+}-doped CdS exhibited a dopant concentration dependency of photoluminescence [243]. $Mg_xZn_{1-x}O$ nanoparticle thin films had band-gap reduction under certain heat treatments [244]. Furthermore, Serpone et al. [245] systematically studied the photoconductivity of TiO_2 colloids doped with Fe^{3+}, Cr^{3+}, and V^{5+} cations. All doped samples successfully showed an extended absorption threshold in the visible range, and the Cr^{3+}-doped sample had the largest red shift, with an absorption edge around 600 nm. By an implantation doping method, Anpo et al. [207, 220, 221] noticed that V^{4+}, Cr^{3+}, and several other transition metal ions remarkably shifted the absorption edge of TiO_2 into the visible region. The electronic structure of doped nanoparticles is the key to understanding the origin of visible light absorption. Recently, a theoretical calculation for the electronic structure of transition metal-doped TiO_2 (rutile) has been made by Umebayashi et al. [246]. They used a super-cell approach with a band calculation to obtain the electronic structures of rutile doped with transition metals (Fe, Cr, V, Mn, Co, Ni). The *ab initio* band calculations were done by the full-potential linearized-augmented-plane-wave (FLAPW) method. The calculated electron density maps (not shown) and the density of states (DOSs) for doped TiO_2 (Fig. 7) indicated that the electronic level introduced by 3d metal doping in the band gap or valence band was due to the t_{2g} state of the dopant. For Fe, Cr, V, and Mn doping, the t_{2g} state lies within the band gap of TiO_2. But the level of Co is located at the top of the valence band, and Ni has contributed only to the valence band formation.

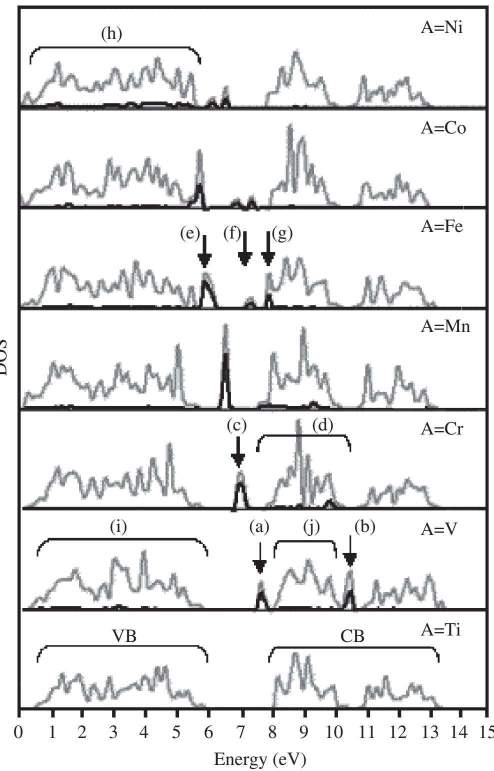


Figure 7. The DOS of the TiO_2 rutile doped with V, Cr, Mn, Fe, Co, or Ni. Gray solid lines: total DOS; black solid lines: dopant's DOS. The states are labeled (a) to (j). Reprinted with permission from [246], T. Umebayashi et al., *J. Phys. Chem. Solids* 63, 1909 (2002). © 2002, Elsevier Science.

4.3.3. Anion Doping

There are fewer reports about anion doping than there are about cation doping. This type of doping uses C, N, F, P, S, etc., to substitute anions of semiconductor nanoparticles and to achieve high photoreactivity and/or visible light response.

Doping F achieves high photoreactivity. The effect of F on the photoreactivity of TiO_2 was first reported by Hattori et al. [247, 248]. Trifluoroacetic acid was added to the starting solution to prepare F-doped nanocrystalline TiO_2 thin films by sol–gel. It was found that F doping increased the absorption coefficient for ultraviolet light because it improved film densification and crystallinity. The photocurrent measurements confirmed a high electron-hole generation rate. F doping also created trapping sites to increase the lifetime of charge carriers. The significantly enhanced overall photoreactivity for the oxidation of 1,3,5,7-tetramethylcyclotetrasiloxane (TMCTS) was mainly attributed to the increased trapping rate of charge carriers on the surface. However, in this case, they did not test the visible light absorption properties. Recently, F-doped TiO_2 prepared by hydrolysis showed that fluoride ions not only suppressed the formation of the brookite phase and prevented phase transition of anatase to rutile, but also exhibited stronger absorption in the UV-visible range with a red shift [249]. The photoreactivity of doped samples was higher than that of Degussa P25 when the molar ratio of NH_4F to H_2O was in the range of 0.5–3.

S doping has also been studied. Aiming at band-gap narrowing of TiO_2, Umebayashi et al. [196] found that doping S into TiO_2 to form Ti-S bonds caused lower-energy absorption compared with a pure sample. The band-gap study was based on the *ab initio* band calculations and the comparison of DOS. It revealed band-gap narrowing, which was caused by the mixing of S 3p states with the valence band of TiO_2 to significantly modify the electronic structures. The detailed effects of S on the electronic structure of TiO_2 were also investigated elsewhere [250–252].

Substitutional N is believed to be most effective for both band-gap narrowing and photoreactivity improvement. Asahi et al. [209, 253] calculated DOSs with substitutional doping of C, N, F, P, or S for O in the anatase TiO_2 by using FLAPW with the local density approximation (LDA), as shown in Figure 8. Their method indicated that substitutional N is more effective than interstitial for band-gap narrowing because its p states can interact with O 2p states to change the band structure. Compared with N, the states introduced by C and P are deep in the gap and have no benefit for the photoreaction. S has band-gap narrowing effects similar to those of N, but it is difficult to accommodate on the O sites because of its large ionic radius. The photodegradation of MB and acetaldehyde gas under irradiation with visible light with a wavelength less than 500 nm showed increased photoreactivity of $TiO_{2-x}N_x$.

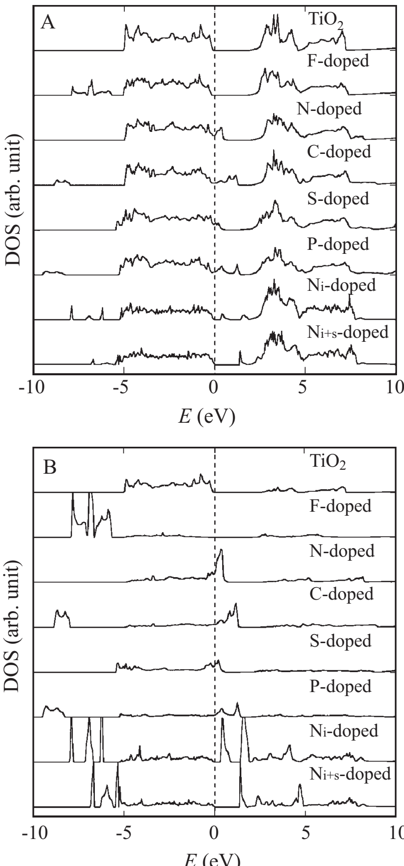


Figure 8. (A) Total DOSs of doped TiO_2 anatase. (B) FLAPW calculation of projected DOSs into the doped anion sites. N_i and N_{i+s} are interstitial N and both interstitial and substitutional N, respectively. The rest of the dopants are in the substitutional sites. Reprinted with permission from [209], R. Asahi et al., *Science* 293, 269 (2001). © 2001, American Association for the Advancement of Science.

4.3.4. Co-Doping

To improve the photocatalytic activity of semiconductor nanoparticles, co-doping methods have been attempted that use multiple dopants instead of a traditional single dopant.

Trapping of both electrons and holes can be achieved by co-doping. Although doping with Zn^{2+} or Fe^{3+} alone had little effect on photoreactivity, TiO_2 co-doped with both was found to noticeably promote the anatase-to-rutile phase transformation [254] as well as to greatly increase the photodegradation rate of phenol, with maximum photoreactivity occurring at 0.5 mol% Zn^{2+} + 1 mol% Fe^{3+} doping [255]. The enhancement was explained based on the charge carrier separation. Photogenerated electrons and holes in TiO_2 were efficiently separated by transferring holes to the valence band of ZnO and electrons to the conduction band of Fe_2O_3 because the valence band edge of ZnO is higher and the conduction band edge of Fe_2O_3 is lower than that of TiO_2. Interestingly, Eu^{3+}- and Fe^{3+}-co-doped TiO_2 nanoparticles also showed increased photoreactivity for degradation of chloroform in solution compared with undoped and monodoped samples, and the optimal concentration was 0.5 at% Eu^{3+} + 1 at% Fe^{3+} [256]. In this case, Fe^{3+} trapped holes and Eu^{3+} served as an electron trap. Simultaneous trapping of holes and electrons increased the number and lifetime of charge carriers on the surface.

Co-doping can largely improve the visible light response and photoreactivity of nanoparticles under visible light irradiation. Single-crystal $SrTiO_3$ doped with equal amounts of La and Cr, such as $SrTiO_3$:$LaCrO_3$, was determined to have an extended absorption range (<560 nm). The strong visible light absorption was attributed to the charge transfer from Cr^{3+} to Ti^{4+} [257]. However, the band-gap changes of these samples were not measured. Recently, Kato and Kudo [258] investigated the Sb^{5+}- and Cr^{3+}-co-doped TiO_2 (rutile) and $SrTiO_3$ systems. Intensive absorption was found in the visible light range. The band gap decreased to 2.2 and 2.4 eV for TiO_2 and $SrTiO_3$, respectively. Moreover, TiO_2:Sb/Cr showed broad photoemission at 828 nm, and $SrTiO_3$:Sb/Cr had a ruby-line emission at 790 nm. Under visible light irradiation, both evolution of O_2 from an aqueous silver nitrate solution with co-doped TiO_2 and evolution of H_2 from an aqueous methanol solution with co-doped $SrTiO_3$ showed increased reaction rates compared with TiO_2 doped with only Cr^{3+} and pure $SrTiO_3$. The increased photoreactivity was due to the suppression of forming recombination centers (Cr^{6+} and oxygen defects) in the lattice by the coexistence of Sb^{5+} and Cr^{3+}, because doping Cr^{3+} alone in the Ti^{4+} sites could cause the appearance of Cr^{6+} and/or oxygen to satisfy the electric charge balance.

4.4. Surface Modifications

4.4.1. Purpose

To increase the overall photocatalytic activity, various selective surface modifications have been developed. Metals, semiconductors, organic materials, and ion complexes are

deposited, mixed, or absorbed on the surface of semiconductor nanoparticles to achieve novel properties by charge transfer and electronic interaction between the surface attachment and the host semiconductor. The benefits of surface modification are (1) efficient charge carrier separation by improved charge transfer between the core semiconductor and its surface attachment, (2) solar energy absorption by surface sensitizer or narrower band-gap semiconductor, (3) change of the selectivity or reaction yield, (4) increased adsorbability of the surface. Thus, high overall efficiency of photocatalytic activity of nanoparticles is expected by surface modifications.

Coupled and capped models are two general surface modified systems. Geometrically, the surface of the host nanoparticles is only partly covered by the attachment in a coupled system, whereas the entire surface is completely surrounded in a capped system. For systems involving thin films, they usually have patterned, pillared, laminated, or intercalated structures or form composite films with different phases to achieve coupled or capped effects.

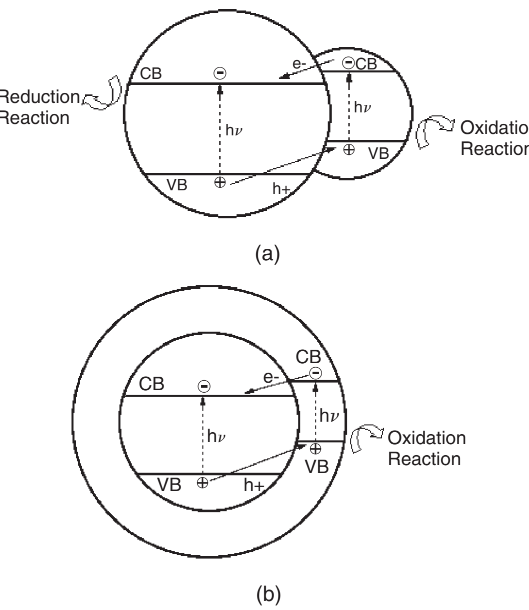


Figure 9. (a) Electron and hole transfer in a coupled semiconductor system. Both electron and hole can have the chance to participate in reactions. (b) Electron and hole transfer in a capped semiconductor system. Whether both electron and hole can have the opportunity to be in reactions is a function of the outside semiconductor thickness.

4.4.2. Composite Semiconductors

Composite semiconductor systems are formed by the deposition of one or more semiconductor layers on the surface of host semiconductor nanoparticles. Typically, these semiconductors are either coupled or capped together. The semiconductors contacting each other must have different band energy levels, so that they can make the charge carrier transfer between them possible. According to this requirement, insulators such as Al_2O_3, SiO_2, and CeO, which have larger band gaps, are also utilized with semiconductors to improve the charge transfer. The composite semiconductor conception was introduced by Nozik with a diode composed of sandwich p- and n-type semiconductors [259].

The mechanism of charge transfer is similar in dual coupled and capped semiconductors, except for the fate of charge carriers. In coupled systems, electrons and holes can be generated in both semiconductors. The electrons in a higher conduction band are injected into the lower conduction band, and holes in a lower valence band transfer to the higher valence band. Both electrons and holes can be utilized for the photoreactions because both semiconductors are exposed to the reactants. The principle of charge transfer in this system is sketched in Figure 9a. However, in capped semiconductor systems the shell semiconductor has sufficient thickness, which is comparable to the radius of the core semiconductor, so that the core and shell semiconductors can maintain their individual identities [260]. The diagram of charge transfer in capped systems is shown in Figure 9b. It can be seen that only one of the photogenerated charge carriers can access the surface reactions while the opposite charge transfers to the inner semiconductor occur. In this case, the selectivity of interfacial transfer is improved, and the oxidation or reduction reaction is enhanced. However, the problem is that this semiconductor system is only suitable for one type of reaction because one of the charge carriers is not utilized. According to the mechanism of composite semiconductors, insulators such as Al_2O_3, SiO_2, and CeO_2, which have very large band gaps, are also utilized with semiconductors in the coupled and capped systems to improve the charge transfer efficiency.

Various dual coupled semiconductor systems have been applied in photocatalysis. Coupled CdS/ZnO [261, 262], Cd_3P_2/ZnO [263], ZnSe/ZnO [89], and ZnS/ZnO [264] were designed to change the photocatalytic activity of ZnO. The CdS-coupled ZnO system showed that charge injection from CdS to ZnO occurred within 18 ps, and an increased photon absorption efficiency was observed [262]. A nearly 10-fold increase in quantum yield for the oxidation of SCN^- with ZnSe-coupled ZnO was obtained [89]. Visible emission of ZnO was quenched by hole transfer to ZnS, but fast radiationless recombination took place at the interface [264]. Likewise, photoreactivity of CdS, ZnSe, and ZnS [264] nanoparticles can be modified by ZnO on their surfaces. AgI was coupled with CdS to compensate for the disadvantage of the large band gap of AgI, and the injection rate of electrons from CdS to AgI was determined to be 2.2×10^7 s^{-1} [265]. The Ag_2S/AgI system was also studied [14]. For TiO_2, Cd/TiO_2 and CdS_3P_2/TiO_2 are commonly selected to enhance its catalytic activity because CdS and CdS_3P_2 have relatively shorter band gap values and respond to lower-energy light. The long-wavelength response of CdS/TiO_2 was earlier found by the evolution of H from alkaline solution with the irradiation of visible light [266]. The photoexcited electron-hole pairs from CdS can quickly transfer to the TiO_2 conduction band at a transfer rate of 5×10^{10} s^{-1} [265]. The electrons in the TiO_2 conduction band and the holes in the CdS valence band are responsible for redox reactions. Enhanced photoreduction of methylviologen in CdS/TiO_2 systems was observed [261]. SnO_2/TiO_2 coupled bilayer systems have been designed particularly for efficient charge separation. This type of nanostructured thin film on optically transparent electrode (OTE) showed an extended photon response and a significantly higher photooxidation rate than

only SnO_2 or TiO_2 for a textile azo dye, naphthol blue black (NBB) [267]. The addressed highest enhancement was seen when the amount of SnO_2 was higher than TiO_2 in the composite. The result was attributed to the efficient charge separation by electrons flowing from TiO_2 to OTE through SnO_2 with the existence of electrochemical bias. However, the disadvantage is that electrons cannot be utilized. Patterned TiO_2 stripe films formed on a SnO_2-film-coated soda lime glass substrate are synthesized to solve that problem. Very high photocatalytic activity was reported for gas-phase oxidation of acetaldehyde [268] and dehydrogenation of liquid methanol [269]. Similarly, both electrons and holes could be used in a bicomponent system of SnO_2 thin films partly covered with TiO_2 on glass slides, and an increased photooxidation rate of phenol solution was achieved [270]. SiO_2/TiO_2 is another system that is used in photocatalysis. Surface bond-conjugated anatase on silica nanoparticles was reported to have enhanced photoreactivity compared with Degussa P25 for the photodegradation of azo dyes, because of increased adsorption of reactant molecules of TiO_2 [271]. The use of SiO_2 substrate for TiO_2 also showed increased photoreactivity for the oxidation of propane [272–274] and the reduction of ammonia [274]. Besides the examples given above, other coupled systems can be seen elsewhere, such as ZnO/TiO_2 [275, 276], CuO/TiO_2 [277], etc.

Some capped dual semiconductors are also used in photocatalysis. TiO_2 (8–10 nm)-capped SnO_2 nanoparticles showed improved charge carrier separation, which was confirmed by laser flash photolysis [260]. The quantum yield of TiO_2/SnO_2 for photooxidation of I^- and SCN^- had a 2–3 factor increase compared with SnO_2 by itself. Moreover, TiO_2-capped SiO_2 exhibits as good a photocatalytic activity as TiO_2 because of the TiO_2 capping effects [260]. Aimed at visible light response, α-Fe_2O_3 capped by TiO_2 was synthesized and studied [278]. To improve the separation of catalysts in treated water, magnetic Fe_3O_4 was used in the photocatalyst TiO_2 to form a TiO_2-capped Fe_3O_4 system [279].

In addition, multiple-layer coupled or capped semiconductor systems are designed. A core semiconductor is used, with its surface covered by several layers of another semiconductor, on which multiple layers of core semiconductor material are then deposited as the outermost shell. The first three-layered structure was reported and described by Weller et al. [280, 281] for CdS/HgS/CdS systems in which core CdS was the Q-particle. It is also known as quantum dot quanta well (QDQW).

4.4.3. Metal/Semiconductors

The addition of metal atoms to the surface of nanoparticles changes the surface characteristics, which in turn alter the photocatalytic activity. In particular, noble metals (Pt, Ag, Au) are most often selected in this approach because they themselves are catalysts. The enhanced photoreactivity of semiconductors by loading of metal was first found for the photodecomposition of H_2O into H_2 and O_2 in Pt/TiO_2 systems by Sato et al. [282].

The principle of charge transfer in metal-deposited semiconductor systems is similar to that of a metal-semiconductor diode. The contact is either Schottky barrier or Ohmic contact, depending on the difference in Fermi energy levels of the two materials before they contact each other. If the Fermi level of the metal is lower or the work function of the metal is larger than that of the semiconductor, electrons will flow from the semiconductor to the metal by diffusion and/or drift (in case of bias) until a steady state is reached. At equilibrium, Fermi levels in two junction parts are the same and the accumulated opposite interior electric field stops the charge from moving. Thus, extra electrons accumulate on the metal side and charge pairs generated in the semiconductor are quickly separated. This type of junction is known as a Schottky barrier. In contrast to the Schottky barrier, Ohmic contact is created if the work function of the semiconductor is larger than that of the metal. It has extra electrons accumulated in the semiconductor conduction band. Figure 10 is a Schottky barrier in a metal-semiconductor photocatalysis system [29]. The metal-semiconductor system used for photocatalysis has a large area of uncovered semiconductor surface with only a small area dispersed with metal atoms. Thus, both electrons and holes can be utilized.

Pt, Ag, and Au are the most frequently used noble metals in metal-semiconductor systems because of their high effectiveness compared with other metals. Ag/TiO_2 was found to be efficient for the reduction of bis(2-dipyridyl)disulfide because the recombination of charge carriers was minimized by moving electrons to the silver [283, 284]. Sol–gel-derived Pt/TiO_2 showed high photoactivity for the oxidation of 4-nitrophenol [285]. Other metals such as Cu, Ni, etc., are also used.

4.4.4. Surface Sensitization

Photosensitization methods are commonly chosen to enhance the photoreactivity of nanoparticles. "Photosensitization" refers to the process in which electrons or holes are generated from certain attachments on the surface of nanoparticles, and electrons are then transferred to the conduction band of a semiconductor that has lower potential energy. The injected electrons then participate in the

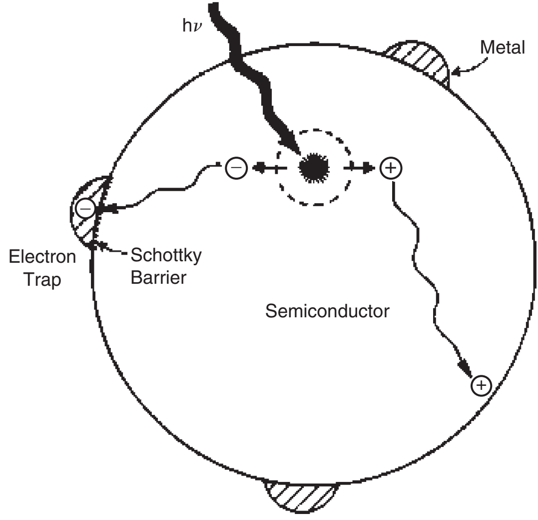


Figure 10. Shottky barrier in a metal-semiconductor photocatalysis system. Reprinted with permission from [29], A. L. Linsebigler et al., *Chem. Rev.* 95, 735 (1995). © 1995, American Chemical Society.

Table 2. Photoreactivity enhancements of typical semiconductor nanoparticles.

TiO_2	ZnO	Fe_2O_3	WO_3	SnO_2
Preferred structure	Doping	Preferred structure	Doping	Surface modification
Anatase [16, 33–35, 173, 174]	Ni^{2+} [226]	α-Fe_2O_3 [170, 171]	Nb_2O_5 [231]	TiO_2 [260]
	Cu^{2+} [225]			CrO_x [346]
Doping	Surface modification	Surface modification	Surface modification	Ag_2S [294]
Al_2O_3 [236]	Al_2O_3 [312]	Al_2O_3 [320]	TiO_2 [337]	CdS [294]
Ca^{2+} [229]	α-Fe_2O_3 [313]	TiO_2 [278, 321]	ZrO_2 [338]	Sb_2S_3 [294]
$V^{3+,4+,5+}$ [47, 220, 221, 245]	WO_3 [313]	TiO_2/SiO_2 [322, 323]	Nb_2O_5 [338]	PbS [294]
$V^{3+,4+,5+}$ [224][a]	In_2O_3 [314]	NH_4Cl [324, 325]	WS_2 [339, 340]	Bi_2S_3 [294]
Cr^{3+} [54, 204, 207, 220, 221, 245]	SnO_2 [290, 315]	$Au + Al_2O_3$ [326]	Pt [341]	$Ru(bpy)_3^{2+} + Pt$ [347]
Cr^{3+} [205, 206, 219, 222, 223][a]	Cd_3P_2 [263]	Au [327, 328]		$Ru(bpy)_3^{2+} + Pt + RuO_2$ [347]
Mn^{5+} [223]	CdS [261, 262, 294, 313]			Pd [348]
Mn^{2+} [222][a]	ZnS [264]			Pt [349]
Fe^{3+} [47, 210–215, 245, 286]	PbS [294]			
Fe^{3+} [45, 54, 217, 218][a]	ZnSe [89]			
Zn^{2+} [286]	Pd [316]			
Sr^{2+} [229]	Ag [316, 317]			
Zr^{4+} [287]				
Nb_2O_5 [230]	**ZnS**	**CdS**	**MoS_2**	**Cu_xS**
Ba^{2+} [229]	Doping	Doping	Doping	Surface modification
La^{3+} [228]	Ni^{2+} [226]	Ag^+ [227]	Alkali [342]	CoPc [350]
Pr^{3+} [228]	Cu^{2+} [225]			(Pc, phthalocyanine)
Nd^{3+} [45, 228]				
Sm^{3+} [228]	Surface modification	Surface modification	Surface modification	
Eu^{3+} [228]	ZnO [264]	Alkyl- [329]	SiO_2 [343]	
W^{6+} [288]	CdS [318]	$H_2Ti_4O_9$ [330]	TiO_2 [344]	
WO_3 [235]	Glutathione [319]	$K_2Ti_4O_9$ [331]	CdS [344, 345]	
SnO_2 [289]		$H_4Nb_6O_{17}$ [330]	MCM-41- [343]	
N [209, 253]		$Cd(OH)_2$ [332]		
F [247–249]		Cu^{2+} [333]		
S [196]				
$Zn^{2+} + Fe^{3+}$ [254, 255]		**CdSe**	**AgI**	**Others**
$Eu^{3+} + Fe^{3+}$ [256]		Surface modification	Surface modification	Al_2O_3
$Sb^{5+} + Cr^{3+}$ [258]		CdS [334]	CdS [265]	AgCl [351, 352]
Surface modification		Au [335]	Ag_2S [14]	Ag [351, 352]
Al_2O_3 [290]		Thiol ligands [336]		
SiO_2 [52, 271–274, 290]				SiO_2
CuO [277]				CuO [353]
ZnO [275, 276]				Fe [354]
SnO_2 [260, 267–270, 291]				Ni [354]
WO_3 [292, 293]				
CdS [261, 265, 294–296]				CeO_2
CdSZnS [297]				ZnO [355]
Pt [282, 285, 298]				Au [356]
Pt/Fe_2O_3 [299]				
Ag [283, 284, 300]				
Au [301, 302]				
Ru [303]				
$Ru(bpy)_2^{2+}$ [9, 304]				
Phthalocyanine [305]				
Xanthene dyes [306]				
Various other dyes [9]				
$SrTiO_3$				
Doping				
NiO [237, 238, 307, 308]				
$La^{3+} + Cr^{3+}$ [257]				
$Sb^{5+} + Cr^{3+}$ [258]				
Surface modification				
Ni [309]				
Pt [310]				
WO_3 [311]				

[a]Negative effect.

chemical reactions. In such cases, the photocatalytic process is indirect. According to the definition of surface sensitization, the surface attachments can be narrow band-gap semiconductors and organic molecules, which can be easily excited with the irradiation of low-energy light. The composite semiconductor systems are described in Section 4.4.2. Semiconductors with organic absorbates or ion species on the surface are the focus of this section. The primary purpose of sensitization is to enhance the light absorption of large-band semiconductors because the most used sensitizers have a broad light response range. Visible light absorption is of special interest.

There are several chemical sensitizers suitable for the surface modification of nanoparticles. Organic dyes are selected most often. The most utilized dyes are $Ru(bpy)_2^{2+}$-related ruthenium organic complexes because of their strong visible light response, easy interaction with semiconductor surfaces, and resistance to ligand substitution [9]. Other dyes, such as anthracene-9-carboxylate, chlorophyll a and b, chlorophyllin, cresyl violet, cyanines, erythrosine B, eosin, phenylfluorone, rhodamines, rose bengal, squaraines, thionine, and transition metal cyanides, etc., with a profound light response in the visible region are also commonly selected to improve the photoreactivity of nanoparticles [9]. Of course, the overall evaluation of the efficiency of these sensitizers is based not only on light absorption but also on the charge injection effectiveness, which is a function of energy levels of the dye and interaction between the dye and the surface of nanoparticles.

Electronic excitation in the sensitizer is the first step in initiating photocatalytic reactions. A molecular excitation diagram is shown in Figure 11 [29]. S_0 is the ground state of electrons (usually at room temperature). S_1 and S_2 are two singlet excited states. T_1 is the triplet excited state. The energy levels among them have a relation of $S_2 > S_1 > T_1 > S_0$. With the irradiation of incident photons, the electrons in molecules are excited from S_0 to S_1 and/or S_2. The excitation from S_0 to T_1 is spin forbidden according to the selection rules. The excited electrons can directly transfer from singlet states to the conduction band of semiconductors or indirectly through the triplet excited state to the conduction band. The rest of the excited electrons will undergo radiation decay to ground state by different mechanisms, such as fluorescence and phosphorescence.

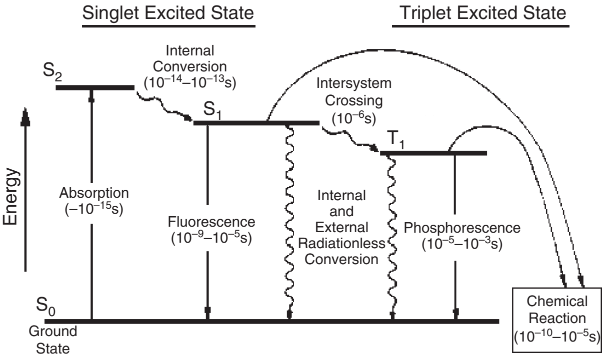


Figure 11. Electronic excitation and decay in a molecule. Reprinted with permission from [29], A. L. Linsebigler et al., *Chem. Rev.* 95, 735 (1995). © 1995, American Chemical Society.

The interaction between the sensitizer and semiconductor surface largely determines the charge injection efficiency. The surface photochemical reactions of nanoparticles affect the behavior of excited sensitizers. Lower photoemission yield occurs in reactive nanoparticles such as TiO_2, which can directly participate in photochemical reactions. Thus, in this situation, the efficiency of charge transfer to the nanoparticles is greater than that in nonactive insulator systems such as silica and alumina, which provide only a two-dimensional environment to affect and control the photochemical reactions. Kamat has published a review of the photochemistry of nonreactive and reactive semiconductor surfaces [184].

4.5. Example List

According to the photoreactivity enhancement methods discussed above, the examples are categorized for easy indexing. Table 2 is a list of the possible doping and surface modifications for typical semiconductor nanoparticles. It should be mentioned that the photocatalytic property in each case is based on a specific reaction system, and it may vary with the change of target chemicals. From this list, the importance of TiO_2 nanoparticles for photocatalytic uses is obvious.

5. PROSPECTS FOR TiO_2 NANOPARTICLES

TiO_2 is the most promising photocatalyst compared with the rest of semiconductors, and thus it is necessary to get a perspective of the likely future trend of TiO_2 nanoparticles. Currently, research is focusing on the enhancement of TiO_2 photocatalytic activity by various methods, as discussed above. However, impediments still exist: (1) the quantum yield of TiO_2 is low and it needs to be further enhanced; (2) the homogeneity of nanoparticle distribution is not very satisfying and should be improved by the development of new synthesis methods; (3) the application of TiO_2 is primarily limited to short-wavelength light irradiation; visible light absorption still poses a challenge.

6. SUMMARY

In this review, we mainly addressed the synthesis methods, photocatalysis principles, and approaches to the enhancement of the catalytic activity of semiconductor nanoparticles. The position of TiO_2 nanoparticles in the area of semiconductor photocatalysis was particularly emphasized. For each section, typical examples were given.

GLOSSARY

Band gap Energy difference between the lowest point of the conduction band and the highest point of the valence band of a semiconductor.

Doping A method to introduce foreign element(s) into the host material. Such methods include ion implantation, thermal diffusion, and conventional chemical reactions.

Heterogeneous photocatalysis A function altering the rate of chemical reactions or their generation under the irradiation of light in the presence of semiconductor photocatalysts.

Metallorganic chemical vapor deposition (MOCVD) A synthetic technique which utilizes chemical reactions among precursors (typically, a metallorganic complex is selected as primary precursor) at a certain growth temperature in a vacuum chamber to condense particles or thin films on substrates.

Photocatalyst A substance that is able to change the rate of chemical reactions by the absorption of light quanta.

Quantum yield A criterion measuring the efficiency of photocatalyst. It is the number of reacted molecules of a given product in a chemical reaction per absorbed photon at a certain given wavelength.

Sensitizer An agent absorbing light and subsequently initiating a photochemical or photophysical changes in the reaction system and itself not being consumed therewith. It is commonly utilized to enhance the properties of other photocatalysts.

Size quantization A phenomenon related to a drastic difference of properties occurring in ultrafine nanomaterials compared with bulk formations due to the spatial confinement of charge carriers when the size decreases to be smaller than the Bohr radius of the first excitation state.

Sol–gel technique A synthetic method that uses chemical reactions from precursors to form a sol (a system composed of particle-phases dispersed in a media liquid phase) and/or gel (a colloidal system constituted as a coherent structure interpenetrated by the dispersion medium) followed by drying and annealing to obtain nanostructured products.

Sputtering A synthetic technique which takes advantage of plasma generated between an anode and a cathode (target being sputtered) in a certain deposition gas in the vacuum system to deposit target materials on a substrate to usually form thin films.

Surface modification A method utilizing and attaching single or multiple second material(s) on the surface of host material to change the electronic, optical, photocatalytic, etc., properties of the host.

REFERENCES

1. S. Ahmed and D. F. Ollis, *Sol. Energy* 32, 597 (1984).
2. M. Schiavello, Ed., "Photoelectrochemistry, Photocatalysis and Photoreactors: Fundamentals and Developments." Reidel, Dordrecht, the Netherlands, 1985.
3. E. Pelizzetti and N. Serpone, Eds., "Homogeneous and Heterogeneous Photocatalysis." Reidel, Dordrecht, the Netherlands, 1986.
4. M. Schiavello, Ed., "Photocatalysis and Environment: Trends and Applications." Kluwer, Dordrecht, the Netherlands, 1988.
5. N. Serpone and E. Pelizzetti, Eds., "Photocatalysis: Fundamentals and Applications." Wiley, New York, 1989.
6. J. R. Norris and D. Meisel, Eds., "Photochemical Energy Conversion." Elsevier, New York, 1989.
7. D. F. Ollis and H. Al-Ekabi, Eds., "Photocatalytic Purification and Treatment of Water and Air." Elsevier, Amsterdam, 1993.
8. W. A. Zeltner and M. A. Anderson, in "Fine Particles Science and Technology" (E. Pelizzetti, Ed.), p. 643. Kluwer, Dordrecht, the Netherlands, 1996.
9. P. V. Kamat and K. Vinodgopal, in "Molecular and Supramolecular Photochemistry" (V. Ramamurthy and K. S. Schanze, Eds.), Vol. 2, p. 307. Dekker, New York, 1998.
10. D. Beydoun, R. Amal, G. Low, and S. McEvoy, *J. Nanoparticle Res.* 1, 439 (1999).
11. K. Kalyanasundaram, in "Energy Resources through Photochemistry and Catalysis" (M. Grätzel, Ed.), p. 217. Academic Press, San Diego, 1983.
12. M. R. Hoffmann, S. T. Martin, W. Choi, and D. W. Bahnemann, *Chem. Rev.* 95, 69 (1995).
13. A. Henglein, *Top. Curr. Chem.* 143, 113 (1988).
14. A. Henglein, *Chem. Rev.* 89, 1861 (1989).
15. Y. Wang and N. Herron, *J. Phys. Chem.* 95, 525 (1991).
16. D. W. Bahnemann, *Isr. J. Chem.* 33, 115 (1993).
17. B. Levy, *J. Electroceram.* 1, 239 (1997).
18. J. Z. Zhang, *J. Phys. Chem. B* 104, 7239 (2000).
19. J. F. Banfield and H. Zhang, in "Nanoparticles and the Environment" (J. F. Banfield and A. Navrotsky, Eds.), Mineralogical Society of America, Washington, DC, 2001.
20. A. N. Goldstein, C. M. Echer, and A. P. Alivisatos, *Science* 256, 1425 (1992).
21. S. H. Tolbert and A. P. Alivisatos, *Science* 265, 373 (1994).
22. S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla, *Phys. Rev. B* 35, 8113 (1987).
23. Y. Wang, *Acc. Chem. Res.* 24, 133 (1991).
24. A. Hagfeldt and M. Grätzel, *Chem. Rev.* 95, 49 (1995).
25. N. Serpone, D. Lawless, and E. Pelizzetti, in "Fine Partricles Science and Technology" (E. Pelizzetti, Ed.), p. 657. Kluwer, Dordrecht, the Netherlands, 1996.
26. H. Weller and A. Eychmüller, in "Advances in Photochemistry" (D. C. Neckers, D. H. Volman, and G. von Bunau, Eds.), Vol. 20. Wiley, New York, 1995.
27. A. J. Hoffman, G. Mills, H. Yee, and M. R. Hoffmann, *J. Phys. Chem.* 96, 5546 (1992).
28. M. A. Fox and M. T. Dulay, *Chem. Rev.* 93, 54 (1993).
29. A. L. Linsebigler, G. Lu, and J. T. Yates, *Chem. Rev.* 95, 735 (1995).
30. R. F. Howe, *Dev. Chem. Eng. Miner. Process* 6, 55 (1998).
31. C. Kormann, D. W. Bahnemann, and M. R. Hoffmann, *J. Photochem. Photobiol., A* 48, 161 (1989).
32. Ch.-H. Fishcher, J. Lillie, H. Weller, L. Katsikas, and A. Henglein, *Ber. Bunsenges. Phys. Chem.* 93, 61 (1989).
33. A. Sclafani, L. Palmisano, and E. Davi, *J. Photochem. Photobiol., A* 56, 113 (1991).
34. A. Vidal, J. Herrero, M. Romero, B. Sanchez, and M. Sanchez, *J. Photochem. Photobiol., A* 79, 213 (1994).
35. K. Y. Jung and S. B. Park, *J. Photochem. Photobiol., A* 127, 117 (1999).
36. A. Fahmi, C. Minot, B. Silvi, and M. Causá, *Phys. Rev. B* 47, 11717 (1993).
37. M. Horn, C. F. Schwerdtfeger, and E. P. Meagher, *Z. Kristallogr.* 136, 273 (1972).
38. A. Fujishima and K. Honda, *Nature* 37, 238 (1972).
39. A. J. Bard, *Science* 207, 139 (1980).
40. B. O'Regan and M. Grätzel, *Nature* 353, 737 (1991).
41. K. Okuyama, Y. Kousaka, N. Tohge, S. Yamamoto, J. J. Wu, R. C. Flagan, and J. H. Seinfeld, *AIChE J.* 32, 2010 (1986).
42. K. Okuyama, J.-T. Jeung, and Y. Kousaka, *Chem. Eng. Sci.* 44, 1369 (1989).
43. K. Okuyama, R. Ushio, Y. Kousaka, R. C. Flagan, and J. H. Seinfeld, *AIChE J.* 36, 409 (1990).
44. Z. Ding, X. J. Hu, P. L. Yue, G. Q. Lu, and P. F. Greenfield, *Catal. Today* 68, 173 (2001).
45. W. Li, S. I. Shah, C.-P. Huang, O. Juang, and C. Ni, *Mater. Sci. Eng., B* 96, 247 (2002).
46. W. Li, S. I. Shah, M. Sung, and C.-P. Huang, *J. Vac. Sci. Technol., B* 20, 2303 (2002).

47. W. Choi, A. Termin, and M. R. Hoffmann, *J. Phys. Chem.* 98, 13669 (1994).
48. C.-C. Wang, Z. Zhang, and J. Y. Ying, *Nanostruct. Mater.* 9, 583 (1997).
49. Z. Zhang, C.-C. Wang, R. Zakaria, and J. Y. Ying, *J. Phys. Chem.* 102, 10871 (1998).
50. N. L. Wu, S. Y. Wang, and I. A. Rusakova, *Science* 285, 1375 (1999).
51. A. Burns, W. Li, C. Baker, and S. I. Shah, *Mater. Res. Soc. Symp. Proc.* 703, 193 (2002).
52. R. van Grieken, J. Aguado, M. J. Lopez-Munoz, and J. Marugan, *J. Photochem. Photobiol., A* 148, 315 (2002).
53. M. I. Litter and J. A. Navio, *J. Photochem. Photobiol., A* 98, 183 (1994).
54. L. Palmisano, V. Augugliaro, A. Sclafani, and M. Shiavello, *J. Phys. Chem.* 92, 6710 (1988).
55. P. Yang, M. Lu, D. Xu, D. Yuan, J. Chang, G. Zhou, and M. Pan, *Appl. Phys. A* 74, 257 (2002).
56. H. Cheng, J. Ma, Z. Zhao, and L. Qi, *Chem. Mater.* 7, 663 (1995).
57. Y. Wang, Y. Hao, H. Cheng, J. Ma, B. Xu, W. Li, and S. Cai, *J. Mater. Sci.* 34, 2773 (1999).
58. Y. Wang, H. Cheng, Y. Hao, J. Ma, W. Li, and S. Cai, *J. Mater. Sci.* 34, 3721 (1999).
59. M. M. Garcia, H. Villavicencio, M. Hernandez-Velez, O. Sanchez, and J. M. Martinez-Duart, *Mater. Sci. Eng., C* 15, 101 (2001).
60. S. Vemury and S. E. Pratsinis, *J. Am. Ceram. Soc.* 78, 2984 (1995).
61. W. H. Zhu and S. E. Pratsinis, *Nanotechnology* 622, 64 (1996).
62. S. Vemury, S. E. Pratsinis, and L. Kibbey, *J. Mater. Res.* 12, 1031 (1997).
63. O. I. Arabi-Katbi, S. E. Pratsinis, P. W. Morrison, and C. M. Megaridis, *Combust. Flame* 124, 560 (2001).
64. M. Howe-Grant, Ed., "Kirk-Othmer Encyclopedia of Chemical Technology," Vol. 24, p. 225. Wiley, New York, 1997.
65. W. Z. Wang, I. Germanenko, and M. S. El-Shall, *Chem. Mater.* 14, 3028 (2002).
66. T. Trindade, P. O'Brien, and N. L. Pickett, *Chem. Mater.* 13, 3843 (2001).
67. K. Rajeshwar, N. R. de Tacconi, and C. R. Chenthamarakshan, *Chem. Mater.* 13, 2765 (2001).
68. J. W. Li, J. D. Chiang, Y. K. Su, and M. Yokoyama, *J. Cryst. Growth* 137, 421 (1994).
69. B. C. Kang, S. B. Lee, and J. H. Boo, *Surf. Coat. Technol.* 131, 88 (2000).
70. K. Shinozaki, M. Sugiura, D. Nagano, T. Kiguchi, N. Wakiya, and N. Mizutani, *J. Ceram. Soc. Jpn.* 110, 416 (2002).
71. D. A. Glocker and S. I. Shah, Eds., "Hand Book of Thin Film Process Technology." IOP Publishing, Bristol, UK, 1995.
72. C. Baker, A. Pradhan, and S. I. Shah, in "Encyclopedia of Nanoscience and Nanotechnology" (Hari Singh Nalwa, Ed.). American Scientific Publishers, Stevenson Ranch, CA, 2003.
73. D. Guerin and S. I. Shah, *J. Vac. Sci. Technol., A* 15, 712 (1997).
74. B. G. Potter and H. Simmons, *J. Appl. Phys.* 68, 1218 (1990).
75. B. G. Potter and H. Simmons, *Phys. Rev. B* 43, 2234 (1991).
76. B. G. Potter, H. Simmons, P. Kumar, and C. J. Stanton, *J. Appl. Phys.* 75, 8039 (1994).
77. S. Facsko, T. Dekorsy, C. Koerdt, C. Trappe, H. Kurz, A. Vogt, and H. L. Hartnagel, *Science* 285, 1551 (1999).
78. S. Facsko, T. Dekorsy, C. Trappe, and H. Kurz, *Microelectron. Eng.* 53, 245 (2000).
79. N. Serpone, D. Lawless, R. Khairutdinov, and E. Pelizzetti, *J. Phys. Chem.* 99, 16655 (1995).
80. L. E. Brus, *J. Chem. Phys.* 79, 5566 (1983).
81. L. E. Brus, *J. Chem. Phys.* 80, 4403 (1984).
82. Y. Nosaka and M. A. Fox, *J. Phys. Chem.* 92, 1893 (1988).
83. G. Rothenberger, J. Moser, M. Grätzel, N. Serpone, and D. K. Sharma, *J. Am. Chem. Soc.* 107, 8054 (1985).
84. D. E. Skinner, D. P. J. Colombo, J. J. Cavaleri, and R. M. Bowman, *J. Phys. Chem.* 99, 7853 (1995).
85. D. P. J. Colombo and R. M. Bowman, *J. Phys. Chem.* 99, 11752 (1995).
86. I. Bedja, S. Hotchandani, and P. V. Kamat, *J. Phys. Chem.* 97, 11064 (1993).
87. K. Vinodgopal, I. Bedja, S. Hotchandani, and P. V. Kamat, *Langmuir* 10, 1767 (1994).
88. I. Bedja, S. Hotchandani, and P. V. Kamat, *J. Phys. Chem.* 98, 4133 (1994).
89. P. V. Kamat and B. Patrick, *J. Phys. Chem.* 96, 6829 (1992).
90. J. J. Cavaleri, D. E. Skinner, D. P. J. Colombo, and R. M. Bowman, *J. Phys. Chem.* 103, 5378 (1995).
91. D. P. J. Colombo, K. A. Rousal, J. Saeh, D. E. Skinner, and R. M. Bowman, *Chem. Phys. Lett.* 232, 207 (1995).
92. D. M. Mittleman, R. W. Schoenlein, J. J. Shiang, V. L. Colvin, A. P. Alivisatos, and C. V. Shank, *Phys. Rev. B* 49, 14435 (1994).
93. S. Nakabayashi and A. Kira, *J. Phys. Chem.* 95, 9961 (1991).
94. O. I. Micic, Y. N. Zhang, K. R. Cromack, A. D. Trifunac, and M. C. Thurnauer, *J. Phys. Chem.* 97, 7277 (1993).
95. C. D. Jaeger and A. J. Bard, *J. Phys. Chem.* 83, 3146 (1979).
96. M. A. Fox and C. C. Chen, *J. Am. Chem. Soc.* 103, 6757 (1981).
97. M. A. Fox, E. Gaillard, and C. C. Chen, *J. Am. Chem. Soc.* 109, 7088 (1987).
98. W. J. Albery, G. T. Brown, J. R. Darwent, and I. E. Saievar, *J. Chem. Soc., Faraday Trans.* 181, 1999 (1985).
99. P. V. Kamat, T. W. Ebbesen, N. M. Dimitrijevic, and A. Nozik, *Chem. Phys. Lett.* 157, 384 (1989).
100. T. Rajh, O. I. Micic, D. Lawless, and N. Serpone, *J. Phys. Chem.* 96, 4633 (1992).
101. P. V. Kamat, K. R. Gopidas, and N. M. Dimitrijevic, *Mol. Cryst. Liq. Cryst.* 183, 439 (1990).
102. M. T. Nenadovic, M. I. Comor, V. Vasic, and O. I. Micic, *J. Phys. Chem.* 94, 6390 (1990).
103. R. F. Howe and M. Grätzel, *J. Phys. Chem.* 89, 4495 (1985).
104. R. F. Howe and M. Grätzel, *J. Phys. Chem.* 91, 3906 (1987).
105. M. Anpo, T. Shima, S. Kodama, and Y. Kubokawa, *J. Phys. Chem.* 91, 4305 (1987).
106. M. Grätzel and R. F. Howe, *J. Phys. Chem.* 94, 2566 (1990).
107. N. Serpone, D. Lawless, and R. Khairutdinov, *J. Phys. Chem.* 99, 16646 (1995).
108. V. Klimov, P. H. Bolivar, and H. Kurz, *Phys. Rev. B* 53, 1463 (1996).
109. J. Z. Zhang, R. H. O'Neil, and T. W. Roberti, *J. Phys. Chem.* 98, 3859 (1994).
110. J. Z. Zhang, R. H. O'Neil, T. W. Roberti, J. L. McGowen, and J. E. Evans, *Chem. Phys. Lett.* 218, 479 (1994).
111. D. P. J. Colombo and R. M. Bowman, *J. Phys. Chem.* 100, 18445 (1996).
112. M. O'Neil, J. Marohn, and G. McLendon, *Chem. Phys. Lett.* 168, 208 (1990).
113. C. Arbour, D. K. Sharma, and C. H. Langford, *J. Phys. Chem.* 94, 331 (1990).
114. M. Grätzel and A. J. Frank, *J. Phys. Chem.* 86, 2964 (1982).
115. P. V. Kamat and M. Fox, in "Lasers in Polymer Science and Technology: Applications" (J. P. Fouassier and J. F. Rabek, Eds.), Vol. II, p. 185. CRC Press, Boca Raton, FL, 1990.
116. Y. Nosaka, N. Ohta, and H. Miyama, *J. Phys. Chem.* 94, 3752 (1990).
117. T. Hirakawa and Y. Nosaka, *Langmuir* 18, 3247 (2002).
118. H. Yoneyama and T. Torimoto, *Shokubai* 40, 544 (1998).
119. R. Rossetti, S. M. Beck, and L. E. Brus, *J. Am. Chem. Soc.* 106, 980 (1984).
120. R. Rossetti and L. E. Brus, *J. Phys. Chem.* 90, 558 (1986).
121. P. V. Kamat, N. M. Dimitrijevic, and R. W. Fessenden, *J. Phys. Chem.* 92, 2324 (1988).
122. N. Getoff and S. Solar, *Radiat. Phys. Chem.* 28, 443 (1986).
123. N. Getoff and S. Solar, *Radiat. Phys. Chem.* 31, 121 (1988).

124. J. Kuruc, M. K. Sahoo, J. Locaj, and M. Hutta, *J. Radioanal. Nucl. Chem.* 183, 99 (1994).
125. U. Stafford, K. A. Gray, and P. V. Kamat, *J. Phys. Chem.* 98, 6343 (1994).
126. C. Nasr, K. Vinodgopal, S. Hotchandani, and P. V. Kamat, *Radiat. Phys. Chem.* 49, 159 (1997).
127. E. J. Hart and A. Henglein, *Radiat. Phys. Chem.* 32, 11 (1988).
128. E. J. Hart, C. H. Fish, and A. Henglein, *Radiat. Phys. Chem.* 36, 511 (1990).
129. K. S. Suslick, *Science* 247, 1439 (1990).
130. N. Serpone, R. Terzian, P. Colarusso, C. Minero, E. Pelizzetti, and H. Hidaka, *Res. Chem. Intermed.* 18, 183 (1992).
131. P. Colarusso and N. Serpone, *Res. Chem. Intermed.* 22, 61 (1996).
132. V. Brezova and A. Stasko, *J. Catal.* 147, 156 (1994).
133. M. A. Grela, M. E. Coronel, and A. J. Colussi, *J. Phys. Chem.* 100, 16940 (1996).
134. P. F. Schwarz, N. J. Turro, S. H. Bossmann, A. M. Braun, A. M. A. Abdel Wahah, and H. Dürr, *J. Phys. Chem. B* 101, 7127 (1997).
135. Y. Nosaka, M. Kishimoto, and J. Nishino, *J. Phys. Chem. B* 102, 10279 (1998).
136. L. Sun and R. J. Bolton, *J. Phys. Chem.* 100, 4127 (1996).
137. Q. Dai, D. Wang, and C. Yuan, *Supramol. Sci.* 5, 469 (1998).
138. K. Ishibashi, A. Fujishima, T. Watanabe, and K. Hashimoto, *J. Photochem. Photobiol., A* 134, 139 (2000).
139. M. L. Sauer and D. F. Ollis, *J. Catal.* 158, 570 (1996).
140. G. Martra, S. Coluccia, L. Marchese, V. Augugliaro, V. Loddo, L. Palmisano, and M. Schiavello, *Catal. Today* 53, 695 (1999).
141. V. Augugliaro, L. Palmisano, M. Schiavello, and A. Sclafani, *J. Catal.* 99, 62 (1986).
142. L. Palmisano, M. Schiavello, A. Sclafani, S. Coluccia, and L. Marchese, *New J. Chem.* 12, 847 (1988).
143. L. A. Dibble and G. B. Raupp, *Environ. Sci. Technol.* 26, 492 (1992).
144. M. R. Nimlos, W. A. Jacob, D. M. Blake, and T. A. Milne, *Environ. Sci. Technol.* 27, 732 (1993).
145. W. H. Glaze, J. F. Kenneke, and J. L. Ferry, *Environ. Sci. Technol.* 27, 177 (1993).
146. U. Stafford, K. A. Cray, P. V. Kamat, and A. Varma, *Chem. Phys. Lett.* 205, 55 (1993).
147. L. A. Dibble and G. B. Raupp, *Catal. Lett.* 4, 345 (1990).
148. L. A. Phillips and G. B. Raupp, *J. Mol. Catal.* 77, 297 (1992).
149. S. A. Larson and J. L. Falconer, in "Photocatalytic Purification and Treatment of Water and Air" (D. F. Ollis and H. Al-Ekabi, Eds.), p. 473. Elsevier, New York, 1993.
150. J. G. Calvert and J. N. Pitts, Eds., "Photochemistry." Wiley, New York, 1967.
151. L. P. Childs and D. F. Ollis, *J. Catal.* 66, 383 (1988).
152. L. Palmisano, V. Augugliaro, R. Campostrini, and M. Schiavello, *J. Catal.* 143, 149 (1993).
153. N. Serpone, R. Terzian, D. Lawless, P. Kennepohl, and G. Sauve, *J. Photochem. Photobiol., A* 73, 11 (1993).
154. L. Davydov, P. G. Smirniotis, and S. E. Pratsinis, *Ind. Eng. Chem. Res.* 38, 1376 (1999).
155. G. P. Fotou and S. E. Pratsinis, *Chem. Eng. Commun.* 151, 251 (1996).
156. K. Karakitsou and X. E. Verykios, *J. Catal.* 152, 360 (1995).
157. V. Augugliaro, V. Loddo, L. Palmisano, and M. Schiavello, *J. Catal.* 153, 32 (1995).
158. W. A. Zeltner, C. G. Hill, and M. A. Anderson, *CHEMTECH* 23, 21 (1993).
159. D. F. Ollis, in "Photochemical Conversion and Storage of Solar Energy" (E. Pelizzetti and M. Shiavello, Eds.), p. 593. Kluwer, Dordrecht, the Netherlands, 1990.
160. L. Davydov, S. E. Pratsinis, and P. G. Smirniotis, *Environ. Sci. Technol.* 34, 3435 (2000).
161. H. Gerischer and A. Heller, *J. Electrochem. Soc.* 139, 113 (1992).
162. C. M. Wang, A. Heller, and H. Gerischer, *J. Am. Chem. Soc.* 114, 5230 (1992).
163. H. Gerischer, *Electrochim. Acta* 38, 3 (1993).
164. O. M. Alfano, M. I. Cabrera, and A. E. Cassano, *J. Catal.* 172, 370 (1997).
165. N. Serpone, *J. Photochem. Photobiol., A* 104, 1 (1997).
166. J. F. Banfield, B. L. Bischoff, and M. A. Anderson, *Chem. Geol.* 110, 211 (1993).
167. A. El Goresy, M. Chen, L. Dubrovinsky, P. Gillet, and G. Graup, *Science* 293, 1467 (2001).
168. R. J. Needs, M. J. Godfrey, and M. Mansfield, *Surf. Sci.* 242, 215 (1991).
169. H. Ibach, *Surf. Sci. Rep.* 29, 193 (1997).
170. J. K. Leland and A. J. Bard, *J. Phys. Chem.* 91, 5076 (1987).
171. M. Gräetzel, J. Kiwi, C. L. Morrison, R. S. Davidson, and A. C. C. Tseung, *J. Chem. Soc., Faraday Trans.* 81, 1883 (1985).
172. S. N. Frank and A. J. Bard, *J. Phys. Chem.* 81, 1484 (1977).
173. L. Gao and Q. Zhang, *Scripta Mater.* 44, 1195 (2001).
174. H. D. Jang, S.-K. Kim, and S.-J. Kim, *J. Nanoparticle Res.* 3, 141 (2001).
175. A. Henglein, *Prog. Colloid Polym. Sci.* 73, 1 (1987).
176. J. P. Wilcoxon, P. P. Newcomer, and G. A. Samara, *J. Appl. Phys.* 81, 7934 (1997).
177. C. Kormann, D. W. Bahnemann, and M. R. Hoffmann, *J. Phys. Chem.* 92, 5196 (1988).
178. L. Kavan, T. Stoto, M. Grätzel, D. Fitzmaurice, and V. Shklover, *J. Phys. Chem.* 97, 9493 (1993).
179. K. Okano, T. Hayashi, and A. Miyamoto, *Jpn. J. Appl. Phys.* 38, 6107 (1999).
180. A. Fojtik, H. Weller, U. Koch, and A. Henglein, *Ber. Bunsenges. Phys. Chem.* 88, 969 (1984).
181. H. Weller, M. H. Schmidt, U. Koch, A. Fojtik, S. Baral, A. Henglein, W. Kunath, K. Weiss, and E. Dieman, *Chem. Phys. Lett.* 124, 557 (1986).
182. A. J. Nozik, F. Williams, M. T. Nenadović, T. Rajh, and O. I. Mićić, *J. Phys. Chem.* 89, 397 (1985).
183. A. V. Nabok, T. Richardson, C. McCartney, N. Cowlam, F. Davis, C. J. M. Stirling, A. K. Ray, V. Gacem, and A. Gibaud, *Thin Solid Films* 327, 510 (1998).
184. P. V. Kamat, *Chem. Rev.* 93, 267 (1993).
185. E. A. Ponomarev, A. Albu-Yaron, R. Tenne, and C. Levy-Clement, *J. Electrochem. Soc.* 144, L277 (1997).
186. J. M. Nedeljković, M. T. Nenadović, O. I. Mićić, and A. J. Nozik, *J. Phys. Chem.* 90, 12 (1986).
187. D. J. Suh, O. O. Park, H.-T. Jung, and M. H. Kwon, *Korean J. Chem. Eng.* 19, 529 (2002).
188. K. M. Reddy, C. V. G. Reddy, and S. V. Manorama, *J. Solid State Chem.* 158, 180 (2001).
189. S. Monticone, R. Tufeu, A. V. Kanaev, E. Scolan, and C. Sanchez, *Appl. Surf. Sci.* 162, 565 (2000).
190. A. Henglein, *Ber. Bunsenges. Phys. Chem.* 101, 1562 (1997).
191. N. Xu, Z. Shi, Y. Fan, J. Dong, J. Shi, and M. Z.-C. Hu, *Ind. Eng. Chem. Res.* 38, 373 (1999).
192. A. J. Hoffman, H. Yee, G. Mills, and M. R. Hoffmann, *J. Phys. Chem.* 96, 5540 (1992).
193. A. P. Rivera, K. Tanaka, and T. Hisanaga, *Appl. Catal., B* 3, 37 (1993).
194. H. Karl, W. Hipp, I. Grosshans, and B. Stritzker, *Mater. Sci. Eng., C* 19, 55 (2002).
195. B. El-Kareh, "Fundamentals of Semiconductor Processing Technology." Kluwer, Boston, 1995.
196. T. Umebayashi, T. Yamaki, H. Itoh, and K. Asai, *Appl. Phys. Lett.* 81, 454 (2002).
197. M. Zhou and H. Luo, *J. Alloy Compd.* 255, 239 (1997).
198. B. Gržeta, E. Tkalčec, C. Goebbert, M. Takeda, M. Takahashi, K. Nomura, and M. Jakšić, *J. Phys. Chem. Solids* 63, 765 (2002).

199. R. Arroyo, G. Cordoba, J. Padilla, and V. H. Lara, *Mater. Lett.* 54, 397 (2002).
200. J. Arbiol, J. Cerda, G. Dezanneau, A. Cirera, F. Peiro, A. Cornet, and J. R. Morante, *J. Appl. Phys.* 92, 853 (2002).
201. D. Morris, Y. Dou, J. Rebane, C. E. J. Mitchell, R. G. Egdell, D. S. L. Law, A. Vittadini, and M. Casarin, *Phys. Rev. B* 61, 13445 (2000).
202. A. L. Efros and M. Rosen, *Annu. Rev. Mater. Sci.* 30, 475 (2000).
203. V. A. Gubanov, C. Boekema, and C. Y. Fong, *Appl. Phys. Lett.* 78, 216 (2001).
204. E. Borgarello, J. Kiwi, M. Gräetzel, E. Pelizzetti, and M. Visca, *J. Am. Chem. Soc.* 104, 2996 (1982).
205. J.-M. Herrmann, J. Disdier, and P. Pichat, *Chem. Phys. Lett.* 108, 618 (1984).
206. W. Mu, J. M. Herrmann, and P. Pichat, *Catal. Lett.* 3, 73 (1989).
207. M. Takeuchi, H. Yamashita, M. Matsuoka, M. Anpo, T. Hirao, N. Itoh, and N. Iwamoto, *Catal. Lett.* 67, 135 (2000).
208. M. Grätzel, "Photocatalysis: Fundamentals and Applications." Wiley, New York, 1989.
209. R. Asahi, T. Morikawa, T. Ohwaki, K. Aoki, and Y. Taga, *Science* 293, 269 (2001).
210. G. N. Schrauzer and T. D. Guth, *J. Am. Chem. Soc.* 99, 7189 (1977).
211. J. Soria, J. C. Conesa, V. Augugliaro, L. Palmisano, M. Schiavello, and A. Sclafani, *J. Phys. Chem.* 95, 274 (1991).
212. J. A. Navio, F. J. Marchena, M. Roncel, and M. A. Del la Rosa, *J. Photochem. Photobiol., A* 55, 319 (1991).
213. K. T. Ranjit and B. Viswanathan, *J. Photochem. Photobiol., A* 108, 79 (1997).
214. J. Arana, O. Gonzalez Diaz, M. Miranda Saracho, J. M. Dona Rodriguez, J. A. Herrera Melian, and J. Perez Pena, *Appl. Catal., B* 32, 49 (2001).
215. J. Arana, O. Gonzalez Diaz, M. Miranda Saracho, J. M. Dona Rodriguez, J. A. Herrera Melian, and J. Perez Pena, *Appl. Catal., B* 36, 113 (2002).
216. M. R. Dhananjeyan, V. Kandavelu, and R. Renganathan, *J. Mol. Catal., A* 151, 217 (2000).
217. A. Di Paola, E. García-López, S. Ikeda, G. Marcì, B. Ohtani, and L. Palmisano, *Catal. Today* 75, 87 (2002).
218. V. Brezová, A. Blažková, L. Karpinsky, J. Grošková, B. Havlinová, V. Jorik, and M. Čeppan, *J. Photochem. Photobiol., A* 109, 177 (1997).
219. J. M. Herrmann, W. Mu, and P. Pichat, *Stud. Surf. Sci. Catal.* 59, 405 (1991).
220. M. Anpo, Y. Ichihashi, M. Takeuchi, and H. Yamashita, in "Science and Technology in Catalysis 1998" (B. Delmon and J. T. Yates, Eds.), p. 305. Kodansha, Tokyo, 1999.
221. H. Yamashita, Y. Ichihashi, M. Takeuchi, S. Kishiguchi, and M. Anpo, *J. Synchrotron Radiat.* 6, 451 (1999).
222. D. Dvoranová, V. Brezová, M. Mazúr, and M. A. Malati, *Appl. Catal., B* 37, 91 (2002).
223. K. Wilke and H. D. Breuer, *J. Photochem. Photobiol., A* 121, 49 (1999).
224. S. T. Martin, C. L. Morrison, and M. R. Hoffmann, *J. Phys. Chem.* 98, 13695 (1994).
225. A. Kudo and M. Sekizawa, *Catal. Lett.* 58, 241 (1999).
226. A. Kudo and M. Sekizawa, *Chem. Commun.* 1371 (2000).
227. A. Kumar and S. Kumar, *Chem. Lett.* 711 (1996).
228. Y. Wang, H. Cheng, L. Zhang, Y. Hao, J. Ma, B. Xu, and W. Li, *J. Mol. Catal., A* 151, 205 (2000).
229. N. I. Al-Salim, S. A. Bagshaw, A. Bittar, T. Kemmitt, A. J. McQuillan, A. M. Mills, and M. J. Ryan, *J. Mater. Chem.* 10, 2358 (2000).
230. H. Cui, K. Dwight, S. Soled, and A. Wold, *J. Solid State Chem.* 115, 187 (1995).
231. H. Wang, P. Xu, and T. Wang, *Thin Solid Films* 388, 68 (2001).
232. Y. Matsumoto, T. Shimizu, A. Toyoda, and E. Sato, *J. Phys. Chem.* 86, 3581 (1982).
233. S. Okazaki and T. Okuyama, *Bull. Chem. Soc. Jpn.* 65, 914 (1983).
234. Y. Gao, *Thin Solid Films* 346, 73 (1999).
235. Y. R. Do, W. Lee, K. Dwight, and A. Wold, *J. Solid State Chem.* 108, 198 (1994).
236. U. Gesenhues, *J. Photochem. Photobiol., A* 139, 243 (2001).
237. K. Domen, S. Naito, M. Soma, T. Onishi, and K. Tamaru, *J. Chem. Soc., Chem. Commun.* 543 (1980).
238. K. Domen, S. Naito, T. Onishi, K. Tamaru, and M. Soma, *J. Phys. Chem.* 86, 3657 (1982).
239. G. Counio, S. Esnouf, T. Gacoin, and J. P. Boilot, *J. Phys. Chem.* 100, 20021 (1996).
240. G. Counio, T. Gacoin, and J. P. Boilot, *J. Phys. Chem.* 102, 5257 (1998).
241. J. F. Suyver, S. F. Wuister, J. J. Kelly, and A. Meijerink, *NanoLetters* 1, 429 (2001).
242. A. A. Bol and A. Meijerink, *J. Phys. Chem. B* 105, 10197 (2001).
243. J. M. Huang and C. J. Murphy, *Mater. Res. Soc. Symp. Proc.* 560, 33 (1999).
244. D. Zhao, Y. Liu, D. Shen, Y. Lu, J. Zhang, and X. W. Fan, *J. Sol-Gel Sci. Technol.* 23, 231 (2002).
245. N. Serpone, D. Lawless, J. Disdier, and J.-M. Herrmann, *Langmuir* 10, 643 (1994).
246. T. Umebayashi, T. Yamaki, H. Itoh, and K. Asai, *J. Phys. Chem. Solids* 63, 1909 (2002).
247. A. Hattori, M. Yamamoto, H. Tada, and S. Ito, *Chem. Lett.* 707 (1998).
248. A. Hattori and H. Tada, *J. Sol-Gel Sci. Technol.* 22, 47 (2001).
249. J. C. Yu, J. U. Yu, W. K. Ho, Z. T. Jiang, and L. H. Zhang, *Chem. Mater.* 14, 3808 (2002).
250. D. Gonbeau, C. Guimon, G. P. Guillouzo, A. Levasseur, G. Meunier, and R. Dormoy, *Surf. Sci.* 254, 81 (1991).
251. E. L. D. Hebenstreit, W. Hebenstreit, H. Geisler, S. N. Thornburg, C. A. Ventrice, Jr., D. A. Hite, P. T. Sprunger, and U. Diebold, *Phys. Rev. B* 64, 115418 (2001).
252. E. L. D. Hebenstreit, W. Hebenstreit, and U. Diebold, *Surf. Sci.* 470, 347 (2001).
253. T. Morikawa, R. Asahi, T. Ohwaki, K. Aoki, and Y. Taga, *Jpn. J. Appl. Phys.* 40, L561 (2001).
254. Z.-H. Yuan and L.-D. Zhang, *Nanostruct. Mater.* 10, 1127 (1998).
255. Z.-H. Yuan, J.-H. Jia, and L.-D. Zhang, *Mater. Chem. Phys.* 73, 323 (2002).
256. P. Yang, C. Lu, N. P. Hua, and Y. K. Du, *Mater. Lett.* 57, 794 (2002).
257. A. Mackor and G. Blasse, *Chem. Phys. Lett.* 77, 6 (1981).
258. H. Kato and A. Kudo, *J. Phys. Chem. B* 106, 5029 (2002).
259. A. J. Nozik, *Appl. Phys. Lett.* 30, 567 (1977).
260. I. Bedja and P. V. Kamat, *J. Phys. Chem.* 99, 9182 (1995).
261. L. Spanhel, H. Weller, and A. Henglein, *J. Am. Chem. Soc.* 109, 6632 (1987).
262. S. Hotchandani and P. V. Kamat, *J. Phys. Chem.* 96, 6834 (1992).
263. L. Spanhel, A. Henglein, and H. Weller, *Ber. Bunsenges. Phys. Chem.* 91, 1359 (1987).
264. J. Rabani, *J. Phys. Chem.* 93, 7707 (1989).
265. K. R. Gopidas, M. Bohorquez, and P. V. Kamat, *J. Phys. Chem.* 94, 6435 (1990).
266. N. Serpone, E. Borgarello, and M. Gratzel, *J. Chem. Soc., Chem. Commun.* 342 (1984).
267. K. Vinodgopal, I. Bedja, and P. V. Kamat, *Chem. Mater.* 8, 2180 (1996).
268. H. Tada, A. Hattori, Y. Tokihisa, K. Imai, N. Tohge, and S. Ito, *J. Phys. Chem. B* 104, 4585 (2000).
269. T. Kawahara, Y. Konishi, H. Tada, N. Tohge, and S. Ito, *Langmuir* 17, 7442 (2001).
270. Y. Cao, X. Zhang, W. Yang, H. Du, Y. Bai, T. Li, and J. Yao, *Chem. Mater.* 12, 3445 (2000).
271. C. Hu, Y. Wang, and H. Tang, *Appl. Catal., B* 30, 277 (2001).

272. T. Tanaka, S. Takenaka, T. Funabiki, and S. Yoshida, *Chem. Lett.* 1585 (1994).
273. S. Takenaka, T. Kuriyama, T. Tanaka, T. Funabiki, and S. Yoshida, *J. Catal.* 155, 196 (1995).
274. T. Tanaka, K. Teramura, T. Yamamoto, S. Takenaka, S. Yoshida, and T. Funabiki, *J. Photochem. Photobiol., A* 148, 277 (2002).
275. G. Marcì, V. Augugliaro, M. J. López-Muñoz, C. Martin, L. Palmisano, V. Rives, M. Schiavello, R. J. D. Tilley, and A. M. Venezia, *J. Phys. Chem.* 105, 1026 (2001).
276. G. Marcì, V. Augugliaro, M. J. López-Muñoz, C. Martin, L. Palmisano, V. Rives, M. Schiavello, R. J. D. Tilley, and A. M. Venezia, *J. Phys. Chem.* 105, 1033 (2001).
277. K. Chiang, R. Amal, and T. Tran, *Adv. Environ. Res.* 6, 471 (2002).
278. M. Penpolcharoen, A. Amal, and V. Chen, "The 26th Australian Chemical Engineering Conference," Queensland, Australia, paper no. 222, 1998.
279. D. Beydoun, R. Amal, G. Low, and S. McEvoy, "World Congress on Particle Technology," Brighton, UK, paper no. 385, 1998.
280. A. Eychmüller, A. Mews, and H. Weller, *Chem. Phys. Lett.* 208, 59 (1993).
281. A. Mews, A. Eychmüller, M. Giersig, D. Schooss, and H. Weller, *J. Phys. Chem.* 98, 934 (1994).
282. S. Sato and J. M. White, *Chem. Phys. Lett.* 72, 83 (1980).
283. H. Tada, K. Teranishi, and S. Ito, *Langmuir* 15, 7084 (1999).
284. H. Tada, K. Teranishi, Y. Inubushi, and S. Ito, *Langmuir* 16, 3304 (2000).
285. G. Facchin, G. Carturan, R. Campostrini, S. Gialanella, L. Lutterotti, L. Armelao, G. Marcì, L. Palmisano, and A. Sclafani, *J. Sol-Gel Sci. Technol.* 18, 29 (2000).
286. Y. Ma, X. T. Zhang, Z. S. Guan, Y. A. Cao, and J. N. Yao, *J. Mater. Res.* 16, 2928 (2001).
287. J. C. Yu, J. Lin, and R. W. M. Kwok, *J. Phys. Chem. B* 102, 5094 (1998).
288. A. Fuerte, M. D. Hernandez-Alonso, A. J. Maira, A. Martinez-Arias, M. Fernandez-Garcia, J. C. Conesa, J. Soria, and G. Munuera, *J. Catal.* 212, 1 (2002).
289. Q. J. Liu, X. H. Wu, B. L. Wang, and Q. A. Liu, *Mater. Res. Bull.* 37, 2255 (2002).
290. C. Anderson and A. J. Bard, *J. Phys. Chem. B* 101, 2611 (1997).
291. L. Y. Shi, C. Z. Li, H. C. Gu, and D. Y. Fang, *Mater. Chem. Phys.* 62, 62 (2000).
292. Y. T. Kwon, K. Y. Song, W. I. Lee, G. J. Choi, and Y. R. Do, *J. Catal.* 191, 192 (2000).
293. K. Y. Song, M. K. Park, Y. T. Kwon, H. W. Lee, W. J. Chung, and W. I. Lee, *Chem. Mater.* 13, 2349 (2001).
294. R. Vogel, P. Hoyer, and H. Weller, *J. Phys. Chem.* 98, 3183 (1994).
295. T. Hirai, K. Suzuki, and I. Komasawa, *J. Colloid Interface Sci.* 244, 262 (2001).
296. M. G. Kang, H. E. Han, and K. J. Kim, *J. Photochem. Photobiol., A* 125, 119 (1999).
297. S. V. Tambwekar, D. Venugopal, and M. Subrahmanyam, *Int. J. Hydrogen Energy* 24, 957 (1999).
298. Y. X. Li, G. X. Lu, and S. B. Li, *Appl. Catal., A* 214, 179 (2001).
299. J. J. Yang, D. X. Li, Z. J. Zhang, Q. L. Li, and H. Q. Wang, *J. Photochem. Photobiol., A* 13, 197 (2000).
300. E. Szabo-Bardos, H. Czili, and A. Horvath, *J. Photochem. Photobiol., A* 154, 195 (2003).
301. C. Y. Wang, C. Y. Liu, X. Zheng, J. Chen, and T. Shen, *Colloid Surf. A* 131, 271 (1998).
302. V. Subramanian, E. E. Wolf, and P. V. Kamat, *Langmuir* 19, 469 (2003).
303. C. Elmasides and X. E. Verykios, *J. Catal.* 203, 477 (2001).
304. J. Kallioinen, G. Benko, V. Sundstrom, J. E. I. Korppi-Tommola, and A. P. Yartsev, *J. Phys. Chem. B* 106, 4396 (2002).
305. V. Iliev, *J. Photochem. Photobiol., A* 151, 195 (2002).
306. X. Z. Li, W. Zhao, and J. C. Zhao, *Sci. China, Ser. B* 45, 421 (2002).
307. K. Domen, S. Naito, T. Onishi, and K. Tamaru, *Chem. Phys. Lett.* 92, 433 (1982).
308. K. Domen, A. Kudo, and T. Onishi, *J. Catal.* 102, 92 (1986).
309. K. Takehira, T. Shishido, and M. Kondo, *J. Catal.* 207, 307 (2002).
310. K. Sayama, K. Mukasa, R. Abe, Y. Abe, and H. Arakawa, *J. Photochem. Photobiol., A* 148, 71 (2002).
311. Y. Ohko, S. Saitoh, T. Tatsuma, and A. Fujishima, *Electrochemistry* 70, 460 (2002).
312. S. Sakthivel, B. Neppolian, M. Palanichamy, B. Arabindoo, and V. Murugesan, *Water Sci. Technol.* 44, 211 (2001).
313. S. Sakthivel, S. U. Geissen, D. W. Bahnemann, V. Murugesan, and A. Vogelpohl, *J. Photochem. Photobiol., A* 148, 283 (2002).
314. A. Kudo and I. Mikami, *Chem. Lett.* 1027 (1998).
315. C. Wang, J. C. Zhao, X. M. Wang, B. X. Mai, G. Y. Sheng, P. Peng, and J. M. Fu, *Appl. Catal., B* 39, 269 (2002).
316. L. Q. Jing, W. M. Cai, X. J. Sun, H. Hou, Z. L. Xu, and Y. G. Du, *Chin. J. Catal.* 23, 336 (2002).
317. C. A. K. Gouvea, F. Wypych, S. G. Moraes, N. Duran, and P. Peralta-Zamora, *Chemosphere* 40, 427 (2000).
318. H. Kisch and P. Lutz, *Photochem. Photobiol. Sci.* 1, 240 (2002).
319. C. L. Torres-Martinez, L. Nguyen, R. Kho, W. Bae, K. Bozhilov, V. Klimov, and R. K. Mehra, *Nanotechnology* 10, 340 (1999).
320. M. Saito, H. Kimura, N. Mimura, J. Wu, and K. Murata, *Appl. Catal., A* 239, 71 (2003).
321. S. Qiao, D. D. Sun, J. H. Tay, and C. Easton, *Water Sci. Technol.* 47, 211 (2003).
322. F. Chen and J. C. Zhao, *Catal. Lett.* 58, 245 (1999).
323. F. Chen, Y. D. Xie, J. C. Zhao, and G. X. Lu, *Chemosphere* 44, 1159 (2001).
324. E. R. Blomiley and E. G. Seebauer, *Langmuir* 15, 5970 (1999).
325. E. R. Blomiley and E. G. Seebauer, *J. Phys. Chem. B* 103, 5035 (1999).
326. H. G. Ahn and D. J. Lee, *Res. Chem. Intermed.* 28, 451 (2002).
327. N. M. Gupta and A. K. Tripathi, *Gold Bull.* 34, 120 (2001).
328. S. Minico, S. Scire, C. Crisafulli, and S. Galvagno, *Appl. Catal., B* 34, 277 (2001).
329. M. Y. Han, W. Huang, C. H. Quek, L. M. Gan, C. H. Chew, G. Q. Xu, and S. C. Ng, *J. Mater. Res.* 14, 2092 (1999).
330. S. Tawkaew, S. Uchida, Y. Fujishiro, and T. Sato, *Mol. Cryst. Liq. Cryst.* 341, 1041 (2000).
331. W. F. Shangguan and A. Yoshida, *J. Phys. Chem. B* 106, 12227 (2002).
332. A. Kumar and D. P. S. Negi, *J. Photochem. Photobiol., A* 134, 199 (2000).
333. A. V. Isarov and J. Chrysochoos, *Langmuir* 13, 3142 (1997).
334. Y. C. Tian, T. Newton, N. A. Kotov, D. M. Guldi, and J. H. Fendler, *J. Phys. Chem.* 100, 8927 (1996).
335. R. Nayak, J. Galsworthy, P. Dobson, and J. Hutchison, *J. Mater. Res.* 13, 905 (1998).
336. J. Aldana, Y. A. Wang, and X. G. Peng, *J. Am. Chem. Soc.* 123, 8844 (2001).
337. T. Ohno, F. Tanigawa, K. Fujihara, S. Izumi, and M. Matsumura, *J. Photochem. Photobiol., A* 118, 41 (1998).
338. C. Martin, I. Martin, V. Rives, G. Solana, V. Loddo, L. Palmisano, and A. Sclafani, *J. Mater. Sci.* 32, 6039 (1997).
339. A. Di Paola, L. Palmisano, M. Derrigo, and V. Augugliaro, *J. Phys. Chem. B* 101, 876 (1997).
340. A. Di Paola, L. Palmisano, and V. Augugliaro, *Catal. Today* 58, 141 (2000).
341. A. Sclafani, L. Palmisano, G. Marci, and A. M. Venezia, *Sol. Energy Mater. Sol. Cells* 51, 203 (1998).

342. K. T. Park and J. Kong, *Top. Catal.* 18, 175 (2002).
343. E. R. Rivera-Munos, D. Lardizabal, G. Alonso, A. Aguilar, M. H. Siadati, and R. R. Chianelli, *Catal. Lett.* 85, 147 (2003).
344. J. H. Choy, S. M. Paek, J. M. Oh, and E. S. Jang, *Curr. Appl. Phys.* 2, 489 (2002).
345. J. K. Lee, W. Lee, T. J. Yoon, G. S. Park, and J. H. Choy, *J. Mater. Chem.* 12, 614 (2002).
346. Y. X. Zhu, I. K. Murwani, C. J. Zhou, E. Kemnitz, and Y. C. Xie, *Catal. Lett.* 85, 205 (2003).
347. K. Gurunathan, P. Maruthamuthu, and M. V. C. Sastri, *Int. J. Hydrogen Energy* 22, 57 (1997).
348. R. Gavagnin, L. Biasetto, F. Pinna, and G. Strukul, *Appl. Catal., B* 38, 291 (2002).
349. K. Liberkova, R. Touroude, and D. Y. Murzin, *Chem. Eng. Sci.* 57, 2519 (2002).
350. E. N. Savinov, G. X. Lu, and V. N. Parmon, *React. Kinet. Catal. Lett.* 48, 553 (1992).
351. Y. Yamashita, N. Aoyama, N. Takezawa, and K. Yoshida, *J. Mol. Catal., A* 150, 233 (1999).
352. Y. Yamashita, N. Aoyama, N. Takezawa, and K. Yoshida, *Environ. Sci. Technol.* 34, 5211 (2000).
353. Z. L. Wang, Q. S. Liu, H. F. Yu, T. H. Wu, and G. J. Wang, *Appl. Catal., A* 239, 87 (2003).
354. M. A. Ermakova and D. Y. Ermakov, *Catal. Today* 77, 225 (2002).
355. R. Li, S. Yin, S. Yabe, M. Yamashita, S. Momose, S. Yoshida, and T. Sato, *Br. Ceram. Trans.* 101, 9 (2002).
356. P. Bera and M. S. Hegde, *Catal. Lett.* 79, 75 (2002).

www.aspbs.com/enn

Semiconductor Nanotransistors

Y. Fu

Chalmers University of Technology, Göteborg, Sweden

CONTENTS

1. Introduction
2. Theoretical Considerations of Devices
3. Devices Based on III–V Semiconductors
4. Devices Based on IV Semiconductors
5. Nitride and Other Systems
6. Other Single-Electron Transistor Applications
7. Fundamental Physics
8. Summary
Glossary
References

1. INTRODUCTION

The majority of current integration circuits, including memories, microprocessors, and signal processors, are manufactured in CMOS technologies. The acronym CMOS stands for the complementary metal-oxide-semiconductor which is constructed by a *p*-channel and an *n*-channel metal-oxide-semiconductor field-effect transistor (MOSFET) on the same substrate. A typical *n*-MOSFET is shown in Figure 1, where SiO_2 serves as the insulator and the aluminum metal plate serves as the gate electrode. A *p*-channel device is obtained by interchanging the *n* and *p* regions. The functioning mechanism of the *n*-MOSFET is to control the electron concentration in the conduction channel by the electric field of the gate bias (thus the so-called field effect) which is insulated from the channel by an insulator. When the gate bias is larger than the threshold bias $V_{\rm th}$, the conduction band edge of the conduction channel becomes low enough (close to the Fermi level E_f; see Fig. 2) that the original *p* material is now populated with electrons (material type inverted). Electrons from the source can then transport to the drain fluently when a drain bias is applied.

Information technology has been rapidly advancing due to the electronic representation, information processing, and communication in a low-cost, high-speed, very compact, and highly reliable fashion [1–3]. The chip component density has been following Moore's law and roughly doubles every other year over the last three decades. Figure 3 shows the number of components per integrated circuit (IC) [4]. As the number of transistors integrated in a circuit continues to increase, discrete device dimensions have begun to reach the nanometer regime. Such a down-scaling progress of individual device components has been tremendous over the past 20 years [5]. A 1.0-μm gate length MOSFET was reported in 1974 by Dennard et al. [6], a 0.1-μm gate length by Sai-Halasz et al. [7] in 1987, a 70-nm gate length by Hashimoto et al. [8] in 1992, a 40-nm gate length by Ono et al. [9] in 1995, and a 30-nm gate length in 1998 by Kawaura et al. [10]. Vertical *p*-MOS transistors also have been extensively investigated, and transistors with channel lengths of 130 and 100 nm based on Si as well as GaAs have been fabricated [11–16]. However, the advancing multimedia infrastructure and service in the future demand further reduction in the chip size [2, 17, 18]. Moreover, the demanding field of optical devices is vast as well [19, 20].

The size-induced metal–insulator transition in mesoscopic conductors was discussed for designing new, even smaller, and simpler transistors [21]. Tunneling-control functional, quantum interband coupled multifunction, resonant electron-transfer, quantized-energy logic/memory, coupled quantum-dot, quantum-wave structure functional, single-electron transistor (SET) logic, and CMOS coupled single-electron devices were studied to develop novel devices for information processing which control quantum phenomena to realize low power consumption, high-speed, and multi-functional operations [22]. Zero-dimensional structures formed by lateral confinement of two-dimensional electron gas were analyzed for the SET [23]. It was suggested that the most promising avenue for downscaling to go beyond CMOS is the quantum dot IC technology [24–36]. The advantages and drawbacks of the low-dimensional quantum devices were discussed [37]. Silicon single-electron tunneling devices were discussed in terms of their graceful assimilation into the current silicon ultra-large-scale integration world [38]. Evaluation of the heterostructures was reported [39, 40]. Three nanoscale devices, that is, CMOS, resonant tunneling diodes, and quantum cellular automata, against

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

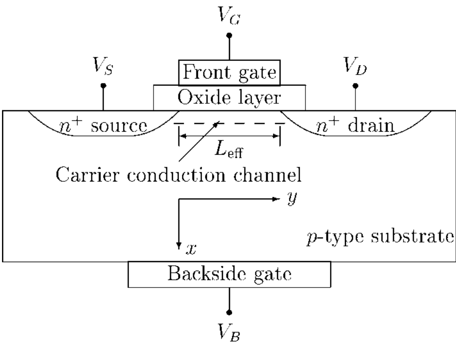


Figure 1. An n-channel metal-oxide-semiconductor field-effect transistor.

a standard computing task were evaluated showing the difficulties in the implantation of the state of the art quantum devices [41]. The application of SETs to high-speed devices will be possible but very difficult [42]. The issue was further discussed by Thornton [43] concerning quantum interference transistors, ballistic transmission devices, resonant tunneling quantum dot devices, Coulomb blockade, and SETs. Bohr [44] and Korotkov [45] described basic device and circuit requirements for electronic logic and memory products for evaluating nanotechnology options such as carbon nanotube field-effect transistors, nanowire FETs, SETs, and molecular devices as possible future replacements for Si MOSFETs. Experiments and simulations show the noise and the uncertainty in determining the geometry of the fabricated circuit [46]. There are technical issues like the very challenging degree of material perfection [33, 47] and necessary design trade-offs in oxide, channel, and storage-dot dimensions [48].

There were also positive observations [49]. Nippon Telegraph and Telephone Corporation has fabricated a prototype SET which allows electric current to be controlled electron-by-electron and has confirmed that the device operates even at temperatures near room temperature [50]. Single-electron charging effects in Si MOS devices were examined and a promise of the application of single-electron technology to more integrated circuitry was concluded [51].

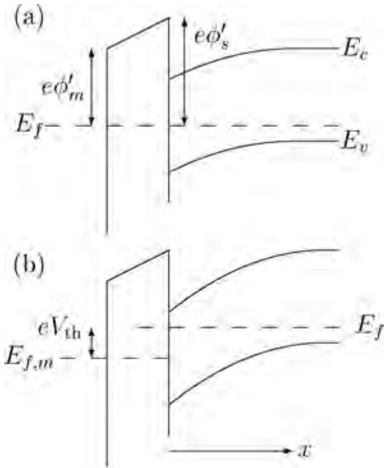


Figure 2. Energy band structures of an n-MOS structure under the gate bias of (a) $V_G = 0$, (b) $V_G = V_{\text{th}}$.

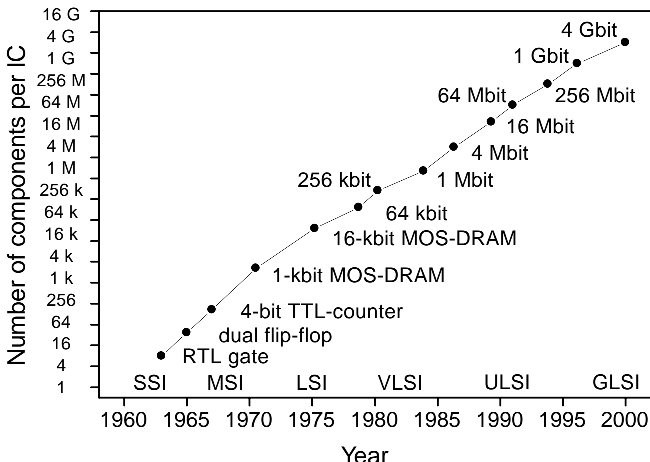


Figure 3. Growth in the number of components per IC. Reprinted with permission from [4], H. Veendrick, "Deep-submircon CMOS IC," p. vii. Kluwer Academic, Dordrecht, 2000. © 2000, Kluwer Academic Publishers.

A similar conclusion was derived by Lee and Park from their research activities in the field [52]. Moreover, we have room-temperature single-electron storage effects that support the pursuit of large-area devices operating on the basis of Coulomb blockade phenomena [53]. Possible are single-shot readout using the radio-frequency SET [54], Si nanocrystal memory cell with room-temperature single-electron effects [55], Si nanocrystal floating gate memories [56], hexagonal binary decision diagram quantum logic circuits using Schottky in-plane and wrap-gate control of GaAs and InGaAs nanowires [57], and silicon single-electron memory and logic devices for room temperature operation [58]. Excellent charge stability drifting by less than $0.01e$ over weeks in Si-based single-electron tunneling transistors (SETTs) was reported, making the devices more robust with respect to voltage-induced instability [59]. A rather bright side of the small transistor is the silicon single-electron charge-coupled device on a silicon-on-insulator (SOI) wafer [60].

Yet the most encouraging positive news came from the advancing semiconductor industry: IBM announced its strained silicon in which electrons experience less resistance and flow up to 70% faster, which can lead to chips that are up to 35% faster, without having to shrink the size of transistors [61]. Meanwhile, Intel now has the technology to make processors 10 times faster than today's speediest Pentium 4, ensuring continued PC performance growth through at least the end of this decade. The transistors are just 70 to 80 atoms in width and 3 atoms thick, which was announced at a semiconductor conference in Kyoto, Japan [62]. The achievement is important because it suggests that semiconductor makers will be able to keep pace with Moore's law until the end of the decade using roughly the same materials and manufacturing techniques used today.

Here we review the progress in the research field of semiconductor three-dimensional-nanoscale transistors. In other words, we concentrate on those structures in which the length, the width, and the depth of the conduction channel are all in the nanoscale. A most prominent device is called the single-electron transistor, which can be metallic as well as

semiconducting [63]. It is to a large degree still a transistor in the conventional sense. It is, however, so small in size that it can handle one single electron at a time [64–66]. By typing the keywords semiconductor + single + electron + transistor in to INSPEC, you can find that the number of publications per year in this field is booming. Such an impressive advance is synchronous with the fast development of nanofabrication technology during the last decade by producing artificial three-dimensional semiconductor nanostructures using laser crystallization [67], zone-melting recrystallization [68], molecular-beam epitaxy [69], metal-organic chemical vapor deposition, and chemical-beam epitaxy [70]. Accurately controlled feature sizes as small as monolayers of atoms in the growth direction for dissimilar semiconductor materials, or heterostructure systems, have been achieved. Nanoscale lithography and patterning by electron-beam (EB) lithography have also been highly developed [71]. Twenty nm resolution of electron lithography for the nanodevices on ultrathin SOI film was reported [72]. Patterning metallic and semiconducting nanostructures using local oxidation mediated via a voltage between a conducting surface and a close-by tip of an atomic force microscope (AFM) was developed especially for III–V [73]. Scanning probe microscope (SPM) lithography, which uses a SPM as a pattern delineation tool, would be one of the most promising candidates for the mass production tool [74].

In the following sections, we go through semiconductor systems of III–V, VI, carbon nanotubes, and others like organic thin films [75].

2. THEORETICAL CONSIDERATIONS OF DEVICES

The first step of the transistor minimization is the reduction of the channel length. As discussed in the previous section, the effective channel length approaches 10 nm [76]. At the same time, the source and drain junction depths are also being drastically reduced. The thickness of the conduction channels in Si-based MOSFET and III–V-based high-electron mobility transistors (HEMTs) is only about 10 nm. On the other hand, the width of the conduction channel must be greatly reduced for the sake of total component miniaturization. Eventually the conduction channel is to take the form of a short quantum wire or, effectively, a three-dimensional quantum dot, with a nanoscale in its width, thickness, and length.

Let us try to elaborate the relevant physics. When transistors are down-scaled and their characteristic dimensions reach the nanometer regime, quantum effects become dominant [77, 78]. The size effects of nanometer structure devices are reviewed including drift velocity overshoot, the hot electron effect, and high field transport, and a single quantum well transistor [79] and then single quantum well devices (e.g., metal-insulator-semiconductor FETs [80, 81], single quantum well transistors [82, 83], and HEMTs [84], and AlGaAs/GaAs heterostructure bipolar transistors [85]) were proposed. For nanoelectronic devices, a spectrum of modeling tools of varying degrees of sophistication is required to meet the needs of the various stages of quantum device development [86]. Reviews of recent developments in experiments and theories of the Coulomb blockade phenomenon were given [87, 88].

Though orthodox-theory-based models can explain the principal physical features of the device characteristics, a large number of observations cannot be quantitatively understood. Theoretically, the reduction in the carrier transport route (conduction channel) results in greatly enhanced direct tunneling between the source and the drain. However, direct tunneling is not likely to be a principal reason for deteriorating the transistor function. A transistor having a 40-nm gate works at room temperature with a threshold bias of 0.4 V [9]. It was shown theoretically that the quantum effects improve the functionality. In one nanoscale MOSFET, the semiclassical method indicated a negative threshold gate bias, whereas quantum confinement effects hold the threshold gate bias at the reasonable circuit level [89]. In addition, the quantum effects also suppress the direct tunneling current significantly [90].

The most prominent effect appearing in a nanoscale transistor is of course the Coulomb blockade effect. When one electron arrives at the conduction channel, its Coulomb potential lifts up the local potential energy so that the later electrons become scattered. For a nanoscale conduction channel, the potential energy changes very significantly so that the scattering is very strong. In other words, the transport of other electrons becomes blockaded. Figure 4 shows a typical drain current as a function of the source-drain and gate voltages.

Theoretically, the Coulomb blockade process is more involved. We consider two electrons. As fermions, the total wave function of the two-electron system is written as

$$\frac{1}{\sqrt{2}}\begin{vmatrix} \psi_i(\mathbf{r}_1) & \psi_i(\mathbf{r}_2) \\ \psi_j(\mathbf{r}_1) & \psi_j(\mathbf{r}_2) \end{vmatrix} \tag{1}$$

because of the antisymmetric property, where $\psi_i(\mathbf{r})$ is the one-electron eigenstate. The matrix element of the Coulomb

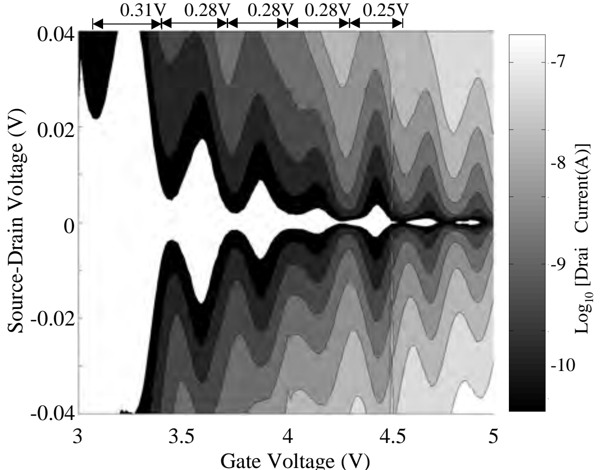


Figure 4. Drain current contour plot of a device at the source–drain voltage and gate voltage plane. The measurement temperature was 20 K. Reprinted with permission from [89], Y. Fu et al., *Superlattices Microstruct.* 28, 177 (2000). © 2000, Elsevier Science.

interaction between the two electrons is

$$V_{ij} = \frac{e^2}{4\pi\epsilon}\int d\mathbf{r}_1\, d\mathbf{r}_2 \frac{|\psi_i(\mathbf{r}_1)|^2|\psi_j(\mathbf{r}_2)|^2}{|\mathbf{r}_1 - \mathbf{r}_2|} - \frac{e^2}{4\pi\epsilon}\int d\mathbf{r}_1\, d\mathbf{r}_2 \frac{\psi_i^*(\mathbf{r}_1)\psi_i(\mathbf{r}_2)\psi_j(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \quad (2)$$

The first integral is the direct Coulomb interaction between two charges while the last one is the exchange energy.

Consider the case that one electron, ψ_i, already occupies the central conduction channel. Another electron ψ_j is in the source. The Coulomb interaction that electron ψ_j experiences is the direct Coulomb interaction, since the exchange energy is zero due to the spatial separation between ψ_i and ψ_j. The conclusion is valid when the number of electrons in the conduction channel is larger than 1.

However, the Coulomb interaction on electron ψ_j changes during its transmission from the source to the conduction channel. Consider that electron ψ_j arrives at the conduction channel. If the two electrons occupy the same energy state (i.e., $i = j$), it is immediately concluded that $V_{ij} = 0$. In other words, each energy state can be occupied by two electrons and the potential energy of the system is unchanged (the sum of the direct Coulomb interaction and the exchange energy is zero).

When $i \neq j$, assume that one state is s-type and the other is p-type. A cylindrical conduction channel can be divided into two identical parts, Ω_1 and Ω_2, according to the sign of the p-state. The spatial integration in the exchange energy becomes

$$\int_{\Omega_1+\Omega_2} = \int_{\Omega_1}\int_{\Omega_1} + \int_{\Omega_1}\int_{\Omega_2} + \int_{\Omega_2}\int_{\Omega_1} + \int_{\Omega_2}\int_{\Omega_2} \quad (3)$$

and notice that

$$\begin{aligned} F(\mathbf{r}_1 \in \Omega_1, \mathbf{r}_2 \in \Omega_1) &= -F(\mathbf{r}_1 \in \Omega_1, \mathbf{r}_2 \in \Omega_2) \\ &= F(\mathbf{r}_1 \in \Omega_2, \mathbf{r}_2 \in \Omega_1) \\ &= -F(\mathbf{r}_1 \in \Omega_2, \mathbf{r}_2 \in \Omega_2) \end{aligned} \quad (4)$$

where

$$F(\mathbf{r}_1, \mathbf{r}_2) = \frac{\psi_i^*(\mathbf{r}_1)\psi_i(\mathbf{r}_2)\psi_j(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \quad (5)$$

It is thus concluded that the exchange energy is zero when one electron occupies the s-state and the other electron occupies the p-state. The two electrons interact with each other via direct Coulomb interaction.

Here we see the necessity of the wave packet model on the electron transport, which makes the theoretical investigation extremely complicated. In the following, however, we review the progress of theoretical work thus achieved.

2.1. Semiclassical and Phenomenological Approaches

The Poisson equation and the master equation based on the semiclassical theory were used to analyze a SET with Schottky tunnel barriers [91]. Through solving a set of device equations, the basic characteristics of a nanoscale MOSFET were simulated and then analyzed [92]. A single-electron simulator was integrated into a SPICE-type conventional circuit simulator to study single-electron multiple tunnel junctions [93]. A kinetic approach based on a master equation models the injection and ejection of electrons into and from the quantum dot [94]. A reduced master equation analysis was discussed to avoid the computational cost of the full master equation method [95]. There was another phenomenological approach based on quantum mechanical considerations [96], and the model was further developed into the Monte Carlo model where the current is computed by counting the sequential tunneling of electrons through the energy level of the dot [97]. Circuit models have also progressed [98]. Using a master equation model and taking into account the discreteness of quantum levels and the finiteness of scattering rates, Saitoh and Hiramoto analyzed the Si SET [99]. The binding of equal numbers of electrons and holes on two islands of the coupled SETs leads to a Coulomb blockade due to the electron–hole pairs [100]. The single-electron tunneling junction was modeled by a nonlinear capacitor based on physics-based Monte Carlo simulations [101]. A simple SET model was obtained based on a numerical solution of the Poisson equation [102].

A semiclassical tunnelling model was developed where the tunnel rates through junctions are determined by directly calculating a golden rule equation [103]. A semiclassical numerical golden rule calculation for SET simulations was reported [104]. Zero-bias current in a SET was calculated based on the semiclassical model [105]. Photoirradiation effects in a single-electron tunnel junction array were calculated by a Monte Carlo method [106]. A self-consistent approach to the Poisson and Schrödinger equations was applied to calculate the confinement potential and determined the conditions for single-electron tunneling [107]. Similar self-consistent calculation was formulated to obtain the charging energy required to add a single electron to a quantum dot which is electrostatically coupled to leads and a gate electrode [108]. Quantum wave transport of carriers was studied numerically in a nanoscale n-Si MOSFET [109]. A linear-response approach to simulate the conductance of semiconductor single-electron transistors at the solid-state level was reported [110] and a multidimensional secant approach calculated self-consistent electron densities in a quantum dot weakly coupled to a macroscopic reservoir [111]. Vertical quantum dots were also studied [112]. Currents and their fluctuations in two capacitively coupled single-electron transistors were studied within the sequential tunneling approach [113]. The electronic transport properties of parallel coupled SETs at strong coupling under asymmetrical voltage bias were investigated theoretically, where it was found that the binding of electrons and holes on the two islands of the coupled SETs is the key element that governs the transport characteristics [114]. An extended Hubbard model with interdot electron–electron interactions was presented to study the conductance through a coupled triple-dot system [115]. Using a simple model of hard wall confinement potential for quantum dots, pillar-shaped quantum dots for single-electron tunneling transistors operating at higher temperatures were suggested [116]. Current-density-functional theory was used to calculate the

electronic structure of two-dimensional quantum dots containing up to 58 electrons [117]. Cotunneling and renormalization effects for the single-electron transistor were studied by a diagrammatic expansion of the reduced density matrix up to fourth order in the tunneling matrix elements [118]. The Monte Carlo method was used to simulate SETs including inelastic macroscopic quantum tunneling of charge [119].

As complicated as the quantum process is in SET, there was an analytical SET model for realistic SET circuits [120], and an analytical I–V model for SETs was set up to be incorporated into a conventional circuit simulator, such as SPICE [121]. A SPICE model for the single-electron tunnel junction was also set up [122]. The macroscopic modeling of silicon devices to include quantum mechanical corrections in the classical transport framework was discussed, and both device and circuit models were provided [123].

2.2. Nanosystem Designs

A silicon effective quantum dot was theoretically designed using oxidation-induced strain [124]. A SET on a SiGe/Si vertical structure transistor was studied with a side contact for easier control of the depletion region by the gate voltage [125]. The single-electron device was controlled using environmental impedance modulation [126].

2.3. Other Relevant Issues

Extended theoretical works were also published discussing quantitative two-dimensional carrier profiling in MOS [127], observation of quantum fluctuations of charge on a quantum dot [128], dynamic input capacitance of SETs, and the effect on charge-sensitive electrometers [129]. Non-Gaussian distribution on nearest-neighbor Coulomb peak spacings in metallic SETs was observed and attributed to impurity-specific parametric charge rearrangements close to the transistor [130].

2.4. Circuit Designs

There are extended works as well aimed at the design issues of the SET circuit (e.g., effects of quantum dot radii and densities on the fabricated memory characteristics) [131]. Considerations of the realization of a GLSI processor based upon nanoelectronic devices [132], massively parallel algorithms for simulating large-scale nanoelectronic networks based on the single-electron tunneling effect [133], and a single-electron device and circuit simulator with a new algorithm to incorporate cotunneling [134] were reported.

3. DEVICES BASED ON III–V SEMICONDUCTORS

3.1. Physical Processes

Individual hopping events in the flow of current (approximately one electron per second) along a single chain of localized states modulate the conductance of a ballistic constriction formed in a GaAs/$Al_xGa_{1-x}As$ split-gate transistor [135]. Besides the well-known Coulomb blockade effect, there are other related physical processes (e.g., Coulomb charging effects in an open quantum dot device) [136]. The Coulomb blockade effect from weak to strong coupling regimes was investigated in a small two-dimensional electron gas island in a AlGaAs/GaAs heterostructure [137].

3.2. Device Fabrications

Advances in the epitaxial synthesis of 10-nm-scale quantum dot (QD) and wire structures were reviewed recently [138]. The basic setup of the devices is the two-dimensional electron gas (2DEG) formed at the GaAs–AlGaAs or other III–V-based heterointerfaces. In general, the thickness of the 2DEG is about 10 nm. Different techniques have been developed to either isolate one small spatial region ($\sim 100 \times 100$ nm^2) of the 2DEG (quantum dot) from the rest 2DEG or first form a narrow quantum wire (about 100 nm wide) and then cut the wire into pieces. The advantage of thus fabricated quantum dots is the precision control over the spatial position of the dot. However, the dot size is large (about $100 \times 100 \times 100$ nm^3). The second category of forming the quantum dots is epitaxially growing a monolayer material on a lattice-mismatched substrate. The monolayer material self-assembles into small quantum dot structures due to the lattice mismatch. The most successful system is self-assembled InAs quantum dots on GaAs which has already been widely used in optoelectronics. The dot size can reach $10 \times 20 \times 20$ nm^3. However, the spatial positions of the dots cannot be properly controlled. Moreover, the dot size is not uniform.

3.2.1. Multiple Gates

With a suitable bias, a metal gate can deplete the carriers underneath. By configurating the geometry of the gate or using multiple gates, quantum dot or quantum wire can be fabricated out of the 2DEG. Multiple-gate quantum dot devices on shallow GaAs–AlGaAs heterostructures were reported [139]. A SET was formed by placing a nanoscale single barrier inside the gap of a conventional split-gate FET [140]. Two tunnel barrier gates for quantum wire and a center gate for a quantum-dot-potential control were applied to fabricate a GaAs-based SET which showed a proper inverter operation at 1.6 K and realized a logic transfer gain of larger than unity (1.3) [141]. Capacitively and resistively coupled single-electron transistors were fabricated using GaAs/AlGaAs 2DEGs and metal Schottky gates [142]. The device configuration is schematically represented in Figure 5. The basic idea is that gates V_1, V_2, V_3, and V_4 deplete the carriers underneath so that the carriers surrounded by the gates become spatially isolated, forming the desired quantum dot. Gates V_{g1} and V_{g2} control the number of carriers in the quantum dot region. A chromium thin-film resistive gate was connected directly to the mesoscopic island formed between two ultrasmall Al/AlO_x/Al tunnel junctions forming a resistive SET [143]. A parallel quantum-dot configuration defined by multiple gates in the two-dimensional electron gas of a GaAs–$Al_xGa_{1-x}As$ heterostructure was reported [144].

3.2.2. Etching and Selective Molecular Beam Epitaxy

The electric field created by a metal line is very diffused so that the quantum dot system fabricated by metal gates is first of all shallow. Second, the space of the depleted

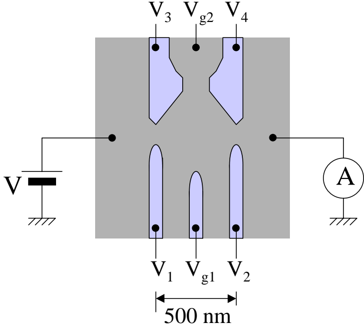


Figure 5. Schematic configuration of a typical multiple-gate SET system with measurement setup. Reprinted with permission from [142], F. Wakaya et al., *Superlattices Microstruct.* 27, 603 (2000). © 2000, Elsevier Science.

area is also wide so that it is difficult to have a good coupling between carriers confined in the quantum dot and in the surrounding area. Etching and selective molecular beam epitaxy (MBE) growth are used to shape the 2DEG into a quantum wire structure with well-defined edges. A typical example is demonstrated by Figure 6, where a quantum wire having a cross section of only 80×10 nm^2 was produced and a GaAs/InGaAs/AlGaAs-based SET thus formed showed clear Coulomb-blockade oscillations [145]. The SET has been designed by three-dimensionally solving Schrödinger and Poisson equations [146].

A SET was produced by etching narrow trenches in a AlGaAs/GaAs heterostructure and defining a channel and six in-plane gates in the two-dimensional electron gas [147]. Using electron-beam lithography and reactive ion etching, conducting wires, rings, and dots with lateral dimensions ~50 nm were fabricated [148]. A parallel dot structure was also fabricated [149]. The SET structure is schematically represented in Figure 6. InGaAs coupled quantum structures were grown by selective MBE on specially designed patterned InP substrates [150]. By mechanically scratching the surface of a GaAs/AlGaAs heterostructure with an AFM, an energy barrier for the two-dimensional electron gas is formed, and thus the SET was formed [151]. The AFM was used for its very accurate movement to position the prefabricated nano-objects into a small gap between two metal electrodes [152]. Fundamental results obtained in an AFM chemically induced direct nanolithography process were presented for fabricating nanoscale devices such as SETs and quantum effect electronic devices [153].

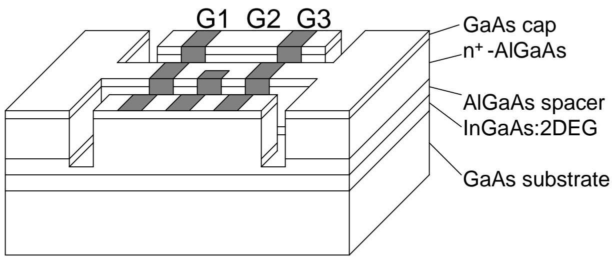


Figure 6. GaAs/InGaAs-based single-electron transistor. Reprinted with permission from [145], Y. Fu et al., *Appl. Phys. Lett.* 78, 3705 (2001). © 2001, American Institute of Physics.

3.2.3. Point Contacts

Side-gated point contact structures in delta-doped layers have been used to form single-electron tunnel junctions [154]. Point contact MOSFETs with extremely narrow channels showed Coulomb blockade oscillations at room temperature, and at low temperatures, negative differential conductances and fine structures are superposed on the device characteristics [155].

3.2.4. Self-Assembly

Self-formed GaAs quantum nanostructures are realized by selective area (SA) metal organic vapor phase epitaxy (MOVPE) [138]. Semiconductor nanostructures are realized by patterning AlGaAs/GaAs heterostructures with an AFM [156, 157]. Quantum dots were fabricated by MBE on prepatterned GaAs substrates [158–160]. Quantum antidot arrays of circular shape by etching and quantum wire transistors were fabricated on InAs/$Al_{0.5}Ga_{0.5}Sb$ heterostructures [161]. Vertical quantum dot single-electron resonant tunneling transistors were reported which move from the few-electron to the several- and many-electron regimes to study the spin-polarized droplet [162]. Negative differential resistance due to the spin-polarized detection of electrons combined with a long spin relaxation time in the dot was reported in the SETs containing a few electrons and spin polarized source and drain contacts were formed in GaAs/GaAlAs heterojunctions using metallic gates [163].

Electronic and optoelectronic properties of self-assembled semiconductor structures were studied [164, 165]. A single self-assembled InAs dot in split-gate delta-doped channel transistor structure was reported [166].

3.2.5. Combination of Techniques

Nanostructures were formed in InGaAs-based modulation doped heterostructures using Ga focused ion beam implantation [167]. A SET was fabricated by current-controlled local oxidation of a two-dimensional electron system [168]. A quantum point contact was defined by lithography with an AFM and subsequent wet-chemical etching [169]. Laser-written *p*-doped lines were used to define a dot with a diameter of about 70 nm and to insulate it from the in-plane gates [170]. SETs with asymmetrical tunnel barriers were proposed [171].

Self-assembling InAs quantum dots were formed in combination with SA-MOVPE on masked (001) GaAs substrates [172]. AFM oxidation was applied for the direct oxidation of a thin InAs layer to fabricate a SET from an InAs/AlGaSb heterostructure [173]. A single self-assembled InAs dot in a split-gate delta-doped channel transistor structure was reported [174]. Experimental and simulated results of a room temperature single-electron transistor formed by an AFM and scanning tunneling microscopy (STM) nano-oxidation process were reported [175].

3.2.6. Traps

The current through transistor switches between trapping and detrapping of a single electron in the trap level was studied [176], and random telegraph signals from these traps using a SETT formed in the GaAs/$Al_xGa_{1-x}As$ heterostructure were studied [177].

3.3. Characterizations

There are a number of very effective ways to characterize the SET devices: nondestructive room temperature characterization using contactless photoreflectance [178], electromodulation and surface photovoltage spectroscopy [179], transport spectroscopy [180], complex impedance spectroscopy and scanning tunneling spectroscopy [181], SET capacitance spectroscopy to measure the flow of charge to and from a quantum dot [182]. Noise of single-electron transistors was measured using a transimpedance amplifier [183]; shot noise in Si-based single-electron transistors was measured in 2–12 MHz at 1.5 K with a lock-in technique [184].

3.4. Applications

A charge transformer was proposed for noise matching single-electron transistor amplifiers to high-capacitance devices [185]. Present status and future prospects of resonant tunneling hot electron transistors (RHETs) were discussed to develop room temperature quantum functional bipolar transistors and quantum box devices based on the RHET technology for the ultrasmall limit of electron devices [186].

A transmission-type radio-frequency SET (rf-SET) based on an AlGaAs/GaAs heterostructure was tested and its application to investigating high-frequency (up to 1 MHz) charge noise in an AlGaAs/GaAs quantum dot [187] was reported. Double quantum dots were used as a highly sensitive submillimeter-wave detector [188].

4. DEVICES BASED ON IV SEMICONDUCTORS

4.1. Theoretical Considerations

The quantum gates of coupled quantum dots were studied theoretically when charging effects can be observed, which shows that the charged states in the qubits can be observed by the channel current of the MOSFET structure [189]. A strain-induced effective Si quantum dot was theoretically proposed [124]. A Si SET was studied by solving a master equation, assuming that the potential barrier has a parabolic potential [190].

4.2. Device Fabrications

A brief survey of different realizations of single-electron devices fabricated in SOI films was presented [191]. For industrially acceptable techniques, a contact pattern mask with nanometer-size slits was fabricated by combining conventional photolithography with anodic oxidation of metal sidewalls [192]. There are quite extended fabrication methods: reducing MOSFET size [193], a SET with a single-island format [194], Si nanoscale point contact channels on the SOI substrate showing room temperature Coulomb blockade [195], 10-nm-scale lithographic techniques to fabricate electrically variable shallow junction MOSFETs, and lateral hot-electron transistors which predict the limitation of MOSFET miniaturization to be around 5 nm in the source-drain tunneling scheme [196]. Since nanoscale-gate-length MOSFET devices require extremely shallow source/drain extension regions with a junction depth of 20–30 nm [2], a 70-nm n-MOSFET fabrication with 12 nm n^+–p junctions was reported by using As_2^+ low energy ($\leq$10 keV) implantation [197]. Due to the limitation of SiO_2 as the gate insulator, an organic self-assembled monolayer of alkyltrichlorosilanes was used as the gate insulator in a hybrid organic/silicon nanometer-size FET having a gate length of 25 nm [198]. MOSFETs with single {111}-facetted silicide/silicon joints have been fabricated on SOI substrates using cobalt silicide for reducing the series resistance [199].

4.2.1. Multiple-Gate Systems

Periodic current oscillations due to the single-electron charging effect have been observed in a Si quantum dot realized in the inversion layer of a Si-MOSFET with a dual-gate structure [200]. Multiple-gate [201] and side-gate systems [202, 203] were also studied. A strong suppression of short-channel effects at room temperature and clear conductance oscillations at liquid helium temperatures on a submicrometer MOSFET equipped with a gate on three sides of the channel were reported [204]. The fluctuations of Coulomb-blockade peak positions were studied in a Si quantum dot defined by patterning the 2DEG of a silicon MOSFET using stacked gates [205]. Metal based SETs were also reported [206, 207]. Ultrasmall quantum dots were formed in SOI material by a metallic top gate and an in-plane gate [208]. A SET was formed with two gate inputs [209]. There were also split-gate Si MOS quantum dot structures [210]. A dual-gate-controlled SET with coupled dot geometry was fabricated on a SOI structure [211].

Later on, modification of the tunneling barrier in a nanocrystalline silicon SET due to different oxidation processes was investigated [212].

4.2.2. Etching and Lithography

EB lithography [213], reactive ion etching (RIE), and thermal oxidation [214] are normally standard in Si-based systems [215]. Si nanodot and nanowire memory transistors with side gates were fabricated using an inorganic EB resist process [216]. Transistors consisting of quantum dots formed using EB nanolithography and RIE show the oscillation of the drain current as a function of the gate voltage at temperatures up to 170 K and drain biases up to 80 mV [217]. Single-charge detection in Si-wire MOSFETs fabricated on a SOI wafer was investigated [218]. Two potential barriers were induced by direct EB irradiation [219]. Atomic fabrication of a silicon based quantum computer was demonstrated [220]. More delicate recipes of SET by combining different techniques including dry etching, inductively coupled plasma sources, and a self-aligned process with a shadow mask were reported [221]. A SET was obtained by bonded and etched back silicon on insulator material [222]. A Si quantum wire with a gate-electrode and a Si SET using SIMOX substrates and electron-beam lithography

were reported where quantized conductance in the wire up to 100 K and periodic conductance oscillation in the SET at room temperature were clearly observed [223, 224]. A similar system was reported elsewhere [225]. A Si SET was fabricated using EB nanolithography combined with an image reversal process and a dry etching technique for room temperature operation [226]. Anisotropic etching was developed to fabricate Si nanostructures for SET applications [227]. SPM-based fabrication techniques of nanostructures by local oxidation of silicon in ambient atmosphere were investigated [228]. Imprint lithography in polymethylmethacrylate on silicon and on gold substrates was reported to have a feature size of sub-10 nm [229]. The inevitable parasitic MOSFETs on Si SETs, which were fabricated in ultrathin Si of a SOI substrate by pattern dependent oxidation (PDOX), were carefully studied and their effects on SETs were found to be reduced by the back gate [230]. A SETT was fabricated by EB lithography, anisotropic RIE, and low temperature oxide deposition [231]. Doped-thin-Si-film SETs were fabricated from a highly doped Si film in a SOI substrate by EB lithography with a high-resolution resist (calixarene) and dry etching with CF_4 gas [232]. A novel method of fabricating a nanoimprinting mold for a single-electron device was proposed which makes use of the peculiarity of anisotropic wet chemical etching [233]. There were also Si nanodevices by optical lithography [234].

4.2.3. Point Contacts

Silicon SETs have been fabricated in the form of point contact MOSFETs with various channel widths using EB lithography and anisotropic etching techniques on SOI substrates [235], which is schematically presented in Figure 7. The dot size, as small as 5.3 nm in a silicon point contact channel, showed a negative differential conductance due to quantum levels up to 25 K [236]. For a similar point contact MOSFET and the gate oxide by CVD instead of thermal oxidation, it was found that the single-electron addition energy is about 259 meV and the dot diameter is less than 4.4 nm [237]. A Si SET with in-plane point contact metal gates was reported [238].

4.2.4. Si Nanocrystals

Rather early there was a report about silicon islands by laser recrystallization [239]. A nanoscale polysilicon dot as the floating gate embedded between the channel and the control gate was used to fabricate a room-temperature silicon SET memory [240–243]. Si SETTs with nanoscale floating dot gate stacked on a Coulomb island by a self-aligned process [244]. A two-terminal nanocrystalline silicon memory device at room temperature was reported [245]. A self-assembled Si quantum dot transistor was fabricated showing room-temperature oscillations in the drain current [246]. Storing electrons in and erasing from Si nanocrystal dots of diameter 8 ± 1 nm deposited by plasma decomposition of SiH_4 were demonstrated in dot-based memory devices at 77 K [247].

A novel memory was based on nanocrystalline Si which acts as a floating gate in large area MOSFETs [248] and a single polysilicon dot was defined by EB lithography over a narrow SOI channel [242, 248, 249]. Si nanocrystals were deposited by very-high-frequency plasma decomposition of silane, resulting in an average dot diameter of 8 nm and a source–drain electrode separation of 30 nm, and a gate electrode is employed so that the charge states in quantum dots can be controlled [250] (see Fig. 8). Si SET memory using a multiple-tunnel junction was also fabricated by EB direct writing [251], and the experimental works have been carefully analyzed theoretically [252–255]. The nanocrystal was then further processed into single-electron nonvolatile memory devices [256]. Microcrystalline silicon thin films were prepared at 300 °C on glass and its application to SET was investigated [257]. A germanium nanocrystal quasi-nonvolatile memory device which consists of a MOSFET with Ge charge traps embedded within the gate dielectric was reported [258]. The electron charging or discharging of the silicon QDs in MOSFETs with the QD layer floating gate were utilized for memory operation [259]. The operation of a Si MOSFET with an array of Ge self-assembled QDs embedded into the active channel showed oscillations caused by Coulomb interaction of holes at $T \leq 200$ K [260]. The device is not only limited to SETs. A double-SiO_2-barrier MOSFET with a poly-Si dot was reported [261]. Si quantum dots fabricated by PDOX in silicon [262], and the origin of the potential profile in such silicon SETs, were investigated by making use of the geometric structure measured by AFM, the bandgap reduction due to compressive stress generated during PADOX

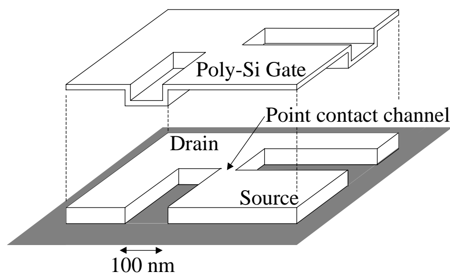


Figure 7. Si point contact MOSFET. Reprinted with permission from [155], H. Ishikuro and T. Hiramoto, *Appl. Phys. Lett.* 71, 3691 (1997). © 1997, American Institute of Physics.

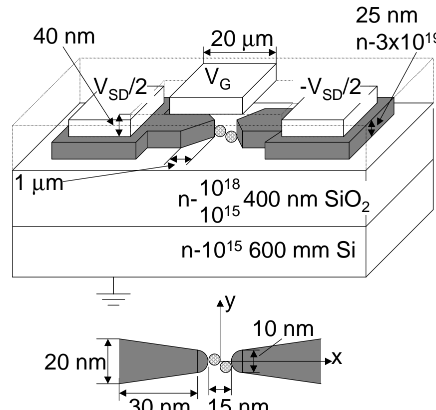


Figure 8. Si nanocrystal based single-electron transistor. Reprinted with permission from [255], Y. Fu et al., *Superlattices Microstruct.* 28, 189 (2000). © 2000, Elsevier Science.

obtained using a first-principles calculation, and the effective potential method [263]. Electrochemically etched silicon was dispersed into a colloid of ultrasmall (about 1 nm) silicon nanoparticles [264]. Nanocrystalline Si dots were also formed by annealing a thin film of SiO_x ($x < 2$) with which room-temperature single-electron narrow-channel memory with silicon nanodots was obtained [265]. Silicon nanocrystal was also used as floating gates to control Coulomb blockade oscillations in silicon SETs [266]. High-spatial-density nanocrystals were formed using a thin layer of amorphous $Si_{0.7}Ge_{0.3}$ deposited on SiO_2 [267]. SETTs were composed of thin polycrystalline silicon films patterned by EB lithography and thermally grown oxidized film [268].

4.2.5. Combination of Dual-Gate and Nanocrystal

Fabrication and low-temperature transport properties of dual-gated SETs were studied based on a QD formed by SA growth of MOVPE [269]. SETs were fabricated in solid phase crystallized polycrystalline silicon films deposited on SiO_2 layers grown on silicon substrates and lateral side-gated nanowires [270]. In a trench (150–600 nm wide) defined inside a large "inversion" gate electrode into which nanocrystalline Si (8 ± 1 nm diameter) was deposited by remote plasma enhanced chemical vapor deposition, a two-gate single-electron memory device with an active area of 1.5×10^{-10} cm^2 was processed to study the Si/SiO_2 interface in SOI nanoscale channels and nanocrystalline Si dots [271].

4.2.6. Impurity and Trap

In 1988, Uren et al. discussed the random telegraph signal in the drain current of submicrometer Si MOSFETs and suggested sequential two-electron capture involving individual Si/SiO_2 interface states which can exist in two or more charge-equivalent, metastable states [272]. The Coulomb energy induced by the transfer of a single electron into an interface trap in MOSFETs was measured by random telegraph signals [273]. Coulomb blockade of a single quantum dot formed by an impurity potential in low-mobility, nanometer size Si MOSFETs was reported [274].

4.3. Characterizations

Low-dimensional p-channel Si MOSFETs with a source/drain width of 100 nm formed by selective focused-ion-beam implantation were fabricated and Coulomb blockade phenomena were examined as a function of channel length [275]. A similar technique was used to fabricate SETs operating at room temperature [276]. Influence of background charges on Coulomb blockade in quantum dots was studied in order to understand the sensitivity of the system to changing background charges due to freezing out or capture of single electrons in the vicinity of the devices in the substrate or the buried oxide of the SOI material [277]. The potentialities of different techniques for detecting signals from barrier structures in a scanning electron microscope were discussed [278]. A silicon nanowire memory utilizes surface charging and its operation analysis by scanning Maxwell-stress microcopy [279]. Oxide trap properties were extracted from the temperature dependence of random telegraph signals in submicrometer MOSFETs [280]. Random telegraph noise was used to study the Coulomb blockade phenomena at room temperature [281]. Si nanostructures were characterized by cross-sectional transmission electron microscopy [282]. There were discussions about strategies for measuring currents from single-electron devices [283]. An asymmetric tunnel barrier in a Si SET was characterized from the strong asymmetry in conductance versus source–drain voltage [284].

4.4. Circuit Applications

From the engineering point of view, it was concluded that memories seem to be one of the most promising large scale SET applications, particularly when granular films are used, in the range of today's process technology [285]. A silicon single-electron memory consisting of three integrated silicon nanowires was demonstrated [286].

On the circuit level, a programmable NAND/NOR logic gate, threshold logic gates, and parallel counters using resonant tunneling transistors were reported [287]. Boolean logic functions were implemented in the so-called quantum-dot cellular automata [288]. A merged SET and MOS transistor logic was introduced for a SET-to-CMOS interface circuit as well as for a basic element for multiple-valued functions [289, 290]. A Si complementary single-electron inverter with two identical SETs was fabricated on a SOI substrate, and input–output transfer with a voltage gain larger than unity is demonstrated at 27 K [291]. A silicon SETT, which shows an inverting voltage gain greater than unity and as high as 3.7, has been fabricated [292]. Voltage gain of Si SETs fabricated by PDOX was investigated and used to evaluate the performance of the n-MOS type inverter with resistive load [293]. There was also vertical PDOX [294]. The SET was used as a Schmitt trigger with excellent noise immunity [295]. Programmable SET logic was demonstrated for low-power intelligent Si large-scale integration [296]. Silicon-based few-electron memory operated with write times as short as 10 ns and voltages of less than 5 V at 4.2 K, and operation persisted to 45 K [297]; then a 3×3-bit Coulomb blockade memory cell array was demonstrated [298]. MOS-integrated metallic single-electron memories were also reported [299].

Multiple-valued memory operation was successfully demonstrated by combining a SET and a MOSFET which is schematically shown in Figure 9 [300], where the source of the MOSFET with a fixed gate bias V_g is connected to the drain (d) of the SET. As discussed previously, the SET drain current I_d increases and decreases periodically as a function of the SET gate bias V_{gs} (the SET source, s, is grounded; see Fig. 9). However, the I_d has such a large dependence on the drain voltage V_{ds} that the peak current is almost proportional to V_{ds}, and the valley current increases more rapidly when the Coulomb blockade breaks. The MOSFET connected to the SET eliminates this large V_{ds} dependence by keeping V_{ds} nearly constant around $V_g - V_{th}$, where V_{th} is the threshold voltage of the MOSFET. This $V_g - V_{th}$ is set low enough to sustain Coulomb blockade. By connecting the SET gate to the MOSFET drain, the peak negative differential resistance (NDR) characteristic in Figure 9b is obtained. If a constant-current source I_0 is connected to the

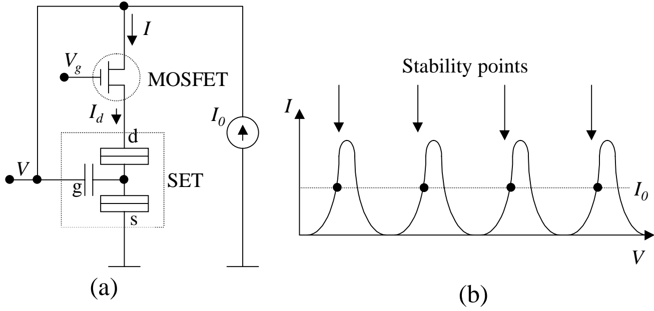


Figure 9. (a) Schematic drawing of the multiple-value NDR device comprising a SET and a MOSFET. (b) Expected I–V characteristics of the device. Reprinted with permission from [300], H. Inokawa et al., *Appl. Phys. Lett.* 79, 3618 (2001). © 2001, American Institute of Physics.

combined SET-MOSFET device, the periodic nature of the I–V characteristics results in a number of stability points, which enables multiple-valued memory operation. The number of peaks is infinity in principle but is in practice limited by the breakdown voltage of the MOSFET drain or SET gate. Another SET application is the so-called exclusive-OR (XOR) function as shown in Figure 10 by utilizing the oscillatory conductance characteristics due to the Coulomb blockade [301].

There was SET-based circuit design [302], and a new strategy of constructing logic trees with SETs and their complementary SETs both working as pull-down devices was reported [303]. In the case of arrays of nanometer-size metallic dots on a SiO_2 substrate, a drastic reduction of the V_{th} dispersion in these highly disordered arrays can be achieved by a special design [304]. A single-electron neuron device was described [305]. The low-charge limit in SET memory devices was discussed [306]. A single-transistor memory structure, with changes in threshold voltage exceeding approximately 0.25 V corresponding to single-electron storage in individual nanocrystals, operating in the sub-3 V range, and exhibiting long term to nonvolatile charge storage, was reported [307]. Concerning the low driving capability of SET devices, it was experimentally demonstrated that the direct transmission of SET signals to the conventional MOSFET current is possible, so that the output impedance can be transformed locally or globally from high to low in the SET-device circuits hybridized with CMOS devices [308]. A cryogenic voltage amplifier using GaAs FETs couple to a superconducting SET inside a dilution refrigerator was developed to extend the maximum output

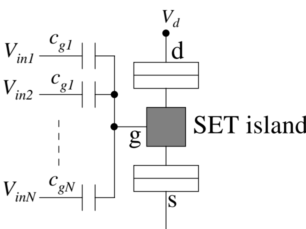


Figure 10. Multi-input-gate SET that implements the XOR function. Reprinted with permission from [301], Y. Takahashi et al., *Appl. Phys. Lett.* 76, 637 (2000). © 2000, American Institute of Physics.

frequency of the SET, normally less than a kilohertz, up to 1 MHz [309]. SET based on a gated silicon nanowire has been used to make a memory together with an electrometer detector to read the memory state [310]. EEPROM based on Si quantum dots on oxide–nitride tunneling dielectrics was fabricated [311].

A design of an optically driven quantum dot quantum computer was proposed based on n-type SETs [312].

Fundamental calculations on the circuit level were conducted: Effects of dot size and dot size distribution on electron number control in silicon floating dot memories at room temperature were investigated by numerical calculation [313]. A functionally complete set of complementary logic circuits based on capacitively coupled single-electron transistors (CSETs) was numerically simulated [314]. A high-speed simulation technique for logic circuits comprising SETs was proposed [315]. A new complementary digital logic based on resistive CSETs was proposed and its basic characteristics were numerically analyzed using the Monte Carlo method [316]. The use of nanoelectronic structures in cellular nonlinear network architectures, for future high-density and low-power CMOS-nanodevice hybrid circuits was simulated [317]. An approach to computation with single-electron circuits was discussed in which information encoded in sensitive circuit signatures was adaptively tuned by adjusting circuit variables to achieve a computationally useful goal [318]. Simulations of SET memory were discussed based on experiments and achievable nanotechnologies [319]. Simulated annealing was proposed as a simulation method for single-electron tunneling devices and circuits, with acceptable computation time when cooling schedule and cooling parameters are appropriately chosen [320].

4.5. Single Hole Transistors

An electron-based SET is normally concentrated. However, studies on hole-based systems were also conducted (e.g., single-hole transistors [321, 322], room-temperature single-hole silicon memory cells [323]). At $4 < T < 100$ K, single-electron and single-hole transport through the lithographically defined dot were reported [324]. Dual-gate quantum dot transistors have been fabricated in n- and p-type strained SiGe on SOI [325].

5. NITRIDE AND OTHER SYSTEMS

5.1. GaN

GaN quantum dots were formed on an AlGaN/SiC substrate by gas-source molecular beam epitaxy using a self-assembling technique with Ga droplets showing Coulomb blockade phenomena with negative conductance at 2.7 K and a clear Coulomb staircase at 200 K [326].

5.2. SiN and SiC

High-density Si pillars were formed using thermally grown SiN nuclei and furnace oxidation [327]. Silicon nitride barrier layers of 2 nm thickness were used to form silicon nanopillars for SET [328]. Multiple-valued memory operation was reported in SiN-based single-electron memory

[329]. Single-electron memory using carrier traps in a silicon nitride layer was reported [330].

Single-hole transistor operation and infrared induced emission were obtained with 6H-SiC diffusion nanostructures [331].

5.3. Carbon Nanotubes

Electronics of single-wall carbon nanotubes was discussed [332, 333]. A 0.4 μm long semiconducting single-walled carbon nanotube reveals single-electron charging at temperatures up to 160 K [334]. Similar works were reported [335]. A SET was fabricated using the single-wall carbon nanotube grown directly on to the Si AFM tip and TiO_x lines as the tunnel junctions which showed room temperature Coulomb oscillation with periods of 1 V [336]. The Coulomb blockade effects were observed in a metal-on-tube metal/multiwall carbon nanotube/metal structure (metal = Ti or Pt/Au) [337], and a multiwalled carbon nanotube between two gold electrodes at the SiO_2 surface [338]. Electrical transport properties of intramolecular *p*–*n*–*p* junctions formed on individual semiconducting carbon nanotubes were reported, and well-defined and highly reproducible SETs with much smaller size than the geometrical length of the nanotube were obtained [339]. Single-electron tunneling through amorphous carbon dots array was reported [340].

5.4. Molecular Devices

A molecular device based on single molecule (Fig. 11) [341–344] or molecular strands [345–347] has been advancing rapidly. The possibility of atom/molecule switching devices, atom relay transistors (ART), and molecular single-electron switching with total dimensions of a few mm and an operation speed of more than terahertz (10^{12}) to supersede the present MOS devices and to establish the new era of atom electronics were discussed [348], and devices based on ART will be important in switching speed performance [349]. Quantum computing devices implemented into the same substrate as the widely used LSI circuits are discussed [350]. A variety of molecular organic semiconductors have been investigated for electronic applications [351].

5.5. Metallic Devices

Single-electron memory utilizing a Schottky in-plane gate quantum wire transistor with nano-Schottky metal dots whose position and size can be precisely controlled was

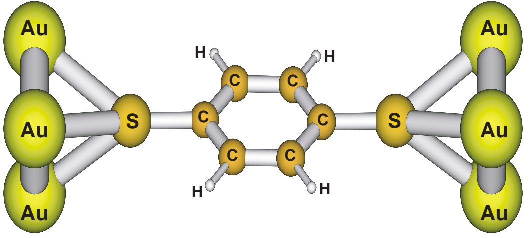


Figure 11. Structure of a single molecular device connected to two atomically sharp gold electron reservoirs. Reprinted with permission from [408], Y. Luo et al., *J. Chem. Phys.* 117, 10283 (2002). © 2002, American Institute of Physics.

realized [352]. The offset charge noise in metallic SET devices fabricated on dielectric substrates was experimentally studied [353, 354]. Background charge rearrangements in metallic SETs were modeled in two-level tunneling systems as a Poisson process [355]. A SET was made from a cadmium selenide nanocrystal [356].

Metal-based SETs were fabricated by the step-edge cutoff process [357]. An aluminum SET was bonded to an InP HEMT for high frequency applications [358]. There were multilevel SET devices with Au–SiO–Al and Al–AlO_x–SiO–Al overlap capacitors [359]. And a self-aligned metallic SETT was fabricated on a separated-by-implanted-oxygen substrate, where an array of 10–50 gold islands of 1–3 nm diameter was isolated between source and drain electrodes on a silicon nanowire [360]. Fabrication of metallic SETTs has been developed using different lithography steps for the preparation of the different layers [361]. A few tens of nm wide oxidized metal line was formed using a STM tip and an AFM cantilever as a cathode, which works as an energy barrier for an electron in SETs [362].

5.6. Ferromagnetic Devices

There was also ferromagnetic dot structure embedded in split-gate quantum wires on AlGaAs/InGaAs/GaAs pseudomorphic HEMT by two-step surface modification with STM [363].

6. OTHER SINGLE-ELECTRON TRANSISTOR APPLICATIONS

Besides circuit applications, the status of the understanding of SET devices with respect to their applicability in metrology was reviewed [364], and the study of SET for metrological application was conducted [365]. Electron pumps consisting of two Coulomb blockade devices have been realized using highly doped SOI material [366]. The behavior of a single-electron pump operated by sinusoidal signals was analyzed [367]. SETs were studied as far-infrared photon detectors [368]. A modified rf-SET technique was proposed to measure the transmission of the rf signal, having advantages such as simpler circuits and clear frequency resonance as compared with the reflection measurement by rf-SETs [369]. Charge sensitivity of rf-SETs was discussed [370]. Constructed with highly doped silicon nanowires, an electron detector can recognize the presence or absence of a single excess electron on a storage node at a temperature of 4.2 K [371]. Quantum dot structures measuring Hamming distance for associative memories were proposed and analyzed [372]. A scanned probe microscope capable of mapping static electric fields and charges with 100-nm spatial resolution and a charge sensitivity of a small fraction of an electron was reported using SETs [373]. Aluminum SETs were used to measure, with extremely high sensitivity, the fluctuation of charge in semiconductor quantum dots [374]. A SET has been used to detect chemical potential variations of the 2DEG [375, 376]. Si nanocrystals in close proximity (1.5–4.5 nm) of a transistor channel lead to plateaus in

Table 1. SETs based on group III–V materials, fabrication process, device characteristics (CB, temperature, and feature size), and reported major applications.

Structure	Heterosystem	Fabrication method	CB temperature and feature size	Application
SET island from 2DEG	GaAs/AlGaAs GaAs/InGaAs/ AlGaAs	multiple and split gates [139–144]	50–100 nm, CB 100 K	inverter operation at 1.6 K and logic transfer gain 1.3 [141], rf-SET [187], submillimeter-wave detector [188]
	GaAs InGaAs InP	etching [145–147], EB lithography and RIE [148, 149], selective MBE [150], AFM [151–153]	cross section 80 × 10 nm^2 [145, 146], lateral 50 nm [148], CB 100 K	
	GaAs/AlGaAs GaAs/InGaAs/ AlGaAs	point contact [154, 155]	room temperature CB	room temperature quantum function bipolar transistor [186]
GaAs and InAs self-formed nanostructures	GaAs GaAs/AlGaAs InAs/AlGaSb	selective area MOVPE [138], patterned AFM [156, 157] and MBE [158–160], Ga focused ion beam implantation [167], local oxidation [168], selective area MOVPE [172], AFM and STM nanooxidation [175]		
Traps	GaAs/AlGaAs	standard HEMT [176, 177]		random telegraph signal [177]

Table 2. SETs based on group IV materials.

Structure	Heterosystem	Fabrication method	CB temperature and feature size	Application
SET island from 2DEG in MOSFET conduction channel	Si MOSFET, SOI	reducing MOSFET size [193], multiple gates [200–211]		SET-MOSFET for multivalue memory [289, 290, 300], SET inverter [291–293], Schmitt trigger [295], few- electron memory with write time of 10 ns and voltage less than 5 V at 4.2 K to 45 K [297], 3 × 3-bit memory cell array [298], XOR function [301], electron detector [364], electrometer [374]
		EB lithography [213], RIE [217], direct EM irradiation [219], anisotropic etching [227, 233], local oxidation [228] and pattern dependent oxidation [230], optical lithography [234]	CB 170 K and drain bias 80 mV [217], CB room temperature [223, 224, 226], sub-10 nm [229]	
		point contact	CB 25 K and 5.3 nm [236], 259 meV charging energy and dot diameter 4.4 nm [237]	
Nanocrystals	Si, SiO_2, and SOI	laser recrystallization [239], EB lithography [242, 248, 249], high-frequency plasma decomposition [250], EB direct writing [251]	dot diameter 8 nm [247], source–drain separation 30 nm [250], hole CB 200 K [260], pattern-dependent oxidation [262], annealing thin SiO_x ($x < 2$) film [265]	room temperature SET memory [240–243, 245, 246], nonvolatile memory [256], EEPROM [311]
Dual-gate and nanocrystal	Si, SiO_2, and SOI	dual-gate and selective area MOVPE [269], remote plasma enhanced chemical vapor deposition [271]	1.5×10^{-10} cm^2 [271]	
Impurity and traps	Si/SiO_2 interface states	Si MOSFET [272–274]	submicrometer [272], nanometer [274]	

threshold voltage at low temperatures due to single-electron effects [248]. The process of evolving from the few-electron regime to the several-electron regime, and then the many-electron regime, was studied in an artificial-atom vertical SET [377]. A SET was used to measure the superconducting energy gap as a function of the external magnetic field for various samples of mesoscopic sizes [378]. The influence of electron-beam writing on the creation of charge trapping centers which cause $1/f$ noise in SETs was discussed [379]. Single-electron tunneling transistors were used as electrometers to verify proper switching of the quantum dot cellular automata cell [380]. A SET electrometer was used to measure the destruction of charge quantization in a single-electron box [381]. Phase-coherent transport was studied in T-shaped quantum dot transistor [382].

7. FUNDAMENTAL PHYSICS

In addition to the device-application-orientated research, fundamental physics investigation utilizing SET devices has also been extensive. Transport in silicon/germanium nanostructures was studied [383]. Strain caused by inserting $Si_{1-x}Ge_x$ can significantly improve the device performance; it is also compatible with existing Si industrial technology [384]. Cotunneling thermopower of SETs was studied [385]. Electron–electron interaction to the high energy tail of the electron energy distribution was studied in nano-MOSFETs [386]. Spring constant and damping constant tuning of nanomechanical resonators using a SET was reported [387]. The lifetime of the emission of a single electron stored in a nanocrystalline Si dot has been studied [388]. Quantum chaos observable in both open and closed systems realized in $GaAs/Al_x/Ga_{1-x}As$ microstructures was reviewed [389]. Dynamical offset charges in SETs were investigated [390]. Kondo physics in the SET with ac driving was studied using a time-dependent Anderson Hamiltonian [391], and measurements of the Kondo effect on SETs were reported [392], reviewed, and compared with theoretical analysis [393]. There were works on the temperature dependence of $k_B T_c$ noise in Coulomb blockade regime [394], quantum storage effects in n-AlGaAs/GaAs heterojunction FETs with embedded InAs QDs between the gate and 2DEG [395] (similar work was reported by [396]), Fano resonances in electronic transport through a SET [397], spin-dependent Coulomb blockade in a SOI-based SET [398], single-spin measurement using SETs to probe two-electron systems [399], and single-electron–hole recombination and photon-pumped current in an asymmetric Si SET [400]. Temperature and voltage dependence of the cotunneling current were analyzed in comparison with inelastic cotunneling theory [401]. The electron–phonon coupling in SETs was studied [402]. Magnetotransport properties of an AlGaAs/GaAs double-barrier resonant tunneling transistor with sub-μm lateral dimensions (where lateral confinement causes quasi-zero-dimensional quantization) were studied [403]. Photocurrent generation in single-electron tunneling transistors was studied theoretically [404]. Confinement enhancement induced by the charge distributed in the floating dot on top of a channel quantum dot in stacked quantum dot transistors was carefully examined [405]. The resistively coupled SET was studied including the Nyquist noise of the coupling resistance [406]. The displacement current in SET was calculated [407].

8. SUMMARY

The rapid advancing information technology demands reductions in the chip size and increases in the density of components per IC. The chip component density has been following Moore's law and doubled roughly every other year over the last three decades. A most prominent discrete component that already reaches nanoscale is the single-electron transistor, which to a large degree functions like a conventional transistor but is so small that it handles a single electron at a time.

The SET has been extensively studied. Its current research and development are largely concentrated on the device design and optimization. There are basically two types of quantum dots in SETs. The first type of quantum dot is defined by either gates or etching, with the advantage of position control and the disadvantage of large dot size. Small dot size is required for high temperature operation. The second type of quantum dot utilizes the self-assembling process of crystalization of one semiconductor on a different one (e.g., Si nanocrystals on SiO_2 and InAs quantum dots on GaAs substrate). The dot size is very small, but the dots are randomly positioned. Precise control over the dot spatial position is required for IC fabrication.

Tables 1 and 2 list details of important III–V and VI semiconductor materials/heterostructures used for fabricating nanoscale transistors, fabrication processes, Coulomb blockade temperature and feature size, and published major applications of these devices.

GLOSSARY

Coulomb blockade When the spatial area that contains the electron charges is small enough, an extra electron charge added to the area significantly increases the Coulomb potential energy of the area so that this area becomes a classical-forbidden region for other electron charges (blockaded).

Moore's law The chip component density of an IC roughly doubles every other year.

Nanofabrication technology To produce low-dimensional feature size in the nanoscale. Molecular beam epitaxy is one example that can grow one-dimensional structure with atomic monolayer (about 0.2 nm) feature size. Current quantum dot self-assembly can produce a three-dimensional quantum dot artificial atom with a volume of 1000 nm^3.

Scale of integration The complexity of a logic system is represented by its component density. A circuit containing between 100 and 1000 logic gates is classified as a large-scale integrated circuit; a circuit with more than 1000 gates (up to 100,000 or more) is classified as a very-large-scale integrated circuit. Processes that produce more than 10^6 devices on an IC are classified as ultra-large-scale integrated circuits.

Single-electron transistor A transistor that is small enough to handle one electron at a time.

REFERENCES

1. "Handbook of Advanced Electronics and Photonics Materials and Devices" (H. S. Nalwa, Ed.), Academic Press, San Diego, 2001.
2. Internal Technology Roadmap for Semiconductors, 2001 Edition.
3. M. N. Yoder, in "Proceedings of the Fourteenth Biennial University/Government/Industry Microelectronics Symposium," 17–20 June 2001, Richmond, VA, 2001, pp. 2–7.
4. H. Veendrick, "Deep-Submircon CMOS IC," p. vii. Kluwer Academic, Dordrecht, 2000.
5. J. Itoh, *Appl. Surf. Sci.* 111, 194 (1997).
6. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, *IEEE J. Solid-State Circuits* SC9, 256 (1974).
7. G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, E. Ganin, S. Rishton, D. S. Zicherman, H. Schmid, M. R. Polcari, H. Y. Ng, P. J. Restle, T. H. P. Chang, and R. H. Dennard, *IEEE Electron Device Lett.* EDL-8, 463 (1987).
8. T. Hishimoto, Y. Sudoh, H. Kurino, A. Narai, S. Yokoyama, Y. Horiike, and M. Koyanagi, in "Ext. Abs. Int. Conf. Solid State Devices Materials," August 1992, p. 490.
9. M. Ono, M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro, and H. Iwai, *IEEE Trans. Electron Devices* 42, 1822 (1995).
10. H. Kawaura, T. Sakamoto, T. Baba, Y. Ochiai, J. Fujita, S. Matsui, and J. Sone, *IEEE Electron Device Lett.* 19, 74 (1998).
11. W. Langen, H. Hardtdegen, H. Luth, and P. Kordos, in "Compound Semiconductors 1994, Proceedings of the 21st International Symposium," pp. 183–188. IOP Publishing, Bristol, UK, 1995.
12. D. Behammer, L. Vescan, R. Loo, J. Moers, A. Muck, H. Luth, and T. Grabolla, *Electron. Lett.* 32, 406 (1996).
13. W. Langen, L. Vescan, R. Loo, H. Luth, and P. Kordos, *Appl. Surface Sci.* 102, 252 (1996).
14. D. Klaes, J. Moers, A. Tonnesmann, S. Wickenhauser, L. Vescan, M. Marso, T. Grabolla, M. Grimm, and H. Luth, *Thin Solid Films* 336, 306 (1998).
15. J. Moers, D. Klaes, A. Tonnesmann, L. Vescan, S. Wickenhauser, M. Marso, P. Kordos, and H. Luth, *Electron. Lett.* 35, 239 (1999).
16. J. Moers, D. Klaes, A. Tonnesmann, V. Vescan, S. Wickenhauser, T. Grabolla, M. Marso, P. Kordos, and H. Luth, *Solid State Electron.* 43, 529 (1999).
17. R. T. Bate, *Sci. Am.* 258, 78 (1988).
18. T. Sugano, *IEICE Trans. Electron.* E76-C, 1029 (1993).
19. Y. Arakawa, *Rev. Laser Eng.* 26, 615 (1998).
20. G. Yusa and H. Sakaki, *Appl. Phys. Lett.* 70, 345 (1997); G. Yusa and H. Sakaki, *Physica E* 2, 734 (1998); G. Yusa and H. Sakaki, *Superlattices Microstruct.* 25, 247 (1999).
21. P. Marquardt and G. Nimtz, in "Proceedings of the First International Conference on Electronic Materials. New Materials and New Physical Phenomena for Electronics of the 21st Century (ICEM '88)," 13–15 June 1988, Tokyo, Japan (T. Sugano, R. P. H. Chang, H. Kamimura, I. Hayashi, and T. Kamiya, Eds.), p. 193.
22. S. Okayama, S. Kimura, A. Asai, and S. Maeda, *Microelectron. Eng.* 30, 17 (1996).
23. K. von Klitzing, *Phys. Scr.* T68, 21 (1996).
24. K. K. Likharev, *IEEE Trans. Magn.* MAG-23, 1142 (1987).
25. J. Randall, G. Frazier, A. Seabaugh, and T. Broekaert, *Microelectron. Eng.* 32, 15 (1996).
26. H. Hasegawa, in "1997 Advanced Workshop on Frontiers in Electronics, WOFE '97 Proceedings" 6–11 January 1997, Puerto de la Cruz, Spain (G. S. Pomrenke, F. Schuermeyer, M. Shur, and J. Xu, Eds.), p. 101.
27. J. Gautier, *Microelectron. Eng.* 39, 263 (1997).
28. L. Dobrzanski, in "12th International Conference on Microwaves and Radar, MIKON-98 Conference Proceedings" 20–22 May 1998, Krakow, Poland, Vol. 3, p. 747.
29. H. L. Hartnagel, in "1998 URSI International Symposium on Signals, Systems, and Electronics, Conference Proceedings," 29 September–2 October 1998, Pisa, Italy, p. 63.
30. J. Camargo da Costa, M. Goossens, J. Hoekstra, C. J. M. Verhoeven, and A. H. M. van Roermund, in "Proceedings of ICMP'99 International Conference on Microelectronics and Packaging," 10–14 August 1998, Curitiba, Brazil (I. J. Chueiri and C. A. dos Reis Filho, Eds.), Vol. 2, pp. 461–473.
31. S. Tiwari, F. Rana, A. Kumar, J. J. Welser, and C. T. Black, in "Proceedings of the Fifth International Symposium on Quantum Confinement: Nanostructures," 2–5 November 1998, Boston, MA, (M. Cahay, D. J. Lockwood, J. P. Leburton, and S. Bandyopadhyay, Eds.), pp. 507–528.
32. L. Geppert, *IEEE Spectrosc* 37, 46 (2000).
33. J. R. Tucker and T.-C. Shen, *Int. J. Circuit Theory Appl.* 28, 553 (2000).
34. A. T. Tilke, F. C. Simmel, R. H. Blick, H. Lorenz, and J. P. Kotthaus, *Progr. Quantum Electron.* 25, 97 (2001).
35. L. Risch, *Mater. Sci. Eng. C* 19, 363 (2002).
36. C. Wasshuber, in "Proceedings of the 2002 Third International Symposium on Quality Electronic Design," 18–21 March 2002, San Jose, CA, pp. 502–507.
37. M. Cahay and S. Bandyopadhyay, *IEEE Potentials* 12, 18 (1993).
38. A. Toriumi, *J. Vac. Sci. Technol. B* 14, 4020 (1996).
39. C. Weisbuch, *Phys. Scr. T* 68, 102 (1996).
40. S. Ahmad, *Def. Sci. J.* 48, 45 (1998).
41. K. Nikolic, D. Berzon, and M. Forshaw, *Nanotechnology* 12, 38 (2001).
42. M. T. Lutwyche and Y. Wada, *J. Appl. Phys.* 75, 3654 (1994).
43. T. J. Thornton, in "Quantum Transport in Ultrasmall Devices," 17–30 July 1994, Il Ciocco, Italy (D. K. Ferry, H. L. Grubin, C. Jacoboni, and A.-P. Jauho, Eds.), pp. 141–169.
44. M. T. Bohr, *IEEE Trans. Nanotechnol.* 1, 56 (2002).
45. A. N. Korotkov, *Int. J. Electron.* 86, 511 (1999).
46. V. Bubanja, K. Matsumoto, and Y. Gotoh, *Jpn. J. Appl. Phys. 1* 40, 87 (2001).
47. T. Hiramoto, H. Ishikuro, and H. Majima, *J. Photopolym. Sci. Technol.* 12, 417 (1999).
48. S. Tiwari, J. J. Welser, and F. Rana, in "1997 Symposium on VLSI Technology: Digest of Technical Papers," 10–12 June 1997, Kyoto, Japan, pp. 133–134.
49. H. Ahmed, *J. Vac. Sci. Technol. B* 15, 2101 (1997).
50. NTT develops prototype nanostructure silicon single-electron transistor operating near room temperature, *Microelectron. J.* 26, VIII (1995).
51. M. Khoury, A. Gunther, D. P. Pivin, Jr., M. J. Rack, and D. K. Ferry, *Jpn. J. Appl. Phys. 1* 38, 469 (1999); D. K. Ferry, M. Khoury, C. Gerousis, M. J. Rack, A. Gunther, and S. M. Goodnick, *Physica E* 9, 69 (2001).
52. E.-H. Lee and K. Park, *Mater. Sci. Eng. B* 74, 1 (2000).
53. E. Kapetanakis, P. Normand, D. Tsoukalas, and K. Beltsios, *Appl. Phys. Lett.* 80, 2794 (2002).
54. G. Johansson, A. Kack, and A. Wendin, *Phys. Rev. Lett.* 88, 046802/1-4 (2002).
55. I. Kim, S. Han, K. Han, J. Lee, and H. Shin, *Jpn. J. Appl. Phys. 1* 40, 447 (2001).
56. A. Kumar, J. J. Welser, S. Tiwari, F. Rana, and K. K. Chan, *Mater. Res. Soc. Symp. Proc.* 638, F2.4.1-8 (2001).
57. H. Hasegawa and S. Kasai, *Physica E* 11, 149 (2001); S. Kasai and H. Hasegawa, in "Device Research Conference Digest," 25–27 June 2001, Notre Dame, IN, pp. 131–132.
58. J. Koga, R. Ohba, K. Uchida, and A. Toriumi, in "International Electron Devices Meeting, Technical Digest," 2–5 December 2001, Washington, DC, pp. 7.1.1–4.
59. N. M. Zimmerman, W. H. Huber, A. Fujiwara, and Y. Takahashi, *Appl. Phys. Lett.* 79, 3188 (2001).
60. A. Fujiwara and Y. Takahashi, in "International Electron Devices Meeting 2000, Technical Digest," 10–13 December 2000, San Francisco, CA, pp. 866–868.

61. http://www.businessweek.com/bwdaily/jun2001/nf2001068_893.htm.
62. http://www.pcworld.com/news/article/0,aid,52347,00.asp.
63. H. Ahmed, *Microelectron. Eng.* 41–42, 15 (1998).
64. M. Bayer, *Nature* 418, 8 August (2002).
65. M. A. Kastner, *Phys. Today* 46, 24 (1993).
66. D. D. Smith, *Int. J. High Speed Electron. Syst.* 9, 165 (1998).
67. N. M. Johnson, D. K. Biegelsen, H. C. Tuan, M. D. Moyer, and L. E. Fennell, *IEEE Electron Device Lett.* EDL-3, 369 (1982).
68. B.-Y. Tsaur, M. W. Geis, J. C. C. Fan, D. J. Silversmith, and R. W. Mountain, *Appl. Phys. Lett.* 39, 909 (1981); M. W. Geis, H. I. Smith, B.-Y. Tsaur, J. C. C. Fan, E. W. Maby, and D. A. Antoniadis, *Appl. Phys. Lett.* 40, 158 (1982); M. W. Geis, C. K. Chen, R. W. Mountain, N. P. Economou, W. T. Lindley, and P. L. Hower, *IEEE Electron Device Lett.* EDL-7, 41 (1986).
69. E. Kasper, *Wiss. Ber. AEG-Telefunken* 52, 147 (1979).
70. H. Luth, *Acta Phys. Polon. A* 90, 667 (1996); Semiconductor nanostructures for device applications, in "Physics of Semiconductor Devices," Vol. 1, pp. 18–25. Narosa, Delhi, 1998; H. Luth, *Appl. Surface Sci.* 130–132, 855 (1998).
71. M. Komuro, *Mater. Res. Soc. Symp. Proc.* 584, 305 (2000).
72. Y. V. Nastaushev, T. Gavrilova, M. Kachanova, L. Nenasheva, V. Kolosanov, O. V. Naumova, V. P. Popov, and A. L. Aseev, *Mater. Sci. Eng. C* 19, 189 (2002).
73. S. Luscher, R. Held, A. Fuhrer, T. Heinzel, K. Ensslin, M. Bichler, and W. Wegscheider, *Mater. Sci. Eng. C* 15, 153 (2001).
74. Y. Wada, *Microelectron. J.* 29, 601 (1998).
75. Y. Noguchi, M. Iwamoto, T. Kubota, and S. Mashiko, *Jpn. J. Appl. Phys. 1* 41, 2749 (2002).
76. J. Appenzeller, R. Martel, P. Solomon, K. Chan, Ph. Avouris, J. Knoch, J. Benedict, M. Tanner, S. Thomas, K. L. Wang, and J. A. del Alamo, *Appl. Phys. Lett.* 77, 298 (2000).
77. K. E. Ismail, P. F. Bagwell, T. P. Orlando, D. A. Antoniadis, and H. I. Smith, *Proc. IEEE* 79, 1106 (1991).
78. V. V. Mitin, Z. S. Gribnikov, and A. N. Korshak, "Physics of Semiconductor Devices," Vol. 1, pp. 134–141. Narosa, Delhi, 1998.
79. C. Hamaguchi, T. Mori, T. Wada, K. Terashima, K. Taniguchi, K. Miyatsuji, and H. Hihara, *Microelectron. Eng.* 2, 34 (1984).
80. K. Seo, Y. Nashimoto, W. P. Hong, M. Elta, P. K. Bhattachary, and K. R. Gleason, in "Proceedings of the IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits," 29–31 July 1985, Ithaca, NY, pp. 102–110.
81. R. A. Kiehl, *IEEE Trans. Microw. Theory Tech.* 37, 1304 (1989).
82. C. Hamaguchi, K. Miyatsuji, and H. Hihara, *Jpn. J. Appl. Phys. 2* 23, L132 (1984).
83. A. Kastalsky, J. H. Abeles, and R. F. Leheny, *Appl. Phys. Lett.* 50, 708 (1987).
84. U. K. Mishra and A. S. Brown, in "GaAs IC Symposium, IEEE Gallium Arsenide Integrated Circuit Symposium, 10th Annual GaAs IC Symposium Technical Digest 1988," 6–9 November 1988, Nashville, TN, pp. 97–100; U. K. Mishra, A. S. Brown, M. J. Delaney, P. T. Greiling, and C. F. Krumm, *IEEE Trans. Microw. Theory Tech.* 37, 1279 (1989).
85. L. P. Ramberg and T. Ishibashi, *J. Appl. Phys.* 63, 809 (1988).
86. J. H. Luscombe and W. R. Frensley, *Nanotechnology* 1, 131 (1990).
87. L. I. Glazman, *J. Low Temp. Phys.* 118, 247 (2000).
88. H. Luth, *Acta Phys. Polon. A* 90, 667 (1996).
89. Y. Fu, M. Karlsteen, and M. Willander, *Superlattices Microstruct.* 22, 405 (1997).
90. Y. Fu and M. Willander, *Superlattices Microstruct.* 26, 289 (1999).
91. H. Fukui, M. Fujishima, and K. Hoh, *Jpn. J. Appl. Phys. 1* 36, 4147 (1997).
92. J. Guo, J.-F. Jiang, and Q.-Y. Cai, *Acta Electron. Sinica* 28, 52 (2000).
93. H.-O. Muller, D. A. Williams, H. Mizuta, Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, *Physica B* 272, 85 (1999).
94. F. Rana, S. Tiwari, and J. J. Welser, *Superlattices Microstruct.* 23, 757 (1998).
95. M. B. A. Jalil and M. Wagner, *J. Appl. Phys.* 88, 869 (2000).
96. K. Natori and N. Sano, *IEICE Trans. Electron.* E82-C, 1599 (1999).
97. K. Natori, T. Uehara, and N. Sano, *Jpn. J. Appl. Phys. 1* 39, 2550 (2000).
98. S. Mahapatra, A. M. Ionescu, and K. Banerjee, *IEEE Electron Device Lett.* 23, 366 (2002).
99. M. Saitoh and T. Hiramoto, *IEICE Trans. Electron.* E84-C, 1071 (2001).
100. M. Shin, S. Lee, and K. W. Park, *J. Korean Phys. Soc.* 39, 581 (2001).
101. M. Hanggi and L. O. Chua, *IEEE Trans. Circuits Syst. I* 48, 1019 (2001).
102. I. I. Abramov and E. G. Novik, *Russ. Microelectron.* 29, 177 (2000); I. I. Abramov and E. G. Novik, *Tech. Phys. Lett.* 26, 726 (2000).
103. H. Nakashima and K. Uozumi, *Int. J. Electron.* 85, 145 (1998).
104. H. Nakashima and K. Uozumi, *J. Vac. Sci. Technol. B* 16, 1252 (1998).
105. H. Nakashima and K. Uozumi, *J. Phys., Condens. Matter* 10, 4291 (1998).
106. M. Tabe, Y. Terao, N. Asahi, and Y. Amemiya, *IEICE Trans. Electron.* E81-C, 36 (1998).
107. S. Bednarek, B. Szafran, and J. Adamowski, *Phys. Rev. B* 61, 4461 (2000).
108. Y. Tanaka and H. Akera, *J. Phys. Soc. Jpn.* 66, 15 (1997).
109. M. Willander, Y. Fu, and O. Nilsson, *Mater. Sci. Forum.* 297–298, 3 (1999).
110. A. Scholze, A. Schenk, and W. Fichtner, in "Proceedings of SISPADS'98 International Conference on Simulation of Semiconductor Processes and Devices," 2–4 September 1998 (K. D. Meyer and Biesemans, Eds.), Leuven, Belgium, pp. 203–206; A. Scholze, A. Schenk, and W. Fichtner, in "1999 International Conference on Simulation of Semiconductor Processes and Devices, SISPAD'99," 6–8 September 1999, Kyoto, Japan, pp. 171–174; A. Scholze, A. Schenk, and W. Fichtner, *IEICE Trans. Electron.* E83-C, 1242 (2000); A. Scholze, A. Schenk, and W. Fichtner, *IEEE Trans. Electron Devices* 47, 1811 (2000).
111. A. Scholze, A. Schenk, and W. Fichtner, in "1998 Sixth International Workshop on Computational Electronics, Extended Abstracts," 19–21 October 1998, Osaka, Japan, pp. 31–33.
112. S. Bednarek, B. Szafran, and J. Adamowski, *Proc. SPIE* 4413, 129 (2001).
113. G. Michalek and B. R. Bulka, *Acta Phys. Polon. A* 100, 431 (2001).
114. M. Shin, S. Lee, K. W. Park, and G.-H. Kim, *Solid State Commun.* 116, 527 (2000).
115. S. D. Lee, K. S. Park, J. W. Park, J. B. Choi, S.-R. E. Yang, K.-H. Yoo, J. Kim, S. I. Park, and K. T. Kim, *Phys. Rev. B* 62, R7735 (2000).
116. A. Matulis, J. O. Fjaerestad, and K. A. Chao, *Int. J. Mod. Phys. B* 11, 1035 (1997).
117. K. Hirose and N. S. Wingreen, *NEC Res. Dev.* 40, 419 (1999).
118. J. Konig, H. Schoeller, and G. Schon, *Phys. Rev. B* 58, 7882 (1998).
119. C. Wasshuber and H. Kosina, *VLSI Des.* 6, 35 (1998).
120. K. Uchida, K. Matsuzawa, J. Koga, R. Ohba, S. Takagi, and A. Toriumi, *Jpn. J. Appl. Phys. 1* 39, 2321 (2000).
121. X. Wang and W. Porod, *Superlattices Microstruct.* 28, 345 (2000); X. Wang and W. Porod, *VLSI Des.* 13, 189 (2001).
122. R. van de Haar, R. H. Klunder, and J. Hoekstra, in "ICECS 2001: Eighth IEEE International Conference on Electronics, Circuits and Systems," 2–5 September 2001, Malta, Vol. 3, pp. 1445–1448.
123. Z. Yu, R. W. Dutton, and R. A. Kiehl, in "Sixth International Workshop on Computational Electronics, Extended Abstracts," 19–21 October 1998, Osaka, Japan, pp. 222–229; Z. Yu, R. W. Dutton, and R. A. Kiehl, *IEEE Trans. Electron Devices* 47, 1819 (2000).

124. K. Shiraishi, M. Nagase, S. Horiguchi, and H. Kageshima, in "Microcrystalline and Nanocrystalline Semiconductors—1998," 30 November–3 December 1998, Boston, MA, pp. 533–538; K. Shiraishi, M. Nagase, S. Horiguchi, H. Kageshima, M. Uematsu, Y. Takahashi, and K. Murase, *Physica E* 7, 337 (2000).
125. C. Lee, *Semicond. Sci. Technol.* 13, A115 (1998); C. Lee, D. S. Choi, H. R. Park, C. S. Kim, and K. L. Wang, *J. Korean Phys. Soc.* 39, S442 (2001).
126. F. Wakaya, F. Yoshioka, H. Higurashi, S. Iwabuchi, Y. Nagaoka, and K. Gamo, *Jpn. J. Appl. Phys. 1* 38, 2812 (1999).
127. P. De Wolf, T. Clarysse, W. Vandervorst, L. Hellemans, P. Niedermann, and W. Hanni, *J. Vac. Sci. Technol. B* 16, 355 (1998).
128. D. Berman, N. B. Zhitenev, R. C. Ashoori, and M. Shayegan, *Phys. Rev. Lett.* 82, 161 (1999).
129. N. M. Zimmerman and M. W. Keller, *J. Appl. Phys.* 87, 8570 (2000).
130. M. Furlan, T. Heinzel, B. Jeanneret, S. V. Lotkhov, and K. S. Ensslin, *Europhys. Lett.* 49, 369 (2000).
131. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, and K. Seki, in "Proceedings of 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, pp. 22–23; T. Ishii, K. Yano, T. Sano, T. Mine, F. Murai, and K. Seki, in "International Electron Devices Meeting 1997, IEDM Technical Digest," 7–10 December 1997, Washington, DC, pp. 171–174; T. Ishii, T. Osabe, T. Mine, F. Murai, and K. Yano, in "International Electron Devices Meeting 2000, Technical Digest," 10–13 December 2000, San Francisco, CA, p. 305.
132. J. C. Da Costa, J. Hoekstra, M. J. Goossens, C. J. M. Verhoeven, and A. H. M. Van Roermund, *Analog Integr. Circuits Signal Process.* 24, 59 (2000).
133. X. Wang, V. P. Roychowdhury, and P. Balasingam, *Parallel Comput.* 22, 1931 (1997).
134. C. Wasshuber and H. Kosina, in "1996 International Conference on Simulation of Semiconductor Processes and Devices, SISPAD '96," 2–4 September 1996, Tokyo, Japan, p. 135.
135. D. H. Cobden, A. Savchenko, M. Pepper, N. K. Patel, D. A. Ritchie, J. E. F. Frost, and G. A. C. Jones, *Phys. Rev. Lett.* 69, 502 (1992).
136. C.-T. Liang, M. Y. Simmons, C. G. Smith, G.-H. Kim, D. A. Ritchie, and M. Pepper, *Jpn. J. Appl. Phys. 1* 40, 1936 (2001).
137. Y. Jin, B. Etienne, D. C. Glattli, C. Pasquier, and F. I. B. Williams, *Superlattices Microstruct.* 12, 367 (1992).
138. H. Sakaki, *Phys. Status Solidi B* 215, 291 (1999).
139. Z. Borsosfoldi, I. A. Larkin, M. Rahman, J. H. Davies, M. C. Holland, J. M. R. Weaver, A. R. Long, and J. G. Williamson, in "Proceedings of 22nd International Conference on the Physics of Semiconductors," 15–19 August 1994, Vancouver, BC (D. J. Lockwood, Ed.), Vol. 3, p. 1887.
140. S. Y. Chou and Y. Wang, in "Extended Abstracts of the 1992 International Conference on Solid State Devices and Materials," 26–28 August 1992, Tsukuba, Japan, p. 738; Y. Wang and S. Y. Chou, *J. Vac. Sci. Technol. B* 10, 2962 (1992).
141. H. Tomozawa, K. Jinushi, H. Okada, T. Hashizume, and H. Hasegawa, *Physica B* 227, 112 (1996); K. Jinushi, H. Okada, T. Hashizume, and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 35, 1132 (1996); S. Kasai, K. Jinushi, I. Tomozawa, and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 36, 1678 (1997); H. Hasegawa, T. Sato, H. Okada, K. Jinushi, S. Kasai, and Y. Satoh, *Appl. Surf. Sci.* 123–124, 335 (1998); H. Okada, H. Fujikura, T. Hashizume, and H. Hasegawa, *Solid-State Electron.* 42, 1419 (1998); Y. Satoh, S. Kasai, J. Jinushi, and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 37, 1584 (1998); Y. Satoh, H. Okada, K. Jinushi, H. Fujikura, and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 38, 410 (1999); S. Kasai, Y. Satoh, and H. Hasegawa, *Physica B* 272, 88 (1999); H. Hasegawa, H. Fujikura, and H. Okada, *Mater. Res. Soc. Bull.* 24, 25 (1999); H. Okada, T. Sato, K. Jinushi, and H. Hasegawa, *Microelectron. Eng.* 47, 285 (1999); S. Kasai, Y. Satoh, and H. Hasegawa, in "Proceedings of the 24th International Symposium on Compound Semiconductors," 22–26 August 1999, Berlin (K. H. Ploog, G. Trankle, and G. Weimann, Eds.), p. 219; S. Kasai and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 40, 2029 (2001).
142. H. Nakashima and K. Uozumi, *Jpn. J. Appl. Phys. 2* 36, L1315 (1997); F. Wakaya, K. Kitamura, S. Iwabuchi, and K. Gamo, *Jpn. J. Appl. Phys. 1* 38, 2470 (1999); F. Wakaya, S. Nakamichi, K. Gamo, S. Mandai, and S. Iwabuchi, *Superlattices Microstruct.* 27, 603 (2000); F. Wakaya, S. Mandai, S. Nakamichi, S. Iwabuchi, and K. Gamo, *Microelectron. Eng.* 53, 195 (2000).
143. Y. Pashkin, Y. Nakamura, and J. Tsai, *Jpn. J. Appl. Phys. 1* 38, 406 (1999).
144. F. Hofmann, T. Heinzel, D. A. Wharam, J. P. Kotthaus, G. Bohm, W. Klein, G. Trankle, and G. Weimann, *Phys. Rev. B* 51, 13872 (1995).
145. Y. Fu, W. Willander, and T. H. Wang, *Appl. Phys. Lett.* 78, 3705 (2001).
146. Y. Fu, T. H. Wang, and M. Willander, *J. Appl. Phys.* 89, 1759 (2001); Y. Fu, M. Willander, and T. H. Wang, *Appl. Phys. A* 74, 741 (2002).
147. H. Pothier, J. Weis, R. J. Haug, K. von Klitzing, and K. Ploog, *Appl. Phys. Lett.* 62, 3174 (1993).
148. T. H. Chang, K. A. Chen, C. H. Yang, M. J. Yang, and D. Park, *J. Vac. Sci. Technol. B* 18, 3493 (2000).
149. T. H. Wang and Y. Aoyagi, *Appl. Phys. Lett.* 78, 634 (2001).
150. H. Fujikura, Y. Hanada, T. Muranaka, and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 38, 421 (1999); H. Fujikura, T. Muranaka, and H. Hasegawa, *Microelectron. J.* 30, 397 (1999); T. Muranaka, H. Okada, H. Fujikura, and H. Hasegawa, *Microelectron. Eng.* 47, 201 (1999); T. Muranaka, H. Fujikura, and H. Hasegawa, in "Proceedings of the 26th International Symposium on Compound Semiconductors," 22–26 August 1999, Berlin (K. H. Ploog, G. Trankle, and G. Weimann, Eds.), p. 187; H. Fujikura, T. Muranaka, and H. Hasegawa, *Physica E* 7, 864 (2000).
151. H. W. Schumacher, U. F. Keyser, U. Zeitler, R. J. Haug, and K. Eberl, *Appl. Phys. Lett.* 75, 1107 (1999); H. W. Schumacher, U. F. Keyser, U. Zeitler, R. J. Haug, and K. Eberl, *Physica E* 6, 860 (2000).
152. L. Montelius, T. Junno, S.-B. Carlsson, M. H. Magnusson, K. Deppert, H. Xu, and L. Samuelson, *Microelectron. Reliab.* 38, 943 (1998).
153. Y. Okada, S. Amano, M. Kawabe, B. N. Shimbo, and J. S. Harris, Jr., in "Proceedings of the IEEE 24th International Symposium on Compound Semiconductors," 8–11 September 1997, San Diego, CA, p. 577; Y. Okada, S. Amano, M. Kawabe, B. N. Shimbo, and J. S. Harris, Jr., *J. Appl. Phys.* 83, 1844 (1998); Y. Okada, S. Amano, Y. Iuchi, M. Kawabe, and J. S. Harris, Jr., *Electron. Lett.* 34, 1262 (1998).
154. K. Nakazato, T. J. Thornton, J. White, and H. Ahmed, *Appl. Phys. Lett.* 61, 3145 (1992).
155. H. Ishikuro and T. Hiramoto, *Appl. Phys. Lett.* 71, 3691 (1997).
156. U. F. Keyser, H. W. Schumacher, U. Zeitler, R. J. Haug, and K. Eberl, *Phys. Status Solidi B* 224, 681 (2001).
157. S. Luscher, A. Fuhrer, R. Held, T. Heinzel, K. Ensslin, and W. Wegscheider, *Appl. Phys. Lett.* 75, 2452 (1999); S. Luscher, A. Fuhrer, R. Held, T. Heinzel, K. Ensslin, W. Wegscheider, and M. Bichler, *J. Low Temp. Phys.* 118, 333 (2000); S. Luscher, A. Fuhrer, R. Held, T. Heinzel, K. Ensslin, M. Bichler, and W. Wegscheider, *Microelectron. J.* 33, 319 (2002).
158. M. Dilger, R. J. Haug, K. Eberl, and K. von Klitzing, *Semicond. Sci. Technol.* 11, 1493 (1996); M. Dilger, R. J. Haug, K. Eberl, A. Kurtenbach, Y. Kershaw, and K. von Klitzing, *Appl. Phys. Lett.* 68, 3132 (1996); R. J. Haug, M. Dilger, T. Schmidt, R. H. Blick, K. v. Klitzing, and K. Eberl, *Physica B* 227, 82 (1996); M. Dilger, K. Eberl, R. J. Haug, and K. von Klitzing, *Superlattices Microstruct.* 21, 533 (1997).
159. J. Christen, F. Bertram, M. Dilger, R. J. Haug, K. Eberl, K. von Klitzing, and D. Bimberg, in "Proceedings of 23rd International

Conference on the Physics of Semiconductors," 21–26 July 1996, Berlin (M. Scheffler and R. Zimmermann, Eds.), Vol. 4, p. 3315; F. Bertram, M. Lipinski, T. Riemann, D. Rudloff, J. Christen, P. Veit, R. Clos, and K. Eberl, *Physica E* 7, 363 (2000).

160. K. Shiralagi, R. Zhang, and R. Tsui, *J. Cryst. Growth* 201, 1209 (1999).
161. S.-I. Osako, T. Sugihara, Y. Yamamoto, T. Maemoto, S. Sasa, M. Inoue, and C. Hamaguchi, *Semicond. Sci. Technol.* 11, 571 (1996).
162. D. G. Austing, T. Honda, Y. Tokura, and S. Tarucha, *Jpn. J. Appl. Phys. 1* 34, 1320 (1995); S. Tarucha, D. G. Austing, and T. Honda, *Superlattices Microstruct.* 18, 121 (1995); D. G. Austing, T. Honda, and S. Tarucha, *Semicond. Sci. Technol.* 11, 388 (1996); D. G. Austing, T. Honda, and S. Tarucha, *Solid-State Electron.* 40, 237 (1996); D. G. Austing, T. Honda, and S. Tarucha, *Jpn. J. Appl. Phys. 1* 36, 4151 (1997); D. G. Austing, T. Honda, and S. Tarucha, *Jpn. J. Appl. Phys. 1* 36, 1667 (1997); D. G. Austing, T. Honda, and S. Tarucha, *Physica E* 2, 583 (1998); T. Honda, S. Tarucha, and D. G. Austing, *IEICE Trans. Electron.* E81-C, 2 (1998); D. G. Austing, Y. Tokura, S. Tarucha, T. H. Oosterkamp, J. W. Janssen, M. W. S. Danoesastro, and J. P. Kouwenhoven, *Physica E* 6, 358 (2000).
163. M. Ciorga, M. Pioro-Ladriere, P. Zawadzki, P. Hawrylak, and A. S. Sachrajda, *Appl. Phys. Lett.* 80, 2177 (2002).
164. H. Jiang and J. Singh, *IEEE J. Quantum Electron.* 34, 1188 (1998).
165. S. K. Jung, S. W. Hwang, D. Ahn, J. H. Park, Y. Kim, and E. K. Kim, *Physica E* 7, 430 (2000).
166. K. Yoh and H. Kazama, *Physica E* 7, 440 (2000).
167. T. Fujisawa, T. Bever, Y. Hirayama, and S. Tarucha, *J. Vac. Sci. Technol. B* 12, 3755 (1994); T. Fujisawa and S. Tarucha, *Appl. Phys. Lett.* 68, 526 (1996).
168. U. F. Keyser, H. W. Schumacher, U. Zeitler, R. J. Haug, and K. Eberl, *Appl. Phys. Lett.* 76, 457 (2000).
169. K. H. Schmidt, M. Versen, C. Bock, U. Kunze, D. Reuter, and A. D. Wieck, *Physica E* 7, 425 (2000).
170. P. Baumgartner, W. Wegscheider, M. Bichler, G. Schedelbeck, R. Neumann, and G. Abstreiter, *Appl. Phys. Lett.* 70, 2135 (1997); P. Baumgartner, W. Wegscheider, M. Bichler, G. Groos, and G. Abstreiter, *Physica E* 2, 441 (1998).
171. Y. Matsumoto, T. Hanajiri, T. Toyabe, and T. Sugano, in "International Electron Devices Meeting, Technical Digest," 8–11 December 1996, San Francisco, CA, p. 429; Y. Matsumoto, T. Hanajiri, T. Toyabe, and T. Sugano, *Jpn. J. Appl. Phys. 1* 35, 1126 (1996).
172. K. Kumakura, J. Motohisa, and T. Fukui, *Physica E* 2, 809 (1998); K. Kumakura, J. Motohisa, and T. Fukui, *Solid-State Electron.* 42, 1227 (1998); T. Fukui, F. Nakajima, K. Kumakura, and J. Motohisa, *Bull. Mater. Sci.* 22, 531 (1999); F. Nakajima, K. Kumakura, J. Motohisa, and T. Fukui, *Jpn. J. Appl. Phys. 1* 38, 415 (1999); F. Nakajima, J. Motohisa, and T. Fukui, *Appl. Surf. Sci.* 162–163, 650 (2000); J. Motohisa, T. Terasawa, T. Kusuhara, F. Nakajima, and T. Fukui, in "Proceedings of the 2001 International Conference on Indium Phosphide and Related Materials, 13th IPRM," 14–18 May 2001, Nara, Japan, p. 370; F. Nakajima, Y. Ogasawara, J. Motohisa, and T. Fukui, *J. Appl. Phys.* 90, 2606 (2001).
173. S. Sasa, T. Ikeda, K. Anjiki, and M. Inoue, *Jpn. J. Appl. Phys. 1* 38, 480 (1999); S. Sasa, A. Ohya, M. Yodogawa, and M. Inoue, in "Proceedings of the 2000 International Conference on Indium Phosphide and Related Materials," 14–18 May 2000, Williamsburg, VA, p. 205; S. Sasa, S. Yodogawa, A. Ohya, and M. Inoue, *Jpn. J. Appl. Phys. 1* 40, 2026 (2001).
174. K. Yoh and H. Kazama, *Physica E* 7, 440 (2000).
175. K. Matsumoto, M. Ishii, K. Segawa, Y. Oka, B. J. Vartanian, and J. S. Harris, in "53rd Annual Device Research Conference Digest," 19–21 June 1995, Charlottesville, VA, p. 46; K. Matsumoto, *IEICE Trans. Electron.* E79-C, 1509 (1996); K. Matsumoto, M. Ishii, K. Segawa, Y. Oka, B. J. Vartanian, and J. S. Harris, *Appl. Phys. Lett.* 68, 34 (1996); K. Matsumoto, *Proc. IEEE* 85, 612 (1997); K. Matsumoto, Y. Gotoh, T. Maeda, J. A. Dagata, and J. S. Harris, in "International Electron Devices Meeting 1998, Technical Digest," 6–9 December 1998, San Francisco, CA, p. 449; S. Miyakawa, R. Kim, J.-I. Shirakashi, K. Taniguchi, K. Matsumoto, and Y. Kamakura, *Electron. Commun. Jpn. 2* 81, 12 (1998); Y. Gotoh, K. Matsumoto, V. Bubanja, F. Vazquez, T. Maeda, and J. S. Harris, *Jpn. J. Appl. Phys. 1* 39, 2334 (2000).
176. T. Sakamoto and K. Nakamura, *Appl. Phys. Lett.* 68, 2861 (1996).
177. T. Sakamoto and K. Nakamura, *Superlattices Microstruct.* 23, 413 (1998).
178. J. Misiewicz, *Proc. SPIE* 2780, 141 (1996).
179. F. H. Pollak, W. Krystek, M. Leibovitch, M. L. Gray, and W. S. Hobson, *IEEE J. Sel. Top. Quantum Electron.* 1, 1002 (1995); F. H. Pollak, *Proc. SPIE* 3944, 408 (2000).
180. R. J. Haug, J. Weis, R. H. Blick, K. Von Klitzing, K. Eberl, and K. Ploog, *Nanotechnology* 7, 381 (1996).
181. U. Simon, *Adv. Mater.* 10, 1487 (1998).
182. M. Koltonyuk, D. Berman, N. B. Zhitenev, R. C. Ashoon, L. N. Pfeiffer, and K. W. West, *Appl. Phys. Lett.* 74, 555 (1999).
183. B. Starmark, P. Delsing, D. B. Haviland, and T. Claeson, *Appl. Supercond.* 6, 837 (1998).
184. S. Sasaki, K. Tsubaki, S. Tarucha, A. Fujiwara, and Y. Takahashi, *Solid-State Electron.* 42, 1429 (1998).
185. K. M. Lewis, C. Kurdak, S. Krishna, and P. Bhattacharya, *Appl. Phys. Lett.* 80, 142 (2002).
186. N. Yokoyama, K. Imamura, M. Takatsu, T. Mori, T. Adachihara, Y. Sugiyama, Y. Sakuma, A. Tackeuchi, and S. Muto, *Philos. Trans. Roy. Soc. London Ser. A* 354, 2399 (1996).
187. T. Fujisawa and Y. Hirayama, *Jpn. J. Appl. Phys. 1* 39, 2338 (2000); T. Fujisawa and Y. Hirayama, *Appl. Phys. Lett.* 77, 543 (2000).
188. O. Astafiev, S. Komiyama, and T. Kutsuwa, *Appl. Phys. Lett.* 79, 1199 (2001).
189. T. Tanamoto, *Physica B* 272, 45 (1999).
190. H. Tamura, Y. Takahashi, and K. Murase, *Microelectron. Eng.* 47, 205 (1999).
191. A. Tilke, A. Erbe, L. Pescini, H. Krommer, R. H. Blick, H. Lorenz, and J. P. Kotthaus, *Superlattices Microstruct.* 27, 597 (2000); A. Tilke, L. Pescini, R. H. Blick, H. Lorenz, and J. P. Kotthaus, *Appl. Phys. A* A71, 357 (2000); A. Tilke, R. H. Blick, H. Lorenz, and J. O. Kotthaus, *J. Appl. Phys.* 89, 8159 (2001).
192. S. Hashioka and H. Matsumura, in "Digest of Papers, 2000 International Microprocesses and Nanotechnology Conference," p. 184.
193. T. Tomaru, T. Ichiguchi, and H. Matsuoka, in "Proceedings of the 22nd International Conference on the Physics of Semiconductors," 15–19 August 1994, Vancouver, BC (D. J. Lockwood, Ed.), Vol. 3, p. 1883.
194. R. Augke, W. Eberhardt, C. Single, F. E. Prins, D. A. Wharam, and D. P. Kern, *Appl. Phys. Lett.* 76, 2065 (2000).
195. T. Hiramoto, H. Ishikuro, T. Fujii, G. Hashiguchi, and T. Ikoma, *Jpn. J. Appl. Phys. 1* 36, 4139 (1997); H. Ishikuro and T. Hiramoto, *Solid-State Electron*. 42, 1425 (1998); H. Ishikuro and T. Hiramoto, *Jpn. J. Appl. Phys. 1* 38, 396 (1999).
196. H. Kawaura and T. Sakamoto, *IEICE Trans. Electron.* E84-C, 1037 (2001).
197. B. Y. Choi, S. K. Sung, B.-G. Park, and J. D. Lee, *Jpn. J. Appl. Phys. 1* 40, 2607 (2001).
198. J. Collet and D. Vuillaume, *Appl. Phys. Lett.* 73, 2681 (1998).
199. T. Ichimori and N. Hirashita, *Jpn. J. Appl. Phys. 2 Lett.* 40, L1019 (2001).
200. H. Matsuoka, T. Ichiguchi, T. Yoshimura, and E. Takeda, *J. Appl. Phys.* 76, 5561 (1994); H. Matsuoka, T. Ichiguchi, T. Yoshimura, and E. Takeda, *Appl. Phys. Lett.* 64, 586 (1994); H. Matsuoka and S. Kimura, *Jpn. J. Appl. Phys. 1* 34, 1326 (1995); H. Matsuoka and S. Kimura, *Appl. Phys. Lett.* 66, 613 (1995).
201. M. Khoury, M. J. Rack, A. Gunther, and D. K. Ferry, *Appl. Phys. Lett.* 74, 1576 (1999); M. Khoury, A. Gunther, D. P. Pivin, Jr., M. J. Rack, and D. K. Ferry, *Jpn. J. Appl. Phys. 1* 38, 469 (1999);

M. Khoury, A. Gunther, M. J. Rack, D. P. Pivin, Jr., and D. K. Ferry, *Microelectron. Eng.* 47, 189 (1999).

202. A. Savin, A. Manninen, J. Kauranen, J. Pekola, M. Prunnila, J. Ahopelto, M. Kamp, M. Emmerling, and A. Forchel, *Mater. Res. Soc. Symp. Proc.* 638, F1.3.1 (2001); A. Manninen, J. Kauranen, J. Pekola, A. Savin, M. Kamp, M. Emmerling, A. Forchel, M. Prunnila, and J. Ahopelto, *Jpn. J. Appl. Phys. 1* 40, 2013 (2001).

203. D. H. Kim, D.-H. Chae, J. D. Lee, B.-G. Park, and H.-G. Lee, *J. Korean Phys. Soc.* 33, S278 (1998); D. H. Kim, D.-H. Chae, J. D. Lee, and B.-G. Park, in "57th Annual Device Research Conference Digest," 28–30 June 1999, Santa Barbara, CA, p. 134; D. H. Kim, K. R. Kim, S. K. Sung, B. H. Choi, S. W. Hwang, D. Ahn, J. D. Lee, and B. G. Park, in "Device Research Conference Digest," 25–27 June 2001, Notre Dame, IN, p. 133; D. H. Kim, S.-K. Sung, K. R. Kim, B. H. Choi, S. W. Hwang, D. Ahn, J. D. Lee, and B. G. Park, in "International Electron Devices Meeting, Technical Digest," 2–5 December 2001, Washington, DC, p. 7.3.1; D. H. Kim, S.-K. Sung, J. S. Sim, K. R. Kim, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang, and D. Ahn, *Appl. Phys. Lett.* 79, 3812 (2001); K. R. Kim, D. H. Kim, J. D. Lee, and B.-G. Park, *J. Korean Phys. Soc.* 40, 140 (2002); K. R. Kim, D. H. Kim, S.-K. Sung, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang, and D. Ahn, *Jpn. J. Appl. Phys. 1* 41, 2574 (2002); D. H. Kim, S.-K. Sung, K. R. Kim, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang, and D. Ahn, *IEEE Trans. Electron Devices* 49, 627 (2002).

204. M. G. Peters, S. G. den Hartog, J. I. Dijkhuis, O. J. A. Buyk, and L. W. Molenkamp, *J. Appl. Phys.* 84, 5052 (1998); M. G. Peters, J. I. Dijkhuis, and L. W. Molenkamp, *Semicond. Sci. Technol.* 14, 1119 (1999).

205. F. Simmel, D. Abusch-Magder, D. A. Wharam, M. A. Kastner, and J. P. Kotthaus, *Phys. Rev. B* 59, 10441 (1999); D. Abusch-Magder, F. Simmel, D. A. Wharam, M. A. Kastner, and J. P. Kotthaus, *Physica E* 6, 382 (2000).

206. W. Chen and H. Ahmed, *J. Vac. Sci. Technol. B* 15, 1402 (1997).

207. K. Hofmann, B. Spangenberg, and H. Kurz, *J. Vac. Sci. Technol. B* 20, 271 (2002).

208. A. Tilke, R. H. Blick, and H. Lorenz, *J. Appl. Phys.* 90, 942 (2001).

209. A. Fukushima, A. Iwasa, and A. Sato, in "Conference on Precision Electromagnetic Measurements Digest, CPEM 2000," 14–19 May 2000, Sydney, Australia (J. Hunter and L. Johnson, Eds.), p. 591.

210. S. M. Goodnice, J. Bird, D. K. Ferry, A. D. Gunther, M. D. Khoury, M. Kozicki, J. Rack, T. J. Thornton, and D. Vasileska-Kafedezka, in "Proceedings of the Ninth Great Lakes Symposium on VLSI," 4–6 March 1999, Ypsilanti, MI (R. J. Lomax and P. Mazumder, Eds.), p. 394.

211. J. W. Park, K. S. Park, B. T. Lee, C. H. Lee, S. D. Lee, J. B. Choi, K.-H. Yoo, J. Kim, S. C. Oh, S. I. Park, K. T. Kim, and J. J. Kim, *Appl. Phys. Lett.* 75, 566 (1999).

212. T. Kamiya, Y. T. Tan, Z. A. K. Durrani, and H. Ahmed, *J. Non-Cryst. Solids* 299–302, 405 (2002).

213. S. Han, T. Park, B. Kim, H. Shin, and K. Lee, in "ICVC '99: 6th International Conference on VLSI and CAD," 26–27 October 1999, Seoul, South Korea, p. 163.

214. Y. Ono, Y. Takahashi, M. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, in "International Electron Devices Meeting 1998, Technical Digest," 6–9 December 1998, San Francisco, CA, p. 123; Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, *IEEE Trans. Electron Devices* 47, 147 (2000).

215. G. Lu, Z. Chen, J. Wang, and W. Ge, *Chin. J. Semicond.* 23, 246 (2002).

216. T. Tsutsumi, K. Ishii, H. Hiroshima, S. Kanemaru, E. Suzuki, and K. Tomizawa, in "Digest of Papers, 2000 International Microprocesses and Nanotechnology Conference," 11–13 July 2000, Tokyo, p. 182; T. Tsutsumi, K. Ishii, E. Suzuki, H. Hiroshima, M. Yamanaka, I. Sakata, S. Kanemaru, S. Hazra, T. Maeda, and K. Tomizawa, *Electron. Lett.* 36, 1322 (2000).

217. E. Leobandung, L. Guo, Y. Wang, and S. Y. Chou, *J. Vac. Sci. Technol. B* 13, 2865 (1995); E. Leobandung, L. Guo, Y. Wang, and S. Y. Chou, *Appl. Phys. Lett.* 67, 938 (1995); E. Leobandung, L. Guo, Y. Wang, and S. Y. Chou, in "53rd Annual Device Research Conference Digest," 19–21 June 1995, Charlottesville, VA, p. 48.

218. A. Fujiwara, and Y. Takahashi, *Jpn. J. Appl. Phys. 1* 41, 1209 (2002).

219. K. S. Park, K. H. Yoo, J. Kim, and J. B. Choi, *J. Korean Phys. Soc.* 39, S173 (2001).

220. M. Y. Simmons, S. R. Schofield, J. L. O'Brien, N. J. Curson, R. G. Clark, T. M. Buehler, R. P. McKinnon, R. Brenner, D. J. Reilly, A. S. Dzurak, and A. R. Hamilton, in "Proceedings of the 2001 First IEEE Conference on Nanotechnology, IEEE-NANO 2001," 28–30 October 2001, Maui, HI, p. 471.

221. E. W. Berg and S. W. Pang, *J. Vac. Sci. Technol. B* 19, 1925 (2001).

222. T. Koster, B. Hadam, K. Hofmann, J. Gondermann, J. Stein, S. Hu, S. Altmeyer, B. Spangenberg, and H. Kurz, *J. Vac. Sci. Technol. B* 15, 2836 (1997).

223. M. Tabe, Y. Takahashi, and K. Murase, *Bull. Res. Inst. Electron. Shizuoka Univ.* 30, 13 (1995); K. Murase, Y. Talahashi, A. Fujiwara, M. Nagase, and M. Tabe, in "Proceedings of ASP-DAC'95/CHDL'95/VLSI'95 with EDA Technofair," 29 August–1 September 1995, Chiba, Japan. p, 697; K. Murase, Y. Takahashi, Y. Nakajima, H. Namatsu, M. Nagase, K. Kurihara, K. Iwadate, S. Horiguchi, M. Tabe, and K. Izumi, *Microelectron. Eng.* 28, 399 (1995); A. Fujiwara, Y. Takahashi, K. Murase, and M. Tabe, *Appl. Phys. Lett.* 67, 2957 (1995); Y. Takahashi, M. Nagase, H. Namatsu, K. Kurihara, K. Iwdate, Y. Nakajima, S. Horiguchi, K. Murase, and M. Tabe, *Electron. Lett.* 31, 136 (1995); Y. Takahashi, A. Fujiwara, M. Nagase, H. Namatsu, K. Kurihara, K. Iwadate, and K. Murase, *IEICE Trans. Electron.* E79-C, 1503 (1996); Y. Takahashi, H. Namatsu, K. Kurihara, K. Iwadate, M. Nagase, and K. Murase, *IEEE Trans. Electron Devices* 43, 1213 (1996).

224. Y. Takahashi, M. Nagase, A. Fujiwara, H. Namatsu, K. Kurihara, K. Iwadate, K. Murase, and Y. Takahashi, in "Proceedings of Quantum Devices and Circuits," 4–7 June 1996, Alexandria, Egypt (K. Ismail, S. Bandyopadhyay, and J. P. Leburton, Eds.), p. 188.

225. B. T. Lee, J. W. Park, K. S. Park, C. H. Lee, S. W. Paik, S. D. Lee, J. B. Choi, K. S. Min, J. S. Park, S. Y. Hahn, T. J. Park, H. Shin, S. C. Hong, K. Lee, H. C. Kwon, S. I. Park, K. T. Kim, and K.-H. Yoo, *Semicond. Sci. Technol.* 13, 1463 (1998).

226. K. Kurihara, H. Namatsu, M. Nagase, and Y. Takahashi, in "Proceedings of the 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, p. 62; Y. Takahashi, A. Fujiwara, M. Nagase, H. Namatsu, K. Kurihara, K. Iwadate, and Murase, *Int. J. Electron.* 86, 605 (1999).

227. T. Hiramoto, H. Ishikubo, K. Saito, T. Fuji, T. Saraya, G. Hashiguchi, and A. Ikoma, *Jpn. J. Appl. Phys. 1* 35, 6664 (1996).

228. P. A. Fontaine, E. Dubois, and D. Stievenard, in "Proceedings of the 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, p. 46.

229. S. Y. Chou, P. R. Krauss, W. Zhang, L. Guo, and L. Zhuang, *J. Vac. Sci. Technol. B* 15, 2897 (1997).

230. A. Fujiwara, Y. Takahashi, H. Namatsu, K. Kurihara, and K. Murase, *Jpn. J. Appl. Phys. 1* 37, 3257 (1998).

231. T. Koester, A. F. Goldschmidtboeing, B. Hadam, J. Stein, S. Altmeyer, B. Spangenberg, H. Kurz, R. Neumann, K. Brunner, and G. Abstreiter, *J. Vac. Sci. Technol. B* 16, 3804 (1998); T. Koester, F. Goldschmidtboeing, B. Hadam, J. Stein, S. Altmeyer, B. Spangenberg, H. Kurz, R. Neumann, K. Brunner, and G. Abstreiter, *Jpn. J. Appl. Phys. 1* 38, 465 (1999).

232. T. Sakamoto, H. Kawaura, and T. Baba, *Appl. Phys. Lett.* 72, 795 (1998); T. Sakamoto, H. Kawaura, and T. Baba, *NEC Res. Dev.* 40, 397 (1999).

233. Y. Hirai, Y. Kanemaki, K. Murata, and Y. Tanaka, *Jpn. J. Appl. Phys. 1* 38, 7272 (1999).

234. D. Tsoukalas, P. Normand, C. Aidinis, E. Kapetanakis, and P. Argitis, *Microelectron. Eng.* 41–42, 523 (1998).

235. H. Ishikuro and T. Hiramoto, in "1997 55th Annual Device Research Conference Digest," 23–25 June 1997, Fort Collins, CO, p. 84; H. Ishikuro and T. Hiramoto, *Appl. Phys. Lett.* 71, 3691 (1997); H. Ishikuro and T. Hiramoto, *Jpn. J. Appl. Phys. 1* 38, 396 (1999).

236. M. Saitoh and T. Hiramoto, in "Proceedings of the 2001 First IEEE Conference on Nanotechnology, IEEE-NANO 2001," 28–30 October 2001, Maui, HI, p. 243; M. Saitoh, T. Saito, T. Inukai, and T. Hiramoto, *Appl. Phys. Lett.* 79, 2025 (2001).

237. M. Saitoh, N. Takahashi, H. Ishikuro, and T. Hiramoto, *Jpn. J. Appl. Phys. 1* 40, 2010 (2001).

238. T. H. Wang and H. W. Li, *Physica B* 301, 169 (2001); T. H. Wang, H. W. Li, and J. M. Zhou, *Appl. Phys. Lett.* 78, 2160 (2001); T. H. Wang, H. W. Li, and J. M. Zhou, *Chin. Phys.* 10, 844 (2001).

239. T. Nishimura, Y. Akasaka, H. Nakata, and A. Ishizu, *Appl. Phys. Lett.* 42, 102 (1983).

240. H. Qin, X. Gu, H. Lu, J. Liu, X. Huang, and X. Chen, *Solid State Commun.* 111, 171 (1999).

241. X. Tang, X. Baie, V. Bayot, F. Van de Wiele, and J. P. Colinge, in "1999 IEEE International SOI Conference Proceedings," 4–7 October 1999, Rohnert Park, CA, p. 46.

242. L. Guo, E. Leobandung, and S. Y. Chou, in "International Electron Devices Meeting, Technical Digest," 8–11 December 1996, San Francisco, CA, 955; L. Guo, E. Leobandung, L. Zhuang, and S. Y. Chou, *J. Vac. Sci. Technol. B* 15, 2840 (1997); L. Guo, E. Leobandung, and S. Y. Chou, *Appl. Phys. Lett.* 70, 850 (1997); L. Guo, E. Leobandung, and S. Y. Chou, *Science* 275, 649 (1997); L. Zhuang, L. Guo, and S. Y. Chou, in "International Electron Devices Meeting 1997, IEDM Technical Digest," 7–10 December 1997, Washington, DC, p. 167; L. Zhuang, L. Guo, and S. Y. Chou, *Appl. Phys. Lett.* 72, 1205 (1998).

243. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, in "International Electron Devices Meeting, Technical Digest," 8–11 December 1996, San Francisco, CA, p. 952; A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *Appl. Phys. Lett.* 70, 1742 (1997).

244. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *Appl. Phys. Lett.* 71, 353 (1997); A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, in "Proceedings of 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, p. 26; A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *J. Vac. Sci. Technol. B* 17, 2163 (1999).

245. A. H. M. Kamal, J. Lutzen, B. A. Sanborn, M. V. Sidorov, M. N. Kozicki, D. J. Smith, and D. K. Ferry, *Semicond. Sci. Technol.* 13, 1328 (1998).

246. B. H. Choi, S. W. Hwang, I. G. Kim, H. C. Shin, Y. Kim, and E. K. Kim, *Appl. Phys. Lett.* 73, 3129 (1998); B. H. Choi, S. W. Hwang, I. G. Kim, H. C. Shin, Y. Kim, and E. K. Kim, *Microelectron. Eng.* 47, 115 (1999).

247. S. Banerjee, S. Huang, T. Yamanaka, and S. Oda, *J. Vac. Sci. Technol. B* 20, 1135 (2002).

248. S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan, *Appl. Phys. Lett.* 68, 1377 (1996); S. Tiwari, F. Rana, K. Chan, S. Leathen, and H. Hanafi, *Appl. Phys. Lett.* 69, 1232 (1996).

249. B. J. Hinds, A. Dutta, Y. Yun, T. Yamanaka, S. Hatanani, and S. Oda, in "58th Device Research Conference Digest," 19–21 June 2000, Denver, CO, p. 151.

250. A. Dutta, M. Kimura, Y. Honda, N. Otobe, A. Itoh, and S. Oda, *Jpn. J. Appl. Phys. 1* 36, 4038 (1997); A. Dutta, S. P. Lee, Y. Hayafune, S. Hatatani, and S. Oda, *Jpn. J. Appl. Phys. 1* 39, 264 (2000).

251. A. Dutta, S. P. Lee, S. Hatatani, and S. Oda, *Appl. Phys. Lett.* 75, 1422 (1999).

252. Y. Fu, M. Willander, A. Dutta, and S. Oda, *Proc. SPIE* 3975, 1027 (2000).

253. A. Dutta, S. Oda, Y. Fu, and M. Willander, *Jpn. J. Appl. Phys. 1* 39, 4647 (2000).

254. Y. Fu, M. Willander, A. Dutta, and S. Oda, *Superlattices Microstruct.* 28, 177 (2000).

255. Y. Fu, M. Willander, A. Dutta, and S. Oda, *Superlattices Microstruct.* 28, 189 (2000).

256. A. Dutta, Y. Hayafune, and S. Oda, *Jpn. J. Appl. Phys. 2* 39, L855 (2000).

257. T. Kamiya, K. Nakahata, Y. T. Tan, Z. A. K. Durrani, and I. Shimizu, *J. Appl. Phys.* 89, 6265 (2001).

258. Y.-C. King, T.-J. King, and C. Hu, *IEEE Trans. Electron Devices* 48, 696 (2001).

259. A. Kohno, H. Murakami, M. Ikeda, S. Miyazaki, and M. Hirose, *Jpn. J. Appl. Phys. 2* 40, L721 (2001).

260. A. I. Yakimov, A. V. Dvurechenskii, V. V. Kirienko, and A. I. Nikiforov, *Appl. Phys. Lett.* 80, 4783 (2002).

261. T. Hatano, Y. Ito, A. Nakajima, and S. Yokoyama, *Jpn. J. Appl. Phys. 1* 40, 2017 (2001).

262. N. Y. Morgan, D. Abusch-Magder, M. A. Kastner, Y. Takahashi, H. Tamura, and K. Murase, *J. Appl. Phys.* 89, 410 (2001).

263. Y. Takahashi, Y. Ono, A. Fujiwara, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, in "Proceedings of Advanced Luminescent Materials and Quantum Confinement," 18–20 October 1999, Honolulu, HI (M. Cahay, S. Bandyopadhyay, D. J. Lockwood, N. Koshida, J. P. Leburton, M. Meyyappan, and T. Sakamoto, Eds.), p. 302; S. Horiguchi, M. Nagase, K. Shiraishi, H. Kageshima, Y. Takahashi, and K. Murase, *Jpn. J. Appl. Phys. 2* 40, L29 (2001).

264. J. M. Therrien, G. Belomoin, and M. H. Nayfeh, *Mater. Res. Soc. Symp. Proc.* 582, H11.4.1 (2001).

265. F. Yun, B. J. Hinds, S. Hatatani, and S. Oda, *Jpn. J. Appl. Phys. 2* 39, L792 (2000).

266. N. Takahashi, H. Ishikuro, and T. Hiramoto, in "International Electron Devices Meeting 1999, Technical Digest," 5–8 December 1999, Washington, DC, p. 371; Y. Shi, K. Saito, H. Ishikuro, and T. Hiramoto, *Jpn. J. Appl. Phys. 1* 38, 2453 (1999); N. Takahashi, H. Ishikuro, and T. Hiramoto, *Appl. Phys. Lett.* 76, 209 (2000).

267. T.-S. Yoon, J.-Y. Kwon, D.-H. Lee, K.-B. Kim, S.-H. Min, D.-H. Chae, D. H. Kim, J.-D. Lee, B.-G. Park, and H. J. Lee, *J. Appl. Phys.* 87, 2449 (2000).

268. A. Ohata, H. Niiyama, T. Shibata, K. Nakajima, and A. Toriumi, *Jpn. J. Appl. Phys. 1* 34, 4485 (1995).

269. J. Motohisa, F. Nakajima, T. Fukui, W. G. van der Wiel, J. M. Elzerman, S. De Franceschi, and L. P. Kouwenhoven, *Appl. Phys. Lett.* 80, 2797 (2002).

270. Y. T. Tan, Z. A. K. Durrani, and H. Ahmed, *J. Appl. Phys.* 89, 1262 (2001).

271. B. J. Hinds, K. Nishiguchi, A. Dutta, T. Yamanaka, S. Hatanani, and S. Oda, *Jpn. J. Appl. Phys. 1* 39, 4637 (2000).

272. M. J. Uren, M. J. Kirton, and S. Collins, *Phys. Rev. B* 37, 8346 (1988).

273. H. H. Mueller, D. Worle, and M. Schulz, *J. Appl. Phys.* 75, 2970 (1994); M. Schulz and H. H. Mueller, *Electron Technol.* 29, 5 (1996); H. H. Mueller and M. Schulz, *J. Appl. Phys.* 79, 4178 (1996).

274. M. Sanquer, M. Specht, L. Ghenim, S. Deleonibus, and G. Guegan, *Phys. Rev. B* 61, 7249 (2000).

275. H. Kondo, S. Baba, K. Izumikawa, M. Sakurai, S. Zaima, and Y. Yasuda, in "Digest of Papers, 1999 International Microprocesses and Nanotechnology Conference," 6–8 July 1999, Yokohama, Japan, p. 126; H. Kondo, K. Izumikawa, M. Sakurai, S. Baba, H. Iwano, S. Zaima, and Y. Yasuda, *Jpn. J. Appl. Phys. 1* 38, 7222 (1999).

276. T. W. Kim, D. C. Choo, J. H. Shim, and S. O. Kang, *Appl. Phys. Lett.* 80, 2168 (2002).

277. M. Skender, R. Straub, F. E. Prins, and D. P. Kern, *Microelectron. Eng.* 57–58, 1023 (2001).
278. E. I. Rau, A. V. Gostev, Z. Shiqiu, D. Phang, D. Chan, D. Thong, and W. Wong, *Mikroelektronika* 30, 243 (2001).
279. T. Matsukawa, S. Kanemaru, M. Masahara, M. Nagao, and J. Itoh, in "2001 International Semiconductor Device Research Symposium Proceedings," 5–7 December 2001, Washington, DC, p. 364.
280. N. V. Amarasinghe, Z. Celik-Butler, and A. Keshavarz, *J. Appl. Phys.* 89, 5526 (2001).
281. H. Kamohara, K. Imai, I. Ma, Y. Cheng, and Hu, in "Proceedings of 1997 Fifth International Conference on VLSI and CAD," 13–15 October 1997, Seoul, South Korea, p. 295.
282. M. V. Sidorov and D. J. Smith, in "Thin Films—Structure and Morphology, Symposium," 2–6 December 1996, Boston, MA, p. 773.
283. A. Hartland, S. W. Chua, and S. P. Giblin, in "Proceedings of BEMC '97," 4–6 November 1997, Teddington, UK, p. 14.
284. A. Fujiwara, Y. Takahashi, and K. Murase, *Microelectron. Eng.* 47, 197 (1999).
285. C. Wasshuber, H. Kosina, and S. Selberherr, *IEEE Trans. Electron Devices* 45, 2365 (1998).
286. N. J. Stone and H. Ahmed, *Microelectron. Eng.* 41–42, 511 (1998).
287. K. Goser and C. Pacha, in "ESSCIRC '98, Proceedings of the 24th European Solid-State Circuits Conference," 22–24 September 1998, The Hague, Netherlands (J. H. Huijsing, A. H. M. van Roermund, and H. Grunbacher, Eds.), p. 18; C. Pacha, O. Kessler, P. Glosekotter, K. F. Goser, W. Prost, A. Brennemann, U. Auer, and F. J. Tegude, *Analog Integr. Circuits Signal Process.* 24, 7 (2000).
288. W. Porod, *Int. J. High Speed Electron. Syst.* 9, 37 (1998).
289. A. Fujiwara, Y. Takahashi, K. Yamazaki, and K. Murase, *NTT Rev.* 10, 114 (1998); Y. Takahashi, A. Fujiwara, K. Yamazaki, H. Namatsu, K. Kurihara, and K. Murase, *Electron. Lett.* 34, 45 (1998); Y. Takahashi, A. Fujiwara, K. Yamazaki, H. Namatsu, K. Kurihara, and K. Murase, *Jpn. J. Appl. Phys. 1* 38, 2457 (1999); A. Fujiwara, Y. Takahashi, K. Yamazaki, H. Namatsu, M. Nagase, K. Kurihara, and K. Murase, *IEEE Trans. Electron Devices* 46, 954 (1999); H. Inokawa, A. Fujiwara, and Y. Takahashi, in "Device Research Conference Digest," 25–27 June 2001, Notre Dame, IN, p. 129; H. Inokawa, A. Fujiwara, and Y. Takahashi, *Jpn. J. Appl. Phys. 1* 41, 2566 (2002).
290. H. Sunamura, H. Kawaura, T. Sakamoto, and T. Baba, *Jpn. J. Appl. Phys. 2* 41, L93 (2002).
291. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, *Appl. Phys. Lett.* 76, 3121 (2000).
292. R. A. Smith and H. Ahmed, *Appl. Phys. Lett.* 71, 3838 (1997).
293. Y. Ono, K. Yamazaki, and Y. Takahashi, *IEICE Trans. Electron.* E84-C, 1061 (2001); K.-T. Liu, A. Fujiwara, Y. Takahashi, K. Murase, and Y. Horikoshi, *Jpn. J. Appl. Phys. 1* 41, 458 (2002).
294. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, *Jpn. J. Appl. Phys. 1* 39, 2325 (2000).
295. Y. Ono and Y. Takahashi, *Jpn. J. Appl. Phys. 1* 41, 2569 (2002).
296. K. Uchida, J. Koga, R. Ohba, and A. Toriumi, in "2002 IEEE International Solid-State Circuits Conference, Digest of Technical Papers," 3–7 February 2002, San Francisco, CA, Vol. 1, pp. 206–460.
297. Z. A. K. Durrani, A. C. Irvine, H. Ahmed, and K. Nakazato, *Appl. Phys. Lett.* 74, 1293 (1999); A. C. Irvine, Z. A. K. Durrani, and H. Ahmed, *J. Appl. Phys.* 87, 8594 (2000).
298. Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, *IEEE Trans. Electron Devices* 47, 2334 (2000).
299. A. Pepin, C. Vieu, H. Launois, M. Rosmeulen, M. Van Rossum, H. O. Mueller, D. Williams, H. Mizuta, and K. Nakazato, *Microelectron. Eng.* 53, 265 (2000).
300. H. Inokawa, A. Fujiwara, and Y. Takahashi, *Appl. Phys. Lett.* 79, 3618 (2001).
301. Y. Takahashi, A. Fujiwara, K. Yamazaki, H. Namatsu, K. Kurihara, and K. Murase, *Appl. Phys. Lett.* 76, 637 (2000).
302. Y. S. Yu, Y. I. Jung, S. W. Hwang, and D. Ahn, in "ICVC '99: Sixth International Conference on VLSI and CAD," 26–27 October 1999, Seoul, South Korea, p. 403; Y. S. Yu, S. W. Hwang, and D. Ahn, *IEEE Trans. Electron Devices* 46, 1667 (1999); Y. S. Yu, B. H. Choi, J. H. Oh, S. W. Hwang, and D. Ahn, *J. Korean Phys. Soc.* 39, S27 (2001).
303. K. Uchida, K. Matsuzawa, and A. Toriumi, *Jpn. J. Appl. Phys. 1* 38, 4027 (1999); A. Toriumi, K. Uchida, R. Ohba, and J. Koga, *Physica B* 272, 522 (1999); K. Uchida, J. Koga, R. Ohba, and A. Toriumi, *IEICE Trans. Electron.* E84-C, 1066 (2001).
304. A. S. Cordan, Y. Leroy, and A. Goltzene, *Mater. Sci. Eng. C* 15, 25 (2001); Y. Leroy, A. S. Cordan, and A. Goltzene, *Mater. Sci. Eng. C* 15, 49 (2001).
305. M. Kirihara and K. Taniguchi, *Jpn. J. Appl. Phys. 1* 36, 4172 (1997).
306. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, and K. Seki, in "International Electron Devices Meeting, Technical Digest," 10–13 December 1995, Washington, DC, p. 525.
307. S. Tiwari, in "Proceedings of Third International Symposium on the Physics and Chemistry of SiO_2 and the Si–SiO_2 Interface," 5–10 May 1996, Los Angeles, CA (H. Z. Massoud, E. H. Poindexter, and C. R. Helms, Eds.), p. 250; S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan, "International Electron Devices Meeting. Technical Digest," 10–13 December 1995, Washington, DC, USA. p. 521.
308. A. Ohata, A. Toriumi, and K. Uchida, *Jpn. J. Appl. Phys. 1* 36, 1686 (1997).
309. S. L. Pohlen, R. J. Fitzgerald, J. M. Hergenrother, and M. Tinkham, *Appl. Phys. Lett.* 74, 2884 (1999).
310. H. Ahmed, in "ESSDERC'99, Proceedings of the 29th European Solid-State Device Research Conference," 13–15 September 1999, Leuven, Belgium, p. 21; H. Ahmed, in "International Electron Devices Meeting 1999, Technical Digest," 5–8 December 1999, Washington, DC, p. 363.
311. I. Kim, S. Han, H. Kim, J. Lee, B. Choi, S. Hwang, D. Ahn, and H. Shin, in "International Electron Devices Meeting 1998, Technical Digest," 6–9 December 1998, San Francisco, CA, p. 111.
312. G. D. Sanders, K. W. Kim, and W. C. Holton, in "57th Annual Device Research Conference Digest," 28–30 June 1999, Santa Barbara, CA, p. 68.
313. H. Wang, N. Takahashi, H. Majima, T. Inukai, M. Saitoh, and T. Hiramoto, *Jpn. J. Appl. Phys. 1* 40, 2038 (2001).
314. R. H. Chen, A. N. Korotkov, and K. K. Likharev, *Appl. Phys. Lett.* 68, 1954 (1996).
315. H. Majima, S. Amakawa, M. Fujishima, and K. Hoh, in "Proceedings of the 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, p. 40; M. Fujishima, S. Amakawa, and K. Hoh, *Jpn. J. Appl. Phys. 1* 37, 1478 (1998); S. Amakawa, K. Kanda, M. Fujishima, and K. Hoh, *Jpn. J. Appl. Phys. 1* 38, 429 (1999).
316. N. Yoshikawa, Y. Jinguu, H. Ishibashi, and M. Sugahara, *Jpn. J. Appl. Phys. 1* 35, 1140 (1996).
317. C. Gerousis, S. M. Goodnick, X. Wang, W. Porod, A. I. Csurgay, G. Toth, and C. S. Lent, in "Proceedings of European Conference on Circuit Theory and Design, ECCTD'99," 29 August–2 September 1999, Stresa, Italy, Vol. 2, p. 835.
318. R. W. Rendell, in "Proceedings of the Fifth International Symposium on Quantum Confinement: Nanostructures," 2–5 November 1998, Boston, MA (M. Cahay, D. J. Lockwood, J. P. Leburton, and S. Bandyopadhyay, Eds.), p. 574.
319. K. Nakazato, in "Proceedings of SISPADS'98 International Conference on Simulation of Semiconductor Processes and Devices," 2–4 September 1998, Leuven, Belgium (K. D. Meyer and Biesemans, Eds.), p. 201.
320. I. Karafyllidis, *Superlattices Microstruct.* 25, 567 (1999).
321. U. Dotsch, U. Gennser, C. David, G. Dehlinger, D. Grutzmacher, T. Heinzel, S. Luscher, and K. Ensslin, *Appl. Phys. Lett.* 78, 341 (2001).

322. H. Ishikuro and T. Hiramoto, in "International Electron Devices Meeting 1998, Technical Digest," 6–9 December 1998, San Francisco, CA, p. 119; H. Ishikuro and T. Hiramoto, *Appl. Phys. Lett.* 74, 1126 (1999).
323. N. T. Bagraev, L. E. Klyachkin, A. M. Malyarenko, and W. Gehlhoff, in "Proceedings of the 23rd International Conference on the Physics of Semiconductors," 21–26 July 1996, Berlin (M. Scheffler and R. Zimmermann, Eds.), Vol. 2, p. 1241; N. T. Bagraev, A. D. Bouravleuv, L. E. Klyachkin, A. M. Malyarenko, and S. A. Rykov, *Phys. Low-Dimens. Struct.* 9–10, 51 (2000); N. T. Bagraev, A. D. Bouravleuv, L. E. Klyachkin, A. M. Malyarenko, and S. A. Rykov, *Proc. SPIE* 4348, 125 (2001).
324. E. Leobandung, L. Guo, Y. Wang, and S. Y. Chou, *J. Vac. Sci. Technol. B* 13, 2865 (1995); E. Leobandung, L. Guo, and S. Y. Chou, *Appl. Phys. Lett.* 67, 2338 (1995); L. P. Rokhinson, L. J. Guo, S. Y. Chou, and D. C. Tsui, *Appl. Phys. Lett.* 76, 1591 (2000).
325. S. Kanjanachuchai, J. M. Bonar, and H. Ahmed, *Semicond. Sci. Technol.* 14, 1065 (1999).
326. K. Kawasaki, D. Yamazaki, A. Kinoshita, H. Hirayama, K. Tsutsui, and Y. Aoyagi, *Appl. Phys. Lett.* 79, 2243 (2001).
327. M. Tabe, T. Yamamoto, Y. Terao, and T. Umeji, in "Proceedings of 1997 IEEE Silicon Nanoelectronics Workshop," 8–9 June 1997, Kyoto, Japan, p. 42.
328. D. M. Pooley, H. Ahmed, H. Mizuta, and K. Nakazato, *J. Appl. Phys.* 90, 4772 (2001).
329. H. Sunamura, H. Kawaura, T. Sakamoto, and T. Baba, in "58th Device Research Conference, Conference Digest," 19–21 June 2000, Denver, CO, p. 153.
330. H. Sunamura, T. Sakamoto, Y. Nakamura, H. Kawaura, J. S. Tsai, and T. Baba, *Appl. Phys. Lett.* 74, 3555 (1999).
331. N. T. Bagraev, L. E. Klyachkin, A. M. Malyarenko, V. L. Sukhanov, A. V. Suvorov, and N. V. Zabrodskaya, in "Proceedings of the 23rd International Symposium on Compound Semiconductors," 23–27 September 1996, St. Petersburg (M. S. Shur and R. A. Suris, Eds.), p. 81.
332. J. Hu, M. Ouyang, P. Yang, and C. M. Lieber, *Nature* 399, 48 (1999).
333. A. T. Johnson, in "1999 IEEE International Solid-State Circuits Conference, Digest of Technical Papers," 15–17 February 1999, San Francisco, CA (J. H. Wuorinen, Ed.), p. 210.
334. J. Kong, C. Zhou, E. Yenilmez, and H. Dai, *Appl. Phys. Lett.* 77, 3977 (2000).
335. Z. Yao, H. W. C. Postma, L. Balents, and C. Dekker, *Nature* 402, 273 (1999).
336. K. Matsumoto, in "Proceedings of WOCSDICE 2000: 24th Workshop on Compound Semiconductor Devices and Integrated Circuits," 29 May–2 June 2000, Aegean Sea, Greece, p. XIII-5–6; K. Matsumoto and K. Gotoh, in "2001 International Semiconductor Device Research Symposium Proceedings," 5–7 December 2001, Washington, DC, p. 354.
337. A. Kanda, Y. Ootuka, K. Tsukagoshi, and Y. Aoyagi, *Appl. Phys. Lett.* 79, 1354 (2001).
338. L. Roschier, J. Penttila, M. Martin, P. Hakonen, M. Paalanen, U. Tapper, E. I. Kauppinen, C. Journet, and P. Bernier, *Appl. Phys. Lett.* 75, 728 (1999).
339. J. Kong, J. Cao, H. Dai, and E. Anderson, *Appl. Phys. Lett.* 80, 73 (2002).
340. N. Miura, T. Numaguchi, A. Yamada, M. Konagai, and J.-I. Shirakashi, *Jpn. J. Appl. Phys. 2* 36, 1619 (1997); N. Miura, T. Numaguchi, A. Yamada, M. Konagai, and J. Shirakashi, *Jpn. J. Appl. Phys. 2* 37, L423 (1998).
341. M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, and J. M. Tour, *Science* 278, 252 (1997).
342. S. J. Tans, R. M. Verschueren, and C. Dekker, *Nature* 393, 49 (1998).
343. S. Frank, P. Poncharal, Z. L. Wang, and W. A. de Heer, *Science* 280, 1744 (1998).
344. C.-K. Wang, Y. Fu, and Y. Luo, *Phys. Chem. Chem. Phys.* 3, 5017 (2001).
345. L. A. Bumm, J. J. Arnold, M. T. Cygan, T. D. Dunbar, T. P. Burgin, L. Jones II, D. L. Allara, J. M. Tour, and P. S. Weiss, *Science* 271, 1705 (1996).
346. R. P. Andres, J. O. Bielefeld, J. I. Henderson, D. B. Janes, V. R. Kolagunta, C. P. Kubiak, W. J. Mahoney, and R. G. Osifchin, *Science* 273, 1690 (1997).
347. J. Chen, M. A. Reed, A. M. Rawlett, and J. M. Tour, *Science* 286, 1550 (1999).
348. Y. Wada, *Optoelectron. Devices Technol.* 10, 205 (1995); Y. Wada, *Microelectron. Eng.* 30, 375 (1996); Y. Wada, *Surf. Sci.* 386, 265 (1997).
349. S. Ahmad, *Def. Sci. J.* (*India*) 48, 45 (1998).
350. T. Tanamoto, *Fortschr. Phys.* 48, 1005 (2000).
351. J. H. Schon, *Synth. Met.* 122, 157 (2001); J. H. Schon, Ch. Kloc, and B. Batlogg, *Synth. Met.* 122, 195 (2001).
352. H. Okada, T. Hashizume, K. Jinushi, T. Kudoh, and H. Hasegawa, in "Proceedings of the 23rd International Conference on the Physics of Semiconductors," 21–26 July 1996, Berlin (M. Scheffler and R. Zimmermann, Eds.), Vol. 2, p. 1237; H. Okada and H. Hasegawa, *Jpn. J. Appl. Phys. 1* 40, 2797 (2001).
353. F.-J. Ahlers, V. A. Krupenin, S. V. Lotkhov, J. Niemeyer, D. E. Presnov, H. Scherer, T. Weimann, H. Wolf, and A. B. Zorin, in "1996 Conference on Precision Electromagnetic Measurements Digest," 17–21 June 1996, Braunschweig, Germany (A. Braun, Ed.), p. 507; A. B. Zorin, F.-J. Ahlers, J. Niemeyer, T. Weimann, H. Wolf, V. A. Krupenin, and S. V. Lotkhov, *Phys. Rev. B* 53, 13682 (1996); H. Wolf, F. J. Ahlers, J. Niemeyer, H. Scherer, T. Weimann, A. B. Zorin, V. A. Krupenin, S. V. Lotkhov, and D. E. Presnov, *IEEE Trans. Instrum. Meas.* 46, 303 (1997).
354. V. A. Krupenin, D. E. Presnov, and M. N. Savvateev, *J. Appl. Phys.* 84, 3212 (1998).
355. M. Furlan, T. Heinzel, B. Jeanneret, and S. V. Lotkhov, *J. Low Temp. Phys.* 118, 297 (2000); H.-O. Muller, M. Furlan, T. Heinzel, and K. Ensslin, *Europhys. Lett.* 55, 253 (2001).
356. D. L. Klein, R. Roth, A. K. L. Lim, A. P. Alivisatos, and P. L. McEuen, *Nature* 389, 699 (1997).
357. S. Altmeyer, A. Hamidi, B. Spangenberg, and H. Kurz, *J. Appl. Phys.* 81, 8118 (1997).
358. J. Pettersson, P. Wahlgren, P. Delsing, D. B. Haviland, T. Claeson, N. Rorsman, and H. Zirath, *Phys. Rev. B* 53, R13272 (1996).
359. E. H. Visscher, S. M. Verbrugh, J. Lindeman, P. Hadley, and J. E. Mooij, *Appl. Phys. Lett.* 66, 305 (1995).
360. E. M. Ford and H. Ahmed, *J. Vac. Sci. Technol. B* 16, 3800 (1998).
361. T. Weimann, H. Scherer, H. Wolf, V. A. Krupenin, and J. Niemeyer, *Microelectron. Eng.* 41–42, 559 (1998).
362. K. Matsumoto, in "Proceedings of the Symposium on Light Emitting Devices for Optoelectronic Applications and 28th State-of-the-Art Program on Compound Semiconductors," 3–8 May 1998, San Diego, CA (H. Q. Hou, R. E. Sah, S. J. Pearton, F. Ren, and K. Wada, Eds.), p. 68.
363. T. Kikutani, N. Aoki, C. U. Hong, H. Hori, and S. Yamada, *Physica B* 249–251, 513 (1998).
364. K. Flensberg, A. A. Odintsov, F. Liefrink, and P. Teunissen, *Int. J. Mod. Phys. B* 13, 2651 (1999).
365. A. Iwasa, A. Fukushima, A. Sato, Y. Sakamoto, and T. Endo, in "Conference on Precision Electromagnetic Measurements Digest," 14–19 May 2000, Sydney, Australia (J. Hunter and L. Johnson, Eds.), p. 593.
366. T. Altebaeumer and H. Ahmed, *J. Appl. Phys.* 90, 1350 (2001); T. Altebaeumer and H. Ahmed, *Jpn. J. Appl. Phys. 1* 40, 80 (2001); T. Altebaeumer and H. Ahmed, *Jpn. J. Appl. Phys. 1* 41, 2694 (2002).
367. A. Iwasa, A. Fukushima, and A. Sato, *Jpn. J. Appl. Phys. 1* 40, 6645 (2001).

368. O. Astafiev, V. Antonov, T. Kutsuwa, and S. Komiyama, in "Device Research Conference Digest," 25–27 June 2001, Notre Dame, IN, p. 145.
369. T. Fujisawa and Y. Hirayama, *Jpn. J. Appl. Phys. 1* 39, 2338 (2000).
370. A. N. Korotkov and M. A. Paalanen, *Appl. Phys. Lett.* 74, 4052 (1999).
371. N. J. Stone and H. Ahmed, *Appl. Phys. Lett.* 77, 744 (2000).
372. T. Morie, T. Matsuura, S. Miyata, T. Yamanaka, M. Nagata, and A. Iwata, *Superlattices Microstruct.* 27, 613 (2000).
373. M. J. Yoo, T. A. Fulton, H. F. Hess, R. L. Willett, L. N. Dunkleberger, R. J. Chichester, L. N. Pfeiffer, and K. W. West, *Science* 276, 579 (1997); M. J. Yoo, T. A. Fulton, H. F. Hess, R. L. Willett, L. N. Dunkleberger, R. J. Chichester, L. N. Pfeiffer, and K. W. West, *Physica E* 3, 8 (1998); H. F. Hess, T. A. Fulton, M. J. Yoo, A. Yacoby, L. N. Pfieffer, and K. W. West, in "Proceedings of the 24th International Conference on the Physics of Semiconductors," 2–7 August 1998, Jerusalem, Israel (D. Gershoni, Ed.), pp. 11–14.
374. D. Berman, N. B. Zhitenev, R. C. Ashoori, H. I. Smith, and M. R. Melloch, *J. Vac. Sci. Technol. B* 15, 2844 (1997); D. Berman, R. C. Ashoori, and H. I. Smith, in "Proceedings of Quantum Devices and Circuits," 4–7 June 1996, Alexandria, Egypt (K. Ismail, S. Bandyopadhyay, and J. P. Leburton, Eds.), p. 217.
375. Y. Y. Wei, J. Weis, K. von Klitzing, and K. Eberl, *Physica E* 1, 135 (1997); Y. Y. Wei, J. Weis, K. von Klitzing, and K. Eberl, *Physica B* 249–251, 496 (1998); J. Weis, Y. Y. Wei, and K. von Klitzing, *Physica E* 3, 23 (1998); J. Weis, Y. Y. Wei, and K. von Klitzing, *Physica B* 256–258, 1 (1998); J. Weis, Y. Y. Wei, and K. von Klitzing, *Microelectron. Eng.* 47, 17 (1999).
376. K. Ishibashi, T. Ida, H. Kotani, Y. Ochiai, T. Sugano, and Y. Aoyagi, *Microelectron. Eng.* 47, 185 (1999).
377. G. Austing, Y. Tokura, T. Honda, S. Tarucha, M. Danoesastro, J. Janssen, T. Oosterkamp, and L. Kouwenhoven, *Jpn. J. Appl. Phys. 1* 38, 372 (1999).
378. H. Sato, S. Katsumoto, and Y. Iye, *Physica B* 249–251, 453 (1998).
379. P. J. Hakonen, J. M. Ikonen, U. Parts, J. S. Penttila, L. R. Roschier, and M. A. Paalanen, *J. Appl. Phys.* 86, 2684 (1999).
380. G. H. Bernstein, I. Amlani, A. O. Orlov, C. S. Lent, and G. L. Snider, *Nanotechnology* 10, 166 (1999).
381. D. S. Duncan, C. Livermore, R. M. Westervelt, K. D. Maranowski, and A. C. Gossard, *Appl. Phys. Lett.* 74, 1045 (1999).
382. C.-T. Liang, J. E. F. Frost, M. Pepper, D. A. Ritchie, G. A. C. Jones, *Solid State Commun.* 105, 109 (1998).
383. M. Holzmann, D. Tobben, and G. Abstreiter, *Appl. Surf. Sci.* 102, 230 (1996).
384. D. J. Paul, B. Coonan, G. Redmond, B. J. O'Neill, G. M. Crean, B. Hollander, S. Mantl, I. Zozoulenko, K.-F. Berggren, J.-L. Lazzari, F. Arnaud d'Avitaya, and J. Derrien, *Thin Solid Films* 336, 130 (1998); D. J. Paul, *Thin Solid Films* 321, 172 (1998); D. J. Paul, *Adv. Mater.* 11, 191 (1999).
385. M. Turek and K. A. Matveev, *Phys. Rev. B* 65, 115332/1 (2002).
386. K. G. Anil, I. Eisele, and S. Mahapatra, *Appl. Phys. Lett.* 78, 2238 (2001).
387. K. Schwab, *Appl. Phys. Lett.* 80, 1276 (2002).
388. B. J. Hinds, T. Yamanaka, and S. Oda, *J. Appl. Phys.* 90, 6402 (2001).
389. A. M. Chang, *Phys. Scr. T* 90, 16 (2001).
390. D. E. Grupp, T. Zhang, G. J. Dolan, and N. S. Wingreen, *Phys. Rev. Lett.* 87, 186805/1 (2001).
391. P. Nordlander, N. S. Wingreen, Y. Meir, and D. C. Langreth, *Phys. Rev. B* 61, 2146 (2000).
392. L. P. Rokhinson, L. J. Guo, S. Y. Chou, and D. C. Tsui, *Phys. Rev. B*, 60, R16319 (1999).
393. D. Goldhaber-Gordon, J. Gores, H. Shtrikman, D. Mahalu, U. Meirav, and M. A. Kastner, *Mater. Sci. Eng. B* B84, 17 (2001).
394. K. Tada, Y. Fujii, and K. Taniguchi, *Jpn. J. Appl. Phys. 1* 40, 2050 (2001).
395. H. Kim and H. Sakaki, *Physica E* 7, 435 (2000).
396. R. H. Blick and H. Lorenz, in "2000 IEEE International Symposium on Circuits and Systems. Emerging Technologies for the 21st Century, Proceedings," 28–31 May 2000, Geneva, Switzerland, Vol. 2, p. 245.
397. J. Gores, D. Goldhaber-Gordon, S. Heemeyer, M. A. Kastner, H. Shtrikman, D. Mahalu, and U. Meirav, *Phys. Rev. B* 62, 2188 (2000).
398. S. D. Lee, K. S. Park, J. W. Park, Y. M. Moon, J. B. Choi, K.-H. Yoo, and J. Kim, *Appl. Phys. Lett.* 77, 2355 (2000).
399. B. E. Kane, N. S. McAlpine, A. S. Dzurak, R. G. Clark, G. J. Milburn, H. B. Sun, and H. Wiseman, *Phys. Rev. B* 61, 2961 (2000).
400. A. Fujiwara, Y. Takahashi, and K. Murase, *Phys. Rev. Lett.* 78, 1532 (1997).
401. Y. Takahashi, S. Horiguchi, A. Fujiwara, and K. Murase, *Physica B* 227, 105 (1996).
402. S. M. Verbrugh, M. L. Benhamadi, E. H. Visscher, and J. E. Mooij, *J. Appl. Phys.* 78, 2830 (1995).
403. A. Forster, M. Griebel, M. Indlekofer, and H. Luth, *Physica E* 2, 502 (1998).
404. O. Tageman, *Low Temp. Phys.* 25, 214 (1999).
405. A. M. Rudin, L. J. Guo, L. I. Glazman, and S. Y. Chou, *Appl. Phys. Lett.* 73, 3429 (1998); L. J. Guo and S. Y. Chou, *Electron. Lett.* 34, 1030 (1998).
406. A. N. Korotkov, *Appl. Phys. Lett.* 72, 3226 (1998).
407. H. S. Lee, Y. S. Yu, and S. W. Hwang, *J. Korean Phys. Soc.* 33, S266 (1998).
408. Y. Luo, C. K. Wang, and Y. Fu, *J. Chem. Phys.* 117, 10283 (2002).

Semiconductor Quantum Dots

Lucjan Jacak, Arkadiusz Wójs, Paweł Machnikowski

Wrocław University of Technology, Wrocław, Poland

CONTENTS

1. Introduction
2. Manufacturing Methods
3. Spectroscopy of Quantum Dots
4. Main Properties of Quantum Dots
5. Applications
Glossary
References

1. INTRODUCTION

The demands for faster operation and less heat dissipation are among the major factors determining the development of modern electronics. The trend toward miniaturization, dominating the technological progress of the past decades, corresponds to these needs. There is, however, a fundamental size limit beyond which the behavior of a physical system is no longer governed by classical laws. This limit, the border of the quantum realm, is reached when the sizes of elements become comparable to atomic dimensions or the number of electrons involved is just a few. In the case of semiconductor heterostructures these ultrasmall elements can be considered as more or less regular boxes for carriers, in particular for electrons, holes, or excitons. As the effective mass of electrons in many semiconductors is considerably smaller than the free electron mass, the quantization effects in these materials are significant for spatial confinement of nanometer scale.

Actually, the size of individual elements making up computer microchips, currently produced by photolithography techniques, is usually not smaller than the shortest wavelength of visible light ($\sim$380 nm). Lithography using ultraviolet or X-rays still encounters strong technical complications and is not yet acceptable for wide commercial use. Also electron-beam or ion-beam lithography, allowing for far higher resolution, is inconvenient for industrial production as they resemble rather rewriting by pencil in comparison to the photocopy-like performance of photolithography. Thus, typical single microstructures (e.g., transistors in microchips) are still rather far from the range of quantum mechanics, and classical intuitions about physical mechanisms are roughly applicable for their description. The same holds true for currently fabricated elements for so-called nanotechnology—news on nanoengines and other ultraprecise mechanical devices concern objects of micrometer rather than really nanometer scale; hence they fall again into the classical range.

On the other hand, the rapid development of the semiconductor technology, including epitaxy techniques for preparation of even single-molecular layers of materials, lithography methods, and advances in controlling self-aggregation processes, results in quite new possibilities for artificial creation of ultrasmall physical systems with controlled properties. The performance of these nanometer-scale objects meets with limitations imposed on the small systems by quantum mechanics and description of the processes taking place on these extremely small scales taken from the macroscopic physics is irrelevant when both their optoelectronic properties as well as mechanical behavior are concerned. This scale of miniaturization seems to be the natural limit for classical microelectronics.

For systems of nanometer dimensions the classical description is very misleading but their quantum properties open the possibilities both for deep understanding of the fundamental laws of nature and for new applications that could not be thought of on the classical grounds. Indeed, the practical realization of quasi-two-dimensional electron gas in semiconductor heterostructures, with electrons confined in a few nanometers wide layer (thickness of order of the electron de Broglie wavelength in semiconductor) resulted in such discoveries as the quantum Hall effect, which revealed sharply the unusual quantum nature of geometrically confined systems. The fractional quantum Hall effect and further experiments with so-called Hall metal [1] suggest even an essential development of quantum theory and lead to exotic new two-dimensional (2D) particles—composite fermions instead of simple 3D electrons.

Since the early 1980s the rapid progress in laboratory techniques of confinement of electrons, first in a plane, as in Hall configurations, and further in remaining directions, has led to creation of more fine objects called quantum wires [2], in which electrons move freely only in one direction, and finally semiconductor quantum dots (QDs) [3–6]

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

with electrons closed in small boxes of dimensions of tens or even a few nanometers. These completely confined electrons in artificially constructed boxes resemble the ordinary atoms; hence frequently QDs are called artificial atoms, and they reproduce many properties from atomic physics. QDs seem to be, however, much more flexible in comparison with atoms—QDs do not contain nuclei inside, so they can confine an arbitrary number of electrons, starting from one to hundreds or thousands closed in small spatial volume. In addition, as opposed to natural atoms, their shapes and various other properties can be modeled at the stage of preparation.

2. MANUFACTURING METHODS

Unlike quantum wells, where the motion of carriers is restricted to a plane through the crystallization of thin epitaxial layers, the creation of quantum wires or dots, which confine the carriers in the space with at least two of three dimensions limited to the range of the de Broglie wavelength, requires far more advanced technology. Currently there are a lot of methods of manufacturing QDs. Let us briefly describe them (see [7–9]).

2.1. Etching

The earliest method of manufacturing QDs was implemented by Reed et al. [3], who etched them in a structure containing two-dimensional electron gas. Subsequent steps of this process are shown in Figure 1a. The surface of a sample containing one or more quantum wells is covered with a polymer mask, and then partly exposed (1). The exposed pattern corresponds to the shape of the created nanostructure. Due to the required high resolution, the mask is not exposed to visible light, but to the electron or ion beam (electron/ion beam lithography). At the exposed areas the mask is removed (2). Later, the entire surface is covered with a thin metal layer (3). Using a special solution, the polymer film and the protective metal layer are removed, and a clean surface of the sample is obtained, except for the previously exposed areas, where the metal layer remains (4). Next, by chemically etching the areas not protected by the metal mask (5), slim pillars are created, containing the cut-out fragments of quantum wells (6). In this way, the motion of electrons, which is initially confined in the plane of the quantum well, is further restricted to a small pillar with a diameter of the order of 10–100 nm. Due to the simplicity of the production of thin, homogeneous quantum wells, GaAs is the most commonly used material for creating dots by means of etching. Figures 1b and 2 (upper) present the pictures of real dots obtained using this method.

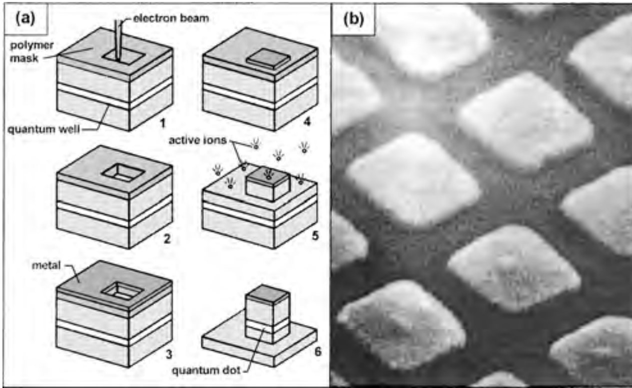
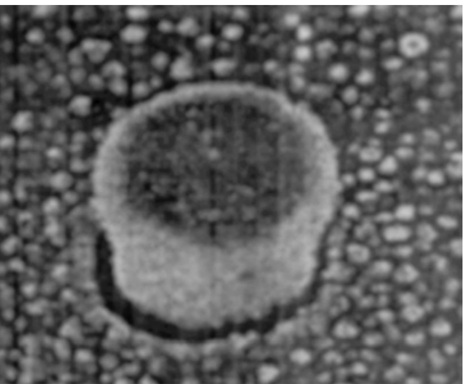


Figure 1. (a) Process of quantum dot etching. Adapted with permission from [219], M. A. Reed, *Sci. Am.* 3, 40 (1993). © 1993, SCI American. (b) Etched quantum dots in GaAs/AlGaAs well, electron scanning microscope picture. Reprinted with permission from [96], T. Smith et al., *Phys. Rev. B* 38, 2172 (1988). © 1988, American Physical Society.

2.2. Modulated Electric Field

Another method consists in the creation of small electrodes over the surface of a quantum well by means of lithographic techniques. Application of an appropriate voltage to the electrodes produces a spatially modulated electric field, which localizes the electrons within a small area. The lateral confinement created in this way has no edge defects, which are characteristic for the etched structures. The process of spreading a thin electrode over the surface of a quantum well may produce both single QDs [10] and large arrays (matrices) of dots [11–14].

Modulation of electric potential, produced by applying a voltage to an electrode, can be realized by the previous preparation (using a lithographic technique) of a regular array of islets of nonmetallic material (e.g., of the barrier material) on the surface of the sample. As a result, the distance between the electrode (overlaying the surface with the islets) and the plane of the quantum well is modulated, and



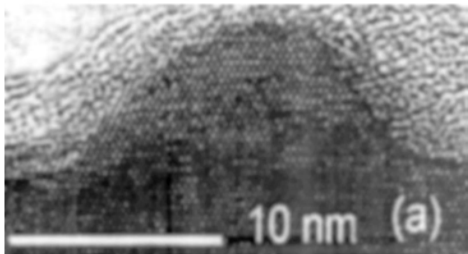


Figure 2. Micrographs of single dots. Upper—etched. Reprinted with permission from [154], S. Tarucha et al., *Phys. Rev. Lett.* 84, 2485 (2000). © 2000, American Physical Society. Lower—self-assembled. Reprinted with permission from [220], P. Fry et al., *Phys. Rev. Lett.* 84, 733 (2000). © 2000, American Physical Society.

the electrons are bound in small areas under the prepared islets. A photograph of a matrix of such dots, together with the profile of the confining potential, is shown in Figure 3a.

Instead of modulating the distance between the electrode and the well, it is also possible to build a pair of parallel, thin electrodes above the well. The lower one can have regularly placed holes which is where QDs are to be created [15, 16]. A voltage applied to the pair of electrodes results in the change of both the dot size and the depth of the confining potential. The potential depth influences the number of confined electrons. However, when an additional electrode is introduced between the quantum-well layer and the doped layer, the number of electrons and the potential depth can be changed independently. Even more complicated systems of microelectrodes are currently also applied [17].

2.3. Interdiffusion between the Barrier and the Quantum Well

QDs can be obtained also by local heating of a quantum well sample with a laser beam [18]. The laser beam guided along a rectangular contour surrounding an unilluminated area of a diameter of 300–1000 nm causes a rapid interdiffusion of atoms between the well and the barriers, which creates a local modulation of the material band structure (i.e., the potential barrier) which surrounds the unilluminated interior of the rectangle. For larger dimensions of the illuminated rectangle the obtained effective potential which confines the electrons is flat inside the rectangle and steep near the edge. With the decrease of the illuminated rectangle, the confined area shrinks. According to the authors of [18], for dimensions near 450 nm the effective potential confining electrons is close to an isotropic parabola. However, it should be mentioned that details of the electron confining potential in a QD of any type cannot be measured directly (except for the geometric dimensions) and are

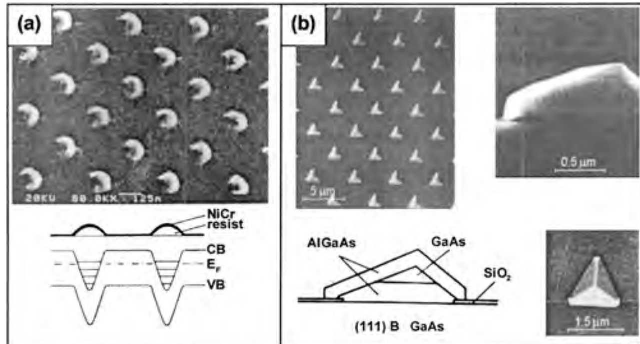


Figure 3. (a) Quantum dots produced on InSb; electrons confined by the electric field (electron scanning microscope picture). Bottom: shape of the electrode and the configuration of band edges (valence and conduction bands). Reprinted with permission from [14], C. A. Sikorski et al., *Phys. Rev. Lett.* 62, 2164 (1989). © 1989, American Physical Society. (b) Quantum dots created on the surface of GaAs in the selective MOCVD growth (scanning electron microscope pictures); configuration of layers in a single dot is given. Width of the electron localization area at the top of the pyramid is ca. 100 nm. Reprinted with permission from [20], T. Fukui et al., *Appl. Phys. Lett.* 58, 2018 (1991). © 1991, American Institute of Physics.

alternatively obtained through an interpretation of various indirect effects, which are related to the electronic structure of a QD.

2.4. Semiconductor Microcrystals

It is also possible to create QDs in the form of semiconductor microcrystals. In the first experiment based on this idea, carried out by Ekimov et al. [19], silicate glass with ca. 1% addition of the semiconducting phase (CdS, CuCl, CdSe, CuBr) was heated for several hours at high temperatures, which led to the formation of appropriate microcrystals of almost equal sizes, depending on the temperature and heating time. Radii of these dots measured in different samples varied in the range of 1.2–38 nm.

2.5. Selective Growth

The next method is the selective growth of a semiconducting compound with a narrower bandgap (e.g., GaAs) on the surface of another compound with a wider bandgap (e.g., AlGaAs) [20]. Restriction of growth to chosen areas is obtained by covering the surface of the sample with a mask (SiO_2) and etching miniature triangles on it. On the surface not covered with the mask the growth is then carried out with the metal-organic chemical vapor deposition method (MOCVD), at a temperature of 700–800 °C. The crystals created have a shape of tetrahedral pyramids and hence, when the first crystallized layers are the layers of the substrate compound (AlGaAs), and only the top of the pyramid is created of GaAs, it is possible to obtain a dot of an effective size below 100 nm. Pictures of such dots, and the configuration of layers GaAs/AlGaAs, are shown in Figure 3b.

2.6. Cleaved Edge Overgrowth

Quantum dots may appear at the intersection of three orthogonal quantum wells [21]. The technology consists of growing subsequent quantum wells on a cleaved edge of a previously existing well [22, 23]. In this way, localization of carriers on a scale of a few nm may be achieved.

2.7. Interface Fluctuations

Thickness fluctuations of a quantum well induce effective potential fluctuation and appearance of localization centers for carriers [24, 25]. Such “natural” quantum dots have low localization energy (~10 meV) and large size fluctuations. Photoluminescence data prove the existence of zero-dimensional density of states, conforming with a model of 1 monolayer thick islands of lateral extension of several tens of nanometers [26].

2.8. Self-Organized Growth

Another very promising method is the self-crystallization of QDs [27]. When lattice constants of the substrate and the crystallized material differ considerably (7% in the case of GaAs and InAs, the most commonly used pair of compounds), only the first deposited monolayers crystallize in the form of epitaxial, strained layers with the lattice constant

equal to that of the substrate. When the critical thickness is exceeded, a significant strain which occurs in the layer leads to the breakdown of such an ordered structure and to spontaneous formation of randomly distributed islets of regular shape and similar sizes, called self-assembled quantum dots (SADs). The shape and average size of islets depend on the strain intensity in the layer related to the misfit of lattice constants, the temperature at which the growth occurs, and the growth rate. The phase transition from the epitaxial structure to the random arrangement of islets is called the Stranski–Krastanow transition [28].

Figure 4 presents subsequent phases of the creation of islets for a pair of compounds with the misfit of lattice constants: InAs and GaAs, where the transition occurs at the 1.8 monolayer deposition. In Figure 2 the micrograph of a single self-assembled dot is presented.

When the process of crystallization is terminated shortly after reaching the phase transition, islets evolve to a quasi-equilibrium state in which they assume the shape of pyramids [29–32] or flat, circular lenses [33–35] formed on a thin layer of InGaAs (the wetting layer).

Raymond et al. [35] report growth of self-assembled dots in the shape of lenses with ~36 nm diameter and ~4.4 nm height (with fluctuations of 5–10%). Marzin et al. [32] obtained dots in the shape of regular pyramids with a square base of ~24 nm side length and ~2.8 nm height (with fluctuations of ~15%), and a distance between neighboring dots of ~55 nm. Ultrasmall pyramidal dots of less than 10 nm dimensions were also obtained in a InAs/GaAs interface [29].

Small sizes of the SADs, homogeneity of their shapes and sizes in a macroscopic sample, perfect crystal structure without edge defects, as well as the convenient growth process, without the necessity of any precise deposition of electrodes or etching, are among their greatest advantages which bring hope regarding their future application in electronics and optoelectronics.

More examples of quantum dots obtained by different techniques from different materials are given in Table 1.

3. SPECTROSCOPY OF QUANTUM DOTS

3.1. Optical Spectroscopy

Three-dimensional confinement of carriers and the resulting δ-shaped density of states manifests itself in the appearance of sharp lines in the optical spectra of quantum dots.

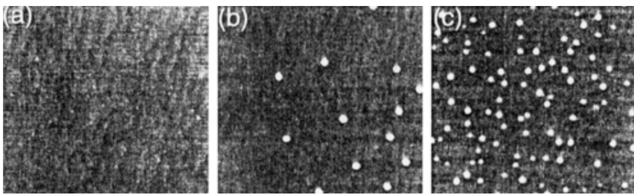


Figure 4. Evolution of islets—self-assembled quantum dots (white circles), in the molecular beam epitaxy growth of InAs on GaAs surface. Subsequent pictures correspond to the increasing coverage of 1.6, 1.7, and 1.8 monolayer; size of presented areas: 1×1 μm^2. Reprinted with permission from [27], P. Petroff et al., *Superlattices Microstructures* 15, 15 (1994). © 1994, Elsevier Science.

In a photoluminescence (PL) measurement, the emission intensity from a sample is measured as a function of the detection energy. This PL radiation from QDs is caused by recombination of trapped electron–hole pairs (i.e., excitons; see Section 4.2).

PL experiments on etched dots [36–39] show a peak at the energy corresponding to the interband transition, increasing with decreasing dot diameter. At higher excitation energy another (redshifted) line appears, due to transition involving a biexciton state. Appearance of further (blueshifted) lines accompanies the saturation of the single-exciton lines and corresponds to recombination from higher excitonic levels. These lines split in nonzero magnetic field due to lifting the degeneracy of the levels [35].

Self-assembled dots have much stronger quantization effects (larger interlevel distance) due to their reduced size. This should lead to more clear photoluminescence features. The PL measurements on such dots (InAs/GaAs [40–44], InGaAs/GaAs [34, 45], InAlAs/AlGaAs [46]) show a distinct feature, shifted by a few hundreds of meV (the exciton binding energy) from the wetting layer. The excited state recombination appears at higher excitation power and is shifted by a few tens of meV, indicating large energy separation from the closest bright exciton state (a state with allowed radiative transition).

The resolution of the PL spectra is limited by the nonuniform size of the dots in the investigated ensemble, leading to inhomogeneous broadening of spectral peaks. This may be overcome by using a modified technique, spot focus cathodoluminescence [47], or by preparing small mesas containing a small number of dots [32, 48]. For decreasing number of dots [33], the initial wide Gaussian peak evolves into a series of densely packed narrow lines separated by an interval much smaller than the calculated [32, 49] interlevel spacing, hence corresponding to the ground-state transition from different dots. On the grounds of such spectra, it is possible to evaluate the homogeneous broadening of spectral lines due to the intrinsic decoherence processes.

The PL spectra not only show the structure of the ground and excited zero-dimensional states but also provide information on relaxation processes. Higher excited states appear before saturation of the ground state transition which indicates a slowdown of relaxation [44, 50].

An alternative experimental technique, of not less importance than PL, is photoluminescence excitation (PLE). In this method, the detection energy is fixed and the luminescence intensity is investigated as a function of the excitation energy. In this way, one can study the capture of carriers and their relaxation to the optically active state (usually ground state). It turns out [51, 52] that PLE spectra of ensembles of dots contain a series of longitudinal optical (LO) phonon resonances instead of the excited dot state. This results from the fact that apart from the existence of an optically active state at the detection energy and a nonvanishing density of states at the excitation energy (which is guaranteed by the size and shape fluctuation in the dot ensemble) an efficient relaxation mechanism must be provided which is the case only around the LO phonon energy [53, 54] (bottleneck effect; see Section 4.6 for a discussion).

Table 1. Selected examples of quantum dots.

Fabrication method	Material	Size	ΔE	Ref.
Cleaved edge overgrowth	GaAs/AlGaAs	$7 \times 7 \times 7$ nm		[22]
Interface fluctuations	GaAs/AlGaAs	~100 nm	6 meV	[26]
Interface fluctuations	GaAs/AlAs	~100 nm		[89]
Interface fluctuations	GaAs/AlGaAs		2 meV	[55]
Electric field	Si	150 nm		[15]
Selective growth	GaAs/AlGaAs	<100 nm		[20]
Selective growth	GaAs/AlGaAs	$25 \times 25 \times 15$ nm		[215]
Etching	InAs/GaAs	40–90 nm	4–9 meV	[38]
Etching	InGaAs/AlGaAs	500 nm	4 meV	[154]
Etching	Si/SiO_2	$7 \times 7 \times 2$ nm	30 meV	[212]
Interdiffusion	GaAs/AlGaAs	300–1000		[18]
Self-organization	InAs/GaAs		80–90 meV	[41]
Self-organization	CdMnSe/ZnSe	8–10 nm		[216]
Self-organization	InAs/GaAs	17–19 nm	60 meV	[188]
Self-organization	InGaAs/GaAs		100	[58]
Self-organization	CdSe/ZnSe	10×1.6 nm	40 meV	[55]
Self-organization	GaN/AlGaN	40×6 nm		[217]
Self-organization	Ge/Si	150×15 nm		[218]

Note: ΔE is the confined energy level separation

3.2. Single Dot Optical Spectroscopy

Improvement in the sample preparation and optical focusing techniques leads to the possibility of single dot spectroscopy [55, 56]. A spatial resolution high enough to probe a single dot may be achieved by near-field optical techniques [57]. Such single-dot experiments provide much more detailed insight into the states of the confined carriers and their interaction with environment.

The single dot optical spectra show features related to the shell structure of the carrier levels [58]. Polarization of the PL signal has been studied [59]. Large self-assembled dots prove to be optically anisotropic [60]. By breaking the dot symmetry with magnetic field it is possible to get information also on dark exciton states [61]. Single dot spectroscopy allows one to study in detail the structure of few-particle states (charged excitons and multiexciton states) [62–69] and the dependence of the spectrum on the number of confined carriers [70]. Additional information about the spectrum (fine structure, few-particle effects) is provided by measurements in magnetic fields (magnetospectroscopy) [71–73]. It turns out that emission from the biexciton ground shell is correlated with the value of the g factor which suggests the effect of dopants on the dot spectrum and its magnetic properties [74, 75]. A nearby defect increases also the coupling to LO phonons and in this way affects the phonon-assisted transitions [76]. Information on the mechanisms of creation of biexciton states has been obtained [77, 78]. Spectra of single "artificial molecules" (coupled dot systems) are also investigated [79].

Homogeneous broadening of single-dot spectral lines yields information on the lifetime of the states and the effects of coupling to the environment, including relaxation mechanisms [57, 80–85]. A direct insight into the system evolution and coherence times is possible by time-resolved spectroscopy methods [56, 86–89]. It was suggested that the linewidth of quantum dots may be to some extent controlled [90].

Single dot experiments on single-photon level yield data on statistics (correlations) of emitted photons [91–94].

3.3. Capacitance Spectroscopy

In capacitance spectroscopy experiments, the capacitance of a quantum dot system is measured as a function of the bias voltage applied. As the capacitance is proportional to the density of states of the dot, the measured differential signal shows oscillations corresponding to the energies of confined carriers [11, 95, 96]. The measured energy differences consist of energy differences of single-particle states and charging energies, related to carrier–carrier repulsion (Coulomb blockade effect).

A more sophisticated variation of this method, single-electron capacitance spectroscopy, allows one to observe ground state energies of multielectron dots by adding electrons one-by-one (with changing the gate bias voltage) to the initially empty dot and measuring the single-electron charging energy (chemical potential) [10, 17, 97, 98].

As the optical methods (far infrared absorption) probe only the center of mass mode of the confined multielectron system (Kohn theorem; see Section 4.3), the capacitance spectroscopy methods are an important tool for investigating the properties of large, multielectron QDs. It is, however, employed also as a complementary method to study the spectral properties of small, self-assembled dots [99–105]. It may be applied to study capacitance associated not only with electrons but also with holes [106, 107]. The time-resolved version of the capacitance spectroscopy has been applied to study the electron escape processes from a InAs quantum dot in an electric field [108], demonstrating the presence of both tunneling and thermally activated processes.

3.4. Transport Spectroscopy

The idea of transport spectroscopy consists in measuring the tunneling current through a dot or a system of dots as a function of the bias voltage applied by external electrodes.

The tunneling current peaks reveal information on the dot spectra and on the charging effects for multicarrier states.

The measurements in a vertical setup (current along the strongest confinement axis) on etched dots in the charge accumulation regime (reverse bias, emitter barrier more transparent than the collector one) [109–114] showed single-electron charging features and revealed information on the many-particle ground state. Other experiments performed on a similar structure, with an additional side electrode, yielded information about many-particle states of a controlled number of confined carriers [115–117].

Spectra of InAs/GaAs self-assembled quantum dots have been obtained by vertical tunneling spectroscopy with forward bias [118, 119]. The results show distinct peaks corresponding to the confined single-electron states in spite of a large number of dots in the mesa (only a few dots from the ensemble contribute).

Vertical tunneling measurements were performed also on single self-assembled dots (Si/SiO_2 [120], InAs/GaAs in a mesa structure [121], ultrasmall Si dots [122], PbS by scanning tunneling spectroscopy [123]).

The spectroscopy in the lateral setup was applied to dots defined by electric field gating in a 2D electron gas to study spectra of the 0D states, Coulomb interactions between carriers, and various tunneling channels [124–128]. Transport measurements in this setup have been performed also on 2D electron gas in close contact with a sheet of self-assembled dots [129, 130].

With transport spectroscopy on various materials it was possible to detect maximum density droplet states in many-electron quantum dots [131], to probe the confined carrier wavefunctions [132], to identify the contribution from different confined states [133] and from elastic and inelastic tunneling [134], to test the coherence of charge transfer [135], and to estimate g factors [136]. By using spin-polarized injection, not only Coulomb blockade but also spin blockade effects may be studied [137].

4. MAIN PROPERTIES OF QUANTUM DOTS

In this section we describe the essential properties of QDs. A more exhaustive description may be found in reviews [7–9, 138, 139].

4.1. Single-Particle Spectrum

Owing to full confinement of carriers inside the dots of dimensions of order of 10–100 nm, their properties correspond to discrete quantum levels separated typically with a few or tens of meV in energy (due to usually smaller effective mass of electron in semiconductors in comparison to free electron mass). The discretization of the energy spectrum enhances sharpness of the density of states. In Figure 5, the densities of states corresponding to the reduction of the dimensionality of the systems are presented.

The sharp density of states of QDs makes these systems very attractive for laser applications. And indeed, the first reported QD laser constructions with application of pyramid-shaped InAs/GaAs ultrasmall self-assembled dots

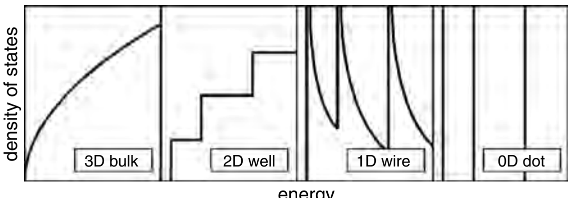


Figure 5. Density of states as a function of energy in systems with different numbers of spatial dimensions: 3D—bulk material, 2D—quantum well, 1D—quantum wire, 0D—quantum dot.

[30, 31] and with lens-shape InAs/InGaAs self-assembled dots [140] revealed better parameters in comparison with even best lasers built on the basis of quantum wells. It concerned both lower threshold injection current and higher temperatures of operating—due to the so-called bottleneck effect corresponding to a significant decrease of the interaction with phonons in the case of confined carriers in comparison to their counterparts in bulk.

At low enough temperatures (several Kelvin) the energy of phonons is too low to excite the electrons in QDs and the strong quantization of energy determines the electronic properties of these systems. The confining potential of QDs is not singular, unlike the Coulomb potential in atoms, and could be expanded in a series with respect to radius—the first term of this expansion is the parabolic potential and it is commonly applied for model investigation of QDs. An advantage of this model potential consists of the possibility of analytical solution of single-electron problem in 2D parabolic confinement even in the presence of perpendicular static magnetic fields. The resulting one-particle states, known as Fock–Darwin levels [141, 142], are presented in Figure 6. The degeneracy of the energy levels in the absence of magnetic field gives rise to the shell structure of the QD spectrum which is present also in real (not necessarily parabolic) systems [100]. In the limit of high magnetic fields

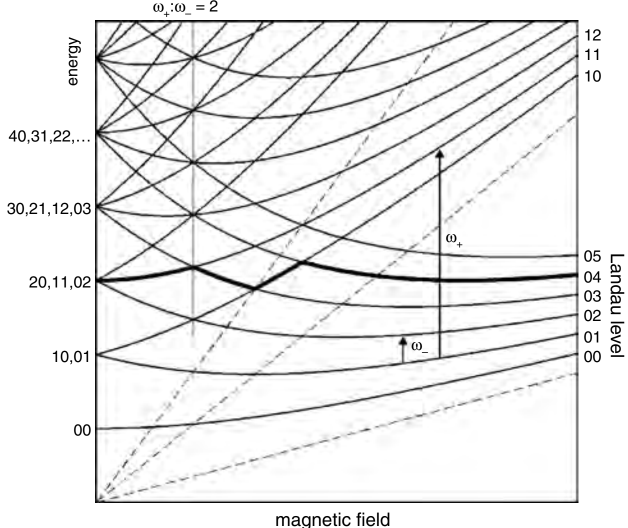


Figure 6. Evolution of the Fock–Darwin energy levels in a magnetic field. Dashed lines—the Landau energy levels; vertical arrows—allowed optical transitions.

the Fock–Darwin states transform into ordinary Landau levels of a free 2D electron in magnetic field, but for lower fields the modifications of Landau levels due to confinement are significant.

In real QDs the lateral confinement potential is a finite well (thus the parabolic approximation holds only for the lowest levels) and for sufficiently high energy originally captured carriers can leave the dot. To analyze such phenomena, the finite well of Gaussian type is usually applied [143], even though for a particular shape of a quantum dot [e.g., pyramid shape for InAs/GaAs self-assembled dots or lens-shaped In(Ga)As/GaAs self-assembled dots] more realistic finite wells are considered. The common feature of all these models is, however, the nonsingular character of the confining potential, which assures that the particularities of the potential well used for modeling do not cause any important qualitative changes.

In a class of semiconductor quantum dots not only electrons but also holes can be trapped (such QDs, like self-assembled InAs/GaAs dots, are sometimes called type I dots, in contrast to the type II dots, e.g., defined by electric field focusing, which have binding potential for only one kind of carrier and repulsive for the other). In contrast to conduction band electrons, a realistic description of valence band holes requires the inclusion of a larger number of sub-bands. The valence band is built from atomic p-type orbitals, and thus its states carry an internal angular momentum equal to unity. In circular dots both the angular momentum length and its projection in the vertical direction are good quantum numbers for a single captured hole, and its states can be assigned to heavy-hole and light-hole sub-bands (also spin-orbit split-off sub-band). In the simplest approach (the effective mass approximation) all sub-bands can be treated independently (it has been suggested that this model may be unexpectedly accurate [144]), but a more accurate description of holes requires taking inter-sub-band mixing into account. The mixing effects increase for larger magnitudes of the hole wave vector and for small dots with a high value of averaged wave vector the sub-band mixing corrections are significant. In self-assembled, strain induced dot structures the inter-sub-band interaction is, however, weakened due to sub-band separations induced by strain inherent to the manufacturing method [145].

4.2. Excitons

Optical excitations in QDs lead to the creation of pairs of particles (electron and hole) which may form charge-neutral states (i.e., excitons). If the confinement potential is attractive for both types of carriers (as it is, e.g., for the self-assembled InAs/GaAs dots) the single-particle shell structure is reflected also in the exciton spectrum, both in the parabolic confinement model [54] and in a more realistic modeling [146]. The hybridization of the nearly degenerate levels forming one shell affects the optical spectra [147] of such two-carrier states. However, unlike in the single-carrier case, the shells are not equidistant due to the decreasing Coulomb energy gain for higher states. The exciton-like complexes and multiexciton states in quantum dots were also investigated [148, 149].

From the point of view of future optoelectronic applications of QDs understanding of their PL properties is very important. The PL spectrum of QDs is much more complicated in comparison with bulk systems or quantum wells. The PL properties of QDs are analyzed very intensively both experimentally and theoretically for various types of QDs, as these phenomena are utilized, for example, in laser constructions [30, 31, 140].

For QDs, instead of a single PL peak as for a quantum well (or bulk), a multipeak structure is usually observed both in the presence and in the absence of a magnetic field [18, 33, 34, 150]. Even for very weakly activated QDs the doublet of PL features occurs. Mutual positions and intensity of particular PL peaks strongly depend on the dot diameter and the magnetic field [18, 42, 151]. There are many approaches for the theoretical investigation of these phenomena [7].

A simplified theoretical model in terms of metastable states can be formulated within the Hartree approach for electric-field-defined dots [152]. The bare lateral potential of this dot is attractive only for either an electron or a hole, contrary to the case of SADs (e.g., InAs/GaAs) where the same empty dot can capture an electron or a hole. If, however, the electric-field-defined dot is filled with an electron, the hole next could also be trapped due to electron–hole attraction. It resembles the similar phenomenon in the case of a charged acceptor or donor in the presence of magnetic field, where a magnetoexciton (i.e., electron–hole pair) can be trapped by the charged defect. In the absence of magnetic field the electron–hole attraction is too weak to overcome repulsion of such a center and the charged donor/acceptor cannot capture an exciton. The mediating role of the magnetic field is clearly connected with its property to enhance localization for both carriers due to diamagnetic effects. In the case of a QD defined by electric field a localized exciton can be formed even in the absence of magnetic field, because the lateral potential of the dot is nonsingular in distinction from a charged donor/acceptor potential. The simultaneous presence of both types of carriers in a small region leads to a strong mutual modification of this nonsingular lateral potential which results in effective confinement for both the electron and the hole in the dot. Moreover, the effective potential can attain the double-well shape, giving rise to a splitting of the excitonic ground state (also unlike a charged donor/acceptor magnetoexciton). Particular properties of effective potentials for an electron and a hole in a QD depend on the geometrical dimension of the QD—cf. Figure 7 (also on the external magnetic field and the activation level) similarly to what is observed in experiments (cf. Fig. 8).

4.3. Many-Particle Effects

In the case of a higher number of carriers captured in a dot, inclusion of the electric interaction for the many-electron system strongly affects the energy levels (e.g., [115, 153, 154]). Due to the fact that the dimensions of QDs are considerably larger than those of atoms, the Coulomb interaction is of much greater importance here. According to the Heisenberg uncertainty principle, the single-particle excitation energy depends on the size L as $\varepsilon \propto 1/L^2$, while the Coulomb interaction energy behaves as $V_C \propto 1/L$. In small

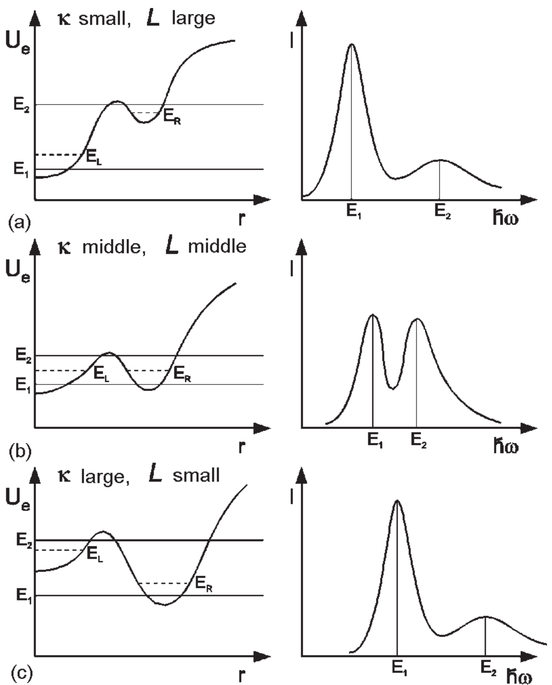


Figure 7. Effective self-consistent potential which acts on the electron U_e (left), and approximated photoluminescence spectrum (right), for three values of κ, that is, the dot size: (a) large dot, $\kappa = 0.4$, (b) medium dot, $\kappa = 0.7$, and (c) small dot, $\kappa = 0.9$ [bare lateral confinement for electron $\sim V \exp(-r^2/L^2)$, $\kappa \sim 1/L^2$].

QDs (e.g., self-assembled dots: $L \sim 20$–40 nm) the relation between ε and V_C is similar to that in atoms, the single-particle levels are relatively strongly separated, and the ground state satisfies the (slightly modified) Hund rules. Thus, while filling the shells of nearly degenerate states at zero magnetic field, ferromagnetic spin alignment is favored. This is, however, not the case for nonzero magnetic field as well as in nonsymmetric (e.g., rectangular) dots, where the shells are split by lifting the degeneracy.

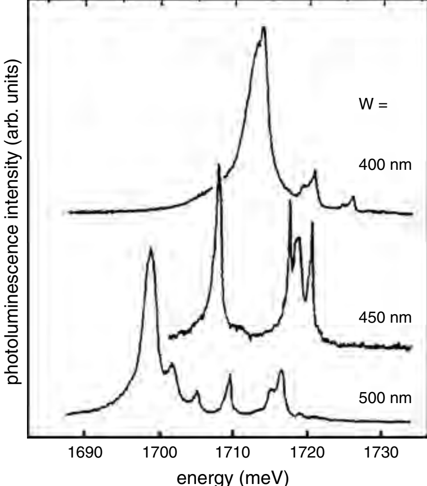


Figure 8. Photoluminescence spectra of single dots with three different diameters; splitting of the PL peak behaves similar to the scheme in Figure 7. Adapted with permission from [151], U. Bockelmann et al., *Solid State Electron.* 37, 1109 (1994). © 1994, Elsevier Science.

On the other hand, in larger dots the interaction between electrons determines the ground state of the system [155–157]. The ground state is a result of the competition between two opposite effects: the Coulomb repulsion favors pushing the electrons apart from one another, whereas the external confinement squeezes electrons in the center of the dot and prevents a large separation between them. This is not the only difference between QDs and ordinary atoms. As the energy scale in atoms is 2–3 orders of magnitude greater than in QDs, the latter are relatively more sensitive to phonons (and also to crystalline disorder) resulting in significant broadening of spectral lines in the QD case (in comparison to atoms). It precludes the usefulness of a simple quantum level picture ("artificial atom" concept) in all the cases when the level separations become comparable with level broadenings.

Under the influence of a magnetic field the electrons are subject to an additional squeezing in the plane perpendicular to the field since the magnetic field term appearing in the Hamiltonian also has the form of parabolic confinement (both for electrons and holes): $1/8m\omega_c^2 r^2$, where ω_c is the cyclotron frequency. In strong magnetic fields, where the magnetic length becomes smaller than the dot diameter (e.g., in a field of magnitude of order of 1 T, the magnetic length is about 25 nm), a dominating single-particle potential is the parabolic Landau potential (of the magnetic field). Thus, following the decrease of the effective dot size with growing magnetic field, the characteristic energy of the electron–electron Coulomb repulsion grows as the square root of the field magnitude. On the other hand, the characteristic kinetic energy scale decreases with an increasing magnetic field. Tuning the magnetic field varies the two energy scales and induces transitions in the ground state.

Detailed theoretical analysis of energy levels including interaction (also in the presence of magnetic field) is done by application of numerical procedures of quasi-exact diagonalization of relevant Hamiltonians or by various versions of Hartree–Fock calculus.

For QDs one can observe a realization of the generalized Kohn theorem—independence of resonant electromagnetic attenuation of electron interaction and number of electrons in the QD. This property corresponds to the separation of the center of mass and relative motion for the electron system in a QD with parabolic confining potential. For far-infrared (FIR) radiation with energy corresponding to interlevel distances in the QD (of order of meV), the wavelength (of order of mm) is far larger than the dot diameter; hence only the center-of-mass dynamics is involved in system response, exactly as for a uniform field. Hence the resonant FIR radiation frequencies correspond only to transitions in the center-of-mass quantum levels of the dot and hence are independent both of interaction and number of electrons. A fair coincidence of experimental data and theoretical predictions for FIR spectroscopy [158, 159]—as presented in Figure 9—is also evidence that the real confinement in QDs is actually close to the parabolic type, for which the separation of the center-of-mass and the relative dynamics occurs.

On the other hand, some features in FIR spectroscopy data, namely the double splitting of frequencies and characteristic anticrossing of resonant lines (cf. Fig. 9a, b, d) occurring for QDs with large number of electrons confined,

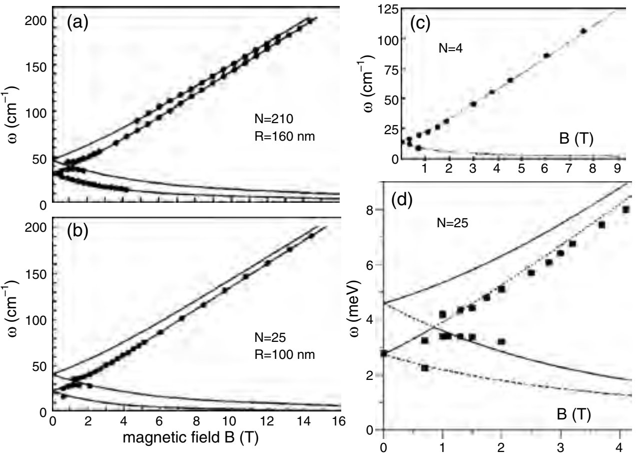


Figure 9. Evolution of the FIR resonance energies for quantum dots in a magnetic field. (a) 210 electrons. (b) 25 electrons. Reprinted with permission from [159], T. Demel et al., *Phys. Rev. Lett.* 64, 788 (1990). © 1990, American Physical Society. (c) 4 electrons. Reprinted with permission from [13], B. Meurer et al., *Phys. Rev. Lett.* 68, 1371 (1992). © 1992, American Physical Society. (d) Dependence of the FIR absorption spectrum of quantum dot containing 25 electrons on the magnetic field. Squares—the experiment. Adapted with permission from [159], *Phys. Rev. Lett.* 64, 788 (1990). © 1990, American Physical Society. Lines—the model with spin–orbit interaction. Adapted with permission from [160], L. Jacak et al., *Physica B* 229, 279 (1997). © 1997, Elsevier Science.

indicate a small breaking of the Kohn theorem (inclusion of spin–orbit interaction [160] or confinement nonparabolicity corrections can explain these effects).

Another simple and unique property of the 2D (i.e., planar) parabolic confinement manifests itself in vanishing of effects caused by in-plane electric field even in the presence of perpendicular magnetic field. In the case of parabolic potential the in-plane electric field can be simply removed by an appropriate gauge transformation, which is, however, not the case for other confinement shapes. Therefore, appearance of electric-field-induced shifts of kinetic angular momentum or magnetization can be treated as an indicator of nonparabolicity of the QD confinement.

The electric interaction between electrons in a QD significantly modifies the relative dynamics of electrons. Very spectacular, although not yet confirmed experimentally, are the so-called magic states of QDs [157, 161–163], revealing the composite fermion nature of 2D electrons in planar QDs in the presence of strong magnetic fields. The crucial role is played here by topological phenomena, similar to the fractional quantum Hall effect. The magic states of strongly interacting electrons in QDs were there identified with the so-called compact states of weakly interacting composite fermions. The transformation of an electron system into a composite-fermion system is performed by attaching to each electron an equal and even number of elementary fluxes of a magnetic field. As shown in the exact numeric calculations [164, 165], the eigenstates of the *interacting* electron system are very close to the respective compact states of *noninteracting* composite fermions. The electronic system described in terms of composite fermions is thus an effectively free (noninteracting) system. It should, however, be stressed that the system of noninteracting composite fermions is essentially a many-body system due to intrinsic nonlocality of topological effects. Each of its quantum states is nonlocal and differs strongly from the antisymmetrized product of single-particle electronic states (i.e., the appropriate Slater determinant).

An interesting property of the confined many-electron system is the observation, in very strong magnetic fields, of characteristic features in the capacitance spectrum at fractional filling factors [16], suggesting a close relation with the fractional Hall effect.

In the recent single-electron capacitance spectroscopy experiment [17], the spatial localization of electron wavefunctions has been observed. It has been shown that at low electron densities, the electrons occupy localized sites in the dot, while for larger number of electrons, the states become delocalized. A similar study has also been done on a two-dot system [166] in magnetic field showing splitting of one low-density droplet into two parts for high fields.

4.4. Coupled Dots

Apart from the properties of a single dot, also systems of two coupled dots (artificial molecules) are of interest for their potential applications. A ground state of two electrons in such a system at zero magnetic field is a singlet. However, the singlet–triplet transition takes place at relatively low magnetic field [167], similar to the case of an electron pair in the same dot [168]. Orbital states of an exciton in a double-dot structure may be mapped onto the entangled states of two spins [169], providing formal grounds for applications in quantum information processing. A coupled two-dot system can be used as an exciton storage device. Even though in single self-assembled QD excitons live only for nanoseconds, in a vertically stacked QD pair obtained by the self-assembled strain-induced technique, the electron–hole pair can be stored even for seconds, provided that the electron and the hole are spatially separated (i.e., they are located in distinct dots of the pair). By application of lateral electrical field it is possible to manipulate this exciton by changing localization of both carriers [170]. Long-living metastable states in QD systems are very desired as candidates for future applications as memory elements, especially combined with ultrafast optical coherent switching techniques [171].

4.5. Magnetic Quantum Dots

The properties of carriers confined in quantum dots may be strongly modified by introducing magnetic ions (e.g., Mn^{2+}) into the crystal lattice. The effects appearing in bulk magnetic or semimagnetic semiconductors (e.g., giant Zeeman splitting, formation of magnetic polarons) have been examined for a few decades (see, e.g., [172]). In a quantum dot, a single confined carrier or exciton interacts with the magnetic environment, reflecting the fundamental aspects of magnetic interaction in a nanometer scale confinement. Magnetic interaction between the carriers and the magnetic ions leads to the formation of a magnetic polaron (i.e., to ferromagnetic alignment of spins in a small region of the crystal). Experimental study of self-organized semimagnetic semiconductor quantum dots [173, 174] revealed some information on the properties of confined magnetic polarons and on the

relaxation processes accompanying their formation. Photoluminescence spectra show that the energy shift related to the formation of the magnetic polaron is of order of a few meV, while the formation times, obtained by time-resolved spectroscopy, are of the order of a hundred picoseconds.

4.6. Carrier–Phonon Interaction in Quantum Dots

QDs are not isolated systems. They are embedded in the macroscopic crystal and interacting with other carriers, crystal defects, and lattice excitations. While the former two kinds of interaction may be eliminated by optimizing the manufacturing conditions, lowering temperature, and decreasing the excitation power, the latter is unavoidable, being the most serious obstacle for many future applications and limiting the usefulness of the "artificial atom" analogy.

There are three major mechanisms of carrier–phonon interaction [175]: (1) Coulomb interaction with the lattice polarization induced by the relative shift of the positive and negative sublattices of the polar compound, described upon quantization by LO phonons; (2) deformation potential coupling which describes the band shifts due to lattice compression and is related to longitudinal acoustical (LA) phonons; (3) Coulomb interaction with piezoelectric field generated by shear crystal deformation described by transversal acoustical (TA) phonons. The latter effect is weak in InAs/GaAs QD systems but may be of more importance for the properties, for example, of GaN dots [176, 177]. Moreover, anharmonic interaction between LO and TA phonons [178, 179] (although other anharmonic channels can be also important) leads to another, indirect, coupling between the localized charges and TA phonons, manifesting itself in multiphonon processes.

It was predicted theoretically [180, 181] that phonon-induced relaxation from excited states should be slow due to the bottleneck effect: the trapped carrier wavefunction is localized on approximately 10 lattice constants and high energy acoustic phonons do not effectively couple to carriers due to the wavelength mismatch, while optical phonons have very weak dispersion around $\mathbf{k} = 0$ so that energy conservation would require a strictly defined energy distance between the carrier levels. Experiments on large ensembles of dots show, however, that the carriers relax from the excited states in a few tens of picoseconds by emitting one [182] or more [53] LO phonons. The cascade relaxation rate is limited by the final step, from the lowest excited to the ground state. The observed LO phonon relaxation, apparently contrary to the bottleneck idea, can be explained by the fact that in the large ensemble of inhomogeneously sized dots some of them satisfy the energy conservation [53].

Another manifestation of the carrier–phonon interaction in confined semiconductor systems (similarly as in bulk) are the phonon replicas in the emission and absorption spectra, corresponding to simultaneous photon and phonon absorption or emission. Such phonon-assisted peaks may be found both in ensemble [183, 184] and single-dot [76] experiments. These peaks are much stronger than could be expected on the grounds of the adiabatic theory and much stronger than in bulk—the relevant so-called Huang–Rhys parameter (i.e., the intensity ratio of the satellite phonon-assisted peak to the central peak) is for surprisingly high for QDs. To explain this increase arguments of a geometrical nature were invoked, referring to relative displacement of electron and hole charges in QDs. Due to localization in the same small region of QD, the electron and hole wavefunctions tend to overlap strongly in the ground state of the charge-neutral exciton, which leads to the diminishing of the polar interaction between the exciton and LO phonons [185]. Nevertheless, for, for example, pyramidal shape QDs the geometry leads to more effective separation of electron and hole charges. However, it is still insufficient to explain the high value of the observed Huang–Rhys factor. It was shown that nonadiabatic effects, included via carrier transitions between ground and excited excitonic states, induced by the lattice dynamics [186] or by taking into account the inertial corrections for confined carriers resulting in enhancement of the Fröhlich constant [187], lead to the increase of the phonon-assisted peak strength in better accordance with experiment.

Apart from these incoherent phonon processes, carrier–phonon interaction induces also coherent effects. The FIR magnetospectroscopy of QDs shows distinct anti-crossings of electron [188] or exciton [189] levels each time they are separated by a multiple of the LO phonon energy (Fig. 10). Such anti-crossings are evidence of the presence of polaron states (i.e., coherent superpositions of carrier and LO phonon states). The appearance of such coherent effects is possible because of the quasi-dispersionless character of the strongly coupled LO phonons [190], assuring LO phonon-induced dephasing times longer than any other relevant process.

Thus, there are two aspects of the carrier–phonon dynamics in QDs: According to the coherent polaron idea, the carrier–LO phonon interaction leads to the reconstruction of the spectrum and to the formation of hybrid states. In this picture, the interaction with phonons induces coherent oscillations between the excited state and the ground state accompanied by a phonon [188, 190]. On the contrary, the LO phonon-induced relaxation [53, 182] assumes a non-invertible transition from the excited state to the ground state, after which the final state dissolves in the phonon continuum. Both these aspects may be consistently described in terms of new quasiparticles, polarons, whose definition includes the coherent effects related to LO phonons.

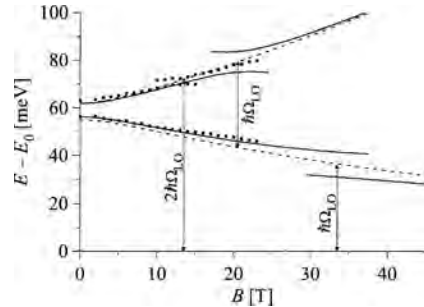


Figure 10. Energy spectrum of the polaron for a single electron in a InAs/GaAs self-assembled dot as a function of the magnetic field. The results of the harmonic approximation (solid lines) successfully reproduce the experimental data [188] (dots). The levels form anti-crossings whenever the electron energy crosses the LO phonon energy (see [191, 221] for discussion and dot parameters).

The polaron levels are shifted with respect to the excitonic ones (the polaron ground state energy is always lower, typically by a few meV, usually not more than 10% of characteristic confinement energy for electrons, and 5% for excitons). Around resonances they show anti-crossings, as observed in the experiment [188]. This effect is present both for charged excitations (excess electrons) and for neutral ones (excitons) [191]. In the latter case the electron and hole wavefunctions overlap to some extent (in self-assembled InAs/GaAs structures) but in a typical dot the Coulomb interaction between the carriers tends to shrink the hole wavefunction, diminishing the charge cancellation effects and strengthening the polar phonon effects. In piezoelectric semiconductors the built-in electric field separates the charges in the growth direction [176, 177] and in the dots created by electric field significant charge separation appears due to the attraction/repulsion of electron/hole by the empty dot. In both these cases the LO phonon effects are increased.

Polarons are not everlasting quasiparticles. They relax due to decomposition of their phonon cloud. These processes, including acoustical phonons, are incoherent due to the existence of a broad, weakly coupled continuum of final states. Such processes are induced both by the carrier–phonon interaction and by the anharmonic phonon–phonon coupling. In the former case, due to the bottleneck effect (i.e., exponential suppression of the coupling to short-wavelength phonons), they are effective only for level spacing around the LO phonon resonance. The latter channel is not affected by the bottleneck and is restricted only by the limited acoustic phonon bandwidth. In a InAs/GaAs SAD, the relaxation typically takes place on the time scale of tens of picoseconds, in agreement with the relaxation experiments [53, 182]. Due to the coherent nature of interaction with LO phonons, any process involving a LO phonon must be actually an at least two-phonon event in which an acoustical phonon participates.

Due to the nanometer scale of confinement, phonon phenomena in QDs differ essentially from those in bulk. The energy scale of the confined carriers is close to the resonance with optical phonons; hence the carrier–LO phonon interaction is in the strong coupling regime leading to considerable modification of the carrier spectrum (i.e., to the polaron effect). The Fröhlich interaction for confined carriers is enhanced by nonadiabatic effects. An electron confined in QD moves with greater quasi-classical velocity than in bulk, thus the inertial local polarization, important for Fröhlich interaction, is enhanced. In contrast to these increased coherent effects, the incoherent carrier–acoustical phonon interaction is diminished due to the typically large interlevel distance compared to the acoustical phonon energy and to the confinement-related bottleneck effect (size incommensurability). These effects are magnified by wide polaron anticrossings in the spectrum induced by the increased Fröhlich coupling and lead to relaxation times of tens of ps (at least one order of magnitude longer than in bulk).

Another phenomenon related to the carrier–phonon interaction is the phonon-induced coherence loss. The recent time-resolved single dot spectroscopy experiment [192] showed a partial decay of the initial signal coherence on a picosecond scale. The coherence loss was then stopped

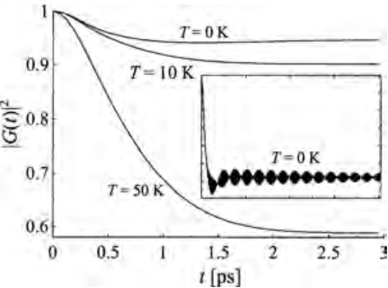


Figure 11. Decoherence of an exciton created in a InAs/GaAs dot by subpicosecond pulse due to interaction with acoustical phonons at different temperatures (theoretical results). The measure of decoherence is the decrease of the correlation function $G(t) = \langle a(t)a^\dagger(0)\rangle$. Inset: Decoherence due to acoustical and optical phonons (at zero temperature). Due to weak optical phonon dispersion the LO phonon beats decay on a much longer time scale.

at a relatively high, temperature dependent level, with a subsequent exponential decay limited by the exciton lifetime (~1 ns) at low temperatures (perhaps influenced also by other effects, like anharmonism). Such a behavior of the system polarization after a short, subpicosecond exciting pulse (modeled by an instant one) was theoretically accounted for, for example, in the single-state (independent boson) model within the linear response regime [185] and then generalized to higher orders and to sequences of pulses [193]. This partial coherence loss is a manifestation of a general dressing phenomenon, where the lattice reaction is a slow process following a fast (nonadiabatic, here of subpicosecond scale) carrier excitation [54]. Due to the slow lattice dynamics, the fast charge excitation must be followed by a relaxation process which may be interpreted as formation of the polarization or deformation dressing (phonon cloud) around the exciton. The time scale of the resulting dephasing, corresponding to the dissipation of an excess energy in the phonon sea, depends on the effective spectral width of the coupled phonon continuum (the most effective in this respect is the LA phonon sea). This process is, in turn, determined by the dot size, similar to the ordinary bottleneck effect. For a typical InAs/GaAs dot, this decoherence time is of the order of a picosecond, while the final degree of the coherence loss strongly depends on temperature and may vary in a wide range (Fig. 11).

5. APPLICATIONS

QDs are promising from the point of view of many state-of-the art and future technology applications. In the rapidly developing new domain of quantum information processing (see, e.g., [194]), a number of proposals based on the unique properties of QDs were put forward. The flexibility and small size of these quantum objects favors their application for definition of so-called qubits—quantum counterparts of classical bits. In such proposals, usually pairs of closely coupled dots (obtained, e.g., by stress-induced spontaneous stacking in the self-organization technology) play the role of two separated and interacting systems in which the entanglement of quantum states could be realized in a controlled manner [169]. Such entangled states, being

nonseparable tensor products of states of subsystems, are crucial for quantum computing. The quantum gate proposals employ dots manufactured in various ways: by electric gate [195], self-organized [196], or multilayered dot systems obtained by an etching technique.

A natural choice for the quantum bit is the spin of a confined electron [195, 197–199]. This choice is favored by long spin decoherence times (of order of microseconds) [171, 200, 201]. In one of the proposals [198, 199] a two-qubit gate is realized by two electrons residing in two coupled dots. The exchange interaction between the electrons may be switched on and off by applying external magnetic field [167], providing an efficient way of generating entanglement and performing two-qubit operations. On the other hand, one-qubit operations are based on the spin coupling with external magnetic field which is very weak in nonmagnetic dots (e.g., InAs/GaAs), leading to slow performance.

Much faster, all-optical operation is possible for a qubit defined by orbital degrees of freedom as, for example, the presence or absence of an exciton in a dot [196, 177]. Conditional two-qubit operations are induced by the dipole interaction when both dots are occupied. Here, however, the information is affected by fast decoherence related mostly to the interaction with phonons.

In order to combine the long decoherence time of the spin with the fast gating provided by strong coupling to the orbital degrees of freedom a new idea was recently put forward, based on the Pauli blocking mechanism [202]. In this proposal, the optically induced excitonic transition leading to dipole interaction (gating) is conditioned on the spin state of an excess electron.

Despite the very rapid progress in nanotechnology the realization of even a small quantum computer is hampered by severe obstacles caused mainly by decoherence processes [203]. So far, coherent control over excitonic states [55] and evidence of Rabi oscillations (an indicator of coherent evolution) [204–207] have been demonstrated. The feasibility of the readout of quantum states is supported by the possibility of converting the state of the dot into currents. In the case of orbital states this was achieved [207] by placing a self-assembled dot in a photodiode. Under external electric field the charge tunnels from the dot contributing to the current. For the spin state, the tunneling current of spin-polarized electrons depends on the state of the localized spin [208].

The quantization of states in a QD leads to the possibility of controlling a current, even of single-electron intensity. This leads to the proposal of QD spin filters and spin memory [209]. The Pauli exclusion principle may also be employed to implement current rectification and spin filtration in doubledot systems [210]. Depending on the polarization of bias voltage applied, either one spin component of the current may flow through the device, or both components are blocked. These proposals, aiming at spintronics applications of QDs, may strongly benefit from application of QDs created in magnetic semiconductor medium (like CdMnTe).

The dot in a diode structure mentioned previously [207] may be used as a highly effective single-electron source. With repeated laser pulses, a photocurrent of controlled intensity is obtained which, potentially, may be spin-polarized by applying polarized laser light. A single-photon source, essential for quantum cryptography applications, is also obtainable in the QD technology. As demonstrated experimentally [211], charging the dots with many charges using a high-intensity laser pulse leads to a single-photon emission at the single-exciton recombination energy.

Other ideas concerning technological applications of QDs, exploring the limits of miniaturization of the classical microelectronics, are related to the single-electron transistor or single-electron memory [212, 213]. In these nanodevices a charge is stored in silicon nanocrystals of a few nanometers diameter. Such a device may be operated at room temperature and can store a single electron for weeks.

The excitonic laser, working within the semiconductor diode scheme and employing recombination of excitons (thus emitting a light corresponding approximately to the forbidden gap of the semiconductor medium), has already been constructed within QD technology [30, 31, 140]. Quantum dots play also a still growing role in the continuously improving constructions of GaN heterostructures for blue lasers.

Another proposal is a version of a far-infrared laser working on transitions between only electronic states of quantum dots (defined by electric field focusing), similar to ordinary molecular lasers. The novelty consists here in a possibility of pumping such a laser by rapid changes of the dot dimension via oscillations of the field which creates the array of dots in the quantum well [214]. The inverse occupation can be achieved due to distinct probability of filling with electrons of upper and lower dot levels. Dots should be, however, created (or changed) very rapidly—nonadiabatically—which is hard to attain in practice.

Even though the practical implementation and, in some cases, even laboratory demonstration of all these novel constructions and ideas needs further technological and theoretical progress, advantages of the next step of miniaturization, toward and beyond the quantum limit, are very attractive. As was noticed in [170], currently the storage of a single bit of information involves millions of atoms, while application of quantum dots would reduce this number to thousands or even hundreds. The semiconductor nanodevices employing QD technology will certainly remain in the mainstream of the technological progress underlying the development of the information society and, possibly, even the forthcoming quantum information breakthrough.

GLOSSARY

Artificial atom See quantum dot.

Artificial molecule A system of closely spaced quantum dots allowing interaction of the localized carriers.

Capacitance spectroscopy An experimental method of inferring on system properties (density of states) based on capacitance changes upon applied voltage.

Confinement potential The effective potential arising from the quantum dot structure, restricting the dynamics of the carriers.

Exciton An electron-hole pair bound by Coulomb attraction (in quantum dots also by the confinement potential). Charged excitons are excitons accompanied by additional charges.

Hole In a semiconductor system, a missing electron in the filled valence band, behaving dynamically as a positive carrier.

Optical spectroscopy An experimental method of inferring on system properties based on luminescence (light emission) upon optical excitation.

Polaron A carrier (or exciton) accompanied by the lattice polarization field induced by carrier-phonon coupling.

Quantum dot (also artificial atom) A modulation by the semiconductor structure (e.g. by composition modification or external electric field) leading to carrier localization on nanometer lengthscale.

Self-assembly (or self-organization) A process in which the strain resulting from the lattice constant misfit between epitaxially grown semiconductor layers is released by aggregation of nanocrystal islands (quantum dots).

Transport spectroscopy An experimental method of inferring on system properties based on the measurement of current flowing through the system while external voltage is applied to the attached electrodes.

REFERENCES

1. S. Das Sarma and A. Pinczuk, "Perspectives in Quantum Hall Effects." Wiley, Chichester, 1997.
2. P. M. Petroff, A. C. Gossard, R. A. Logan, and W. Wiegmann, *Appl. Phys. Lett.* 41, 635 (1982).
3. M. A. Reed, R. T. Bate, K. Bradshaw, W. M. Duncan, W. M. Frensley, J. W. Lee, and H. D. Smith, *J. Vac. Sci. Technol. B* 4, 358 (1986).
4. J. Cibert, P. M. Petroff, G. J. Dolan, S. J. Pearton, A. C. Gossard, and J. H. English, *Appl. Phys. Lett.* 49, 1275 (1986).
5. H. Temkin, G. J. Dolan, M. B. Panish, and S. N. G. Chu, *Appl. Phys. Lett.* 50, 413 (1987).
6. K. Kash, A. Scherer, J. M. Worlock, H. G. Craighead, and M. C. Tamargo, *Appl. Phys. Lett.* 49, 1043 (1986).
7. L. Jacak, P. Hawrylak, and A. Wojs, "Quantum Dots." Springer-Verlag, Berlin, 1998.
8. D. Bimberg, M. Grundmann, and N. Ledentsov, "Quantum Dot Heterostructures." Wiley, Chichester, 1999.
9. "Self-Assembled InAs/GaAs Quantum Dots" (M. Sugawara, Ed.). Academic Press, San Diego, 1999.
10. R. C. Ashoori, *Nature* 379, 413 (1996).
11. W. Hansen, T. P. Smith, K. Y. Lee, J. A. Brum, C. M. Knoedler, J. M. Hong, and D. P. Kern, *Phys. Rev. Lett.* 62, 2168 (1989).
12. A. Lorke, J. P. Kotthaus, and K. Ploog, *Phys. Rev. Lett.* 64, 2559 (1990).
13. B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* 68, 1371 (1992).
14. C. Sikorski and U. Merkt, *Phys. Rev. Lett.* 62, 2164 (1989).
15. J. Alsmeier, E. Batke, and J. P. Kotthaus, *Phys. Rev. B* 41, 1699 (1990).
16. W. Hansen, T. P. Smith, K. Y. Lee, J. M. Hong, and C. M. Knoedler, *Appl. Phys. Lett.* 56, 168 (1990).
17. N. B. Zhitenev, M. Brodsky, R. C. Ashoori, L. N. Pfeiffer, and K. W. West, *Science* 285, 715 (1999).
18. K. Brunner, U. Bockelmann, G. Abstreiter, M. Walther, G. Böhm, G. Tränkle, and G. Weimann, *Phys. Rev. Lett.* 69, 3216 (1992).
19. A. I. Ekimov, A. L. Efros, and A. A. Onushchenko, *Solid State Commun.* 56, 921 (1985).
20. T. Fukui, S. Ando, and Y. Tokura, *Appl. Phys. Lett.* 58, 2018 (1991).
21. M. Grundmann and D. Bimberg, *Phys. Rev. B* 55, 4054 (1997).
22. W. Wegscheider, G. Schedelbeck, A. Abstreiter, M. Rother, and M. Bichler, *Phys. Rev. Lett.* 79, 1917 (1997).
23. G. Schedelbeck, W. Wegscheider, M. Bichler, and A. Abstreiter, *Science* 278, 1792 (1998).
24. J. Christen and D. Bimberg, *Phys. Rev. B* 42, 7213 (1990).
25. J. Christen, M. Krahl, and D. Bimberg, *Superlattices Microstructures* 7, 1 (1990).
26. K. Brunner, G. Abstreiter, G. Böhm, G. Tränkle, and G. Weinmann, *Appl. Phys. Lett.* 64, 3320 (1994).
27. P. M. Petroff and S. P. Denbaars, *Superlattices Microstructures* 15, 15 (1994).
28. I. N. Stranski and L. von Krastanow, *Akad. Wiss. Let. Mainz Math. Natur. K1 IIb* 146, 797 (1939).
29. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 52, 11969 (1995).
30. N. Kirstaedter, N. N. Ledentsov, M. Grundmann, D. Bimberg, V. M. Ustinov, S. S. Ruvimov, M. V. Maximov, P. S. Kop'ev, Z. I. Alferov, U. Richter, P. Werner, U. Gösele, and J. Heydenreich, *Electron. Lett.* 30, 1416 (1994).
31. M. Grundmann, *Physica E* 5, 167 (2000).
32. J. Y. Marzin, J. M. Gérard, A. Izraël, D. Barrier, and G. Bastard, *Phys. Rev. Lett.* 73, 716 (1994).
33. S. Fafard, R. Leon, D. Leonard, J. L. Merz, and P. M. Petroff, *Phys. Rev. B* 50, 8086 (1994).
34. S. Fafard, R. Leon, D. Leonard, J. L. Merz, and P. M. Petroff, *Phys. Rev. B* 52, 5752 (1995).
35. S. Raymond, S. Fafard, P. J. Poole, A. Wójs, P. Hawrylak, S. Charbonneau, D. Leonard, R. Leon, P. M. Petroff, and J. L. Merz, *Phys. Rev. B* 54, 11548 (1996).
36. R. Steffen, A. Forchel, T. Reinecke, T. Koch, M. Albrecht, J. Oshinowo, and F. Faller, *Phys. Rev. B* 54, 1510 (1996).
37. R. Steffen, T. K. J. Oshinowo, F. Faller, and A. Forchel, *Surf. Sci.* 361/362, 805 (1996).
38. A. Forchel, R. Steffen, T. Koch, M. Michel, M. Albrecht, and T. L. Reinecke, *Semicond. Sci. Technol.* 11, 1529 (1996).
39. M. A. Michel, A. Forchel, and F. Faller, *Appl. Phys. Lett.* 70, 393 (1997).
40. M. Grundmann, N. N. Ledentsov, J. Christen, J. Böhrer, D. Bimberg, S. S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heidenreich, V. M. Ustinov, A. Egorov, A. E. Zhukov, P. S. Kop'ev, and Z. I. Alferov, *Phys. Status Solidi B* 188, 249 (1995).
41. M. Grundmann, N. N. Ledentsov, O. Stier, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, and Z. I. Alferov, *Appl. Phys. Lett.* 68, 979 (1996).
42. S. Fafard, D. Leonard, J. L. Merz, and P. M. Petroff, *Appl. Phys. Lett.* 65, 1388 (1994).
43. R. Mirin, A. Gossard, and J. Bowers, *Electron. Lett.* 32, 1732 (1995).
44. K. Mukai, N. Ohtsuka, and M. Sugawara, *Jpn. J. Appl. Phys.* 35, L262 (1996).
45. R. Leon, P. M. Petroff, D. Leonard, and S. Fafard, *Science* 267, 1966 (1995).
46. P. Castrillo, D. Hessmann, M.-E. Pistol, S. Anand, N. Carlsson, W. Seifert, and L. Samuelson, *Appl. Phys. Lett.* 67, 1905 (1995).
47. M. Grundmann, J. Christen, N. N. Ledentsov, J. Böhrer, D. Bimberg, S. S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heidenreich, V. M. Ustinov, A. Y. Egorov, A. E. Zhukov, P. S. Kop'ev, and Z. I. Alferov, *Phys. Rev. Lett.* 74, 4043 (1995).
48. D. Hessmann, P. Castrillo, M.-E. Pistol, C. Pryor, and L. Samuelson, *Appl. Phys. Lett.* 69, 749 (1996).
49. J.-Y. Marzin and G. Bastard, *Solid State Commun.* 92, 437 (1994).
50. H. Lipsanen, M. Sopanen, and J. Ahopelto, *Phys. Rev. B* 51, 13868 (1995).
51. N. N. Ledentsov, M. Grundmann, N. Kirstaedter, J. Christen, R. Heitz, J. Böhrer, F. Heinrichsdorff, D. Bimberg, S. S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V. M. Ustinov, A. Egorov, M. V. Maximov, P. S. Kop'ev, and Z. I. Alferov, in "Proceedings of 22nd International Conference on the Physics of

Semiconductors," Vancouver, Canada (D. J. Lokwood, Ed.), Vol. 3, p. 1855. World Scientific, Singapore, 1995.
52. R. Heitz, M. Grundmann, N. N. Ledentsov, L. Eckey, M. V. N. Kirstaedter, D. Bimberg, V. M. Ustinov, A. Egorov, A. E. Zhukov, P. S. Kop'ev, and Z. I. Alferov, *Appl. Phys. Lett.* 68, 361 (1996).
53. R. Heitz, M. Veit, N. N. Ledentsov, A. Hoffmann, D. Bimberg, P. S. Kop'ev, and Z. I. Alferov, *Phys. Rev. B* 56, 10435 (1997).
54. L. Jacak, P. Machnikowski, and J. Krasnyj, *Eur. Phys. J. D* 22, 319 (2003).
55. N. N. Bonadeo, J. Erland, D. Gammon, D. S. Katzer, D. Park, and D. G. Steel, *Science* 282, 1473 (1998).
56. V. Zwiller, M.-E. Pistol, D. Hessman, R. Cederstrom, W. Seifert, and L. Samuelson, *Phys. Rev. B* 59, 5021 (1999).
57. Y. Toda, O. Moriwaki, M. Nishioka, and Y. Arakawa, *Phys. Rev. Lett.* 82, 4114 (1999).
58. P. Hawrylak, G. A. Narvaez, M. Bayer, and A. Forchel, *Phys. Rev. Lett.* 85, 389 (2000).
59. V. Zwiller, L. Jarlskog, M.-E. Pistol, C. Pryor, P. Castrillo, W. Seifert, and L. Samuelson, *Phys. Rev. B* 63, 233301 (2001).
60. M. Sugisaki, H.-W. Ren, S. V. Nair, K. Nishi, S. Sugou, T. Okuno, and Y. Masumoto, *Phys. Rev. B* 59, (1999).
61. M. Bayer, O. Stern, A. Kuther, and A. Forchel, *Phys. Rev. B* 61, 7273 (2000).
62. V. D. Kulakovskii, G. Bacher, R. Weigand, T. Kummell, A. Forchel, E. Borovitskaya, K. Leonardi, and D. Hommel, *Phys. Rev. Lett.* 82, 1780 (1999).
63. E. Dekel, D. Gershoni, E. Ehrenfreund, J. M. Garcia, and P. M. Petroff, *Phys. Rev. B* 61, 11009 (2000).
64. A. Hartmann, Y. Ducommun, E. Kapon, U. Hohenester, and E. Molinari, *Phys. Rev. Lett.* 84, 5648 (2000).
65. J. J. Finley, P. W. Fry, A. D. Ashmore, A. Lemaître, A. I. Tartakovskii, R. Oulton, D. J. Mowbray, M. S. Skolnick, M. Hopkinson, and P. P. A. Maksym, *Phys. Rev. B* 63, 161305 (2001).
66. J. J. Finley, A. D. Ashmore, A. Lemaître, D. J. Mowbray, M. S. Skolnick, I. E. Itskevich, P. A. Maksym, M. Hopkinson, and T. F. Krauss, *Phys. Rev. B* 63, 073307 (2001).
67. M. Lomascolo, A. Vergine, T. K. Johal, R. Rinaldi, A. Passaseo, R. Cingolani, S. Patane, M. Labardi, M. Allegrini, F. Troiani, and E. Molinari, *Phys. Rev. B* 66, 041302 (2002).
68. D. V. Regelman, E. Dekel, D. Gershoni, E. Ehrenfreund, A. J. Williamson, J. Shumway, A. Zunger, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. B* 64, 165301 (2001).
69. E. S. Moskalenko, K. F. Karlsson, P. O. Holtz, B. Monemar, W. V. Schoenfeld, J. M. Garcia, and P. M. Petroff, *Phys. Rev. B* 64, 085302 (2001).
70. D. Hessman, J. Persson, M.-E. Pistol, C. Pryor, and L. Samuelson, *Phys. Rev. B* 64, 233308 (2001).
71. M. Bayer, G. Ortner, O. Stern, A. Kuther, A. A. Gorbunov, A. Forchel, P. Hawrylak, S. Fafard, K. Hinzer, T. L. Reinecke, S. N. Walck, J. P. Reithmaier, F. Klopf, and F. Schafer, *Phys. Rev. B* 65, 195315 (2002).
72. K. Hinzer, P. Hawrylak, M. Korkusinski, S. Fafard, M. Bayer, O. Stern, A. Gorbunov, and A. Forchel, *Phys. Rev. B* 63, 075314 (2001).
73. L. Besombes, K. Kheng, L. Marsal, and H. Mariette, *Phys. Rev. B* 65, 121314 (2002).
74. M. Bayer, A. Kuther, F. Schafer, J. P. Reithmaier, and A. Forchel, *Phys. Rev. B* 60, R8481 (1999).
75. M. Bayer, A. Kuther, A. Forchel, A. Gorbunov, V. B. Timofeev, F. Schafer, J. P. Reithmaier, T. L. Reinecke, and S. N. Walck, *Phys. Rev. Lett.* 82, 1748 (1999).
76. A. Lemaître, A. D. Ashmore, J. J. Finley, D. J. Mowbray, M. S. Skolnick, M. Hopkinson, and T. F. Krauss, *Phys. Rev. B* 63, 161309 (2001).
77. F. Findeis, A. Zrenner, G. Bohm, and G. Abstreiter, *Phys. Rev. B* 61, R10579 (2000).
78. G. Chen, T. H. Stievater, E. T. Batteh, L. Xiaoqin, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, and L. J. Sham, *Phys. Rev. Lett.* 88, 117901 (2002).
79. I. Shtrichman, C. Metzner, B. D. Gerardot, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. B* 65, 081303 (2002).
80. P. G. Blome, M. Wenderoth, M. Hubner, R. G. Ulbrich, J. Porsche, and F. Scholz, *Phys. Rev. B* 61, 8382 (2000).
81. L. Besombes, K. Kheng, L. Marsal, and H. Mariette, *Phys. Rev. B* 63, 155307 (2001).
82. K. Matsuda, K. Ikeda, T. Saiki, H. Tsuchiya, H. Saito, and K. Nishi, *Phys. Rev. B* 63, 121304 (2001).
83. M. Bayer and A. Forchel, *Phys. Rev. B* 65, 041308 (2002).
84. C. Kammerer, C. Voisin, G. Cassabois, C. Delalande, P. Roussignol, F. Klopf, J. P. Reithmaier, A. Forchel, and J. M. Gérard, *Phys. Rev. B* 66, 041306 (2002).
85. C. Kammerer, G. Cassabois, C. Voisin, C. Delalande, P. Roussignol, A. Lemaître, and J. M. Gérard, *Phys. Rev. B* 65, 033313 (2002).
86. E. Dekel, D. V. Regelman, D. Gershoni, E. Ehrenfreund, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. B* 62, 11038 (2000).
87. C. Santori, G. S. Solomon, M. Pelton, and Y. Yamamoto, *Phys. Rev. B* 65, 073310 (2002).
88. T. H. Stievater, L. Xiaoqin, D. G. Steel, D. Gammon, D. S. Katzer, and D. Park, *Phys. Rev. B* 65, 205319 (2002).
89. T. Guenther, C. Lienau, T. Elsaesser, M. Glanemann, V. M. Axt, T. Kuhn, S. Eshlaghi, and A. D. Wieck, *Phys. Rev. Lett.* 89, 057401 (2002).
90. R. Oulton, J. J. Finley, A. D. Ashmore, I. S. Gregory, D. J. Mowbray, M. S. Skolnick, M. J. Steer, S.-L. Liew, M. A. Migliorato, and A. J. Cullis, *Phys. Rev. B* 66, 045313 (2002).
91. C. Santori, M. Pelton, G. Solomon, Y. Dale, and Y. Yamamoto, *Phys. Rev. Lett.* 86, 1502 (2001).
92. C. Becher, A. Kiraz, P. Michler, A. Imamoglu, W. V. Schoenfeld, P. M. Petroff, Z. Lidong, and E. Hu, *Phys. Rev. B* 63, 121312 (2001).
93. D. V. Regelman, U. Mizrahi, D. Gershoni, E. Ehrenfreund, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. Lett.* 87, 257401 (2001).
94. A. Kiraz, S. Falth, C. Becher, B. Gayral, W. V. Schoenfeld, P. M. Petroff, Z. Lidong, E. Hu, and A. Imamoglu, *Phys. Rev. B* 65, 161303 (2002).
95. T. P. Smith, H. Arnot, J. M. Hong, C. M. Knoedler, S. E. Laux, and H. Schmid, *Phys. Rev. Lett.* 59, 2802 (1987).
96. T. P. Smith, K. Y. Lee, C. M. Knoedler, J. M. Hong, and D. P. Kern, *Phys. Rev. B* 38, 2172 (1988).
97. R. C. Ashoori, H. L. Störmer, J. S. Weiner, L. N. Pfeifer, S. J. Pearton, K. W. Baldwin, and K. West, *Phys. Rev. Lett.* 68, 3088 (1992).
98. R. C. Ashoori, H. L. Störmer, J. S. Weiner, L. N. Pfeifer, K. W. Baldwin, and K. West, *Phys. Rev. Lett.* 71, 613 (1993).
99. H. Drexler, D. Leonard, W. Hansen, J. P. Kotthaus, and P. M. Petroff, *Phys. Rev. Lett.* 64, 1943 (1994).
100. M. Fricke, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, *Europhys. Lett.* 36, 197 (1996).
101. B. T. Miller, W. Hansen, S. Manus, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, *Phys. Rev. B* 56, 6764 (1997).
102. P. N. Brunkov, S. G. Konnikov, V. M. Ustinov, A. E. Zhukov, A. Egorov, M. V. Maximov, N. N. Ledentsov, and P. S. Kop'ev, *Semicond.* 30, 492 (1996).
103. W.-H. Chang, T. M. Hsu, N. T. Yeh, and J.-I. Chyi, *Phys. Rev. B* 62, 13040 (2000).
104. P. N. Brunkov, A. Patane, A. Levin, L. Eaves, P. C. Main, Y. G. Musikhin, B. V. Volovik, A. E. Zhukov, V. M. Ustinov, and S. G. Konnikov, *Phys. Rev. B* 65, 085326 (2002).
105. H. Pettersson, R. J. Warburton, J. P. Kotthaus, N. Carlsson, W. Seifert, M. E. Pistol, and L. Samuelson, *Phys. Rev. B* 60, R11289 (1999).
106. G. Medeiros-Ribeiro, D. Leonard, and P. M. Petroff, *Appl. Phys. Lett.* 66, 1767 (1995).

107. A. J. Chiquito, Y. Pusep, M. S, J. C. Galzerani, and M. Nt, *Phys. Rev. B* 61, 4481 (2000).
108. C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, R. Heitz, M. Grundmann, N. D. Zakharov, D. Bimberg, and P. Werner, *Phys. Rev. B* 60, 14265 (1999).
109. B. Su, V. J. Goldman, and J. E. Cunningham, *Phys. Rev. B* 46, 7644 (1992).
110. B. Su, V. J. Goldman, and J. E. Cunningham, *Science* 255, 313 (1992).
111. M. Tewordt et al., *Phys. Rev. B* 45, 14407 (1992).
112. P. Guéret, N. Blanc, R. Germann, and H. Rothuizen, *Phys. Rev. Lett.* 68, 1896 (1992).
113. P. C. Main et al., *Physica B* 189, 125 (1993).
114. T. Schmidt et al., *Phys. Rev. B* 51, 5570 (1995).
115. S. Tarucha, D. G. Austing, T. Honda, R. J. van der Hage, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 77, 3613 (1996).
116. L. P. Kouvenhoven, T. H. Oosterkamp, S. Tarucha, D. G. Austing, and T. Honda, *Physica B* 249–251, 191 (1998).
117. S. Tarucha, T. Honda, D. G. Austing, Y. Tokura, K. Muraki, T. H. Oosterkamp, J. W. Janssen, and L. P. Kouvenhoven, *Physica E* 3, 112 (1998).
118. I. E. Itskevich, T. Ihn, A. Thornton, M. Henini, T. J. Foster, P. Moriarty, A. Nogaret, P. H. Beton, L. Eaves, and P. C. Main, *Phys. Rev. B* 54, 16401 (1996).
119. M. Narihiro, G. Yusa, Y. Nakamura, T. Noda, and H. Sakaki, *Appl. Phys. Lett.* 70, 105 (1997).
120. M. Fukuda, K. Nakagawa, S. Miyazaki, and M. Hirose, *Appl. Phys. Lett.* 70, 2291 (1997).
121. K. H. Schmidt, M. Versen, U. Kunze, D. Reuter, and A. D. Wieck, *Phys. Rev. B* 62, 15879 (2000).
122. L. P. Rokhinson, L. J. Guo, S. Y. Chou, and D. C. Tsui, *Phys. Rev. B* 60, R16319 (1999).
123. Z. Hens, D. Vanmaekelbergh, E. J. A. J. Stoffels, and H. van Kempen, *Phys. Rev. Lett.* 88, 236803 (2002).
124. P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, *Phys. Rev. Lett.* 66, 1926 (1991).
125. P. L. McEuen, E. B. Foxman, J. Kinaert, U. Meirav, M. A. Kastner, N. S. Wingreen, and S. J. Wind, *Phys. Rev. B* 45, 11419 (1992).
126. A. T. Johnson, L. P. Kouwenhoven, W. de Jong, N. C. van der Vaart, C. J. P. M. Harmans, and C. T. Foxon, *Phys. Rev. Lett.* 69, 1592 (1992).
127. J. Weis, R. J. Haug, K. von Klitzing, and K. Ploog, *Phys. Rev. B* 46, 12837 (1992).
128. J. Weis, R. J. Haug, K. v. Klitzing, and K. Ploog, *Phys. Rev. Lett.* 71, 4019 (1993).
129. H. Sakaki, G. Yusa, T. Someya, Y. Ohno, T. Noda, H. Akiyama, Y. Kadoya, and H. Noge, *Appl. Phys. Lett.* 67, 3444 (1995).
130. N. Horiguchi, T. Futatsugi, Y. Nakata, and N. Yokoyama, *Appl. Phys. Lett.* 70, 2294 (1997).
131. T. H. Oosterkamp, J. W. Janssen, L. P. Kouwenhoven, D. G. Austing, T. Honda, and S. Tarucha, *Phys. Rev. Lett.* 82, 2931 (1999).
132. A. Patane, R. J. A. Hill, L. Eaves, P. C. Main, M. Henini, M. L. Zambrano, A. Levin, N. Mori, C. Hamaguchi, Y. V. Dubrovskii, E. E. Vdovin, D. G. Austing, S. Tarucha, and G. Hill, *Phys. Rev. B* 65, 165308 (2002).
133. A. L. Roest, J. J. Kelly, D. Vanmaekelbergh, and E. A. Meulenkamp, *Phys. Rev. Lett.* 89, 036801 (2002).
134. S. De Franceschi, S. Sasaki, J. M. Elzerman, W. G. Van Der Wiel, S. Tarucha, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 86, 878 (2001).
135. L. P. Rokhinson, L. J. Guo, S. Y. Chou, D. C. Tsui, E. Eisenberg, R. Berkovits, and B. L. Altshuler, *Phys. Rev. Lett.* 88, 186801 (2002).
136. A. S. G. Thornton, T. Ihn, P. C. Main, L. Eaves, and M. Henini, *Appl. Phys. Lett.* 73, 354 (1998).
137. M. Ciorga, A. S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, S. Jullian, Y. Feng, and Z. Wasilewski, *Phys. Rev. B* 61, R16315 (2000).
138. T. Chakraborty, "Quantum Dots." Elsevier, Amsterdam, 1999.
139. U. Woggon, "Optical Properties of Semiconductor Quantum Dots." Springer-Verlag, Berlin, 1997.
140. S. Fafard, K. Hinzer, S. Raymond, M. Dion, J. McCaffrey, and S. Charbonneau, *Science* 274, 1350 (1996).
141. V. Fock, *Z. Phys.* 47, 446 (1928).
142. C. G. Darwin, *Proc. Cambridge Philos. Soc.* 27, 86 (1930).
143. J. Adamowski, M. Sobkiewicz, B. Szafran, and S. Bednarek, *Phys. Rev. B* 62, 4234 (2000).
144. M. Califano and P. Harrison, *Phys. Rev. B* 61, 10959 (2000).
145. C. Pryor, *Phys. Rev. B* 57, 7190 (1998).
146. A. Wójs, P. Hawrylak, S. Fafard, and L. Jacak, *Phys. Rev. B* 54, 5604 (1996).
147. P. Hawrylak, G. A. Narvaez, M. Bayer, and A. Forchel, *Phys. Rev. Lett.* 85, 389 (2001).
148. A. Hartmann, Y. Ducommun, E. Kapon, U. Hohenester, and E. Molinari, *Phys. Rev. Lett.* 84, 5648 (2000).
149. R. Rinaldi, S. Antonaci, M. DeVittorio, R. Cingolani, U. Hohenester, E. Molinari, H. Lipsanen, and J. Tulkki, *Phys. Rev. B* 62, 1592 (2000).
150. M. Bayer, A. Schmidt, A. Forchel, F. Faller, T. L. Reinecke, P. A. Knipp, A. A. Dremin, and V. D. Kulakovskii, *Phys. Rev. Lett.* 74, 3439 (1995).
151. U. Bockelmann, K. Brunner, and G. Abstreiter, *Solid State Electron.* 37, 1109 (1994).
152. L. Jacak, J. Krasnyj, M. Korkusiński, and A. Wójs, *Phys. Rev. B* 57, 9069 (1998).
153. S. Tarucha, D. G. Austing, S. Sasaki, Y. Tokura, W. v. Wiel, and L. P. Kouwenhoven, *Appl. Phys. A* 71, 367 (2000).
154. S. Tarucha, D. G. Austing, Y. Tokura, W. G. v. Wiel, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 84, 2485 (2000).
155. G. W. Bryant, *Phys. Rev. Lett.* 59, 1140 (1987).
156. P. Hawrylak and D. Pfannkuche, *Phys. Rev. Lett.* 70, 485 (1993).
157. P. A. Maksym and T. Chakraborty, *Phys. Rev. Lett.* 65, 108 (1990).
158. D. Heitmann, *Physica B* 212, 201 (1995).
159. T. Demel, D. Heitmann, P. Grambow, and K. Ploog, *Phys. Rev. Lett.* 64, 788 (1990).
160. L. Jacak, J. Krasnyj, and A. Wójs, *Physica B* 229, 279 (1997).
161. R. B. Laughlin, *Phys. Rev. B* 23, 5632 (1981).
162. R. B. Laughlin, *Phys. Rev. Lett.* 50, 1395 (1983).
163. R. B. Laughlin, *Phys. Rev. B* 27, 3383 (1983).
164. J. K. Jain and T. Kawamura, *Europhys. Lett.* 29, 321 (1995).
165. R. K. Kamilla and J. K. Jain, *Phys. Rev. B* 52, 2798 (1995).
166. M. Brodsky, N. B. Zhitenev, R. C. Ashoori, L. N. Pfeiffer, and K. W. West, *Phys. Rev. Lett.* 85, 2356 (2000).
167. Y. Tokura, D. G. Austing, and S. Tarucha, *J. Phys: Cond. Matt.* 11, 6023 (1999).
168. M. Wagner, U. Merkt, and A. V. Chaplik, *Phys. Rev. B* 45, 1951 (1992).
169. M. Bayer, P. Hawrylak, K. Hinzer, S. Fafard, M. Korkusinski, Z. R. Wasilewski, O. Stern, and A. Forchel, *Science* 291, 451 (2001).
170. T. Lundstrom, W. Shoenfeld, H. Lee, and P. M. Petroff, *Science* 286, 2312 (1999).
171. D. D. Awschalom and J. M. Kikkawa, *Phys. Today* 6, 33 (1999).
172. "Diluted Magnetic Semiconductors" (J. K. Furdyna and J. Kossut, Ed.). Academic Press, Boston, 1988.
173. A. A. Maksimov, G. Bacher, A. McDonald, V. D. Kulakovskii, A. Forchel, C. R. Becker, G. Landwehr, and L. W. Molenkamp, *Phys. Rev. B* 62, R7767 (2000).
174. J. Seufert, G. Bacher, M. Schreibner, A. Forchel, S. Lee, M. Dobrowolska, and J. K. Furdyna, *Phys. Rev. Lett.* 88, 027402 (2002).
175. G. D. Mahan, "Many-Particle Physics." Kluwer, New York, 2000.
176. U. Hohenester, R. Di Felice, and E. Molinari, *Appl. Phys. Lett.* 75, 3449 (1999).
177. S. De Rinaldis, I. D'Amico, E. Biolatti, R. Rinaldi, R. Cingolani, and F. Rossi, *Phys. Rev. B* 65, 081309 (2002).

178. F. Vallée and F. Bogani, *Phys. Rev. B* 43, 12049 (1991).
179. F. Vallée, *Phys. Rev. B* 49, 2460 (1994).
180. H. Benisty, C. M. Sotomayor-Torres, and C. Weisbuch, *Phys. Rev. B* 44, 10945 (1991).
181. U. Bockelmann and G. Bastard, *Phys. Rev. B* 42, 8947 (1990).
182. I. V. Ignatiev, I. E. Kozin, V. G. Davydov, S. V. Nair, J.-S. Lee, H.-W. Ren, S. Sugou, and Y. Masumoto, *Phys. Rev. B* 63, 075316 (2001).
183. R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, and D. Bimberg, *Phys. Rev. Lett.* 83, 4654 (1999).
184. R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, and D. Bimberg, *Physica E* 7, 398 (2000).
185. B. Krummheuer, V. M. Axt, and T. Kuhn, *Phys. Rev. B* 65, 195313 (2002).
186. V. M. Fomin, V. N. Gladin, J. T. Devrese, E. P. Pokatilov, S. N. Balaban, and S. N. Klimin, *Phys. Rev. B* 57, 2415 (1998).
187. L. Jacak, J. Krasnyj, and W. Jacak, *Phys. Lett. A* 304, 168 (2002).
188. S. Hameau, Y. Guldner, O. Verzelen, R. Ferreira, and G. Bastard, *Phys. Rev. Lett.* 83, 4152 (1999).
189. O. Verzelen, R. Ferreira, and G. Bastard, *Phys. Rev. Lett.* 88, 146803 (2002).
190. O. Verzelen, S. Hameau, Y. Guldner, J. M. Gérard, R. Ferreira, and G. Bastard, *Jpn. J. Appl. Phys.* 40, 1941 (2001).
191. L. Jacak, J. Krasnyj, D. Jacak, and P. Machnikowski, *Phys. Rev. B* 65, 113305 (2002).
192. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, *Phys. Rev. Lett.* 87, 157401 (2001).
193. A. Vagov, V. M. Axt, and T. Kuhn, *Phys. Rev. B* 66, 165312 (2002).
194. J. Preskill, *Quantum Information and Computation*, http://www.theory.caltech.edu/preskill/ph229.
195. T. Tanamoto, *Phys. Rev. A* 61, 022305 (2000).
196. E. Biolatti, R. Iotti, P. Zanardi, and F. Rossi, *Phys. Rev. Lett.* 85, 5647 (2000).
197. M. S. Sherwin, A. Imamoğlu, and T. Montroy, *Phys. Rev. A* 60, 3508 (1999).
198. D. Loss and D. P. DiVincenzo, *Phys. Rev. A* 57, 120 (1998).
199. G. Burkard, D. Loss, and D. P. DiVincenzo, *Phys. Rev. B* 59, 2070 (1999).
200. J. M. Kikkawa and D. D. Awschalom, *Phys. Rev. Lett.* 80, 4313 (1998).
201. T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, *Nature* 419, 278 (2002).
202. E. Pazy, E. Biolatti, T. Calarco, I. D'Amico, P. Zanardi, F. Rossi, and P. Zoller, Spin-based optical quantum gates via Pauli blocking in semiconductor antum dots, cond-mat/0109337, 2001.
203. R. Landauer, *Philos. Trans. Roy. Soc. London* 353, 367 (1995).
204. T. H. Stievater, X. Li, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, C. Piermarocchi, and L. J. Sham, *Phys. Rev. Lett.* 87, 133603 (2001).
205. H. Kamada, H. Gotoh, J. Temmyo, T. Takagahara, and H. Ando, *Phys. Rev. Lett.* 87, 246401 (2001).
206. H. Htoon, T. Takagahara, D. Kulik, O. Baklenov, A. L. Holmes, Jr., and C. K. Shih, *Phys. Rev. Lett.* 88, 087401 (2002).
207. A. Zrenner, E. Beham, S. Stuffer, F. Findeis, M. Bichler, and G. Abstreiter, *Nature* 418, 612 (2002).
208. M. Ciorga, A. S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, Y. Feng, and Z. Wasilewski, *Physica E* 11, 35 (2001).
209. P. Recher, E. V. Sukhorukov, and D. Loss, *Phys. Rev. Lett.* 85, 1962 (2000).
210. K. Ono, D. G. Austing, Y. Tokura, and S. Tarucha, *Science* 297, 1313 (2002).
211. P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoğlu, *Science* 290, 2282 (2000).
212. L. Guo, E. Leobandung, and S. Y. Chou, *Science* 275, 649 (1997).
213. S. Tiwari, F. Rana, H. Hanafi, A. Harstein, and E. F. Crabbe, *Appl. Phys. Lett.* 68, 1377 (1996).
214. L. Jacak, J. Krasnyj, D. Jacak, and L. Bujkiewicz, *Phys. Rev. A* 65, 063813 (2002).
215. Y. Nagamune, H. Watabe, M. Nishioka, and Y. Arakawa, *Appl. Phys. Lett.* 67, 3257 (1994).
216. Y. Oka, J. Shen, K. Takabayashi, N. Takahashi, H. Mitsu, I. Souma, and R. Pittini, *J. Lumin.* 83–84, 83 (1999).
217. S. Tanaka, S. Iwai, and Y. Aoyagi, *Appl. Phys. Lett.* 69, 4096 (1996).
218. G. Abstreiter, P. Schittenhelm, C. Engel, E. Silveira, A. Zrenner, D. Meertens, and W. Jäger, *Semicond. Sci. Technol.* 11, 1521 (1996).
219. M. A. Reed, *Sci. Am.* 3, 40 (1993).
220. P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Baker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafaji, J. P. R. David, A. G. Cullis, G. Hill, and J. C. Clark, *Phys. Rev. Lett.* 84, 733 (2000).
221. L. Jacak, J. Krasnyj, D. Jacak, and P. Machnikowski, *Phys. Rev. B* 67, 035303 (2003).

www.aspbs.com/enn

III/V Semiconductor Quantum Dots

M. Guzzi, S. Sanguinetti

Universitá di Milano Bicocca, Milan, Italy

M. Gurioli

Universitá di Firenze, Sesto Fiorentino, Italy

CONTENTS

1. Introduction
2. Growth
3. Electronic Structure
4. Photoluminescence
5. Applications
   Glossary
   References

1. INTRODUCTION

The ability to control or tailor the conduction and valence band profiles in semiconductor nanostructures—denominated bandgap engineering—has completely changed the physics and technology of semiconductors in the last few decades. An enormous amount of studies, realizations, and achievements have been accumulated during almost 30 years of research on semiconductor nanostructures. The fundamental quantum mechanical properties of condensed matter have been deeply tested. A large variety of new physical phenomena, such as the quantum Hall effect [1], the resonant tunnelling [2], the quantum confined Stark effect [3], and the Coulomb blockade [4], have been demonstrated. Electronic nanodevices using a very small number of electrons or even a single electron have been proposed or realized [5]. Extremely efficient lasers and emitting diodes based on nanostructures covering nearly the full optical spectrum are commercially available.

The band engineering revolution is fully based on the technology of semiconductor heterostructures, which are synthesized crystalline systems composed of more than one coherently grown material. The variation of the composition across the heterostructures, which leads to a spatial localization of the electrons and holes, is used to control the carrier motion and therefore to modify the electronic and optical properties of the system. The band alignment at the heterojunction between two different compounds, which is one of the most relevant parameters for band engineering, determines the spatial position of the lowest energy states for both electrons and holes. Since the spatial overlap of electrons and holes is an unavoidable constraint for all the electro-optical devices, the largest interest is in heterostructures where the lowest energy state for both kinds of carriers is spatially localized in the material with lower bandgap. In fact, this direct alignment holds for the best studied heterojunction, $In_xGa_{1-x}As/Al_yGa_yAs$; in the following we will focus our attention to this case.

Whenever the carrier localization, at least in one spatial direction, becomes comparable or smaller than the de Broglie wavelength of carriers, quantum mechanical effects occur. In this limit the optical and electronic properties of the heterostructure change as a function of the size and the system is called a *nanostructure*. As the size is reduced the electronic states are shifted toward higher energy and the oscillator strength becomes concentrated into few transitions. In order to fix the order of magnitude, let us mention that in the case of most of the semiconductor systems the limit of the *nanoworld* is of the order of few tens of nanometers.

The physics of the quantum size effect relies on the Heisenberg uncertainty principle between the spatial position and kinetic momentum of a quantum particle. It is not possible to measure both the momentum and position of a particle to an arbitrary precision. The product of the standard deviation in space and momentum satisfies the uncertainty relation:

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (1)$$

This equation means that the smaller is the carrier localization in the nanostructure, the larger is the spread in the momentum p, or, better said for semiconductor systems, in the crystal momentum $\hbar k$. The energy may still be well defined, but the momentum is not well defined. In bulk systems, for states around the edge of conduction and valence

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

band, the dependence of the energy on the wavevector k is quadratic,

$$E = \frac{(\hbar k)^2}{2m^*} \tag{2}$$

where m^* is the carrier effective mass. Following this equation, the spread in the momentum $\hbar k$ gives a minimum kinetics energy to the localized particle. This is in contrast with the classical physics, where the lowest energy state in whatever potential corresponds to no kinetic energy. The uncertainty principle of quantum mechanics imposes a positive zero-point energy, which is approximately inversely proportional to the square of the nanostructure size. Therefore the energy of the ground state of electrons and holes in semiconductor nanostructures not only depends on the materials but also on the dimension of the confinement region. The quantum size dependence of the confinement energy is the key to the tunability of the optical transitions in nanostructures.

Nanostructures are classified by the number of dimensions in which the carriers are confined or, alternatively, free to move. In case of confinement in only one spatial direction, the nanostructure is named a quantum well (QW). The carrier motion is frozen in one dimension but electrons and holes can still freely move over the other two directions. Therefore the QW is a quasi two-dimensional (2D) system. A structure which provides carrier confinement in two directions, allowing the motion along the remaining dimension, is called quantum wire (QWR) and it is a quasi 1D system. In the case of confinement in all three spatial coordinates, the nanostructure is denominated quantum dot (QD). QDs are 0D systems since the carrier motion is completely frozen.

Obviously the physics of the nanostructures strongly depends on their dimensionality. This property is elucidated by the concept of electronic density of states (DOS). In a semiconductor structure a given energy usually corresponds to a large number different electronic states resulting from the carrier motion. In a bulk material where the motion can occur in three different directions the density of states increases proportionally to the square root of the energy. In quantum wells the motion in the plane gives a staircase DOS, where each step is associated with a new state in the confining potential. In quantum wires a continuum of states is still present, but strong resonances appear in the DOS associated with the states in the confining potential. Finally in quantum dots only discrete energy states are allowed and the DOS is therefore a comb of delta functions. Quantum nanostructures and their DOSs are illustrated in Figure 1.

The possibility to concentrate the DOS in a reduced energy range is extremely important for a large variety of fundamental topics and device applications. It is at the base of the quantum Hall effect in QW, of the quantization of the conductance in QWR, and of the single electron tunnelling in QDs. In the case of lasers the presence of a continuum DOS leads to losses associated with the population of states that do not contribute to the laser action. Conversely, the concentration of the DOS produces a reduction of the threshold current and enhances the thermal stability of the device operation. Clearly this property is optimized in QD structures. Due to the three-dimensional carrier

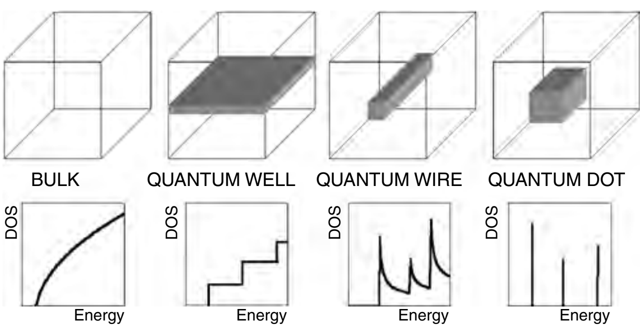


Figure 1. Schematic view and graph of quantum nanostructures and of their density of states.

confinement and the resulting discrete energy spectrum, semiconductor QDs can be regarded as artificial atoms. Nevertheless a QD is a crystalline system which is composed of a large number of unit cells (nearly 10^5). On one side this justifies the description of QD physics by the condensed matter approach (band structure, effective masses, etc.). On the other side this simple consideration suggests a large linear and nonlinear optical susceptibility of QDs, thus explaining the huge interest in QDs in electro-optical device applications.

Historically the research on semiconductor nanostructures was initiated by the discovery of superlattices (a periodic stack of QWs) by Esaki and Tsu in the 1970s [6]. They predicted Bloch oscillations and negative resistance for carriers moving in the superlattices. Few years later Dingle et al. [7] demonstrated that a strong enhancement of the excitonic optical transitions can be obtained in QWs. Rapidly the QWs become the tool for producing semiconductor systems with designed electronic and optical properties. A large variety of new optical phenomena in QWs was demonstrated: the enhancement of the excitonic resonances [8], the quantum confined Stark effect [9], the presence of large optical nonlinearity [10]. These peculiar properties rapidly produced a new generation of electro-optical devices, such as quantum well lasers, optical switches, modulators, and bistable devices [11]. The success of the QW structures is strictly related to the technological progress made in the growth of layered structures. The requirements of controlling the materials composition at the nanometer scale and of maintaining high crystal quality are well satisfied by the epitaxial growth. Several types of epitaxial growth have been developed so far allowing the deposition of defect-free layers down to an atomic level of control.

Very soon it was also realized that the possibility to confine carriers in additional spatial directions will further improve the performances of nanostructure devices. In 1982 Arakawa and Sakaki [12] theoretically demonstrated that the QWR and QD lasers will be far superior to QW lasers. This prediction stimulated a huge effort to produce QWR and QD nanostructures. However, the realization of QWRs and QDs requires the control of the nanostructure geometry not only along the epitaxial growth direction but also along the lateral directions. This achievement has been pursued by different approaches. The first attempts were based on standard semiconductor processing such as lithography for designing the nanostructure geometry and etching for

reducing the lateral dimensions [13]. These methods, in spite of various relevant advantages and some successful results, suffer from the low quality of the processed interfaces and therefore were not very useful for device applications. A completely different route to the production of semiconductor QDs is to use a colloidal chemical synthesis of nanometer crystallites of semiconductors [14]. The colloidal QDs are typically surrounded by dielectrics such as glasses, polymers, or organic solvents. The strength of this approach consists in the control and tuning of the QDs. The main disadvantage is related to the difficulty of integrating the colloidal synthesis with the standard technology of semiconductor devices. The breakthrough in the achievement of QD devices has occurred in the early 1990s with the discovery of the spontaneous formation of QDs by the deposition of highly strained layers during the epitaxial growth [15] or by droplet epitaxy [16]. The self-assembled QDs are coherently grown and show very high crystal quality. There is no processing related damage and the technology of the growth of self-assembled QDs is almost identical to the technology of multilayered deposition developed for the QW devices. These conditions have contributed to a rapid growth of interest for self-assembled QDs, opening the door to the intriguing world of zero-dimensional physics and devices.

This chapter is mainly dedicated to self-assembled III/V QDs (in particular $In_xGa_{1-x}As/Al_yGa_{1-y}As$ systems) which are nowadays one of the major areas of semiconductor research. We will focus our attention to the electronic and optical properties of the semiconductor QDs and in particular on the photoluminescence (PL) characterization of QDs. The analysis of the wavelength, time, and polarization dependence of optical spectra from nanostructures allows one to obtain information on the carrier–photon interaction (optical transition energy, oscillator strength, and nonlinear response), the electronic structure (level positions, density of states, and wavefunction overlap), the carrier dynamics (energy, momentum, coherence, and spin relaxation). Other relevant advantages of optical spectroscopy, with respect to other characterization methods, are its noninvasive nature and the lack of time-consuming sample preparation. Many different spectroscopic techniques have been used for the investigation of semiconductor QDs. Among others let us mention absorption, PL, PL excitation, modulated reflectance, microprobe and near field spectroscopies, magneto-optical measurements, and four wave mixing. Here we will focus our attention on the PL based techniques, which are extremely useful and very common tools, providing ample information ranging from the statistical distributions of QD size to quantum correlations of electronic states in a single QD.

This chapter is organized as follows. Section 2 breifly describes the main aspects of the growth of QD nanostructures, whose electronic structure is discussed in Section 3. A detailed description of the QD photoluminescence is reported in Section 4, where both steady state PL (Section 4.1) and time resolved PL (Section 4.2) are considered. Finally, Section 5 presents some relevant applications and future developments of QD technology.

2. GROWTH

A huge effort has been devoted in recent years to the development and understanding of self-assembling techniques for the fabrication of QD structures. As already remarked, the self-assembling approaches to QD production have the advantage that the structures are formed in the growth environment and no processing is needed either before or after the growth. From the point of view of optoelectronic applications, this is an important advantage. Growth modes which proceed via nucleation at the growth surface of islands, capable of quantum confinement, are the heteroepitaxy of strained layers (Stranski–Krastanow or Volmer–Weber [17]) and droplet epitaxy [16].

The QD growth environment is that of the usual epitaxial techniques [18], like molecular beam epitaxy (MBE) [19]. A more detailed description of the QD growth modes can be found in [20–23].

2.1. Dot Formation via Strained Layer Epitaxy

2.1.1. General Aspects

Three epitaxial growth modes have been classified (Fig. 2): the layer-by-layer growth in lattice matched systems [Frank–van der Merwe (FM)], the island growth mode [Volmer–Weber (VW)], and the layer by layer, followed by island nucleation, growth mode [Stranski–Krastanow (SK)] in lattice mismatched systems [17]. All three growth modes have been observed experimentally, depending on the interface free energy and on the lattice mismatch, in the epitaxy of semiconductors.

In the growth of strained III/V epilayers the SK mode dominates. This gives rise to a growth characterized by nanometer size islands, thus able to induce electronic confinement. Various combinations of III/V semiconductors based on P [24–26], Sb [27, 28] and As [29–32] have shown a SK

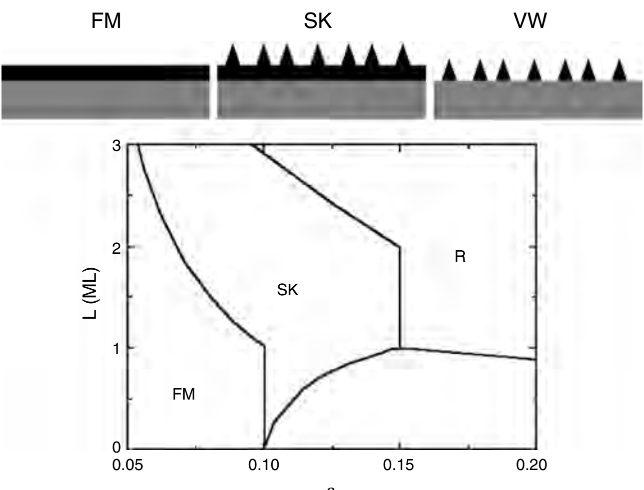


Figure 2. Equilibrium phase diagram in function of the coverage L and the lattice mismatch ϵ calculated using the model reported in [33]. Illustration of the morphology of the surface in the three coherent growth modes (FM, SK, and VW). The R phase is incoherent (large relaxed islands and presence of dislocations).

growth mode. The more widely studied system for optoelectronic applications is $In_xGa_{1-x}As/Al_yGa_{1-y}As$, where the maximum lattice mismatch, between the epilayer of InAs and the substrate of GaAs, is 7%.

A schematic of the SK growth for the typical InAs/GaAs system is reported in Figure 3. In short, the SK growth mode proceeds as follows: after the epitaxial growth of a thin pseudomorphic layer, a two-dimensional layer which "wets" the surface [thus called the "wetting layer" (WL)], the accumulated strain is relieved by the formation of nanometer-scale islands. Thus, a deep change is produced in the characteristics of the epilayer, whose growth switches from layer by layer to islands. This is called the SK transition and happens at a precise epilayer coverage, called critical coverage L_c (see Fig. 3). For the InAs/GaAs system $L_c = 1.5$ monolayers (MLs) [32]. The SK transition, according to [33], can be described as a second order phase transition.

At the early stage of QD formation, the islands can be free of defects. This type of growth produces coherent islands necessary for the fabrication of QDs. Near the critical thickness L_c increasing the coverage gives rise to an increase of the island density, eventually of larger size, without destroying the coherent crystalline growth. On the contrary, when thick layers are grown the islands start to coalesce [34] and defects (like dislocations) are introduced in the structure [30].

The SK formed islands show a pyramidal shape [35], eventually truncated and asymmetric respect to the [110] and [$\bar{1}$10] directions [36]. The exact shape of the QD depends on the surface free energy of the developed facets [37].

A statistical noise affects the size of the grown islands. Typical fluctuation of the self-assembled QD size around the mean is about 10% [38]. A certain degree of control of the shape, size, density, and overgrowth of the dots can be obtained through a suitable variation of the growth conditions. Lateral ordering because of strain fields, preferential nucleation at step edges, and growth on patterned substrates have been reported [20]. Nevertheless the achievement of quantitative control of size and uniformity of QDs is one of the open issues in QD science and technology.

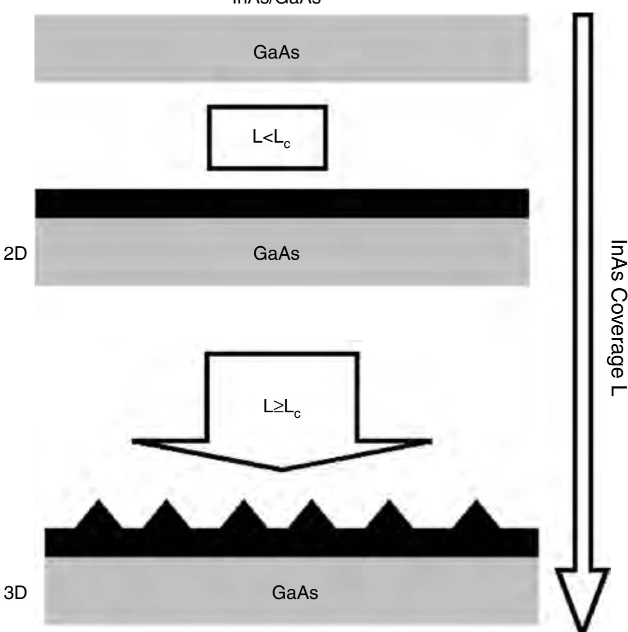


Figure 3. Schematic of the SK growth process.

Multilayer depositions of InAs separated by thin GaAs spacers result in the formation of stacked, vertically aligned, InAs QD columns (Fig. 4). The self-organization of the QDs into vertical columns is driven by the strain fields of the buried islands of the previous QD layer [39, 40]. By subsequent QD layer growth, the QD size becomes more and more uniform within each layer, giving rise to a large reduction of the size distribution. Moreover, the interdot in-plane spacing is made more uniform by stacking, eventually resulting in some kind of lateral ordering [41]. These systems show interesting properties associated with the strain induced by the spatial correlation among different layers.

2.1.2. Theoretical Models

A complete understanding of the mechanisms determining the island nucleation is still a matter of debate. The description is complicated by the coexistence, during the growth, of equilibrium and nonequilibrium effects [33, 42]. Nevertheless, the theoretical equilibrium studies have been quite successful in explaining the key mechanisms of island formation in strained heteroepitaxial systems [33, 43]. By including the effects of in-island strain relaxation and island formation in the determination of the free energy of the heteroepitaxial system, these studies predicted the existence in the equilibrium phase diagram of a range of material constants (namely lattice mismatch and coverage) where stable islands can exist on the surface.

Phenomenologically, the heteroepitaxial growth of a strained film of material A on top of a substrate of material B may proceed in different growth ways, depending on materials and growth conditions: (a) pseudomorphic layer by layer growth, (b) small size islands, or (c) extremely large and defected islands. Daruka and Barabasi modelled this phenomenology by recognizing that the effective distribution of the material A among the three different growth types should be determined by the free energy balance of the whole system. The free energy density of a number L of monolayers of epitaxial strained material A is modelled as

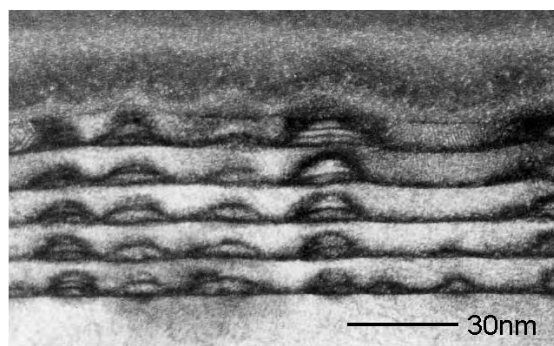


Figure 4. Cross-sectional transmission electron microscope image of a structure consisting of five embedded layers of InAs QDs, separated by 10 nm thick GaAs spacers [110]. Courtesy of G. Salviati.

the sum of the three terms, L dependent, pertaining to the allowed growth modes [33, 44].

The presence of strain drives the system towards an island growth, which, because of symmetry breaking, allows the epitaxial material to relax the strain. Two effects counteract this tendency: the interaction with the substrate, which stabilizes the otherwise unstable (because of accumulation of strain energy) pseudomorphic layer by layer growth at low coverage and low lattice mismatch, and the increase of surface energy due to the expansion of the surface imposed by the island growth.

In the case of typical III–V semiconductor heteroepitaxy, the deposited material A grows, at low mismatch, in the FM mode, thus showing that the interaction with the substrate and the energy cost connected with the increasing free surface due to island nucleation quench the possible strain relaxation permitted by an island mode. Conversely, the VW mode characterizes the growth at high lattice mismatch. Here, the reduction of free energy permitted by the symmetry breaking controls the growth mode. At intermediate lattice mismatch (between 5% and 15%) and coverages larger than unity, the growth mode is dominated by the SK mode. In the Daruka and Barabasi model [33] the insurgency of the island growth mode after an initial layer by layer growth at low coverages is due to the fading interaction of the epilayer with the substrate and to a surface energy cost for the formation of the islands lower than the strain relaxation energy gain connected with the island growth. As the binding energy of the epilayer with the substrate is no longer able to stabilize the pseudomorphic layer (this happens at a certain critical coverage L_c), the extra material A will aggregate in well defined islands, thus showing an SK-type growth. The predicted phase diagram is shown in Figure 2.

In addition to strain energy, binding energy with the substrate, and island surface energy, also stress discontinuities at the island surface edges, surface stress cross-terms, and island–island repulsive interaction have to be taken into account [33]. Substantial refinement of the theoretical model of Daruka and Barabasi [33] can be found in [37] where the surface energies for both the island facets and the wetting layer have been calculated using density-functional theory, and the elastic energy in both the islands and the substrate has been obtained from continuum elasticity theory.

2.2. Dot Formation in the Droplet Epitaxy Mode

New paths for obtaining QDs with novel characteristics, which possibly will give devices with even higher performances, may be provided by the droplet epitaxy (DE) method for the self-assembling [16, 45–50]. The DE method does not rely on the lattice mismatch between the substrate and the epilayer in order to nucleate the islands, thus allowing one to self-assemble QDs in strain-free materials.

The DE process is performed in an MBE system. The DE growth proceeds on a substrate taken at relatively low temperatures (around 180 °C [47]) and consists of two steps. First, a deposition, with the As pressure in the growth chamber depleted, of metallic Ga (or In and Ga in the case of InGaAs dots) which, due to the low growth temperature, leads to the formation of metallic droplets on the top of the growth substrate. Following the deposition of the Ga (or InGa) droplets a high flux As_4 molecular beam is irradiated on the surface leading to the formation of the dots. Subsequent annealing is performed in order to improve the structure.

Structural and electronic differences characterize the DE QDs with respect to standard SK QDs. From a microscopic point of view, the DE QDs are nanocrystal inclusions in the barrier matrix. Their typical dimensions are of the order of 20–30 nm at the base and 12 nm height. It is important to remark that it is possible to grow QDs with or without a WL connecting the QDs (Fig. 5). These latter systems are then a closer approximation to ideal quantum dots than SK QDs.

3. ELECTRONIC STRUCTURE

The accurate prediction of the electronic structure of a semiconductor-embedded self-assembled QD is a highly nontrivial task. Yet it is a crucial prerequisite for understanding the basic optical properties of such systems. Theoreticians who tried to address the problem faced a series of complications. The first one is connected with the low amount of precise structural data available on the "really" grown structures. In addition to QD shape and size, also strain relaxation and composition gradients (for example, In segregation in InAs/GaAs QDs) play fundamental roles in determining the electronic structure of the QDs. Moreover, due to the self-assembling process in the SK mode a residual, thin, 2D QW is present at the base of the QDs, which may affect the QD electronic states. From the theoretical point of view, another problem is given by the low symmetry, no more than C_{2v}, which poses serious difficulties in dealing with excited states such as exciton formation.

The typical procedure followed by all the authors for the electronic structure calculation is to factorize the problem of finding the many-body electronic states in a system with a highly deformed structure in a series of steps:

(a) calculation of the strain profile in the QD material via classical models,
(b) determination of the single electron states in the QD,
(c) inclusion of the many-body effects.

The possibility to factorize the problem in these three levels is given by two approximations, which are justified by the nature of the QD system. The first approximation relies on the observation that, in the self-assembled III–V QDs, the excitonic radius is always larger than the QD dimensions. This means that, in determining the electronic structure of the QD, the major role is played by the confinement energy, leaving to the Coulomb interaction between the carriers only a minor correction, usually treated in a perturbative way.

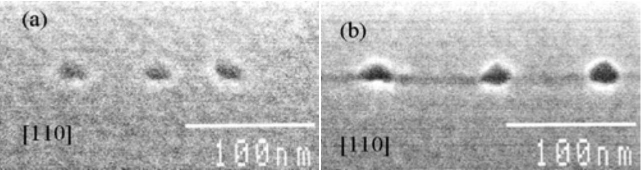


Figure 5. Scanning electron cross-section micrographs of DE QDs without WL (a) and with WL (b). After [50].

The second is that the elastic approximation for the calculation of the relaxed structure is valid (i.e., the electron–phonon interaction is small).

3.1. Strain Calculation

A large strain is present in the SK QDs, which strongly couples to carriers through band structure and piezoelectric effect. Therefore, it is essential to consider the influence of strain on the properties of the dot itself and its surroundings, because the wavefunctions and energies are very sensitive to the underlying strain distribution. As a matter of fact, in the SK QDs the impact of strain on the optical transitions is comparable to that of the carrier confinement induced by the band offset due to the variation of the chemical composition at the heterojunctions [35]. A relevant point is that the strain in the QD is insensitive to the WL primarily because the WL is so thin and also because it is biaxial strained to match the substrate lattice. So the WL is usually excluded from strain field calculations. Several methods can be used to calculate the strain. They can be divided in two main categories: the continuum elastic theory [35, 51, 52] and the atomistic approaches [53–55].

3.1.1. Continuum Elastic Theory

Within the continuum elastic theory (CET) approach, the strain field $\epsilon_{ij}(\mathbf{r})$ for a given structure is obtained by minimizing the total elastic energy U_{CET} [56],

$$U_{\mathrm{CET}} = \frac{1}{2} \sum_{i,j,k,l} \int_V C_{ijkl}(\mathbf{r}) \epsilon_{ij}(\mathbf{r}) \epsilon_{kl}(\mathbf{r}) d(\mathbf{r}) \tag{3}$$

where C_{ijkl} are the crystal compliances. The usually reported procedure to minimize U_{CET} is via numerical (finite differences) methods [35, 52, 57]. A more intriguing way to achieve the minimization of U_{CET} is described by the isotropic elasticity (IE) theory [51, 58]. This approach considers the elastic properties of the QD material as homogeneous and isotropic. This is a quite strong approximation, because the elastic properties of III–V semiconductors are significantly anisotropic. Moreover, the sensitivity of some physical properties (such as effective masses) to strain suggests that anisotropic effects could be important in semiconductor materials. Nevertheless, as shown by [51, 59] the IE is able to reveal the main features of the strain field distribution in QD materials. This is due to the low symmetry of the QD shape, which is the major source of the strain anisotropy in QD structures [59]. IE shows that the elastic field which minimizes U_{CET} can be analytically derived from a scalar potential that obeys a Poisson-like equation with the lattice mismatch acting as charge density. The strain field is analogous to the electric field, thus providing an intuitive way to visualize the distortion around the dot [51].

3.1.2. Atomistic Approach

The elastic energy, in the case of the atomistic approach U_{AT}, is written as a sum of the single atomic contributions. These are in turn determined by the appropriate interatomic potentials which depend on the relative positions of the nearest neighbor atoms. Because III–V semiconductors show covalent bonds with only some partial charge transfer, the more suitable interatomic potential is given by the valence force field (VFF) [60, 61]. The VFF model is a microscopic theory which includes the bond stretching and bond bending contributions to the total energy. The equilibrium atomic positions are then derived through numerical minimization of U_{AT}, thus allowing the determination of the displacement vector field and, in turn, of the strain field [56]. This type of approach becomes necessary when the exact atomic positions are needed for the electronic structure calculation, as in the pseudopotential approach [54, 55].

The strain field calculated with the two approaches has been found to agree reasonably well [52]. The main discrepancy is near the dot boundary, where large strain variation is found in the atomistic approach with respect to the CET [62].

Let us now describe the calculated strain field of an InAs pyramidal QD embedded in GaAs within the CET approach. The local variation of the hydrostatic and biaxial strain [56] is visualized through a linescan passing in the dot center in Figure 6. These strain components define the potential profile for the electrons and the splitting of the light and heavy hole states, respectively. The inner part of the QD contains nearly homogeneous hydrostatic strain, which is almost absent in the barrier. This means that the materials dilation is also confined in the QD. On the contrary, the biaxial strain is transferred significantly from the dots to the barrier and exhibits a distinct minimum in the QD. Thus the deformation of the electronic profile extends deeply in the barrier. The exact shape of the QD does not play a major role in determining the strain field distribution [59], although large changes in the QD aspect ratio do play a role [63]. Additionally, shear strain is present in the structure, which turns out, in QD grown on (100) oriented substrates, to be significantly close to the pyramidal edges. This strain component is responsible for the insurgency of a piezoelectric polarization charge. In fact, the piezoelectric polarization induced by the shear strain tensor components

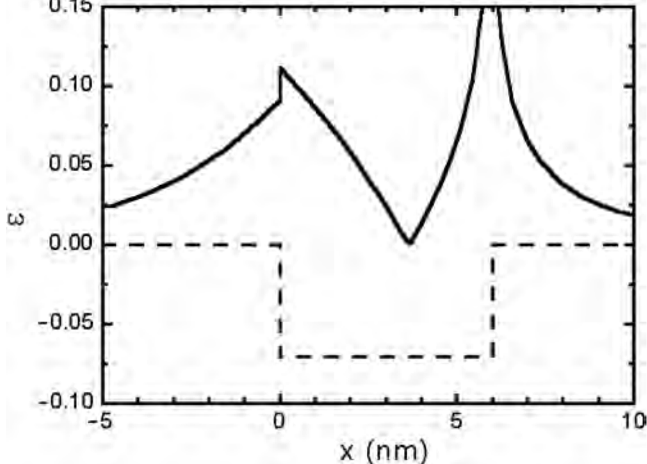


Figure 6. Strain, calculated via IE approach [51], as a function of the coordinate x of the dot axis, for a pyramidal quantum dot with 6 nm height, 12 nm base, and a lattice mismatch 7.2% inside. The dotted line indicates the material dilation, while the continuous line shows the deviation of strain from the purely hydrostatic form.

is given by [56]

$$P_k = e_{1,4}(\epsilon_{ij} + \epsilon_{ji}) \tag{4}$$

where $i \neq j \neq k$ and $e_{1,4}$ is the piezoelectric constant. The divergence of the polarization P_k gives rise to piezoelectric charges spatially distributed around the dot thus contributing to the QD potential profile.

3.2. Energy Level Scheme

3.2.1. Single Electron States

The more widespread method for calculating the electronic energy level scheme in QDs is the effective mass approximation (EMA) [35, 52, 53, 57, 63, 64]. In this approach, the states of the dots are determined by expanding their wavefunctions on a basis containing the zone center (Γ) bulk valence band states (from two to six states, including spin) and the bulk conduction band (two states, including spin),

$$\Psi_i^{\mathrm{EMA}}(\mathbf{r}) = \sum_{n=1}^{N_B} f_n^{(i)}(\mathbf{r}) u_{n,\Gamma}(\mathbf{r}) \tag{5}$$

where $u_{n,\Gamma}(\mathbf{r})$ are the bulk Bloch functions at Γ and $f_n^{(i)}(\mathbf{r})$ are the envelope functions satisfying the equation

$$H_{\mathrm{EMA}} f = Ef \qquad f = [f_1 f_2 \cdots f_{N_B}] \tag{6}$$

The H_{EMA} Hamiltonian contains the contributions coming from the band structure of the heterostructure, the strain field, and the piezoelectric field. The calculation of the strain field is performed usually by solving the elasticity continuum theory problem with numerical methods. Comparison of the changes in the obtained electronic band structure when using the IE analytical solutions or the VFF atomistic method is reported in [65] and in [52], respectively.

Early calculations by Grundmann et al. [35] for InAs/GaAs QDs made use of a three-dimensional EMA single band Scröedinger equation for the electrons and holes moving in the strain-induced potential corresponding to the band edges. The effective masses were different in the island and barrier materials, but the effective mass dependence on the strain was not included.

This approach has been subjected to a number of criticisms, mainly dictated by the model prediction of a single electronic bound state in QD, in poor agreement with the experimental findings. As pointed out in [63], the narrow InAs bandgap leads to significant band mixing, resulting in large strain induced variations of the effective masses. Based on a pseudopotential calculation, the authors of [63] set the electron effective mass $m_{\mathrm{eff}} = 0.04 m_0$, which is the value predicted for bulk InAs under the average hydrostatic strain in the island. This gives two electronic confined levels, in better agreement with the experimental findings.

As matter of fact, the calculation of the QD band structure by solving the decoupled electron and hole problems is not correct in the case of SK QDs. In general, in SK QDs strong mixing of the bands is promoted by the low bandgap of the constituent semiconductors, by the very large strain field, and also by strong spatial variation of the strain field. More reliable results are claimed when using EMA methods with a larger set of bands in the wavefunction expansion given in Eq. (5) [52, 53, 64, 66].

A very different picture comes from the GaAs/AlGaAs QDs grown by droplet epitaxy. Here no particular strain is expected due to close resemblance of the lattice constant of the QD and the material. In this case, a simple one band EMA is sufficient in order to obtain theoretical predictions very close to the observed transitions [67].

Due the confined nature of the QD states, a good reproduction of the band structure in the proximity of the Γ point becomes necessary. In their calculations Stier et al. [52, 68] selected an EMA set of parameters which gave the best available agreement with the bulk band structures calculated from the pseudopotentials within the largest possible region around the Γ-point.

The piezoelectric potential is a crucial ingredient for obtaining the correct description of the QD electronic properties. In the most widely studied case of QDs grown on (100) substrates, the shear strain at the QD edges gives rise to a quadrupolar piezoelectric charge distribution which reduces the symmetry of the QD (in the case of QDs of pyramidal shape the symmetry is lowered from C_{4v} to C_{2v}) [35]. The presence of a piezoelectric field only slightly alters the ground state energies (few meV) while it forces the electron states to prefer the [110] direction and the holes to be more attracted toward the $[1\bar{1}0]$ directions. The effect is more marked on the hole states. This symmetry breaking reduces the overlap between the electron and hole states and has strong effects on the polarization dependence of the optical transition matrix [52, 68]. Its effect is shown in Figure 7 where such reduction can be appreciated both on the band edge profiles and on the ground state wavefunctions.

In Figure 8 the electron and hole energies in pyramidal InAs/GaAs QDs for different sizes, calculated using the CET strain model and the eight band ($N_B = 8$) EMA method [52, 68]. The presence of the WL at the base of the QD was taken into account. The discrete nature of the electronic levels is clearly perceptible. This gives rise, as in the case of atoms, to a well defined shell structure of the electronic levels.

For pyramidal QD with a base b smaller than 10 nm only 1 electronic state is predicted, but for $b = 20$ nm there can be as many as 11 [52]. The ground state of electrons (see Fig. 7) is always s-like, with a shape slightly elongated along the [110] direction. Excited states have p-like symmetries. No intermixing between the QD electronic states and the WL has been found. Holes are generally confined at the bottom of the pyramid. The shape of their wavefunctions strongly depends on the dot size due to pronounced band mixing effects. The wavefunctions of the excited hole states extend into the WL, thus coupling with the 2D quantum well states.

More recently, empirical pseudopotential methods [69] able to calculate the electronic structure of QDs containing $\approx 10^6$ atoms became available [55]. In this pseudopotential approach the wavefunction, $\psi_i(r)$, is a solution of the Schrödinger's equation

$$\left\{-\frac{1}{2}\nabla^2 + \sum_{n\alpha} \hat{v}_\alpha(r - R_{n\alpha})\right\}\psi_i(r) = \epsilon_i \psi_i(r) \tag{7}$$

where $\hat{v}_\alpha(r - R_{n\alpha})$ is the pseudopotential of the atoms of type α and $R_{n\alpha}$ is the relaxed position of the nth atom of type α. The atomic pseudopotentials $\{v_\alpha\}$ are strain

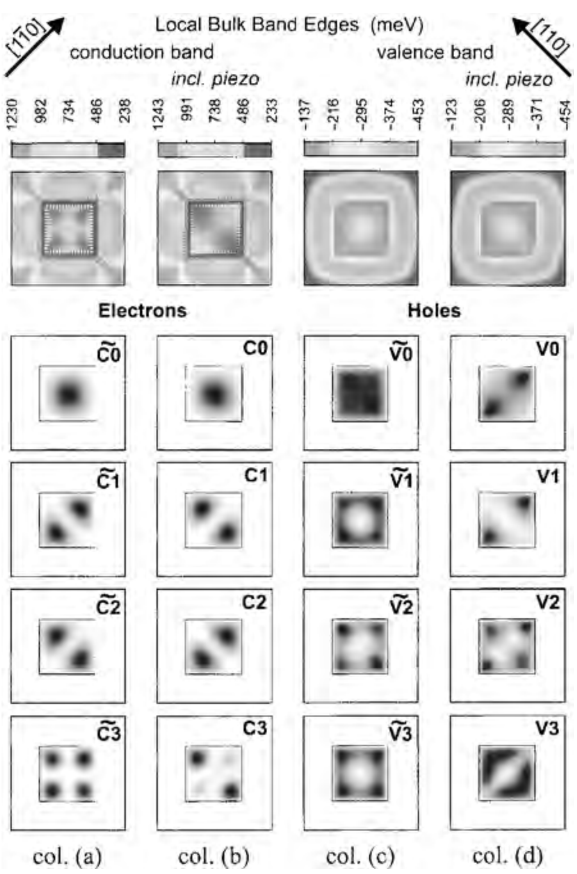


Figure 7. Top row: electron (cols. a,b) and hole (cols. c,d) band edge profiles in a (001) cross section plane 1.0 nm above the WL with and without piezoelectric effect for a square based InAs/GaAs QD. The energies refer to that of the unstrained InAs. Other rows: corresponding probability density of bound states along the same section plane. After [68].

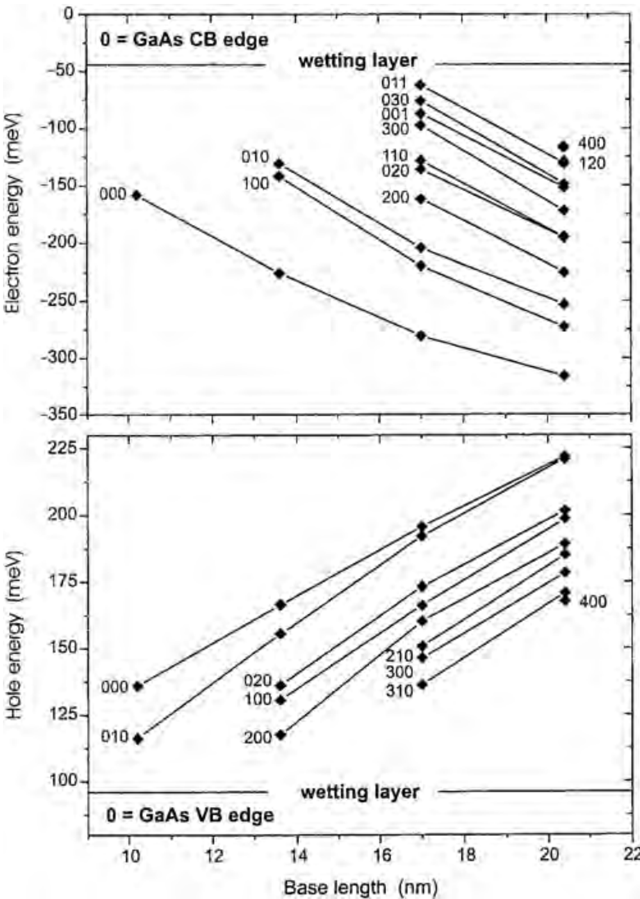


Figure 8. Electron and hole energies in pyramidal InAs/GaAs QDs for different sizes, calculated using CET strain model. The lines connect levels having the same wavefunction symmetry. After [68].

dependent and are carefully fitted [70] to measured bulk bandgaps, effective masses, and first-principles calculations of the band offsets [71] and deformation potentials [72]. The relaxed atomic positions are determined by minimizing the VFF energy, as outlined in the previous section. The main disadvantage of this method is the requirement of large computer time, although some attempts toward a more treatable solution of the problem have been taken [73]. In addition, the pseudopotential approach, in the present version, is unable to treat the piezoelectric field contributions.

The advantage of the pseudopotential method is the ability to take into account the whole band structure of the material when calculating the electronic structure of the confined system. This becomes a crucial point when treating with semiconductors where other minima in the **k** space may contribute to ground state wavefunction [74]. In principle, also the EMA method could give an accurate description of the band structure effect, provided that the number of the Γ functions N_b involved in the calculation [Eq. (5)] is sufficiently large. However, in the EMA calculations the use of more than eight bands becomes rather cumbersome. As shown by Wang et al. [74], $N_b = 8$ becomes, in some cases, an insufficient base for a correct description of the QD electronic structure.

Eventually, the electronic level schemes of a "well behaved" InAs/GaAs QD system, calculated via a pseudopotential method [75] or via the eight-band EMA [52], largely agree [68].

The oscillator strength of the electron–hole transitions can be theoretically determined from the single electron level scheme [52, 63]. In the linear response approximation the resulting spectrum is equivalent to the absorption. A comparison of the oscillator strength spectra with the emission spectra must be performed cautiously because population [76] or many-body effects [77] may substantially change both the energy and the intensity of the transitions.

The electron–hole transition spectra exhibit [52] a large number of lines which reflect the presence of several excited electron and hole states. Some electron states yield a significant oscillator strength with more than one hole state so they contribute to more than one line. Therefore straightforward attribution of a line to a given transition, except the ground state one, is impossible because the peaks are made of a significant number of equal strength electron—hole transitions with no particular dominant transition [52, 78].

A significant decrease of the optical strength of the ground state transitions with increasing dot size has been found, owing to the increasing piezoelectric field which, by symmetry breaking of the hole state wavefunction, reduces the overlap with the electron [52]. Moreover, the

piezoelectric induced symmetry breaking induces a polarization anisotropy between the two directions [110] and [1$\bar{1}$0].

3.2.2. Multiparticle States

As soon as the QD is populated with two or more carriers, the single particle description of the electronic properties is no longer valid. As a matter of fact, the Coulomb energy, the exchange, and the correlation effects between the carriers lead to the formation of distinct many-body states with different energies. Thus a demanding challenge in the QD modeling arises when calculating the optical transitions because, in addition to the shape, size, and composition of the dot, they are also dependent on the population of the dot itself. Multiparticle states have a strong influence not only on the QD optical properties, but also on the way in which a QD is charged. The QD state filling by carrier injection has been calculated, taking into account the many-body effects, by Fonseca et al. [79] who showed that, even in the presence of large state filling, the QD shows a distinct electronic shell structure and that the state filling follows Hund's rule.

Generally speaking, the many-body effects on the QD electronic structures are usually divided into three different contributions, namely: (1) a perturbative part, reflecting carrier–carrier direct Coulomb and exchange energies obtained from fixed single-particle orbitals, (2) a self-consistency correction arising when the single particle orbitals are allowed to adjust to the presence of carrier–carrier interaction, and (3) a correlation correction [80]. This picture is correct since self-assembled III/V semiconductor QDs are always in the strong confinement regime.

The presence of more than one electron–hole (e–h) pair inside the QD strongly modifies the QD radiative emission, leading to the so-called multiexciton transitions [77, 81, 82]. The energy of the photon emitted by the recombination of a single e–h pair depends on the number of carriers filling the QD. The essential physics of the multiexciton transitions can be understood by considering the fundamental recombination from the lowest electron level and the highest hole level if other electron–holes pairs are present in the dot as spectators [77, 81, 83]. Let us consider only the Hartree–Fock energy (thus neglecting correlation effects) of a single configuration of electrons and holes [77, 83]. By indicating with J_{ij} the direct Coulomb interaction and with K_{ij} the exchange interaction between carriers in levels i and j, the recombination energy $E_{eh}^{1\to 0}$ of the single exciton is given by

$$E_{eh}^{1\to 0} = (\epsilon_{e_1} - \epsilon_{h_1}) - J_{eh} \tag{8}$$

where ϵ_{e_1} and ϵ_{h_1} are the single-particle levels. Then the electron–hole recombination in the presence of $N-1$ spectator electrons and holes is [83]

$$E_{eh}^{N\to N-1} = \left[\sum_{i=2}^{N}(J_{e_1e_i} - J_{e_1h_i}) + \sum_{i=2}^{N}(J_{h_1h_i} - J_{h_ie_1})\right] + - \left[\sum_{i=2}^{N} K_{e_1e_i} + \sum_{i=2}^{N} K_{h_1h_i}\right] + E_{eh}^{1\to 0} \tag{9}$$

where e_1 (h_1) is the electron (hole) that recombines and e_i (h_i) are the spectator electrons (holes). Thus the e_1–h_1 transition is shifted in energy with respect to the single exciton transition energy by a quantity depending on the number of spectator charges present in the QD. The sign of this shift depends on the exact material characteristics and on the charge state of the QD [84]. Accurate many-body calculations of ground state transition energies as a function of the number of excitons in the InAs/GaAs QD system can be found in [77, 83]. In this case, multiexcitons transitions occur at lower energy (few meV) with respect to single exciton transition energy. Experimental evidence of multiexciton transition in QDs has been reported by using single QD spectroscopy (see Section 4.1.8).

3.3. Dot–Dot Correlation

As reported in Section 2.1.1, repeated depositions of InAs on thin GaAs spacers result in the formation of stacked, vertically aligned, InAs QDs in a GaAs matrix. The self-ordering of QD columns produces strong quantum mechanical modifications of the electronic structure. Whenever the barrier thickness is thin enough to allow tunneling, a carrier delocalization occurs and the carrier wavefunctions become extended over many QD sites along the column. The coupling between adjacent QDs leads to the splitting of the levels [85, 86] and eventually minibands can open in a QD superlattice [87].

The fascinating concept of self-aligned QD molecule has been introduced [88, 89] and several interesting aspects have been addressed. In connection with the growing interest in quantum information processes, the use of self-aligned QD molecules as quantum gates has been recently proposed and the possibility of entangled states of the electron–hole complex in QD molecules has been reported [90].

It is usually assumed that the hole coupling is weaker than the electronic one, due to the heavier hole masses [40, 86]. However, a detailed study of the electron and hole confinement energies in stacked self-assembled InAs QDs as a function of the layer separation is reported in [91]. It is shown that very little coupling is observed in case of electron levels, while the hole levels are strongly affected by the coupling, which definitively lifts the hole degeneracy. Of consequence, this results in a different character of the electron and hole wavefunctions. It is also shown that the electron wavefunction in a nine-layer structure with an interlayer separation of about 10–15 nm remains localized within each QD, whereas a complete delocalization over the entire dot column is predicted for the hole wavefunction. Only for smaller interlayer separations is the electron wavefunction also delocalized over the coupled QDs.

Carrier delocalization along the QD column modifies the recombination kinetics in stacked QDs, as well [91–93]. An increase of the emission lifetime in stacked QDs has been observed [91, 92]. The interpretation relies on the reduction of the oscillator strength of the exciton inside the QD column [91], also due to the different delocalization between the wavefunctions of electrons and holes [91, 94].

4. PHOTOLUMINESCENCE

The study of the optical properties of semiconductor structures is a very wide topic. It covers a large variety of experimental techniques and it allows one to investigate many

physical aspects ranging from the intimate nature of quantum mechanics to the characterization of impurity contamination and defect content. The optical methods are very often complementary of the electrical measurements, which are the other powerful techniques for the investigation of electronic properties in semiconductors. They offer the advantages of the spectroscopic techniques, that is, the possibility of discriminating over different contributions. A further advantage of many optical methods is the requirement of very small, if any, sample preparation which therefore makes it a noninvasive technique. Among the optical spectroscopies, PL is certainly the most widely used, due to its simplicity and powerfulness.

The PL consists of the spectroscopic analysis of the radiation emitted by a system which has been excited via absorption of light with energy above the absorption edge. The radiation is generated by the radiative recombination of the photoexcited electron–hole pairs. The analysis of the photoluminescence spectrum as a function of different parameters, such as temperature, excitation energy, excitation intensity, and external fields, provides general information on the electronic properties and on the quality of semiconductors and semiconductor heterostructures. Generally speaking, the PL spectroscopies can be divided into steady state (or CW) and time resolved. We will treat separately these two experimental techniques.

4.1. CW Photoluminescence

The CW-PL is a well established technique for the study of semiconductor structures and we refer to [95–98] for general reviews on different types of recombination in semiconductors and also to experimental details of the CW-PL apparatuses. Here we will review only a few experimental aspects that are of peculiar importance for the study of QD structures and then we will refer to the main topics that can be addressed by CW-PL in the study of QD structures.

4.1.1. Experimental

The measurement of QD-PL spectra is commonly performed using conventional optical systems. The exciting source is a laser with energy higher than the energy gap of the confining barrier and the exciting laser beam is focused on a spot of diameter of a few tenths of a millimeter. In this case the QD-PL band is almost unstructured and broad some tens of meV, due to the spread of the QD dimensions. Thus, no special resolution is required for the detection.

More demanding are the optical systems needed to measure the PL spectra of a single or of a few QDs. In order to resolve the single QD emission high spectral and spatial resolution is required. The intrinsic homogeneous linewidth of the fundamental optical transition at low temperature has been found to be as small as 2 μeV. [99]. The need for this high spectral resolution is not very common in the PL study of semiconductor structure and it is related to the atomic-like nature of the QD electronic structures. High spectral resolution apparatuses are obtained by the use of diffractive spectrometers with high focal length [100] and/or inserting a Michelson interferometer in the detection path of a photoluminescence system [101]. At the same time very high spatial resolution is needed for isolating a single QD emission. A submicrometer spatial resolution is required for single QD spectroscopy. This has been realized following different methods. The simplest way is to use a micro-PL apparatus that can allow diffraction limited resolution. To fix the orders of magnitude, let say that a 1 μm spatial resolution allows one to detect one single QD only if the areal density is lower than 10^8 dot cm^{-2}, which is slightly smaller than the achievable QD densities. Fortunately, due to the inhomogeneous size distribution of QDs, the emissions of individual QDs are spectrally separated and even for 10^9 dot cm^{-2} it is possible to isolate a single QD emission with micro-PL techniques after a careful choice of the point on the sample. Nevertheless the control of the areal density down to 10^9 dot cm^{-2} or even lower is not a very simple task for the epitaxial growth of self-aggregated QDs [82]. To further improve the spatial resolution near field optics [102] or sample processing with very small mesas (hundreds of nm) [101] have been used.

Usually, the PL of quantum structures in general and of QDs in particular is excited using photons whose energy is higher than the energy of the bandgap of the barrier of the structure. However, excitation spectroscopy is frequently used for characterizing the complex electronic structure of QDs. Excitation spectroscopy means the study of PL emission as a function of the energy of the absorbed photons. Two main spectroscopic methods have been used so far: resonant PL (RPL) and PL excitation (PLE). PLE measurements are made by monitoring the intensity of a given emission energy (usually within the inhomogeneous broadened QD-PL band) while the excitation energy spans the QD density of states. Peaks in the PLE spectrum are interpreted as resonances in the QD absorption spectrum due to excited states. RPL measurements refer to PL spectrum taken for a given excitation energy which is usually below the absorption edge of the bidimensional WL. The RPL spectrum shows a spectral narrowing due to the selection of a sub-ensemble of QDs which have excited states in resonance with the laser energy. The relationship between RPL and PLE has been recently discussed in [103].

4.1.2. General Aspects

The analysis of the CW photoluminescence spectra of 0D strained epitaxial heterostructures provides important information on the electronic and morphological properties of these systems [21]. In this section we introduce the main aspects of the CW photoluminescence of QDs which will be discussed in more detail in the following sections.

A first example is reported in Figure 9 for the InAs/GaAs system; PL spectra measured for increasing thickness of the deposited layer evidence the morphological changes induced by the coverage variation [104, 105]. At low coverages ($\Theta <$ 1.5 ML) a high energy structure peaked at about 1.45 eV appears, which smoothly shifts to lower energies and whose intensity increases for increasing coverage. This structure, which is attributed to the 2D confined states of the WL, vanishes at coverages just beyond the critical thickness L_c. At this coverage the PL spectra provide some evidence of more confined states due to the islands or QDs. The corresponding PL band is peaked at about 1.215 eV, thus showing that the QD localization energy with respect to the WL

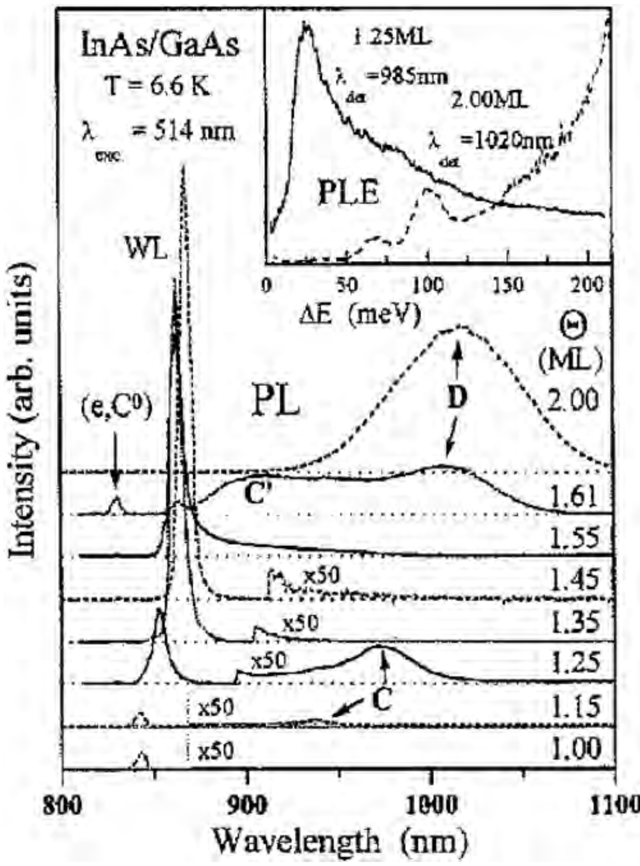


Figure 9. PL and PLE (inset) spectra measured at low temperature ($T = 6.6$ K) of InAs/GaAs samples with different InAs coverages θ; the excitation power density is 5 W/cm^2. C and C′ indicate large and small quasi-3D clusters, D fully developed QDs. For the PLE spectra $\Delta E = E_{\text{excitation}} - E_{\text{detection}}$. Reprinted with permission from [105], T. Ramachandran et al., *J. Crystal Growth* 175/176, 216 (1997). © 1997, Elsevier Science.

ground state is about 200–250 meV. After the SK transition, the narrow peak due to the WL disappears from the spectra measured at low and intermediate excitation power densities, as shown in Figure 9. The WL optical transition, however, still appears increasing the excitation intensities due to the filling of the lower energy confined states [106].

Another interesting feature of the PL spectra consists in providing information on the morphological characteristics of the QDs [107]. Low temperature PL spectra measured on different InAs/GaAs QD structures grown with increasing InAs coverage show that, slightly after the SK transition, the QD PL band slowly shifts toward the red with increasing the coverage. This is proof of the tendency toward the increase in the QD size with the amount of deposited matter (Fig. 10). Analogous results are reported also in Figure 2 of [108].

The low temperature PL spectra measured at low exciting power densities, well before the appearance of the effects related to the excited states, clearly reflect the QD size distribution. The presence of a structured band is usually an indication of multimodal QD size distribution, as evidenced by the correlation between the PL band shape and the size distributions deduced from the morphological

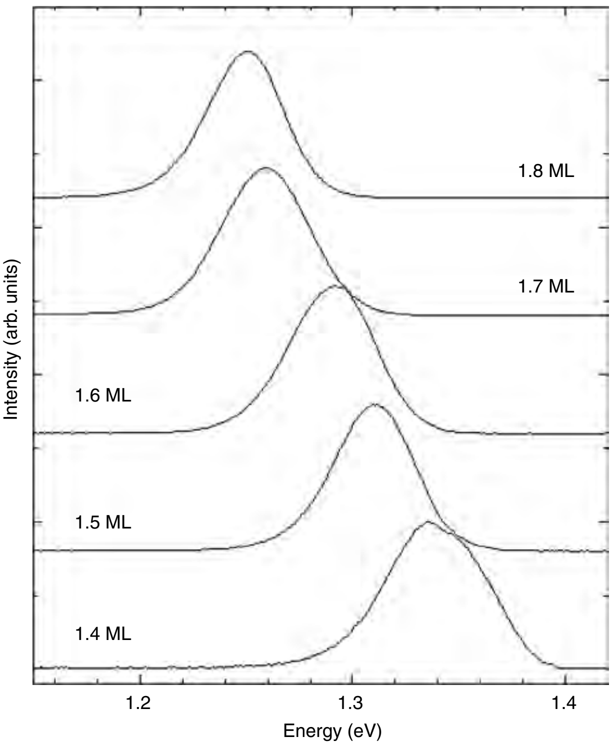


Figure 10. Photoluminescence spectra of InAs QDs grown on (100) substrate with different coverages measured at $T = 2$ K with an excitation power density of 20 W/cm^2. After [121].

analysis of QDs [109]. An example is reported in Figure 11 for InAs/GaAs QDs grown on (311)A and (100) surfaces. Figure 11 shows, for each surface, the histograms of the QD lateral size together with the respective low temperature PL spectra. A well defined bimodal distribution is observed

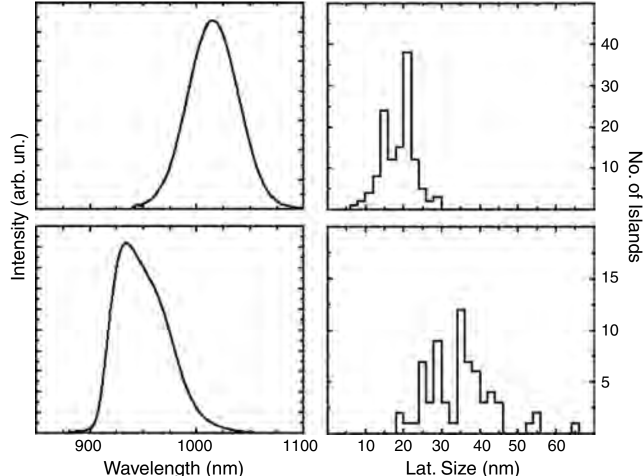


Figure 11. Comparison between the 2 K PL spectra and the lateral size histograms of InAs/GaAs QDs grown on (100) and (311)A substrates. In the left panels the PL spectra are reported: top spectrum (100) sample, bottom spectrum (311)A sample. In the right panels the size histograms are reported: top histogram (100) sample, bottom histogram (311)A sample. Reprinted with permission from [109], S. Sanguinetti et al., *Micron* 31, 309 (2000). © 2000, Elsevier Science.

for the QDs grown on GaAs (311)A and the corresponding PL spectrum shows an asymmetric lineshape with two clearly visible structures. In contrast to the (311)A case, a broad QD size distribution characterizes the (100) sample with a corresponding broad and unstructured band in the PL spectrum.

In analogy with the QW case [98], also in the case of QDs the conduction band and the valence band confinement barriers play an important role in determining the energy of the optical transition. In Figure 12 the blueshift of the PL spectra of InAs/$Al_yGa_{1-y}As$ self-assembled QDs for different Al mole fractions in the barrier are reported [110–112]. The PL spectra show a broad and intense band due to the electron–hole recombination in the dots, whose peak energy shifts from about 1.13 to about 1.55 eV increasing the Al mole fraction y from $y = 0.0$ to $y = 0.8$. The PL band blueshift is likely due to the increase of the potential confining barrier. However, following [110], also the formation of smaller dots when the barrier is formed by the alloy and to the Al incorporation in the dots during the (AlGa)As overgrowth contribute to the blueshift. The two last effects seem to dominate for the higher Al concentrations.

The photoluminescence is one of the few spectroscopic techniques which provide information on the excited electronic states. Both PLE spectra and RPL spectra have been applied to the study of excited states in QDs. One of the most interesting results is the evidence of the so-called quantum size effect. The analysis of the PLE spectra of self-organized InAs/GaAs QDs has shown that the energy of the resonances in the PLE spectra depends on the choice of the detection energy inside the inhomogeneously broadened PL band [103, 113, 114]. This means that, as expected from simple quantum mechanical considerations, the excited state transition energy is a function of the ground state transition energy. These results have been supported by an eight-band EMA calculations for pyramidal QDs [114]. The results of

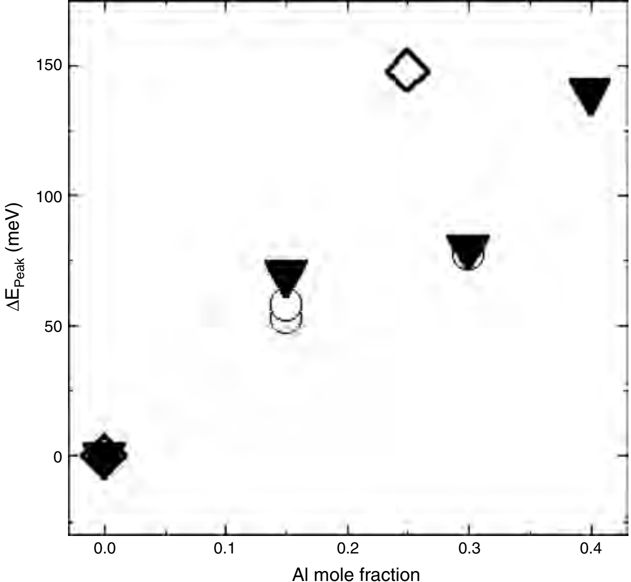


Figure 12. Blueshift the QD PL band as the Al content is increased. Different symbols refer to data taken from [111] (circles), [110] (triangles), and [112] (diamonds).

RPL spectra on InGaAs/GaAs QDs confirm these conclusions [103]. Evidence of the size dependence of the excited state spectrum has been reported also in the case of stacked InAs/GaAs QDs [115].

4.1.3. Study of the Quantum Dot Morphology

As already discussed in the previous sections, the QD morphology strongly influences the CW PL spectra and, of consequence, the PL spectra provide interesting information on the morphological properties of QDs.

In the SK growth regime (see Sections 2.1 and 4.1.2) it is usually assumed that at coverages larger than the critical thickness L_c a layer by layer to island transition sets in the growth process and QDs start to grow superimposed to the interconnecting WL. Obviously L_c is one of the most relevant parameter for the SK QD growth, and, as discussed in Section 4.1.2, it can be estimated from PL measurements. However, different values of the critical thickness for surface elastic relaxation have been reported, ranging from less than 1 to 1.8 MLs. As a matter of fact it is rather difficult to exactly define experimentally the critical thickness, because the determined value of L_c strongly depends on the sensitivity of the technique used [116]. The concept of L_c has been critically commented on in [116], where it has been shown that the critical thickness does depend, even if in a limited range, on the growth conditions and in particular on the growth temperature. The description of the initial stages of QD growth is not univocal, thus evidencing the complexity of the problem. For instance, at coverages just below the critical one, the evidence of some kind of QD "precursors," the so-called "quasi three-dimensional islands," has been reported [104]. These structures thus behave in an intermediate way between WL and QDs. In addition the study of InAs/GaAs QDs grown on nonconventionally oriented surfaces only partially confirms the results of Figure 9 [107]. In fact at coverages lower than the critical thickness weak bands appear on the low energy side of the WL peak whose energy does not seem to change with the coverage (Fig. 13). As discussed in [107], this suggests that in these systems the QD self-assembling starts well before a truly

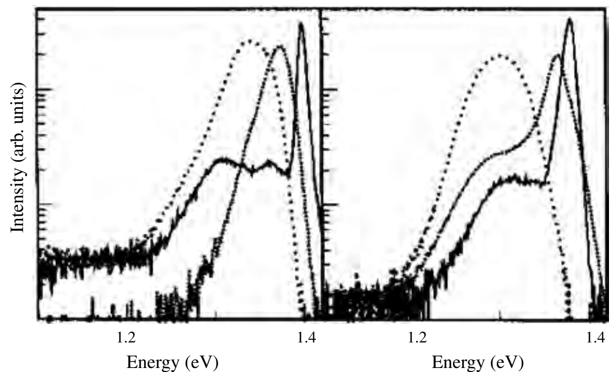


Figure 13. Photoluminescence spectra of InAs QDs grown on (511)B and (311)B substrates with different coverages measured at $T = 2$ K with an excitation power density of 20 W/cm^2. Reprinted with permission from [107], S. Sanguinetti et al., *Mater. Sci. Eng. B* 74, 239 (2000). © 2000, Elsevier Science.

three-dimensional growth mode takes place and the QD size seems not to change appreciably across the SK transition. Thus the island nucleation on high index planes seems to proceed through a direct nucleation of QDs with a well defined size and aspect ratio rather than through the formation of quasi three-dimensional islands as shown for the (100) growth in [104].

The relationships between the QD optical properties and morphology are discussed in [116], comparing the features of PL spectra measured on several tens of MBE samples grown in different laboratories and under different growth conditions. This analysis was stimulated by the observation that PL spectra of QDs grown with the same coverage L are characterized by spectral bands with different peak energies E_p and lineshapes. The peak energy of the PL bands E_p measured at low temperature as a function of the thickness of the deposited layer L for a number of InAs/GaAs QD samples is reported in Figure 14 from [116]. The authors highlight an excellent agreement between their results (Fig. 14) and literature data, thus proving that the (E_p, L) plane is intrinsically divided into the three regions shown in Figure 14.

The interpretation of these data is as follows. The energy of the PL emissions of samples with L lower than about 1.5 MLs (region I) shows a weak linear dependence on L (about 40 meV/ML) and is some tens of meV lower that the WL recombination. The emission in this region can be attributed to the three-dimensional precursors of QDs, that is, lens-shaped structures with a quite low aspect ratio. These emissions are seen only in samples grown at low temperature (420 °C). For L values around the critical thickness L_c, namely in the 1.5–2.0 MLs range (region II), E_p shows a strong dependence on L (about 600 meV/ML). In this region QDs develop and their aspect ratios and E_p are intermediate between those of precursors and those of well developed QDs. The high slope of the E_p on L dependence can be attributed to the change in the dimension and shape of QDs. Finally, at L values larger than about 2 MLs, samples grown in different conditions give rise to independent sets of emission energies, with the result that discrete values of energy can be observed at a given L value. The rate of change of E_p with L decreases to the value observed in region I. As suggested by morphological high resolution cross-sectional transmission electron microscopy characterization, a different lateral faceting seem to characterize the QDs in the samples connected with the different lines in region III.

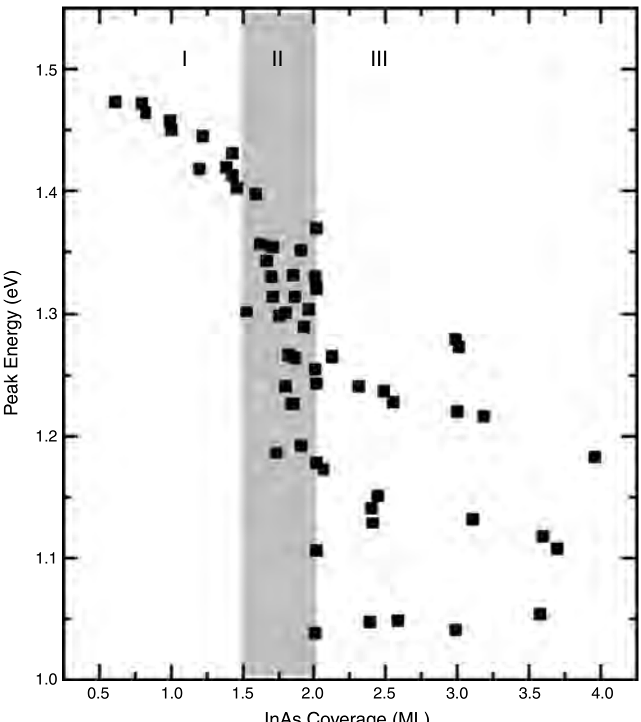


Figure 14. Peak emission energies of self-assembled InAs/GaAs QDs vs InAs coverage L. The data are taken from [116] and references therein. The three regions discussed in the text are indicated. Courtesy of M. Capizzi.

Obviously, the size also influences the energy of the QD excited states. The dependence of the energy of the QD excited states on the QD size has been analyzed in InGaAs/GaAs samples [103]. The dependence of the resonance energy (determined by the PLE and RPL spectra and attributed to excited state transitions) on the energy of the QD ground state allows the authors to evidence the combined role of quantum size and quantum shape effects in determining the energy of QD excited states.

The QD morphology can also be modified by a suitable choice of the GaAs substrate orientation. An early study of strained InGaAs epilayers grown on non-(100)-oriented substrates showed that this is a promising method for the self-aggregation of nanostructures [117]. Indeed the self-organization of InGaAs QDs on different GaAs surfaces intermediate between (100) and (111) has been subsequently demonstrated and the PL of these QDs has been studied [118–120].

A careful analysis of the QD self-assembling process in these QD systems has shown that the island assembling critical thickness in InAs QDs is affected in a controlled way by the substrate orientation. The determination of the onset of the island nucleation via PL measurements in samples grown on (100), (311)B, and (511)B oriented substrates is reported in [121]. In this paper it is shown that the critical coverage on the (311)B surface is larger than that on the (511)B substrate, which in turn is larger than that on the (100) surface. In [121] it is concluded that the driving force of the critical thickness dependence on the substrate orientation is the reduction of the in-island strain relaxation in high Miller index substrates. Furthermore, modifications of the internal energy induced by island shape and faceting seem to be less relevant than strain relaxation effects [121].

Another powerful tool in order to manipulate the QD morphology and, in turn, the QD energy levels is the rapid thermal annealing (RTA) technique [122–125]. During the RTA treatment the QD material undergoes to a thermal annealing in argon atmosphere for few tens of seconds and for temperature ranging from 700 to 950 °C [124, 126]. The presence of group III vacancies in the InGaAs/GaAs QD material should stimulate [127], under RTA, interdiffusion of the In and Ga atoms at the interface between the QD and the barrier [126]. This should result in a change in the

QD composition, thus reducing the confining potential and an effective increase in the dot size [126]. The study of the morphological modifications of the QDs after RTA has been performed mainly by PL measurements, a well established method for the characterization of the interdiffusion effects in nanostructures [128, 129]. As expected (Fig. 15) an increasing blueshift in the QD emission, caused by a change in the QD composition, and a narrowing QD PL line, due to a larger mean size of the QDs, is observed for increasing RTA temperature [123, 124, 126].

4.1.4. Optical Anisotropy

Although quantum dots are often referred to as artificial atoms, their geometrical shape is quite often characterized by a much stronger confinement along some peculiar crystallographic directions. This morphological in-plane anisotropy of the QDs can be directly monitored by analyzing the polarization properties of the PL emission both from the top and from the edge of the samples.

A clear indication of optical anisotropy in QD emission has been reported in [130] where natural QDs, formed by large interface fluctuations in GaAs/AlGaAs quantum wells, are studied. A scanning tunnel microscope (STM) analysis shows that these islands are asymmetric with an elongated shape along the [−110] crystal axis. As a consequence of this reduction of symmetry, the PL spectra are polarized. A discussion of the origin of the observed polarization behavior is reported in [130].

The presence of optical anisotropy is of the utmost relevance also for laser application. In [131] anisotropic self-assembled $In_{0.5}Ga_{0.5}As/Al_{0.25}Ga_{0.75}As$ QDs, whose PL spectra show optical polarization, are used to control the polarization mode of a QD vertical cavity self-emitting laser.

QDs with very large morphological asymmetry can be growth on (N11) oriented substrates. InAs quantum dashes have been grown on GaAs (211)B substrates by increasing the growth temperature [132]. Arrow-like InAs QDs have been obtained by growth on (311)A GaAs substrates. The peculiar shape of the (311)A InAs QDs is determined by the properties of the reconstruction of the surface on which nanostructures are grown. The (311)A surface reconstruction, at typical MBE growth temperatures, discussed in detail in [133–135], gives rise to a corrugated surface with channels running along the [$\bar{2}$33] direction.

The shape of MBE InAs QDs grown on the (311)A oriented surface reflects this peculiar reconstruction. The STM analysis (Fig. 16) of QDs grown with 1.8 ML of InAs on (311)A and (100) substrates placed side by side provides evidence of a marked shape differences. The QDs grown on (100) oriented substrates have a shape ranging from round domes to square based pyramids, whereas the (311)A QDs display an arrow-head-like faceted shape. The arrow orientation points along the [$\bar{2}$33] direction, which is related to the (311)A surface reconstruction.

The QDs grown on two faces have comparable PL efficiency and a strict mutual relationship is observed between the QD shape and their optical response [136]. In particular, no PL polarization anisotropy was observed on the (100) grown sample. On the contrary, on the (311)A sample the maximum intensity was observed with a Polaroid analyzer parallel to the [$\bar{2}$33] direction, the same direction of the arrow QD tips (see Fig. 16). The polarization ratio ρ (i.e., the fraction of light polarized along [$\bar{2}$33]) was

$$\rho = \frac{I_{[\bar{2}33]} - I_{[01\bar{1}]}}{I_{[\bar{2}33]} + I_{[01\bar{1}]}} = 0.13 \pm 0.02 \tag{10}$$

where $I_{[\bar{2}33]}$ and $I_{[01\bar{1}]}$ are the integrated PL intensities along the two orthogonal directions [$\bar{2}$33] and [01$\bar{1}$], respectively. An explanation of the observed polarization anisotropy has been proposed taking into account the nonspherical QD shape as well as of the degeneracy of the valence bands in bulk III/V semiconductors [137].

It should be mentioned that the polarization anisotropy of QD emission is connected to asymmetries in the hole confining potential introduced or by anisotropic shapes [136, 137] or by quadrupolar piezoelectric effects (see Section 3.2.1) [52]. Symmetric QDs show isotropic oscillator strength [137].

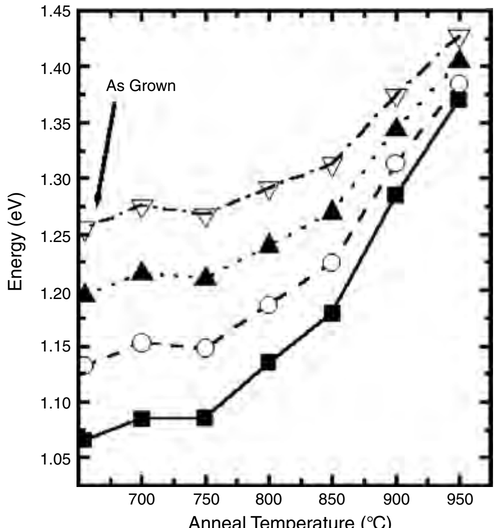


Figure 15. Photoluminescence peak energies of the ground state and of the first three excited state emissions of as-grown and rapid thermally annealed InAs/GaAs self-assembled quantum dots. Data taken from [124].

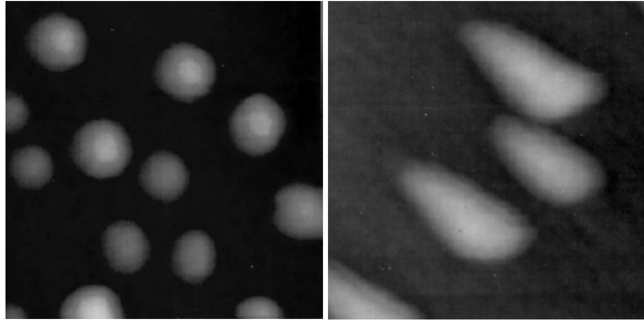


Figure 16. STM pictures (100 × 100 nm^2) of InGa/GaAs QDs grown on (100) (left panel) and (311)A (right panel) surfaces. Courtesy of P. Moriarty.

Lowering the symmetry of the QD leads to a substantial mixing between "light-hole" (LH) and "heavy-hole" (HH) states, hence to a polarization asymmetry of the oscillator strength.

An interesting study addressing the polarization properties of the interband optical dipole transition in InAs/GaAs QDs is reported in [138]. The reported results show, in fact, that a moderate LH–HH mixing characterizes the ground hole state.

4.1.5. Photoluminescence Thermal Quenching

The temperature dependence of the QD PL shows some complex characteristics. In fact, an unusual temperature dependence of the integrated intensity, the energy of the maximum, and the width of the PL bands has been reported in InAs/GaAs QDs by several groups [139–144]. In particular, the peak energy of the PL bands exhibits a sigmoidal behavior and, at low temperature, the PL band linewidth decreases with increasing temperature. Such features cannot be explained using the rate equation models commonly adopted to describe the temperature dependence of the PL in confined structures, even if one introduces two activation energies to account for the temperature quenching of the PL bands. For instance, in [110] it is reported that the temperature dependence of the PL integrated intensity of InAs/GaAs/$Al_xGa_{1-x}As$ QDs with different Al mole fractions x cannot be described simply in terms of an Arrhenius plot with one or more activation energies. Moreover, the general description in terms of a carrier thermal evaporation from the QD ground state is complicated by the presence of the WL connecting the QDs. The use of a simple thermal escape model has been questioned and the critical role of the carrier capture and relaxation, as well as of their dependence on temperature, in determining the QD radiative recombination efficiency has been pointed out [115]. In particular, a detailed study of the temperature dependence of the PL decay measured at different wavelengths across the PL band is reported in [144]. The results strongly suggest that excitons or carriers, thermally activated from the QD states to the WL states, move in the WL and finally are recaptured into other QDs. During this process carriers or excitons in QDs with higher transition energy are effectively promoted to the WL states, whereas the population of lower energy QDs tends to increase, due to retrapping. The experimental temperature dependence of the PL decay times is well explained quantitatively in terms of a rate equation model which considers the effects of both carrier emission and retrapping [144].

Recently, a steady-state thermal model for the temperature dependence of the PL spectra in InAs/GaAs QDs has been introduced which takes into account all the most important mechanisms invoked for explaining the peculiar evolution of the InAs QD PL band with temperature, namely, carrier thermal escape and retrapping, QD inhomogeneous broadening, random population of the QD ground states, lack of direct coupling between different QDs, saturation effects, and increasing carrier thermalization for increasing temperature [145]. The model is based on a system of rate equations connecting the main components of the system: the QD states, the wetting layer states, and the GaAs barrier states.

The model accurately fits the integrated intensity dependence on temperature, well reproduces the redshift of the peak position, and reasonably describes the full width at half maximum (FWHM) temperature dependence, as shown in Figure 17. The substantial agreement between model simulations and experimental data suggests a description of the carrier quenching process in QDs in terms of thermal excitation of correlated e–h pairs from the QD to the WL state, which acts as a transit channel to other QDs or to quenching states inside and/or outside the GaAs barrier. The non monotonic dependence on T of the linewidth and peak energy is accounted for in terms of an increasing carrier thermalization within the QD disordered system and in the WL.

A general conclusion that can be drawn from the T-dependence of the QD PL is that the QD interconnection via the WL has a strong influence on the InAs/GaAs QD optical properties [139, 140]. In particular, following the model presented in [145], the carrier escape toward the WL is mainly responsible for the thermal quenching of the InAs/GaAs QD emission. The possibility to control the electronic levels of the WL is therefore of the utmost relevance for the optimization of QD structures toward device applications at room temperature.

This objective has be pursued by different approaches. In particular, it has been recently demonstrated experimentally by means of PL measurements of self-assembled $In_{0.5}Ga_{0.5}As/Al_xGa_{1-x}As$ QD structures grown by atomic layer MBE that the energy separation between the fundamental QD state and the WL electronic levels can be tuned by increasing the barrier bandgap [146]. The energy separation between 0D and 2D levels strongly increases with the Al content, due to the larger penetration into the barrier of the wavefunctions of the WL states with respect to the QD states, and thus the confinement effect related to the increase of the barrier potential is larger for the 2D-WL levels. Furthermore, the use of ternary alloys for both the QD

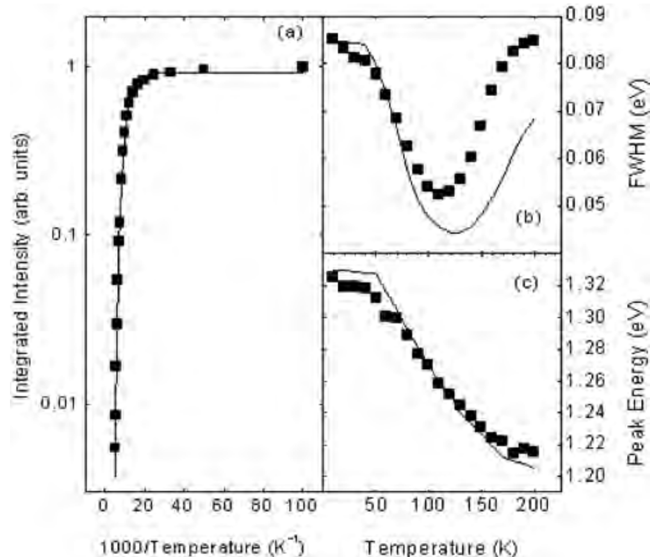


Figure 17. Integrated intensity (a), FWHM (b) and peak position (c) of the PL band of InAs/GaAs QDs vs temperature. The full lines are calculated following the model discussed in [145].

layers and barrier layers allows one to separately tune the QD and the WL emission energies [146]. An alternative way to control independently the energy of the WL states and of the QD states has been proposed: InGaAs QDs have been grown on an AlInAs WL whose energy is matched to the GaAs barrier and embedded in a GaAs matrix [147]. In this system the energy of the WL, increased with respect to the conventional InGaAs/GaAs structures, reduces the thermal carrier redistribution processes, as shown by the temperature dependence of the PL spectra. A strong reduction of the PL thermal quenching has been found when increasing the energy separation between QD and WL electronic levels [110–112, 146, 147].

4.1.6. Piezoelectric Effect in InAs Quantum Dots

The zinc blende III–V semiconductors are piezoelectric because of the lack of inversion symmetry. As a consequence a built-in piezoelectric field, due to the shear strain, can be present. In (001) QD structures the shear strain at the QD edges gives rise to a quadrupolar piezoelectric charge distribution which reduces the QD symmetry [35]. As discussed in Section 4.1.4, this has a quite small effect on the ground state energy transition. On the other side, the piezoelectric field forces the electron states to prefer the [110] direction and the holes to be more attracted toward the [$1\bar{1}0$] directions. This reduces the overlap between the electron and hole states, leading to an intrinsic dipole moment [52, 68].

Piezoelectric effects are relevant also in enhancing the interaction with longitudinal optical (LO) phonons. The polar exciton–LO phonon coupling strength, described by the Huang–Rhys factor S, is dependent on the difference between the probability densities of the electron and hole. In the case of self-assembled QDs, a strong difference between the electron and hole wavefunctions is expected. The charge separation between electron and holes, driven by the piezoelectric potential in the QD, provides an intrinsic way of enhancing polar exciton–phonon coupling [148, 149]. This has been demonstrated in [148] where low energy sidebands of the main exciton recombination peak in InAs/GaAs QDs have been observed. These sidebands, attributed to phonon-assisted exciton recombinations, allow the determination of the Huang–Rhys factor S, which is about 5 times larger than in bulk InAs, thus proving the enhancement of the polar-exciton–LO-phonon interaction in QDs. It should be remarked that the determined value of the Huang–Rhys factor ($S = 0.012$) in InAs/GaAs QDs nicely agrees with the theoretically predicted value. Finally, it is reported that the experimental S values in InAs/GaAs QDs exhibit a clear trend as a function of the QD emission energy: samples emitting at high energy are characterized by high S values, whereas in samples emitting at low energy the S values decrease with the emission energy [150].

Much stronger piezoelectric effects are expected on QDs grown on ($N11$) oriented GaAs substrates where a dipolar distribution of the piezoelectric charges is predicted [151]. Whereas piezoelectric effects in InGaAs/GaAs quantum wells grown on ($N11$) substrates have been largely investigated, leading to a well-established understanding of the problem [152, 153], only very recently has this problem been investigated in QDs. The reverse quantum confined stark (QCS) effect has been evidenced in a series of ($N11$) InAs/GaAs QDs by means of CW PL measurements [154], the direct QCS effect has been investigated by means of photocurrent spectroscopy in a (311)B InGaAs/GaAs QD [155], and finally, the dynamic QCS effect has been analyzed by means of time-resolved PL [156].

The first evidence of piezoelectric effects in ($N11$) QDs was reported in [154]. The PL band of QDs grown on ($N11$) oriented substrates ($N = 2, 3$, and 5) with the two A and B terminations shifts toward the blue for increasing the excitation power density [154]. The blueshift (Fig. 18) shows a clear dependence on $1/N$ with a marked asymmetry about the (100) orientation, being larger on the A faces than on the B faces. The PL integrated intensity of these QDs measured at $T = 2$ K and low excitation power density shows a reduction for large values of $1/N$. Also in this case an asymmetry is found between A and B terminations: the PL efficiency of ($N11$)A samples drops by more than two orders of magnitude when the coverage and thus the QD size is increased, while on the ($N11$)B side the PL intensity reduction is definitely less pronounced.

This phenomenology can be interpreted in the light of the strain induced piezoelectric field, whose presence is expected in self-assembled QDs grown on ($N11$) GaAs substrates. The injection of photogenerated carriers produces a reduction of the internal field and then the QD PL band

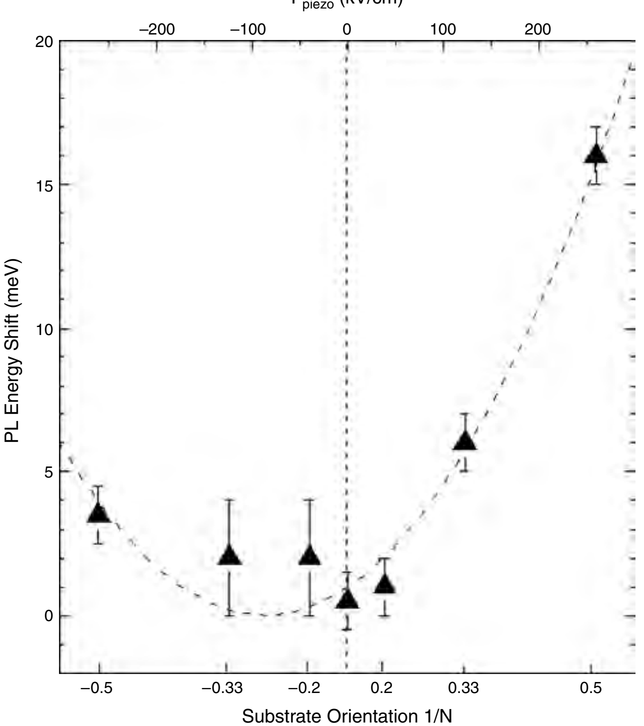


Figure 18. Maximum PL blueshift induced by the increase in the exciting intensity as a function of the substrate orientation. The markers indicate different InAs coverages: 1.4 (squares), 1.8 (triangles), 2.3 (stars), and 3.8 MLs (circles). Reprinted with permission from [151], S. Sanguinetti et al., *Microelectron. J.* 33, 583 (2002). Elsevier Scicence.

shifts toward the blue (photoinduced reverse QCS shift) for increasing the excitation power density. The observed dependence of the PL blueshift on the substrate orientation agrees with the expected QCS shift due to the built-in piezoelectric fields in InAs (N11) QDs. Finally, the asymmetry with respect to A or B termination can be explained by assuming that in the (N11) QDs a permanent dipole similar to the (100) case is present [157]. Confirmation of this picture has been found by time-resolved investigation (Section 4.2.6).

4.1.7. Electronic Coupling Effects in Multilayer Quantum Dots

A key aspect in the technological applications of self-assembled QDs is connected to the achievement of both high spatial density and good size uniformity of the nanostructures. Actually, the state-of-the-art of QD lasers is represented by multilayer structures, where layers of self-assembled QDs are grown in close sequence. As discussed in Section 2.1.1, multilayer depositions of InAs separated by thin GaAs spacers result in the formation of stacked, vertically aligned, InAs QD columns.

Spectroscopic evidence of the interdot coupling is shown in Figure 19 where PL peak energies and Full width at half maximum (FWHM) as a function of the number of vertically stacked InAs/GaAs QDs layers are reported [40]. The band of the single layer sample is broad, reflecting a large QD size variation. Increasing the number of QD layers, the peak energy of the PL band shifts to lower energies and the FWHM decreases. The reduction of the PL band broadening is clear evidence of the reduction of the spread in QD size distribution. The stacking of QD layers produces a relevant ordering in the QD size due to the fact that the QD nucleation is driven by the strain of the buried layer. Thus the spontaneous vertical ordering of QDs in multilayer structures is very important also for controlling the QD size distribution. Nevertheless, more evidence is reported in

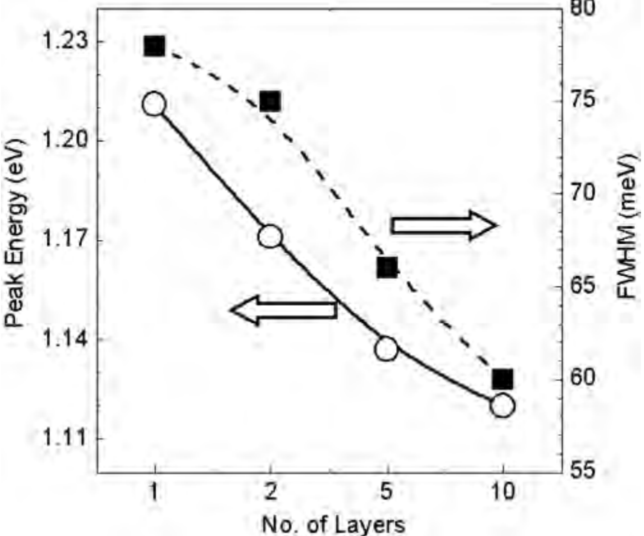


Figure 19. Low temperature photoluminescence peak position (circles) and FWHM (squares) as a function of the number of vertically stacked QD layers. The lines are guides for the eye. Data taken from [40].

[158] leading to the conclusion that in the InAs/GaAs system the stacking of QD layers determines a deterioration of the GaAs barrier. In fact, increasing the number of layers and decreasing the interlayer spacing induce a decrease of the carrier transfer efficiency from the GaAs to the QDs. Efficient nonradiative quenching channels of the QD emission have also been observed. This deterioration of the GaAs interdot layers should be connected with the strain accumulated in the structure [158].

The spectral shift of the QD PL shown in Figure 19 has been interpreted in terms of carrier delocalization induced by the vertical electronic coupling of QDs in columns. This reduces the ground state confinement energy of each dot column since the carrier wavefunction is spread over many QD sites. The coupling between adjacent QDs leads to the splitting of the levels [85, 86] and eventually minibands can open in a QD superlattice [87]. The interplay of Coulomb correction and carrier delocalization complicates the analysis of the optical transition of stacked QDs. In [159] it has been shown that the different delocalization of electrons and holes can also produce a blueshift of the PL from vertically aligned QDs due to the reduction of the e–h Coulomb interaction. A more detailed analysis with analogous results is reported in [160], where the low temperature PL of vertically aligned InAs/GaAs QD structures with 10 layers and spacers varying between 7.7 and 12.5 nm at steps of 0.2 nm is measured.

4.1.8. Photoluminescence of Single Quantum Dots

The main part of the experimental results published on PL of QDs refers to large ensembles of nanostructures. Indeed the large areal dot density, together with the relatively low spatial resolution of the conventional optical techniques, usually makes a huge number of QDs contribute to the optical spectrum. Of consequence, large inhomogeneously broadened bands characterize, even at low temperature, the PL spectra. The bands usually display a Gaussian shape, thus showing that the broadening is due to statistical fluctuations of the QD size around a mean value determined by the growth conditions (coverage, growth temperature, etc.). The best results refer to PL bands whose FWHM is of the order of few tens of meV. We could mention a FWHM of about 15 meV obtained in a InAs/GaAs heterostructure grown on a (511)A oriented substrate [119] and FWHM of 22 meV in InGaAs/GaAs DE grown structures [49].

Such a large width of the PL bands limits in a significant way the potentialities of the QD applications and, furthermore, partially masks the intrinsic properties of QDs, sometimes making their study complex. An example where the discrete nature of the transitions contributing to the large inhomogeneous band is evidenced is reported in Figure 20. The measurements have been performed on MBE grown self aggregated $In_{0.5}Ga_{0.5}As/GaAs$ structure using a conventional, broad area, PL or a confocal micro-PL apparatus capable of micrometer spatial resolution. The broad area PL spectrum displays the typical Gaussian shape, with a FWHM of about 50 meV; reducing the diameter of the spot via micro-PL measurements, the spectrum is resolved in a series of narrow lines (Fig. 20), whose FWHM is about 40–60 μeV [100].

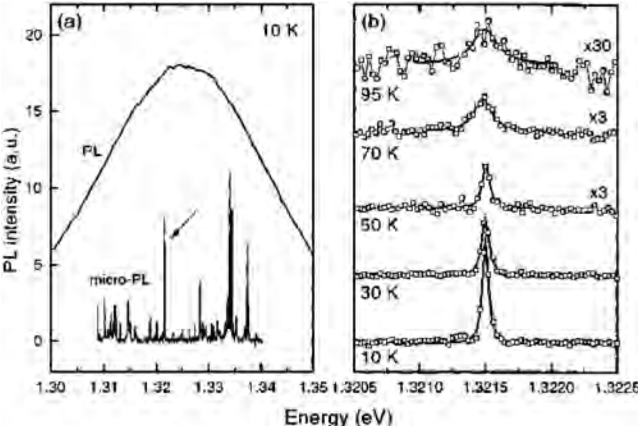


Figure 20. (a) Photoluminescence and microphotoluminescence spectra of InGaAs/GaAs QDs. (b) A single line [marked with an arrow in (a)] measured at different temperatures. The spectra have been vertically displaced for clarity and horizontally shifted for $T > 10$ K to compensate the change in bandgap. Reprinted with permission from [100], K. Leosson et al., *Phys. Status Solidi B* 221, 49 (2000). © 2000, Wiley Science.

These interesting results have stimulated an intense activity aimed at developing experimental techniques for the measurement of the optical properties of a small number of QDs or of a single QD. In particular, the development of near-field spectroscopy (see, e.g. [102]) has allowed the spectroscopic study of single QDs.

A typical example of low temperature PL spectra of a single self-assembled GaAs/$Al_{0.3}Ga_{0.7}As$ QD for different excitation intensities is shown in Figure 21. At the lowest excitation intensity, a single narrow spectral line, whose linewidth is limited by the spectral resolution of the apparatus (0.7 meV), is observed. Increasing the excitation intensity, weak lines appear some meV below the narrow line. Higher energy lines may appear some tens of meV above the first line [81] and can be attributed to the excited states of the QDs. The low energy lines provide evidence of the existence of multiexciton complexes. In [81, 82] these attributions are supported also by the time evolution of the PL spectra and by the results of a theoretical multiexciton model. A detailed analysis of the multiexciton states of single self-assembled QDs is presented in [161, 162], where emission spectra obtained at different excitation intensities are compared with a theoretical many-body model.

The linewidth of the fundamental interband transition of single InGaAs QDs has recently been measured as a function of the temperature using a high spectral resolution apparatus obtained inserting a Michelson interferometer in the detection path of a microphotoluminescence system [101]. The very high spectral resolution obtained in this way (0.2 μeV) showed that the FWHM strongly varies from QD to QD, due to extrinsic effects connected to the electrostatic environment of QDs. The data also show that the interaction with acoustic phonons dominates at low temperature, producing a linear variation of the FWHM for $T \leq 40$ K. At higher temperature an exponential variation is observed, attributed to the interaction with optical phonons.

In [101] the temperature dependence of the homogeneous linewidth of the excited states of single InAs/GaAs QDs is also studied, evidencing an efficient acoustic phonon broadening.

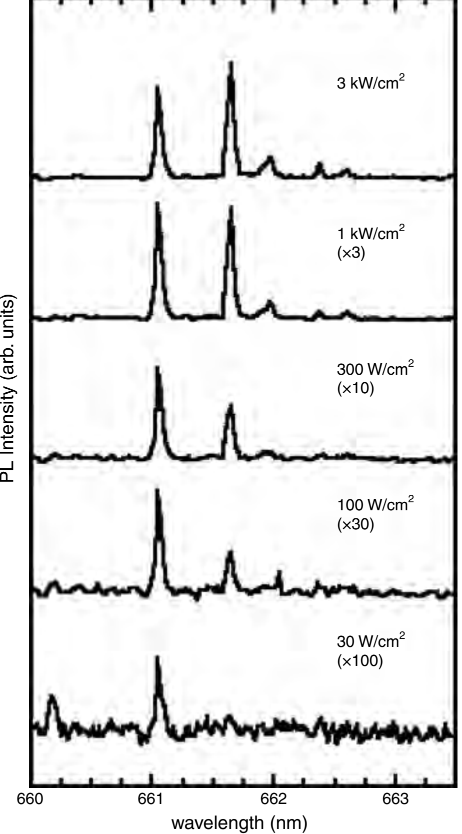


Figure 21. PL spectra of single self assembled GaAs/AlGaAs QD measured at $T = 15$ K for different excitation powers. Courtesy of T. Kuroda.

Interesting effects on single QD PL, the so-called "random telegraph noise" [163] and "spectral diffusion" [164], have been recently reported. These are assumed to be extrinsic effects, usually attributed to defects surrounding the QD and to their state of charge.

The "random telegraph noise," or blinking, consists of a QD switch among two or three levels, with a corresponding change in the PL intensity. The switching rate has a typical scale of seconds or even minutes and increases, increasing the temperature or the excitation intensity. Apparently the emission energy does not change appreciably between the on and the off state [165]. For the explanation of these data a model has been proposed involving a photoactivated defect, having two fairly long–lived configurations. In one configuration the defect captures carriers from the QD, which of consequence has a low emission efficiency. In the other configuration the carrier capture is less efficient and the QD is characterized by a high PL efficiency [163].

The "spectral diffusion" manifests itself as an inhomogeneous broadening of the PL of single QDs at low temperature, determined by the electric fields due to fluctuations in the charge distributions surrounding the dot, with an unexpected temperature behavior [164]. In InP/GaInP QDs the single QD band at low temperature is characterized by a FWHM of some meV; increasing the temperature the band

abruptly changes, switching into several narrow lines and then, at even higher temperatures, to a single sharp peak of very narrow width. The interpretation of these data, as discussed in [164], is as follows. The sharp line in the high temperature spectra is due to the unperturbed fundamental transition of the OD and the broadening of this emission at low temperature is attributed to the interaction of the QD electron and hole states with charges trapped or released by the sites surrounding the QD, leading to fluctuations in the charge configuration. The mode of QD emission depends on the position of the Fermi level determining the charge of these sites. The presence of spectral diffusion effects may not allow one to measure the QD homogeneous broadening using single QD PL. This problem can be overcome by using four-wave mixing techniques [99].

4.1.9. External Applied Fields

The spectroscopic techniques permitting the study of single QDs have opened great opportunities for the analysis of the effect of external fields on the electronic states of zero-dimensional structures.

A recent paper on $Al_xIn_{1-x}As/Al_yGa_{1-y}As$ self-assembled QDs gives some insight on the effect of an electric field on the electronic properties of these systems [166]. Microphotoluminescence measurements have been performed on single QDs in the presence of an external field oriented along the growth direction (the field was considered positive when the field lines were oriented from top to substrate); the field, both positive and negative, has been varied in a wide interval. The most interesting result is an asymmetric Stark shift of the emission energy as a function of the applied field. In general, the single QD lines blueshift for increasing positive fields and this shift reaches a maximum; on the contrary, the emission is redshifted for negative applied fields. The results of theoretical calculations are quantitatively consistent with the experimental findings; in particular, the theory shows that the blueshift is mainly due to a strong increase in the electron confinement energy because the field displaces the electron toward the apex of the QD. Another interesting effect of the application of the electric field is that lines appear and disappear in the spectrum as the field is tuned. The effect is probably related to the modification of the emission selection rules [166].

The spectroscopy of single QDs in magnetic fields up to 8 T is discussed in [167], where a magneto-optical study of excitons, biexcitons, and charged exciton complexes in $Al_{0.36}In_{0.64}As/Al_{0.33}Ga_{0.67}As$ is presented. In particular, as the magnetic field is increased, the single exciton splits into σ^+ and σ^- polarized states, the splitting between the two states increasing linearly with the magnetic field. A slight diamagnetic shift of the order of 0.1 meV in the range 0 to 8 T is observed for the single exciton. Furthermore, the splitting of the biexciton is identical to that of the single exciton, despite the fact that the biexciton energy is not expected to split in a magnetic field, due to the singlet nature of the biexciton. In fact the final state of the biexciton transition is an exciton and thus the splitting of the biexciton emission is given by the Zeeman splitting of the exciton [167].

The use of magnetic field allows study of the presence of dark excitons in QDs. The concept of dark exciton refers to fine electronic structure of the electron–hole complex. Generally speaking, the e–h complex consists of a multiplet of states with different total angular momenta. The states are split (usually by few hundreds of μeV) due to the exchange interaction. Only the states with angular momentum equal to 1 can recombine radiatively; the other states are not coupled to the light field and are denominated dark exciton states. The presence of dark states has a strong influence on the recombination kinetics of bright excitons. In fact, dark states, due to their long lifetime, act as a reservoir for the bright exciton recombination.

A determination of the dark exciton states is quite difficult, first because a direct access by optical spectroscopy is allowed only if a symmetry breaking relaxes the conservation of the angular momentum. This occurs, for example, by using a magnetic field, which produces a mixture of dark and bright exciton states. A further complication in the study of dark excitons is related to the very small splitting between states of different angular momentum, which, in semiconductor structures, is usually much smaller than the inhomogeneous broadening. However, this last limitation is overcome in single QD spectroscopy. A spectroscopic study of dark exciton states in InGaAs single QDs under magnetic field has been reported in [168]. By varying the strengths and the orientation of the external magnetic field, the authors were able to resolve the full detail of the fine structure of the QD excitons, determining both the electron and hole g factors and the electron–hole exchange splitting.

4.2. Time-Resolved Photoluminescence

Ultrashort laser pulses below 100 fs are nowadays routinely generated and used for providing a sort of instantaneous optical excitation of semiconductor systems. The availability of laser sources with a large range of wavelengths, repetition rate, pulse duration, and energy has made it possible to explore a large variety of physical phenomena by means of time-resolved spectroscopy. After the excitation by an ultrashort pulse of a semiconductor in thermodynamic equilibrium, this undergoes several stages of relaxation before it returns to the initial condition. The investigation of these relaxation processes allows one to get information on the intrinsic time constant ruling the carrier dynamics. Time-resolved spectroscopic techniques with the appropriate time resolution are then largely exploited in order to follow, in real time, the relaxation paths of the photoexcited carriers. For a general review we refer to [169].

4.2.1. Experimental

Ultrafast Lasers Ultrashort laser technology is based on the mode locking method, in which the longitudinal modes of the laser cavity are locked in phase, in active systems with large gain bandwidth. Synchronously pumped organic dye lasers with a mode-locked Nd:YAG or Ar^+ ion laser, invented at the end of the 1970s, were the first efficient tool for ultrashort laser pulse generation. Enormous progress in tunability, stability, and quality of the laser pulses has been achieved with the development of solid state lasers and in particular with Ti:sapphire lasers. Generation of pulses shorter than 10 fs has been realized directly from the laser

oscillator without the need for laser amplification and compression. Optical parametric oscillators or amplifiers have recently increased the tunability range of ultrafast lasers.

PL Detection Several detection techniques have been so far developed for time-resolved PL spectroscopies. The most widely used method makes use of a streak camera. The PL photons, focused on a photocathode, extract photoelectrons that are deflected by time modulated electrostatic fields to be streaked, after multichannel plate amplification, across a phosphorescent screen. Therefore photoelectrons emitted at different times impinge onto the screen at different positions and the information on the time is transferred into a spatial distance. Picosecond resolution can be achieved in state-of-the art streak camera apparatuses. Higher time resolution, limited only by the pulse duration, can be achieved by nonlinear techniques like the parametric up-conversion. In this case the PL signal is combined in a nonlinear crystal, such as KTP or $LiIO_3$, with part of the laser pulse suitably delayed. Frequency sum generation can occur only during the laser pulse gate and its intensity as a function of the delay of the laser pulse directly gives the PL time evolution.

4.2.2. General Aspects

The carrier relaxation in semiconductors can be classified into three main temporally overlapping regimes.

Coherent Regime The excitation in the semiconductor created by a ultrashort laser pulse has a well defined phase relationship within the excitation and with the electromagnetic field of the laser pulse. In this situation, the system optical response is driven by the interaction of the field with the coherent macroscopic polarization induced by the laser pulse. This coherent regime shows several interesting phenomena, which are strictly related to the basic quantum mechanics in semiconductors. The time scale on which the coherent interaction relaxes is usually very short in semiconductor structures due to the scattering processes leading to a coherent loss. Typically, in semiconductor systems the dephasing processes are very fast (few picoseconds or even less). On the contrary, long dephasing times (in the nanosecond timescale) have been recently observed in semiconductor QDs.

Cooling Regime In this regime the photoexcited carriers lose their excess of energy, mainly by emission of phonons. In the first stages (few picoseconds) of this regime the distribution of the photoexcited carriers is very likely nonthermal (i.e., cannot be described by a temperature) Carrier–carrier scattering processes are very often responsible for the carrier redistribution within the accessible states leading to a thermalized distribution of the carriers. The related temperature can be higher than the lattice temperature for a long time scale (few tens of picoseconds). This stage is indeed denominated as a hot-carrier regime. In this regime the carriers lose their excess of energy mainly by emission of phonons. At the end of the intraband relaxation processes the photoexcited carriers eventually reach a thermalized distribution in equilibrium with the lattice vibrations. In the case of semiconductor nanostructures within the cooling regime the thermalization processes include also the carrier capture into the nanostructures because they represent the lowest energy states of the system.

Recombination Regime The carriers and the phonons are in thermal equilibrium with each other. However, there is an excess of electrons and holes with respect to the thermal distribution due to the laser pulse excitation. The photoinjected carrier population decays by either radiative or nonradiative recombination with a typical time constant of 100 ps–1 ns, in direct semiconductor nanostructures.

It should be stressed that often the relaxation processes in the different regimes occur simultaneously. For example, the electron–hole pair recombination can also occur during the cooling regime leading to the so-called hot PL. Nevertheless the description in terms of different relaxation regimes provides a useful framework for discussing the carrier dynamics in semiconductors.

4.2.3. Carrier Dynamics in Quantum Dots

Several experimental techniques have been so far developed for investigating different aspects of the carrier relaxation. Here we will focus our attention on the time-resolved PL originated by the thermalized distribution of electron–hole pairs (or excitons) in semiconductor QDs. Since the electronic levels in QDs are separated by tens of meV, low temperature PL measurements mainly probe the fundamental optical transition, unless state filling effects take place under large excitation power density. Generally speaking, the PL time evolution after short pulse excitation can be characterized by two time constants, that is, the rise time and the decay time. The rise time describes the delay after the excitation that occurs for the carriers to reach the levels from where the emission takes place and it contains information on the carrier capture and energy relaxation processes. The decay time is a direct measurement of the electron–hole pair recombination time by either radiative or nonradiative mechanisms.

It should be also noted that the carrier capture and recombination in QDs are random processes which are not adequately described by standard rate equation models. A theory of random population has indeed been developed for a correct description of the carrier dynamics in QDs [76]. Monte Carlo models have been developed for taking into account the random nature of the carrier capture in the QD recombination kinetics [170].

4.2.4. Carrier Relaxation and Capture in Quantum Dots

The nature of the mechanisms underlying the carrier relaxation in QD materials has been largely debated in the last years [171–175] and it is not yet completely understood. The relaxation processes in QDs attract much attention since they involve fundamental physical aspects of the zero-dimensional semiconductor systems, and also due to their relevance for device applications. In QD lasers the carriers are injected into the barriers embedding the QDs. After energy dissipation processes, the carriers are captured by the QDs and then relax to the fundamental lasing state. The efficiency of the relaxation cascade directly affects the device performances, such as threshold current, temperature

stability, and so on. The achievement of very fast capture rate is therefore a relevant aspect of the device optimization study. At the same time the interband phonon scattering in quantum confined electron gases is expected to show strong dimensionality effects [176]. Bockelmann and Bastard proposed a model calculating the theoretical emission rates for longitudinal acoustic (LA) phonons as a function of the lateral dimension L of the nanostructure in 0D, 1D, and 2D systems, corresponding to QDs, QWRs, and QDs, respectively. The model is based on weak phonon-coupling assumptions and Fermi's golden rule was used to estimate the carrier relaxation rate. In this model the initial electron state energy E_i was chosen in order to have the same energy difference from the fundamental level for the QD, QWR, and QW as illustrated in Figure 22. The initial electron states are the first excited level for the QD and the edge of the first excited subband for the QWR and the QW. For lateral sizes exceeding 200 nm the three scattering rates decrease and become very close, since the differences vanish when the confinement size becomes very large. In the presence of quantum confinement effects ($L < 100$ nm), a clear reduction of the LA phonon scattering efficiency is instead predicted when decreasing the dimensionality of the electron gas. In particular the LA phonon scattering rate in QDs is extremely small and exhibits strong oscillations. The oscillations are manifestations of the discrete atomiclike energy spectrum of the QD. Even more dramatic are the effects of the QD constraints on the allowed inelastic carrier scattering via LO phonons. In particular, single LO phonon emission is forbidden unless the energy separation between the QD levels exactly matches the LO phonon energy. The predicted inefficient carrier relaxation via phonon emission in QDs is usually referred to as phonon bottleneck.

From the experimental side, the characteristic carrier relaxation time from the barrier states to the fundamental QD level is found to be very fast, that is, of the order of few tens of picoseconds, and becomes even shorter for large injection of carriers. Several mechanisms have been invoked to explain the lack of phonon bottleneck. In the low injection regime, the fast relaxation time (30–70 ps) has been interpreted as a consequence of the presence of a continuum tail of defect states to which the carriers easily relax the excess energy [174]. Alternatively, resonant multiphonon processes can mediate the relaxation between the localized states [177]. The formation of vibronic (polaron) states, which might provide wide energy windows for efficient relaxation, has been also proposed [178]. In summary, a clear understanding of the origin of the fast carrier capture and relaxation in QDs has not yet been reached. On the contrary, there is evidence of the major role played by Auger-like processes in the carrier relaxation under large injection. This is shown in Figure 23 where the experimental PL rise time (equivalent to the relaxation time) is reported as a function of the excitation density and wavelength [179]. Increasing the carrier injection, the carrier–carrier interaction speeds up the energy relaxation rate. This effect has been assigned to Auger–like processes where an electron (or a hole) occupying an excited level of the dot relaxes to a lower state giving the excess of energy to another electron (or hole) [179, 180]. In particular, the Auger processes involving carriers localized in the 2D WL states have been claimed to increase the carrier relaxation rate inside the QD [181, 182]. Nevertheless recent experimental data unambiguously show that very similar phenomenology occurs also in GaAs/AlGaAs QD grown without WL, pointing out the relevance of internal Auger mechanisms in the QD carrier relaxation [183].

4.2.5. Spin Relaxation in Quantum Dots

The discrete nature of the electronic DOS of semiconductor QDs modifies the carrier spin relaxation dynamics with respect to the bulk or to quantum well structures [184]. The lack of available energy states is expected to inhibit both elastic spin flip mechanisms and inelastic phonon scattering. The possibility of very long spin relaxation time has been exploited for proposing the QDs as candidates for the implementation of spintronics [185].

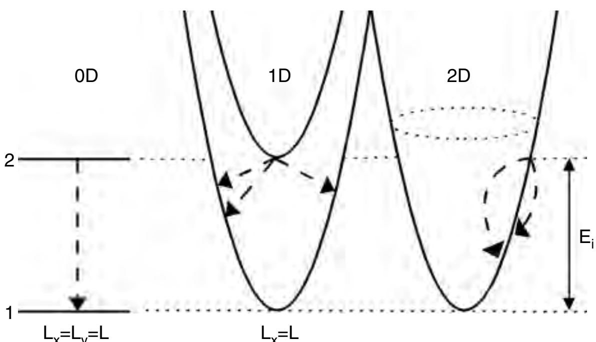


Figure 22. Schematic dispersion curves in the case of 0D, 1D, and 2D confinement.

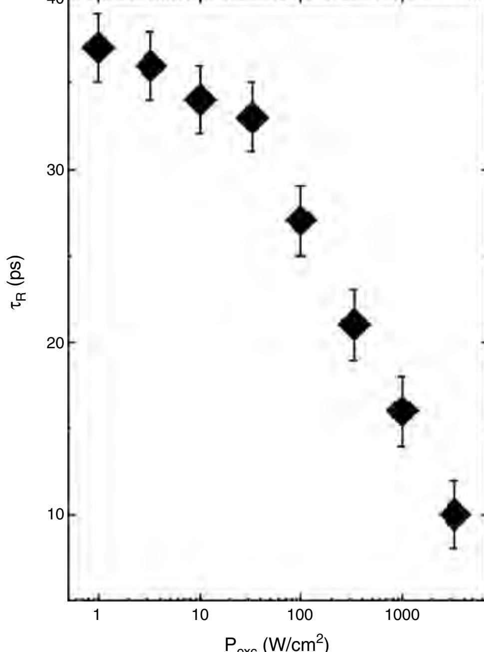


Figure 23. Dependence of the QD ground state PL rise time as a function of the laser excitation density in GaAs/AlGaAs QDs. After [183].

These predictions have stimulated large interest on the spin relaxation in QDs. In the study of natural QDs formed by interface fluctuations in GaAs QWs, it has been shown, by means of CW-PL experiment, that the rate of spin flipping was lower than the radiative recombination rate [130]. The same conclusions were obtained in [186] in the case of self-aggregated InAs QDs. On the contrary, longer spin relaxation times, almost twice the recombination time, have been observed in InGaAs quantum discs [187]. It should be noted that in all these studies the spin dynamics has been addressed via nonresonant excitation of the QD structures. This means that the carriers were photoinjected into the barriers. Therefore the observed polarization dynamics was the result of all the spin relaxation processes occurring during the relaxation path.

Recently the study of the QD spin relaxation has been performed by using strictly resonance conditions [188]. The laser excitation energy was exactly coincident with the detection energy. The use of up-conversion methods for detecting the QD-PL allowed not only a very good time resolution but also the elimination of the strong initial contribution due to Rayleigh scattering. The QD emission was analyzed both in the linear and in the circular polarization channel. The result was that spin relaxation was completely frozen on the carrier relaxation time scale. The discrepancy with the previous study was assigned to the relevant role of resonant injection in reducing carrier occupation of higher energy states which, via Coulomb exchange mechanisms, may produce a spin flip of the QD ground state [188].

4.2.6. Carrier Recombination in Quantum Dots

The measurement of the QD-PL lifetime τ_{PL} directly gives the recombination time of the electron–hole pairs inside the QD. Obviously τ_{PL} depends both on the radiative recombination time τ_{rad} and on the nonradiative recombination time τ_{nrad} via the well known relation $\tau_{\mathrm{PL}} = (1/\tau_{\mathrm{rad}} + 1/\tau_{\mathrm{nrad}})^{-1}$. It is commonly assumed that self-aggregated QDs are almost defect-free and that the parasitic nonradiative recombination channels do not play a major role in the low temperature PL emission. Therefore the experimental τ_{PL} is used as a direct measurement of the intrinsic QD radiative lifetime [189, 190], even if the presence of the nonradiative recombination affecting preferentially the small QDs has been reported in the case of high QD density [177]. The radiative decay rate $1/\tau_{\mathrm{rad}}$ can be evaluated by means of Fermi's golden rule in the dipole approximation [191]. In the limit of strong confinement conditions, where the carrier quantization energies are much larger than the exciton binding energy, $1/\tau_{\mathrm{rad}}$ is simply proportional to the atomiclike lifetime [192],

$$\frac{1}{\tau_{\mathrm{rad}}} = \frac{e^2 E_p \omega n}{2 m_0 \epsilon_0 h c^3} \phi^2 \tag{11}$$

where $\hbar\omega$ is the QD transition energy, m_0 and e are the electron vacuum mass and charge, E_p is the Kane energy, c is the speed of light in vacuum, and n is the index of refraction. The factor ϕ^2 is the overlap of the electron and hole wavefunctions. This expression predicts two relevant features. The first is the increase of the radiative recombination rate with the QD transition energy. Smaller QDs are expected to have shorter radiative lifetime. The second feature is that the radiative lifetime is proportional to the overlap of the electron–hole wavefunctions. The predicted dependence of the QD-PL lifetime on the QD size has been recently reported in a set of annealed InAs/GaAs QDs [190]. The thermally activated interdiffusion effect produces a large blueshift of the fundamental QD optical transition. Therefore the annealing induces a reduction of the effective QD size. As a consequence, the ground state radiative lifetimes are found to vary from 800 to 490 ps as the QD size is decreased. This variation is larger than the reduction predicted by the variation of the optical transition energy. The difference has been attributed to the increase of the overlap of the electron–hole wavefunctions [190].

The dependence of the radiative lifetime on the overlap of the electron–hole wavefunctions has been used to monitor the relative separation of the photogenerated carriers. Let us discuss two examples.

The first is related to the QDs grown on high index Miller substrates. As discussed in Section 4.1.6, the presence of piezoelectric fields in these QDs produces a spatial separation of the electron and hole wavefunctions which is expected to produce a reduction of the radiative recombination probability. The investigation of the carrier dynamics in InAs/GaAs (N11) QD performed in [156] by time-resolved (TR) measurements confirms this picture. The study of the time evolution of different spectral components of the PL band of (211)A QDs for increasing emission wavelengths evidences a progressive lengthening of the PL decay time. Of consequence, the PL emission peaks shift toward lower energy increasing the delay time after the exciting pulse, with minor variations of PL shape and linewidth (Fig. 24). In the long time limit the QD PL spectrum tends to recover the lineshape of the CW PL spectrum measured at low power density. Both the dynamic redshift of the PL band and the wavelength dependence of the PL decay times can be interpreted in terms of dynamic screening of the internal electric fields in the QDs. The electrons and holes injected by the optical excitation are spatially separated by the field resulting in a photoinduced screening of the internal field. Thus, in a TR-PL experiment, when increasing the delay time after the pulse excitation, an effective increase of the internal electric field is expected and the dynamic redshifts of the PL band can be attributed to the increase of the quantum confined Stark shift of the QD optical transition. The PL decay time dependence on the wavelength also reflects the same physical mechanism [156]. Note also that a large intrinsic nonlinearity in the carrier recombination kinetics [156] is associated with the described phenomenology. A detailed discussion of the influence of intrinsic internal fields on the recombination kinetics of high coverage (N11) InAs/GaAs quantum dots is reported in [193].

A second example is the possibility to tune the radiative lifetime in multilayer QD structures. The strain fields from the preformed islands induce vertical alignment in closely stacked QD layers and a self-organization of the QDs in vertical column occurs (see Sections 2.1.1 and 4.2). Whenever the barrier thickness is thin enough to allow tunneling, the carrier wavefunction becomes extended over many QD sites

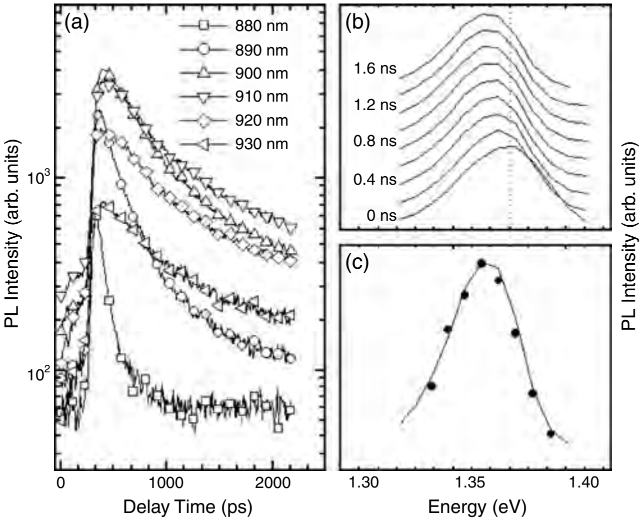


Figure 24. (a) PL intensity decay of the sample (211)A measured at various wavelengths: 880 (squares), 890 (circles), 900 (up-triangles), 910 (down-triangles), 920 (diamonds), and 930 nm (left-triangles). (b) Time-resolved spectra of the (211)A QDs. The step of the time delay between each spectrum is 200 ps. Each spectrum is normalized to its maximum. (c) Comparison between the time-resolved spectra of the (211)A QDs for an excitation power of 200 W cm^{-2} at 1.6 ns delay (corresponding to a reduction of nearly 1/10 of the PL integrated intensity at zero delay—continuous line) and the time-resolved spectrum for an excitation power of 20 W cm^{-2} at 0 ns delay (circles). Reprinted with permission from [151], S. Sanguinetti et al., *Microelectron. J.* 33, 583 (2002). Elsevier Science.

along the column. However, the tunnel probability is larger for electrons than for holes. Therefore there exists a range of barrier thicknesses where the electrons are delocalized over the QD column and the hole are confined within a single QD. This induces a strong reduction of the electron–hole overlap and therefore an increase of the radiative recombination time [91].

An increase of the emission lifetime in stacked QDs increasing the number of layers has been observed [91, 92]. Data from [91] are reported in Figure 25. The emission

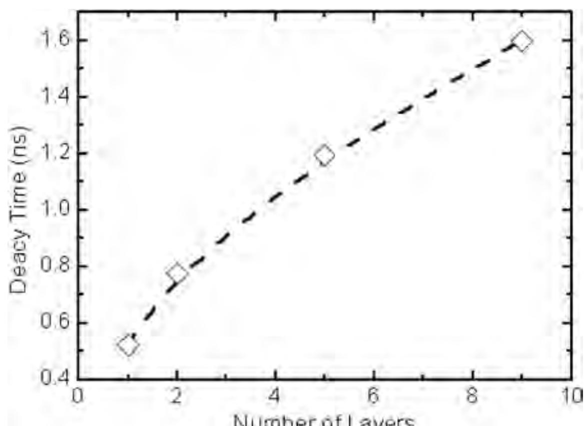


Figure 25. Dependence of the PL decay time at $T = 10$ K on the number of stacked QD layers (the solid line is only a guide for the eye). Courtesy of M. Colocci.

lifetime increases with the interlayer separation [160]. The interpretation relies in the reduction of the oscillator strength of the exciton inside the QD column [91], also due to the different delocalization between the wavefunctions of electrons and holes [91, 94].

Finally the excitation with ultrashort laser pulses has been used for the characterization of the QD excited state structure. [177, 189, 194, 195]. State filling in QDs, which is due to the exclusion principle, is effective when only a few carriers populate the lower states. Therefore the population of the excited states is easily achieved by photoinjection of carriers inside the barriers at excitation power exceeding few hundred W cm^{-2}, also due to the hindering of the interlevel relaxation. QD state filling has lead to the observation of a large number of excited state optical transitions. The measurement of the emission decay time of each transition was used to evaluate the excited state radiative lifetime, which was estimated to be of the same order of magnitude as the ground state lifetime [189].

5. APPLICATIONS

Significant technological improvements in the growth of nanostructures have been realized in the last decade. This opened the route to a new generation of QD devices taking advantage of the unique QD features arising from carrier quantization effects in all the three spatial directions. On one side, QDs have attracted large interest as possible efficient replacements of other quantum heterostructures for standard optoelectronic devices [196]. On the other side, a more general prospect has been recently addressed, aiming at exploiting the specific features of QDs for novel devices in the revolutionary field of quantum communication and quantum computation [197, 198]. Here we will touch on only two topics: the QD laser and QD based quantum cryptography.

5.1. Quantum Dot Lasers

Semiconductor lasers are nowadays key components in several widely used technological products, such as laser printers and compact disc players. At the beginning of the 1980s it was demonstrated that the use of quantum wells as the active layer in semiconductor lasers should result in more efficient devices. Even better performances can be obtained for lasers with quantum dot active layers [12]. This can be explained as follows. Stimulated recombination of electron–hole pairs occurs inside the nanostructure, where the confinement of carriers and also of the optical mode enhances the radiation–matter interaction. In particular, the change in the electronic density of states as a function of the nanostructure dimensionality is relevant, as discussed the Introduction. One relevant figure of merit of an active system is the optical gain spectrum [199]. Figure 26 reports a comparison between the optical gain of bulk and low dimensional semiconductors. The optical gain is enhanced in the case of QDs. It is clear that a QD laser will allow a reduction in threshold currents and an increase in differential gain. Other benefits of quantum dot active layers include less temperature dependence than other semiconductor lasers,

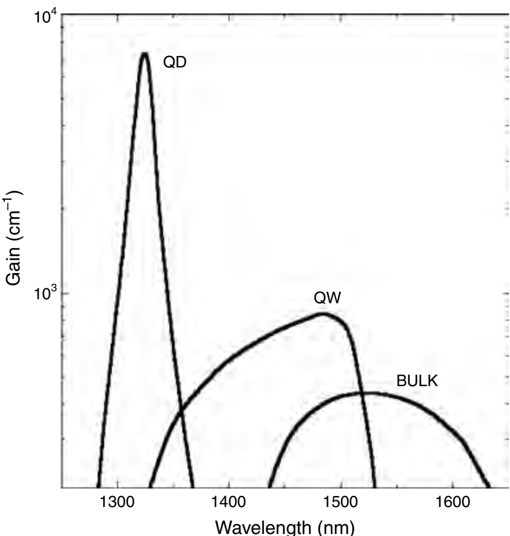


Figure 26. Calculated gain spectra at 300 K for a QD (a cube of 10 nm side), a QW (width of 10 nm), and bulk of $Ga_{0.47}In_{0.53}As/InP$. Data from [199].

reduced degradation at elevated temperatures, and larger tuning range.

However, the advantages in operation depend also on the uniformity of QD size. A large distribution of QD dimensions broadens the density of states, reducing the optical gain. Thus, the challenge in realizing QD lasers with operation superior to other semiconductor lasers was that of forming high density, high quality, and uniform quantum dots in the active layer. The first demonstration of a quantum dot laser with high threshold current density was reported by Kirstaedter et al. in 1994 [200]. QD lasers with improved operation were later achieved by increasing the density of the QD structures, stacking successive, strain-aligned rows of quantum dots and therefore obtaining vertical as well as lateral coupling of the QDs [196]. Nowadays 1300 nm GaAs-based and 1550 nm InP-based QD lasers are commercially available with performances not achievable with QW lasers. Nevertheless a number of critical issues on QD technology, involving the uniformity of size, the controllable achievement of higher quantum dot density, and a better understanding of loss mechanisms are still under study and will very likely improve the device characteristics in the next years.

5.2. Quantum Cryptography

The principles of quantum mechanics appear to offer novel degrees of freedom in the development of technological applications. Assuming to be able to organize and control a complex system governed by the law of quantum mechanics, it will be possible to engineer devices with revolutionary features. Among them quantum computational [197] and quantum cryptography [198] devices have been proposed to be realizable by means of QD technology. While quantum computing is still a dream, the realization of quantum cryptography seems to be very near. Let us briefly discuss the latter. Cryptography is the science of devising code to allow two users of a secret communication channel. The information generally takes the form of a string of bits that can be decoded only by the use of either a public-key encryption or secret-key encryption. Roughly speaking, in both cases the security of the communication channels is nowadays based on the assumed difficulty of factoring of extremely large prime numbers. Quantum information exchange is expected to provide a way of agreeing to a secure key without making this assumption [198]. The basic elements of quantum encryption are the measurements of a quantum state, typically the polarization state of a single photon. The quantum state is decided by the sender and then measured by the recipient. Due to the uncertainty principle, conjugate variables cannot be measured simultaneously. For example the photon polarization can be expressed in the linear or circular bases, but measuring the polarization in one specific basis will destroy the conjugates. Thus if the receiver do not know the basis used by the sender, the observation of the receiver can destroy the sender's information without understanding it.

The key technology for implementing quantum cryptography is the realization of a true single photon source. This cannot be realized simply by attenuating a standard light source since it has been predicted that even a very low rate of two–photon events makes the cryptography completely insecure. Recently the goal of a single photon source has been realized with QDs embedded in micropillar structures [201–203]. The QD layer placed inside a high Q planar microcavity, formed by two distributed Bragg reflectors, was fabricated monolithically using epitaxial growth techniques. Then the sample was patterned by electron beam lithography and etched through the microcavity leading to arrays of microposts with submicrometer diameters. This structure provides simultaneously the presence of very few QDs in the micropillar and fully three-dimensionally confined optical modes. Due to the quantum electrodynamical Purcel effect, the QD emission is enhanced and has preferential output direction [201]. It has been shown that, under these conditions, the QD generates one and only one photon from the fundamental optical transition for every excitation laser pulse [202, 203]. This demonstrates the realization of a quantum-dot single photon turnstile device.

GLOSSARY

Epitaxy The deposition of a crystalline substance, usually in a thin layer, upon the surface of a single crystal with a similar crystalline structure. The process does not involve a chemical reaction between the two materials.

Exciton Electron-hole pair bound by electrostatic interaction free to move in the semiconductor crystal. The exciton has no net electrical charge but it transports energy.

Heterostructure Synthetised crystalline system composed of more than one coherent material.

Phonon Quantum of lattice vibrational energy. The phonon can be viewed as a wave packet with particle-like properties.

Photoluminescence Emission of light as a result of absorption of electromagnetic radiation. The emitted light energy is usually lower that of the absorbed light.

Quantum structures Nanostructures capable of electronic quantum confinement in one (quantum well), two (quantum wire) or three (quantum dot) dimensions.

Self-assembling Generation of complex structures via atom organization driven by thermodynamics and kinetic processes.

Strain Relative deformation of elastic, plastic, and fluid materials under applied forces.

Quantum computer Device that employs properties described by quantum mechanics to enhance computational capabilities.

REFERENCES

1. M. Jannsen, "Introduction to the Theory of the Integer Quantum Hall Effect." VCH, Weinheim, 1994.
2. H. Mizuta and T. Tanoue, "The Physics and Applications of Resonant Tunnelling Diodes." Cambridge, Univ. Press, Cambridge, UK, 1995.
3. G. Bastard, "Wave Mechanics Applied to Semiconductor Heterostructures." Les Editions de Physique, Les Ulis, 1988.
4. K. Mullen, E. Ben-Jacob, R. Jaklevic, and Z. Schuss, *Phys. Rev. B* 37, 9 (1988).
5. D. Goldhaber–Gordon, H. Shtrikman, D. Mahalu, D. Abusch–Magder, U. Meirav, and M. A. Kastner, *Nature* 391, 156 (1998).
6. L. Esaki and R. Tsu, *IBM J. Res. Dev.* 14, 61 (1970).
7. R. Dingle, W. Wiegmann, and C. H. Henry, *Phys. Rev. Lett.* 33, 827 (1974).
8. G. Bastard, E. E. Mendez, L. L. Chang, and L. Esaki, *Phys. Rev. B* 26, 1974 (1982).
9. G. Bastard, E. E. Mendez, L. L. Chang, and L. Esaki, *Phys. Rev. B* 28, 3241 (1983).
10. D. S. Chemla, D. A. B. Miller, P. W. Smith, A. C. Gossard, and W. Wiegmann, *IEEE J. Quantum Electron.* QE–20, 265 (1984).
11. C. Weisbuch and B. Vinter, "Quantum Semiconductor Structures." Academic Press, Boston, 1991.
12. Y. Arakawa and H. Sakaki, *Appl. Phys. Lett.* 40, 939 (1982).
13. H. Qiang, F. H. Pollak, Y. S. Tang, P. D. Wang, and C. M. S. Torres, *Appl. Phys. Lett.* 64, 2830 (1994).
14. A. P. Alivisatos, *Science* 271, 933 (1996).
15. D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, and P. M. Petroff, *Appl. Phys. Lett.* 63, 3203 (1993).
16. N. Koguchi, S. Takahashi, and T. Chikyow, *J. Cryst. Growth* 111, 688 (1991).
17. E. Bauer, *Z. Kristallogr.* 110, 372 (1958).
18. I. V. Markov, "Crystal Growth for Beginners." World Scientific, Singapore, 1995.
19. M. A. Herman and H. Sitter, "Molecular Beam Epitaxy." Springer-Verlag, Berlin, 1996.
20. R. Nötzel, *Semicond. Sci. Technol.* 11, 1365 (1996).
21. D. Bimberg, M. Grundmann, and N. N. Ledentsov, "Quantum Dots Heterostructures." Wiley, New York, 1999.
22. M. Sugawara, "Self-Assembled InGaAs/GaAs Quantum Dots," Vol. 60 of *Semiconductors and Semimetals.* Academic Press, San Diego, 1999.
23. W. Weinberg, C. M. Reaves, B. Nosho, R. Pelzel, and S. P. DenBaars, in "Nanostructured Materials and Nanotechnology" (H. S. Nalwa, Ed.), Ch. 6, p. 207. Academic Press, San Diego, 2000.
24. J. F. Carlin, F. Houdrá, A. Rudra, and W. Llegems, *Appl. Phys. Lett.* 59, 3018 (1991).
25. J. Ahopelto, H. Lipsanen, M. Sopanen, T. Koljonen, and H. E. M. Niemi, *Appl. Phys. Lett.* 65, 1662 (1994).
26. A. Kurtenbach, K. Eberl, and T. Shitara, *Appl. Phys. Lett.* 66, 361 (1995).
27. B. R. Bennett, R. Magno, and B. V. Shanabrook, *Appl. Phys. Lett.* 68, 605 (1996).
28. F. Hatami, N. N. Ledentsov, M. Grundmann, J. Böhrer, F. Heinrichsdorff, M. Beer, D. Bimberg, S. S. Ruvimov, P. Werner, U. Gosele et al., *Appl. Phys. Lett.* 67, 656 (1995).
29. L. Goldstein, F. Glas, J. Y. Marzin, M. N. Charasse, and G. LeRoux, *Appl. Phys. Lett.* 47, 1099 (1985).
30. S. Guha, A. Madhukar, and K. Rajkumar, *Appl. Phys. Lett.* 57, 2110 (1990).
31. C. Snyder, B. G. Orr, D. Kessler, and L. M. Sander, *Phys. Rev. Lett.* 66, 3032 (1991).
32. D. Leonard, K. Pond, and P. M. Petroff, *Phys. Rev. B* 50, 11 687 (1994).
33. I. Daruka and A. L. Barabasi, *Phys. Rev. Lett.* 79, 3708 (1997).
34. R. Leon and S. Fafard, *Phys. Rev. B* 58, R1726 (1998).
35. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 52, 11969 (1995).
36. K. Georgsson, N. Carlsson, L. Samuelson, W. Seifert, and L. R. Wallenberg, *Appl. Phys. Lett.* 67, 2981 (1995).
37. L. G. Wang, P. Kratzer, N. Moll, and M. Scheffler, *Phys. Rev. B* 62, 1897 (2000).
38. J. M. Moison, F. Houzay, F. Barthe, L. Leprince, E. André, and O. Vatel, *Appl. Phys. Lett.* 64, 196 (1994).
39. J. Tersoff, C. Teichert, and M. G. Lagally, *Phys. Rev. Lett.* 76, 1675 (1996).
40. G. S. Solomon, J. A. Trezza, A. F. Marshall, and J. S. Harris, *Phys. Rev. Lett.* 76, 952 (1996).
41. N. N. Ledentsov, V. A. Shchukin, M. Grundmann, N. Kirstaedter, J. Böhrer, O. Schmidt, D. Bimberg, V. M. Ustinov, A. Y. Egorov, A. E. Zhukov et al., *Phys. Rev. B* 54, 8743 (1996).
42. H. T. Dobbs, D. D. Vvedensky, Z. Zangwill, J. Johansson, N. Carlsson, and W. Seifert, *Phys. Rev. Lett.* 79, 897 (1997).
43. V. A. Shchukin, N. N. Ledentsov, P. S. Kop'ev, and D. Bimberg, *Phys. Rev. Lett.* 75, 2968 (1995).
44. I. Daruka and A. L. Barabasi, *Appl. Phys. Lett.* 72, 2102 (1998).
45. N. Koguchi and K. Ishige, *Jpn. J. Appl. Phys.* 32, 2052 (1993).
46. N. Koguchi, K. Ishige, and S. Takahashi, *J. Vac. Sci. Technol. B* 11, 787 (1993).
47. K. Watanabe, N. Koguchi, and Y. Gotoh, *Jpn. J. Appl. Phys.* 39, L79 (2000).
48. T. Mano, K. Watanabe, S. Tsukamoto, H. Fujioka, M. Oshima, and N. Koguchi, *Jpn. J. Appl. Phys.* 38, L1009 (1999).
49. T. Mano, K. Watanabe, S. Tsukamoto, N. Koguchi, H. Fujioka, M. Oshima, C. Lee, J. Leem, H. J. Lee, and S. K. Noh, *Appl. Phys. Lett.* 76, 3543 (2000).
50. T. Tateno, K. Watanabe, S. Sanguinetti, S. Tsukamoto, M. Wakaki, Y. Gotoh, and N. Koguchi, in "Growth of Well Defined Nanostructures" (N. Koguchi, Ed.), p. 298. National Reserch Institute for Metals, Tsukuba, Japan, 2001.
51. J. Davies, *J. Appl. Phys.* 84, 1358 (1998).
52. O. Stier, M. Grundmann, and D. Bimberg, *Phys. Rev. B* 59, 5688 (1999).
53. H. Jiang and J. Singh, *Phys. Rev. B* 56, 4696 (1997).
54. J. Kim, L. W. Wang, and A. Zunger, *Phys. Rev. B* 57, R9408 (1997).
55. A. Zunger, *Mater. Res. Soc. Bull.* 23, 35 (1998).
56. J. Nye, "'Physical Properties of Crystals." Clarendon, Oxford, 1964.
57. C. Pryor, M. E. Pistol, and L. Samuelson, *Phys. Rev. B* 56, 10404 (1997).
58. J. R. Downes, D. A. Faux, and E. P. O'Reilly, *J. Appl. Phys.* 81, 6700 (1997).
59. A. D. Andreev, J. R. Downes, D. A. Faux, and E. P. O'Reilly, *J. Appl. Phys.* 86, 297 (1999).
60. P. Keating, *Phys. Rev.* 145, 637 (1966).
61. R. Martin, *Phys. Rev. B* 1, 4005 (1970).
62. C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger, *J. Appl. Phys.* 83, 2548 (1998).

63. M. A. Cusack, P. R. Briddon, and M. Jaros, *Phys. Rev. B* 56, 4047 (1997).
64. C. Pryor, *Phys. Rev. B* 57, 7190 (1998).
65. M. Tadic, F. M. Peeters, and K. L. Janssens, *Phys. Rev. B* 65, 165333 (2002).
66. H. Jiang and J. Singh, *Appl. Phys. Lett.* 71, 3239 (1997).
67. S. Sanguinetti, K. Watanabe, T. Kuroda, F. Minami, Y. Gotoh, and N. Koguchi, *J. Crystal Growth* 242, 321 (2002).
68. O. Stier, "Electronic and Optical Properties of Quantum Dots and Wires." Wissenshaft and Technik Verlag, Berlin, 2001.
69. M. L. Cohen and J. R. Chelikowsky, "Electronic Structure and Optical Properties of Semiconductors." Springer-Verlag, Berlin, 1988.
70. T. Mattila, L. W. Wang, and A. Zunger, *Phys. Rev. B* 59, 15270 (1999).
71. S. H. Wei and A. Zunger, *Appl. Phys. Lett.* 72, 2011 (1998).
72. A. Franceschetti, S. H. Wei, and A. Zunger, *Phys. Rev. B* 50, 17797 (1994).
73. L. W. Wang and A. Zunger, *Phys. Rev. B* 59, 15806 (1999).
74. L. W. Wang, A. J. Williamson, A. Zunger, H. Jiang, and J. Singh, *Appl. Phys. Lett.* 76, 339 (2000).
75. L. W. Wang, J. Kim, and A. Zunger, *Phys. Rev. B* 59, 5678 (1999).
76. M. Grundmann and D. Bimberg, *Phys. Rev. B* 55, 9740 (1997).
77. E. Dekel, D. Gershoni, E. Ehrenfreund, J. M. Garcia, and P. M. Petroff, *Phys. Rev. B* 61, 11009 (2000).
78. W. Sheng and J. P. Leburton, *Appl. Phys. Lett.* 80, 2755 (2002).
79. L. R. C. Fonseca, J. L. Jimenez, J. P. Leburton, and R. M. Martin, *Phys. Rev. B* 57, 4017 (1998).
80. J. Shumway, A. Franceschetti, and A. Zunger, *Phys. Rev. B* 59, 15819 (2001).
81. E. Dekel, D. V. Regelman, D. Gershoni, E. Ehrenfreund, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. B* 62, 11038 (2000).
82. T. Kuroda, S. Sanguinetti, M. Gurioli, K. Watanabe, F. Minami, and N. Koguchi, *Phys. Rev. B* 66, 121302 (2002).
83. A. J. Williamson, A. Franceschetti, and A. Zunger, *Europhys. Lett.* 53, 59 (2001).
84. J. J. Finley, A. D. Ashmore, A. Lemaitre, D. J. Mowbray, M. S. Skolnick, I. E. Itskevich, P. A. Maksym, M. Hopkinson, and T. F. Krauss, *Phys. Rev. B* 63, 073307 (2001).
85. T. Schmidt, R. J. Haug, K. v. Klitzing, A. Förster, and H. Lüth, *Phys. Rev. Lett.* 78, 1544 (1997).
86. S. Fafard, M. Spanner, J. P. McCaffrey, and Z. R. Wasilewski, *Appl. Phys. Lett.* 76, 2268 (2000).
87. Y. S. Joe, D. S. Ikeler, R. M. Cosby, A. M. Satanin, and C. S. Kim, *J. Appl. Phys.* 88, 2704 (2000).
88. G. Schedelbeck, W. Wegscheider, M. Bichler, and G. Abstreiter, *Science* 278, 1792 (1997).
89. T. H. Oosterkamp, T. Fujisawa, W. G. van der Wiel, K. Ishibashi, R. V. Hijman, S. Tarucha, and L. P. Kouwenhoven, *Nature* 395, 873 (1998).
90. M. Bayer, P. Hawrylak, K. Hinzer, S. Fafard, M. Korkusinski Z. R. Wasilewski, O. Stern, and A. Forchel, *Science* 291, 451 (2001).
91. M. Colocci, A. Vinattieri, L. Lippi, F. Bogani, M. Rosa-Clot, S. Taddei, A. Bosacchi, S. Franchi, and P. Frigeri, *Appl. Phys. Lett.* 74, 564 (1999).
92. R. Heitz, M. Veit, A. Kalburge, Q. Xie, M. Grundmann, P. Chen, N. Ledentsov, A. Hoffmann, A. Madhukar, D. Bimberg, et al., *Physica E* 2, 578 (1998).
93. A. Crist, H. Giessen, W. W. Rüle, K. Korona, J. Kuhl, M. Zundel, Y. Manz, and K. Eberl, *Phys. Status Solidi* (*b*) 221, 59 (2000).
94. M. Hayne, R. Provoost, M. K. Zundel, Y. M. Manz, K. Eberl, and V. V. Moshchalkov, *Phys. Rev. B* 62, 10324 (2000).
95. H. Bebb and E. Williams, in "Transport and Optical Phenomena" (R. Willardson and A. Beer, Eds.), Vol. 8 of *Semiconductors and Semimentals*, p. 181. Academic Press, New York, 1972.
96. P. Dean, *Progr. Cryst. Growth Charact.* 5, 89 (1982).
97. L. Pavesi and M. Guzzi, *J. Appl. Phys.* 75, 4779 (1994).
98. M. Herman, D. Bimberg, and J. Christen, *J. Appl. Phys.* 70, R1 (1991).
99. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, *Phys. Rev. Lett.* 87, 157401 (2001).
100. K. Leosson, J. Jensen, J. Hvam, and W. Langbein, *Phys. Status Solidi B* 221, 49 (2000).
101. C. Kammerer, C. Voisin, G. Cassabois, C. Delalande, P. Roussignol, F. Klopf, J. P. Reithmaier, A. Forchel, and J. M. Gérard, *Phys. Rev. B* 66, 041306 (2002).
102. K. Matsuda, K. Ikeda, T. Saiki, H. Tsuchiya, H. Saito, and K. Nishi, *Phys. Rev. B* 63, 121304 (2001).
103. M. Bissiri, G. Baldassarri Höger von Högersthal, M. Capizzi, P. Frigeri, and S. Franchi, *Phys. Rev. B* 64, 245337 (2001).
104. R. Heitz, T. R. Ramachandran, A. Kalburge, Q. Xie, I. Mekhametzhanov, P. Chen, and A. Madhukar, *Phys. Rev. Lett.* 78, 4071 (1997).
105. T. Ramachandran, R. Heitz, N. Kobayashi, A. Kalburge, W. Yu, P. Chen, and A. Madhukar, *J. Crystal Growth* 175/176, 216 (1997).
106. R. Heitz, F. Guffarth, I. Mukhametzhanov, M. Grundmann, A. Madhukar, and D. Bimberg, *Phys. Rev. B* 62, 16881 (2000).
107. S. Sanguinetti, G. Chiantoni, E. Grilli, M. Guzzi, M. Henini, A. Polimeni, A. Patané, L. Eaves, and P. Main, *Mater. Sci. Eng. B* 74, 239 (2000).
108. M. Colocci, F. Bogani, L. Carraresi, R. Mattolini, A. Bosacchi, S. Franchi, P. Frigeri, S. Taddei, and M. Rosa-Clot, *Appl. Phys. Lett.* 70, 3140 (1997).
109. S. Sanguinetti, G. Chiantoni, A. Miotto, E. Grilli, M. Guzzi, M. Henini, A. Polimeni, A. Patané, L. Eaves, and P. Main, *Micron* 31, 309 (2000).
110. A. Polimeni, A. Patané, M. Henini, L. Eaves, and P. C. Main, *Phys. Rev. B* 59, 5064 (1999).
111. P. Altieri, S. Sanguinetti, M. Gurioli, E. Grilli, M. Guzzi, P. Frigeri, S. Franchi, and G. Trevisi, *Mater. Sci. Eng. B* 88, 234 (2002).
112. Y. S. Kim, U. H. Lee, D. Lee, S. J. Rhee, Y. A. Leem, H. S. Ko, D. H. Kim, and J. C. Woo, *J. Appl. Phys.* 87, 241 (2000).
113. F. Adler, M. G. A. Bauknecht, D. Haase, P. Ernst, A. Dornen, F. Scholz, and H. Schweizer, *J. Appl. Phys.* 83, 1631 (1998).
114. R. Heitz, O. Stier, I. Mukhametzhanov, A. Madhukar, and D. Bimberg, *Phys. Rev. B* 62, 11017 (2000).
115. R. Heitz, A. Kalburge, Q. Xie, M. Grundmann, P. Chen, A. Hoffmann, A. Madhukar, and D. Bimberg, *Phys. Rev. B* 57, 9050 (1998).
116. M. Grassi Alessi, M. Capizzi, A. S. Bhatti, A. Frova, Martelli, P. Frigeri, A. Bosacchi, and S. Franchi, *Phys. Rev. B* 59, 7620 (1999).
117. P. O. Vaccaro, K. Fujita, and T. Watanabe, *Jpn. J. Appl. Phys.* 36, 1948 (1997).
118. K. Nishi, R. Mirin, D. Leonard, G. Medeiros-Ribeiro, P. M. Petroff, and A. C. Gossard, *J. Appl. Phys.* 80, 3466 (1996).
119. S. Fortina, S. Sanguinetti, E. Grilli, M. Guzzi, M. Henini, A. Polimeni, and L. Eaves, *J. Cryst. Growth* 187, 126 (1998).
120. W. Jiang, H. Xu, B. Xu, W. Zhou, Q. Gong, D. Ding, J. Liang, and Z. Wang, *J. Vac. Sci. Technol. B* 19, 197 (2001).
121. S. Sanguinetti, G. Chiantoni, E. Grilli, M. Guzzi, M. Henini, A. Polimeni, A. Patané, L. Eaves, and P. C. Main, *Europhys. Lett.* 47, 701 (1999).
122. A. O. Kosogov, P. Werner, U. Gösele, N. N. Ledentsov, D. Bimberg, V. M. Ustinov, A. Y. Egorov, A. E. Zhukov, P. S. Kopev, N. A. Bert et al., *Appl. Phys. Lett.* 69, 3072 (1996).
123. S. Fafard, Z. R. Wasilewski, C. N. Allen, D. Picard, M. Spanner, J. P. McCaffrey, and P. G. Piva, *Phys. Rev. B* 59, 15368 (1999).
124. T. M. Hsu, Y. S. Lan, W. Chang, N. T. Yeh, and J. Chyi, *Appl. Phys. Lett.* 76, 691 (2000).
125. Y. Berhane, M. O. Manasreh, H. Yang, and G. J. Salamo, *Appl. Phys. Lett.* 78, 2196 (2001).
126. S. Malik, C. Roberts, R. Murray, and M. Pate, *Appl. Phys. Lett.* 71, 1987 (1997).

127. D. G. Deppe, L. J. Guido, N. Holonyak, K. C. Hsieh, R. D. Burnham, R. L. Thornton, and T. L. Paoli, *Appl. Phys. Lett.* 49, 510 (1986).
128. G. P. Kothiyal and P. Bhattacharya, *J. Appl. Phys.* 63, 2760 (1988).
129. I. Shtrichman, D. Gershoni, and R. Kalish, *Phys. Rev. B* 56, 1509 (1997).
130. D. Gammon, E. S. Snow, B. V. Shanabrook, D. S. Katzer, and D. Park, *Science* 273, 87 (1996).
131. H. Saito, K. Nishi, S. Sugou, and Y. Sugimoto, *Appl. Phys. Lett.* 71, 590 (1997).
132. S. P. Guo, H. Ohno, A. Shen, F. Matsukura, and Y. Ohno, *Appl. Phys. Lett.* 70, 2378 (1997).
133. M. Wassermeier, J. Sudijono, M. D. Johnson, K. T. Leung, B. G. Orr, L. Däweritz, and K. Ploog, *Phys. Rev. B* 51, 14 721 (1995).
134. D. J. Chadi, *J. Vac Sci. Technol. B* 3, 1167 (1985).
135. G. Brocks, P. J. Kelley, and R. Car, in "20th Int. Conf. On the Physics of Semiconductors" (E. Anastassakis and J. Joannopoulos, Eds.). World Scientific, Singapore, 1991.
136. M. Henini, S. Sanguinetti, S. Fortina, E. Grilli, M. Guzzi, G. Panzarini, L. Andreani, M. Upward, P. Moriarty, P. Beton et al., *Phys. Rev. B* 57, R6815 (1998).
137. S. Sanguinetti, S. Castiglioni, E. Grilli, M. Guzzi, G. Panzarini, L. C. Andreani, and M. Henini, *Jpn. J. Appl. Phys.* 38, 4676 (1999).
138. S. Cortez, O. Krebs, P. Voisin, and J. M. Gérard, *Phys. Rev. B* 63, 233306 (2001).
139. L. Brusaferri, S. Sanguinetti, E. Grilli, M. Guzzi, A. Bignazzi, F. Bogani, L. Carraresi, M. Colocci, A. Bosacchi, P. Frigeri, et al., *App. Phys. Lett.* 69, 3354 (1996).
140. D. I. Lubyshev, P. P. Gonzalez-Borrero, E. Marega, E. Petitprez, N. LaScala, and P. Basmaji, *Appl. Phys. Lett.* 68, 205 (1996).
141. H. Lee, W. Yang, and P. C. Sercel, *Phys. Rev. B* 55, 9757 (1997).
142. Z. Y. Xu, Z. D. Lu, X. P. Yang, Z. L. Yuan, B. Z. Zheng, J. Z. Xu, Y. Wang, J. Wang, and L. L. Chang, *Phys. Rev. B* 54, 11528 (1996).
143. S. Fafard, S. Rymond, G. Wand, R. Leon, D. Leonard, S. Charbonneau, J. L. Merz, P. M. Petroff, and J. E. Bowers, *Surface Science* 361/362, 778 (1996).
144. W. Yang, R. R. Lowe-Webb, H. Lee, and P. C. Sercel, *Phys. Rev. B* 56, 13314 (1997).
145. S. Sanguinetti, M. Henini, M. Grassi-Alessi, M. Capizzi, P. Frigeri, and S. Franchi, *Phys. Rev. B* 60, 8276 (1999).
146. M. Gurioli, S. Testa, P. Altieri, S. Sanguinetti, E. Grilli, M. Guzzi, G. Trevisi, P. Frigeri, and S. Franchi *Physica E* 17, 19 (2003).
147. Y. Zhang, C. Huang, X. Ye, B. Xu, D. Ding, J. Wang, Y. Li, F. Liu, and Z. Wang, *Chin. Phys. Lett.* 18, 1411 (2001).
148. R. Heitz Mukhametzhanov, O. Stier, A. Madhukar, and D. Bimberg, *Phys. Rev. Lett.* 83, 4654 (1999).
149. R. Heitz, I. Mukhametzhanov, O. Stier, A. Madhukar, and D. Bimberg, *Physica E* 7, 398 (2000).
150. M. Bissiri, G. H. V. H. Baldassarri, A. S. Bhatti, M. Capizzi, A. Frova, P. Frigeri, and S. Franchi, *Phys. Rev. B* 62, 4642 (2000).
151. S. Sanguinetti, M. Gurioli, and M. Henini, *Microelectron. J.* 33, 583 (2002).
152. P. Vaccaro, K. Tominaga, M. Hosoda, K. Fujita, and T. Watanabe, *Jpn. J. Appl. Phys.* 34, 1362 (1995).
153. J. D. Bruno and R. L. Tober, *J. Appl. Phys.* 85, 2221 (1999).
154. S. Sanguinetti, M. Gurioli, E. Grilli, M. Guzzi, and M. Henini *Appl. Phys. Lett.* 77, 1982 (2000).
155. A. Patané, P. Levin, A. Polimeni, F. Schindler, P. C. Main, L. Eaves, and M. Henini, *Appl. Phys. Lett.* 77, 2979 (2000).
156. M. Gurioli, S. Sanguinetti, and M. Henini, *Appl. Phys. Lett.* 78, 931 (2001).
157. P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksyn et al. *Phys. Rev. Lett.* 84, 733 (2000).
158. S. Sanguinetti, M. Padovani, M. Gurioli, E. Grilli, M. Guzzi, A. Vinattieri, M. Colocci, P. Frigeri, and S. Franchi, *Appl. Phys. Lett.* 77, 1307 (2000).
159. S. Taddei, M. Colocci, A. Vinattieri, F. Bogani, S. Franchi, P. Frigeri, L. Lazzarini, and G. Salviati *Phys. Rev. B* 62, 10220 (2000).
160. M. Gurioli, S. Sanguinetti, S. Lozzia, E. Grilli, M. Guzzi, P. Frigeri, S. Franchi, M. Colocci, A. Vinattieri, and S. Taddei, *Phys. Status Solidi A* 190, 577 (2002).
161. E. Dekel, D. Gershoni, E. Ehrenfreund, D. Spektor, J. M. Garcia, and P. M. Petroff, *Phys. Rev. Lett.* 80, 4991 (1998).
162. L. Landin, M. S. Miller, M. Pistol, C. E. Pryor, and L. Samuelson, *Science* 280, 262 (1998).
163. M. E. Pistol, P. Castrillo, D. Hessman, J. A. Prieto, and L. Samuelson, *Phys. Rev. B* 59, 10725 (1999).
164. P. G. Blome, M. Wenderoth, M. Hubner, R. G. Ulbrich, J. Porsche, and F. Scholz, *Phys. Rev. B* 61, 8382 (2000).
165. N. Panev, M. E. Pistol, V. Zwiller, L. Samuelson, W. Jiang, B. Xu, and Z. Wang, *Phys. Rev. B* 64, 045317 (2001).
166. S. Raymond, J. P. Reynolds, J. L. Merz, S. Fafard, Y. Feng, and S. Charbonneau, *Phys. Rev. B* 58, R13415 (1998).
167. K. Hinzer, P. Hawrylak, M. Korkusinski, S. Fafard, M. Bayer, O. Stern, A. Gorbunov, and A. Forchel, *Phys. Rev. B* 63, 075314 (2001).
168. M. Bayer, O. Stern, A. Kuther, and A. Forchel, *Phys. Rev. B* 61, 7273 (2000).
169. J. Shah, "Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures" Springer-Verlag, Berlin 1996.
170. P. Buckle, P. Dawson, S. A. Hall, X. Chen, M. J. Steer, D. J. Mowbray, M. S. Skolnick, and M. Hopkinson, *J. Appl. Phys.* 86, 2555 (1999).
171. H. Benisty, C. M. Sotomayor-Torres, and C. Weisbuch, *Phys. Rev. B* 44, 10 945 (1991).
172. A. L. Efros, V. A. Kharchenko, and M. Rosen, *Solid State Commun.* 93, 281 (1995).
173. D. F. Schroeter, D. J. Griffiths, and P. C. Sercel, *Phys. Rev. B* 54, 1486 (1996).
174. Y. Yoda, O. Moriwaki, M. Nishioka, and Y. Arakawa, *Phys. Rev. Lett.* 82, 4114 (1999).
175. R. Ferreira and G. Bastard, *Appl. Phys. Lett.* 74, 2818 (1999).
176. U. Bockelmann and G. Bastard, *Phys. Rev. B* 42, 8947 (1990).
177. R. Heitz, M. Veit, N. N. Ledentsov, A. Hoffmann, D. Bimberg, V. M. Ustinov, P. S. Kop'ev, and Z. I. Alferov, *Phys. Rev. B* 56, 10435 (1997).
178. O. Verzelen, R. Ferreira, and G. Bastard, *Phys. Rev. B* 62, 4809 (2000).
179. D. Morris, N. Perret, and S. Fafard, *Appl. Phys. Lett.* 75, 3593 (1999).
180. B. Ohnesorge, M. Albrecht, J. Oshinowo, and A. Forchel, *Phys. Rev. B* 54, 11532 (1996).
181. U. Bockelmann and T. Egeler, *Phys. Rev. B* 46, 15 574 (1992).
182. A. V. Uskov, F. Adler, H. Schweizer, and M. H. Pilkuhn, *J. Appl. Phys.* 81, 7895 (1997).
183. S. Sanguinetti, K. Watanabe, T. Tateno, M. Wakaki, N. Koguchi, T. Kuroda, F. Minami, and M. Gurioli, *Appl. Phys. Lett.* 81, 613 (2002).
184. A. Khaetskii and Y. Nazarov, *Physica E* 6, 470 (2000).
185. R. Fiederling, M. Keim, G. Reuscher, W. Ossau, G. Schmidt, A. Waag, and L. W. Molenkamp, *Nature* 402, 787 (1999).
186. M. Bayer, A. Kuther, A. Forchel, A. Gorbunov, V. B. Timofeev, F. Schäfer, J. P. Reithmaier, T. L. Reinecke, and S. N. Walck, *Phys. Rev. Lett.* 82, 1748 (1999).
187. H. Gotoh, H. Ando, H. Kamada, A. Chavez–Pirson, and J. Temmyo, *Appl. Phys. Lett.* 72, 1341 (1998).

188. M. Paillard, X. Marie, P. Renucci, T. Amand, A. Jbeli, and J. M. Gérard, *Phys. Rev. Lett.* 86, 1634 (2001).
189. S. Raymond, S. Fafard, P. J. Poole, A. Wojs, P. Hawrylak, S. Charbonneau, D. Leonard, R. Leon, P. M. Petroff, and J. L. Merz, *Phys. Rev. B* 54, 11548 (1996).
190. S. Malik, E. C. Le Ru, D. Childs, and R. Murray, *Phys. Rev. B* 63, 155313 (2001).
191. U. Bockelmann, *Phys. Rev. B* 48, 17 637 (1993).
192. M. Paillard, X. Marie, E. Vanelle, T. Amand, V. K. Kalevich, A. R. Kovsh, A. E. Zhukov, and V. M. Ustinov, *Appl. Phys. Lett.* 76, 76 (2000).
193. S. Sanguinetti, M. Gurioli, and M. Henini, *Eur. Phys. J. B* 27, 75 (2002).
194. P. Castrillo, D. Hessman, M. Pistol, S. Anand, N. Carlsson, W. Seifert, and L. Samuelson, *Appl. Phys. Lett.* 67, 1905 (1995).
195. M. Brasken, M. Lindberg, M. Sopanen, H. Lipsanen, and J. Tulkki, *Phys. Rev. B* 98, R15993 (1998).
196. D. Bimberg, M. Grundmann, and N. N. Ledentsov, *Mater. Res. Soc. Bull.* 23, 31 (1998).
197. A. Steane, *Rep. Prog. Phys.* 61, 117 (1998).
198. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* 74, 145 (2002).
199. M. Asada, Y. Miyamoto, and Y. Suematsu, *IEEE J. Quantum Elelctron.* QE22, 1915 (1986).
200. N. Kirstaedter, N. Ledentsov, M. Grundmann, D. Bimberg, V. Ustinov, S. Ruvimov, M. Maximov, P. Kop'ev, Z. Alferov, U. Richter, et al. *Electron. Lett.* 30, 1416 (1994).
201. J. M. Gérard, B. Sermage, B. Gayral, B. Legrand, E. Costard, and V. Thierry–Mieg, *Phys. Rev. Lett.* 81, 1110 (1998).
202. P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu, *Science* 290, 2282 (2000).
203. C. Santori, M. Pelton, G. Solomon, Y. Dale, and Y. Yamamoto, *Phys. Rev. Lett.* 86, 1502 (2001).

www.aspbs.com/enn

Semiconductor Quantum Wires

Xing-Quan Liu, Xue-Lun Wang

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

C. Jagadish

The Australian National University, Canberra, Australia

M. Ogura

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

CONTENTS

1. Introduction
2. Growth of Quantum Wires
3. Optical Properties of Quantum Wires Structures
4. Quantum Wires Devices
5. Summary
Glossary
References

1. INTRODUCTION

The superior performances of quantum well (QWL) semiconductor devices over bulk devices have stimulated considerable interest in lower dimensional structures such as quantum wires (QWRs) and quantum dots (QDs) [1]. Carrier confinement to one or zero dimensions is expected to give rise to sharp peaks in the density of states (DOS). Figure 1 shows the DOS profile in energy of quantum nanostructures. The sharp peaks of the DOS can result in many interesting properties. For example, when used for optoelectronic devices, the sharp DOS peaks can lead to increased exciton binding, enhanced optical nonlinearities, narrower gain spectra, and higher differential gain.

The mature modern epitaxial techniques, such as metal-organic vapor-phase epitaxy (MOVPE) and molecular-beam epitaxy (MBE), provide the possibility of the fabrication of nanometer-sized semiconductor structures, including QWLs, QWRs, and QDs. In this chapter, we mainly review the advances in the growth, optical properties, and devices of semiconductor QWRs.

Several types of QWRs are realized by using III–V, II–VI, and SiGe semiconductor heterostructures, and also by using polymers and colloids. Here, we concentrate mainly on GaAs- and InGaAs-related QWRs produced by epitaxy, which are the two most mature kinds of semicondutor QWRs. Compared with others, these two kinds of QWRs have been well investigated, and several electronic devices have been realized, including QWR lasers, photodetectors, modulators, and transistors.

2. GROWTH OF QUANTUM WIRES

The current fabrication techniques of semiconductor QWRs make it difficult to obtain symmetric strong confinement from both dimensions. Most of the methods are based on the modified heteroepitaxial techniques, that is, QWRs are formed using the edges, corners, or intersections of two QWL layers. Therefore, one dimension confinement is from the potential barriers of heterostructures, while the other confinement is from the differences of the quantization energy, which is weaker. One aim to fabricate high-quality QWRs is to obtain large confinements from both sides. Two superior epitaxial methods of semiconductor QWRs are MOVPE and MBE.

2.1. V-Groove Quantum Wires

V-groove QWRs are one type of the most mature and promising QWR structures. The idea of V-groove QWRs was first proposed by Kapon and co-workers [3–5]. The idea is to grow QWL structures on a nonplanar substrate

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (763–773)

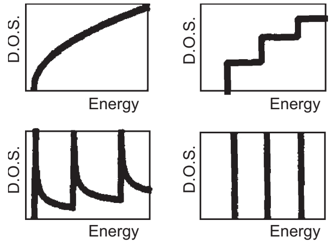


Figure 1. Density of states of semiconductor bulk, quantum wells, quantum wires, and quantum dots.

to achieve lateral confinement by QWL thickness variation. V grooves are formed by two {111} facets. Figure 2 demonstrates the processing of the V grooves on GaAs (001) substrates. A (001) GaAs substrate wafer is processed using standard photolithography to obtain 2 μm wide photoresist-masked stripes with a 2 μm blank region between two stripes. Wet chemical etching is used to etch this patterned wafer. Because of the selectivity of the chemical etching to the different crystal facets, the {111} facets will be exposed to the surface. The V grooves are then formed by two {111} crystal facets, as demonstrated in Figure 2(a). Two kinds of chemical solutions are usually used to produce the V-groove patterned substrates: H_3PO_4:H_2O_2:H_2O (3:1:1), or

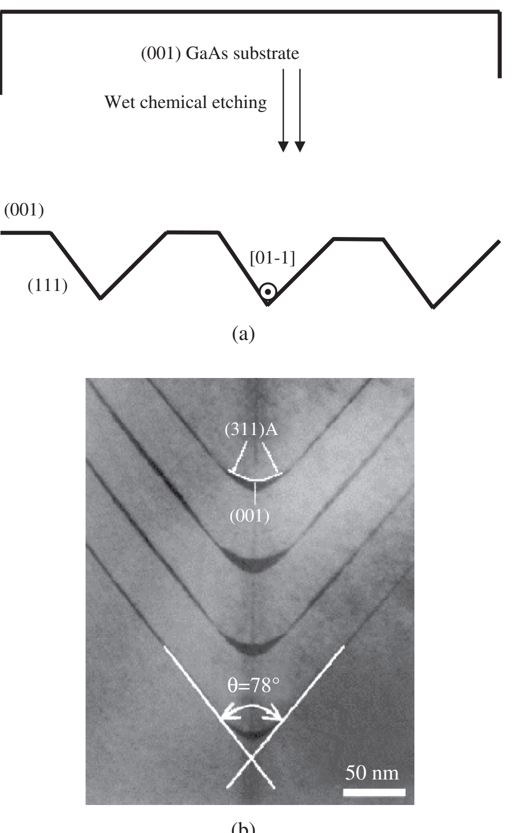


Figure 2. (a) Schematic of preparation of V-groove pitches. (b) TEM image of multiple AlGaAs/GaAs V-groove QWRs grown using flow rate modulation epitaxy method. Reprinted with permission from [16], X. L. Wang and M. Ogura, *J. Cryst. Growth* 221, 556 (2000). © 2000, Elsevier Science.

NH_3:H_2O_2 (1:5). The etching temperature can be at 0 °C or at room temperature. 4 μm pitch, [01–1]-oriented V grooves are formed after chemical etching of a 4 μm pitch photoresist patterned substrate wafer. To obtain a high density of the V grooves, the holographic lithography technique has been used. One short-wavelength laser beam interferes with its reflected beam on a substrate surface, which is coated with photoresist to obtain a high density of stripe patterns. This patterned substrate is also etched with the chemical solution. A 250 nm period V-groove array has been reported [6]. The growth on these V grooves will be self-limiting growth. When a GaAs or InGaAs QWL structure is grown on the patterned substrates, the well thickness at the V groove bottom is much thicker than that at the lateral side (so-called sidewall QWLs). Therefore, confinement can be obtained from the sidewall QWLs. At the bottom of the V grooves, a crescent-shaped QWR will be formed. Figure 2 shows a schematic illustration and a cross-section TEM image of a multiple GaAs/AlGaAs V-groove QWR structure grown on a 4 μm pitch V-groove grating substrate using the flow-rate-modulated epitaxy (FME) method [7]. The crescent shape is formed by different crystal facets, which is demonstrated in the schematic diagram. Because the diffusion length of the Al atom on the GaAs substrate is much smaller than those of Ga and In atoms, therefore the AlGaAs layer will not show much of a self-limiting profile; then the growth of the AlGaAs layer can maintain the V-groove shape. This property provides a chance to grow a thick AlGaAs buffer layer to recover the etching-induced damages on the surface without changing the V shape, and also to recover the V shape after the QWR layer.

2.1.1. Growth of V-Groove Quantum Wires with Conventional Epitaxy

AlGaAs/GaAs QWR MOCVD is preferred to grow V-groove QWR structures rather than MBE, which has been predicted in theory [8–9]. Here, conventional epitaxy means growing a QWL structure on the V-groove substrate as a conventional process of QWL growth on a planar substrate without interruption. Figure 3(a) shows a TEM image of multiple GaAs/AlGaAs QWRs grown with conventional epitaxy using low-pressure MOVPE on a 4 μm pitch V-groove patterned substrate. During the growth, a 0.1 μm GaAs buffer layer was grown before a 1 μm $Al_{0.5}Ga_{0.5}As$ layer. On the top of this thick $Al_{0.5}Ga_{0.5}As$ layer, a nominal multiple QWL structure was grown. All layers were grown at 750 °C. Multiple QWRs were then obtained at the V-groove bottoms. The thickness difference between the QWR and the lateral QWL is about three times. A dark vertical band in the TEM image is formed at the bottom of the V grooves as a low Al composition AlGaAs band due to the longer migration length of a Ga atom than that of an Al atom, and this low Al composition AlGaAs band forms a vertical quantum well (VQWL). Figure 3(b) shows the room-temperature PL spectra by using the micro-PL method. The dotted line is a Gaussian fitting. The sample is silicon doped in the wire region, which is for the purpose of photodetector devices using intersubband transition.

The lateral confinement of the QWR is always a limitation to obtaining the one-dimensional properties. To increase the

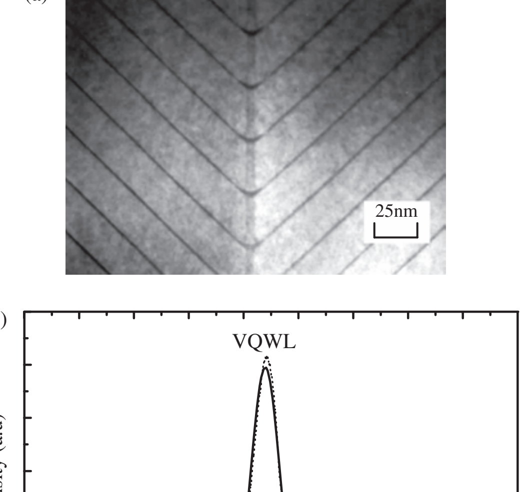


Figure 3. (a) TEM image of a multiple QWR structure grown by conventional MOVPE epitaxy. (b) Room-temperature micro-PL spectrum.

lateral confinement, the sublevel energy difference between a sidewall QWL and QWR should be enlarged. Liu et al. used dual self-aligned geometry to selectively implant the sidewall QWL, and then used thermal annealing to recover the damage. The sublevels of the sidewall QWL are then lifted from the interface intermixing [10, 11]. Figure 4(a) and (b) shows the CL images of the implanted and as-grown QWR structures. After implantation, the sidewall QWLs are disabled, and the CL from QWRs are clearly shown.

InGaAs/AlGaAs QWRs Indium atoms have a larger diffusion length than gallium and aluminum atoms. In the meantime, the InGaAs has a smaller bandgap than GaAs. Therefore, incorporation of indium in the wire region will benefit the lateral confinement. Figure 5 shows a TEM image of the cross section of an $In_{0.2}Ga_{0.8}As/Al_{0.5}Ga_{0.5}As$ QWR structure. The growth selectivity of the QWR to the sidewall QWL is obviously improved compared with GaAs/AlGaAs QWR structures. One difficulty for the growth of the $In_{0.2}Ga_{0.8}As/Al_{0.5}Ga_{0.5}As$ QWR structure is the difference of the growth temperature of the $In_{0.2}Ga_{0.8}As$ layer and the $Al_{0.5}GaAs$ barrier layer, which could be more than 150 °C. Asymmetric $Al_{0.5}Ga_{0.5}As$–$In_{0.25}Ga_{0.75}As$–GaAs QWR was successfully fabricated. Strain effects and the DOS were theoretically investigated [12].

High-Density QWRs To increase the planar density of QWRs, holographic lithography can be used to obtain a high density of stripes, then the high density of V grooves by using wet chemical etching. Figure 6 demonstrates the high density of an InGaAs/GaAs V-groove QWR structure. The distance between two QWRs is about 320 nm. This kind of high-density QWR structure will benefit optical devices, such as lasers and detectors.

2.1.2. Growth of V-Groove Quantum Wires Using FME

To achieve larger lateral confinement, one direct way is to increase the thickness difference of the sidewall QWL and the QWR layers, which needs stronger self-limited growth. Wang et al. used FME to achieve a larger lateral confinement [13–15]. Figure 7 shows the diagram of the flow rate control for the FME growth. We can see that Ga will have a larger diffusion length on the sample surface with arsine flow off. The TEM image in Figure 2 is a typical FME QWR [16]. It is clearly shown that the thickness ratio of the QWR to the sidewall QWL is increased compared with the QWR grown with conventional epitaxy (see Fig. 3). The enhanced lateral confinement can be seen in the optical properties, which will be described in the next part. Another key limitation to achieve real one-dimensional quantum structures is the strong interface effect, which is much stronger than that of QWLs because of the dramatically increased interface-to-volume ratio. Most of the QWRs are quantum box structures, which come from the monolayer fluctuation. The FME method dramatically improved the QWR interfaces [15].

TBA (tertiarybutylarsine) as Arsenic Source Wang et al. found that, when the TBA is used as an arsenic source in FME growth, the QWR interfaces were obviously

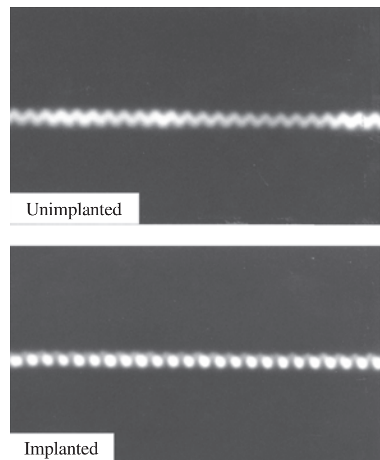


Figure 4. CL spectra of the as-grown (labeled as unimplanted) and the implanted samples.

Figure 5. TEM image of the cross section of an $In_{0.2}Ga_{0.8}As/Al_{0.5}Ga_{0.5}As$ QWR.

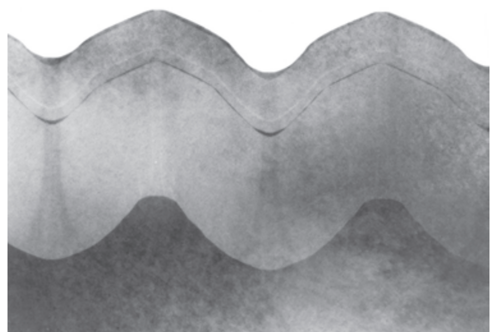


Figure 6. TEM image of high-density GaAs/$In_{0.2}Ga_{0.8}As$ QWRs grown on high-density V-groove substrates prepared by using holographic lithography. The distance between two QWRs is 320 nm.

improved compared with the conventional arsine source [16]. A comparison of V-groove structures by using arsine and TBA as an arsenic source in FME growth was demonstrated by AFM observation. Clear improvements were observed.

QWR Superlattice Buffer Layer One of the methods to improve the QWR interface is to use a QWR superlattice as a buffer layer. Microphotoluminescence (micro-PL) measurements showed that the number of exciton emissions within the excitation spot dramatically decreased (to be discussed in Section 3).

2.2. T-Shape Quantum Wires

The concept of the T-shaped QWR was originally proposed and investigated by Chang et al. [17]. The QWR is formed by using the cleaved-edge-overgrowth method, which is realized by cleaving a grown QWL structure in the MBE growth chamber followed by QWL growth over the cleaved edge. Two QWLs form a T shape; at the cross-section line, a QWR can be formed [18–20]. Figure 8 shows a schematic diagram of a T-shaped QWR. An $In_{0.17}Ga_{0.83}As$ T-QWR with $Al_{0..3}Ga_{0.7}As$ barriers is fabricated by the cleaved-edge-overgrowth method. The MQW layer consists of ten periods of $In_{0.17}Ga_{0.83}As$ QW and $Al_{0.3}Ga_{0.7}As$ barriers, and the total width of the MQW region is about 1 μm. Two 5 μm wide regions of the overgrowth QW were formed on both sides of the QWR region. This structure enables the unambiguous assignment of the PL peaks to QWR, QW1, and QW2 (after [20]). The T-shaped QWR has been extensively studied, particularly for the basic 1-D excitonic properties [21–23].

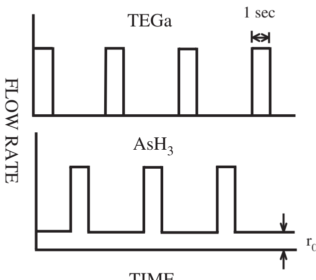


Figure 7. Schematic of the FME flow sequence during the epitaxy of QWR region.

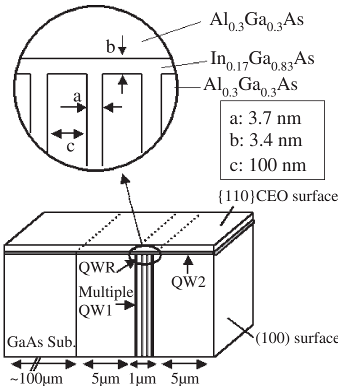


Figure 8. Structure of $In_{0.17}Ga_{0.83}As$ T-QWR with $Al_{0..3}Ga_{0.7}As$ barriers fabricated by the cleaved-edge-overgrowth method. The MQW layer consists of ten periods of $In_{0.17}Ga_{0.83}As$ QW and $Al_{0.3}Ga_{0.7}As$ barriers, and the total width of the MQW region is about 1 μm. Two 5 μm wide regions of the overgrowth QW were formed on both sides of the QWR region. This structure enables the unambiguous assignment of the PL peaks to QWR, QW1, and QW2. Reprinted with permission from [20], M. Yoshita et al., *J. Appl. Phys.* 83, 3777 (1998). © 1998, American Institute of Physics.

2.3. Sidewall Quantum Wires

Using the selectivity in (Al,Ga)As growth on patterned GaAs(311)A substrates, another kind of GaAs sidewall QWR can be fabricated by using MBE growth [24–27]. This concept is based on the intentional introduction of distinct lateral thickness variations of narrow quantum wells (QWs) in the x and/or y direction, in addition to the abrupt potential barrier in the z direction, which can be grown epitaxially with great perfection. These lateral thickness variations translate directly into bandgap variations via the transverse quantum size effect, that is, there is a strong variation of carrier confinement energy with well thickness. As a result, in addition to the vertical confinement in the z direction, there is a lateral 1-D or 2-D potential "well" in the (x; y) plane for electrons and holes at the thicker part of the (tapered) quantum well. By combining the lithographic patterning with self-organized growth during MBE growth, this kind of QWR can be obtained. Figure 9 shows a schematic illustration of the formation of lateral quasiplanar single (a) and multiple (c) GaAs QWRs in an (Al,Ga)As matrix at the fast-growing sidewall of shallow [01–1] mesa stripes on patterned GaAs(311)A. The one-dimensional confinement depicted by the real-space energy band diagram in (b) arises from the strong variation of the GaAs layer thickness along [−233].

2.4. Other Kinds of Quantum Wires

Quite a few other kinds of methods are used to fabricate the QWR structures, such as vicinal-substrate QWRs, etched QWRs, and so on. A vicinal-substrate QWR is achieved by making use of the step flow mode of growth on vicinal surfaces with MBE or MOCVD. The vicinal surfaces with small misorientation angles have an atomic step array with a well-defined periodicity in the 10 nm range. In this growth mode, the precursor atoms will preferentially adsorb at the edge of the steps. Based on this mode, the QWR structures are

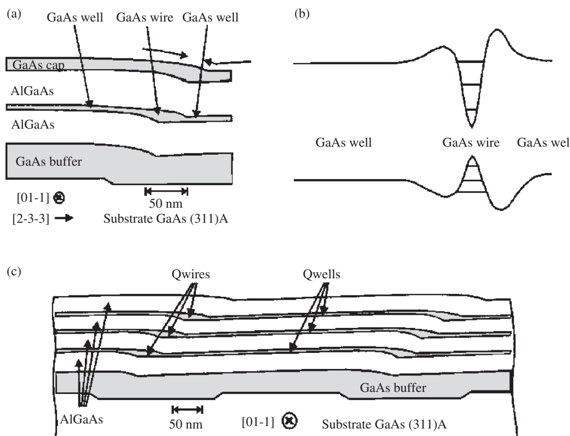


Figure 9. Schematic illustration of the formation of lateral quasiplanar single (a) and multiple (c) GaAs QWRs in (Al,Ga)As matrix at the fast-growing sidewall of shallow [01–1] mesa stripes on patterned GaAs(311)A. The one-dimensional confinement depicted by the real-space energy band diagram in (b) arises from the strong variation of the GaAs layer thickness along [−233]. Reprinted with permission from [27], K. H. Ploog and R. Nötzel, *Phys. E* 11, 78 (2001). © 2001, Elsevier Science.

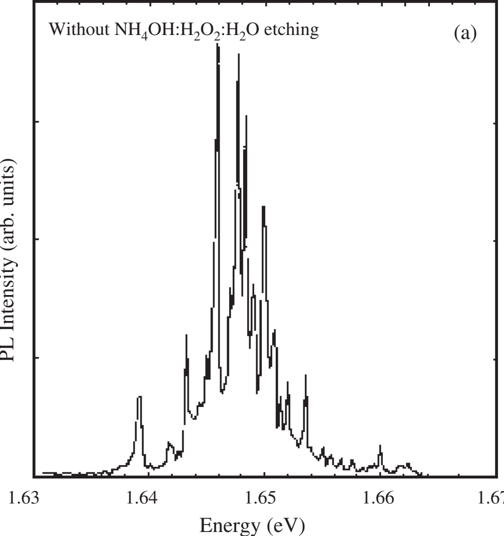
formed by deposition of a superlattice structure on the tilted substrates, in which the lateral confinement is from the composition modulation [28–31].

Etched QWR is achieved by postgrowth etching of a QWL structure. The lateral confinement is from the vacuum potential when the stripe size is narrow enough. This requires a narrow stripe. To achieve nanometer stripes, electron-beam lithography or holographic lithography are necessary, followed by wet or dry etching. One great problem of this QWR is the strong lateral surface effects resulting from chemical etching, particularly for the optical properties [32–34].

3. OPTICAL PROPERTIES OF QUANTUM WIRES STRUCTURES

3.1. Light Emission from Real One-Dimensional Quantum Wires

PL is a direct way to study the one-dimensional properties of QWR. Also, it is a basic property (light emission) for application of the QWR to optoelectronic devices. As described in the first part, most of the QWR structures are quantum-box-like structures because of the monolayer fluctuation at the interface between the wire region and the barriers. The properties can be seen from micro-PL measurements at low temperature with low excitation power. Figure 10 shows low-temperature micro-PL spectra [15]. The structures are $Al_{0.45}Ga_{0.55}As$/GaAs single QWRs before and after improvements. The excitation laser spot diameter is about 1 μm. From Figure 11, we see that, within this 1 μm range, there are many quantum boxes before the interfaces are improved. Each individual sharp PL peak is from one quantum box. When the interface is improved, the quantum box number decreases dramatically, which means that the single QWR section is longer. Scanning near-field optical microscope (SNOM) experiments give a clear indication of a single QWR section. Figure 9 shows the PL emission from an 800 nm long single QWR section [35]. Within 800 nm, only one exciton emission line can be observed, which means that, within 800 nm, there is no interface fluctuation, 800 nm is very close to an ideal one-dimensional QWR. This promising quantum structure provides a good chance to investigate true one-dimensional quantum phenomena.



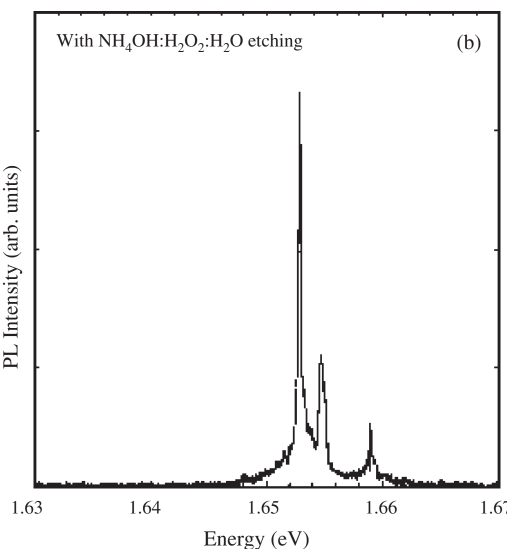


Figure 10. Micro-PL spectra of QWRs (a) without and (b) with interface improvement achieved by chemical etching. Reprinted with permission from [15], X.-L. Wang et al., *J. Cryst. Growth* 213, 19 (2000). © 2000, Elsevier Science.

3.2. Thermal Stability of Quantum Wires Compared with Quantum Well Structures

The QWR was predicted to show higher thermal stability because of the sharp peak of the DOS. However, the increased ratio of the interface to the volume enhanced

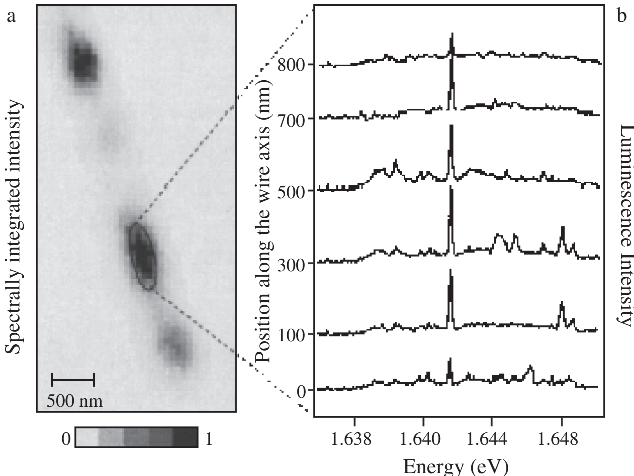


Figure 11. Real space image of PL intensity and the PL spectra corresponding to the long QWR section. Reprinted with permission from [35], A. Crottini et al., *Phys. Rev. B* 63, 121313 (2001). © 2001, American Physical Society.

the interface effects, and therefore decreased of the thermal stability of the QWR devices. Using a TBA source and special processing before the growth, the interface was tremendously improved, as discussed in Section 3.1. In the meantime, the nonradiative centers in the QWR structures are also dramatically reduced. Figure 12 shows the improvement of the thermal stability by using TBA as an arsenic source [36]. The intensity was increased about 1000 times at room temperature. Temperature-dependent lifetimes give further proof, as shown in Figure 13, in which the temperature of the maximum lifetime of the QWRs grown with TBA is about 100 K higher than that of those QWRs grown with arsine. The nonradiative centers are obviously reduced. We made a systematic comparison with QWL structures. Figure 14 shows the temperature-dependent PL intensities of 2, 5, and 9 nm single QWRs and single QWLs [37]. The PL intensities of a 9 nm QWR are larger than those of a 9 nm QWL at temperatures lower than 270 K. Our latest

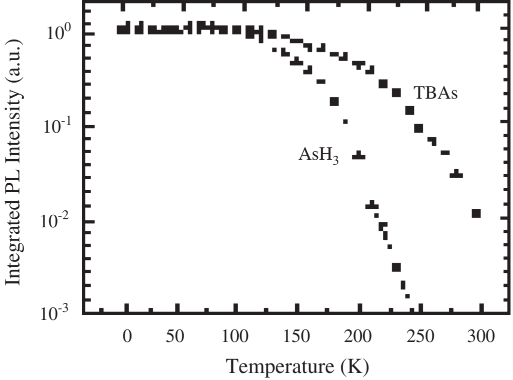


Figure 12. Temperature dependence of the integrated PL intensity of the TBAs (solid circle) and the AsH_3 (hollow triangle) samples, respectively. Reprinted with permission from [36], X. Q. Liu et al., *Appl. Phys. Lett.* 79, 1622 (2001). © 2001, American Institute of Physics.

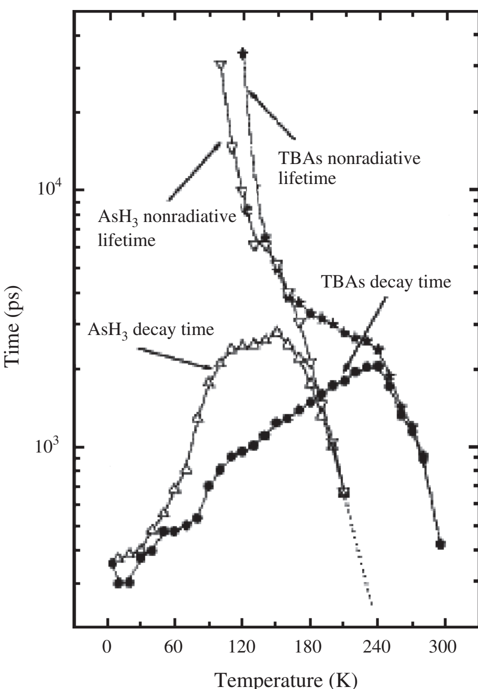


Figure 13. Temperature dependence of decay times and nonradiative lifetimes. The solid circle and star lines are the decay times and nonradiative lifetimes, respectively, of the TBA sample. The hollow up-triangle and down-triangle lines are the decay times and nonradiative lifetimes, respectively, of the AsH_3 sample. The dotted line is the extension of the nonradiative lifetime line to higher temperatures. Reprinted with permission from [36], X. Q. Liu et al., *Appl. Phys. Lett.* 79, 1622 (2001). © 2001, American Institute of Physics.

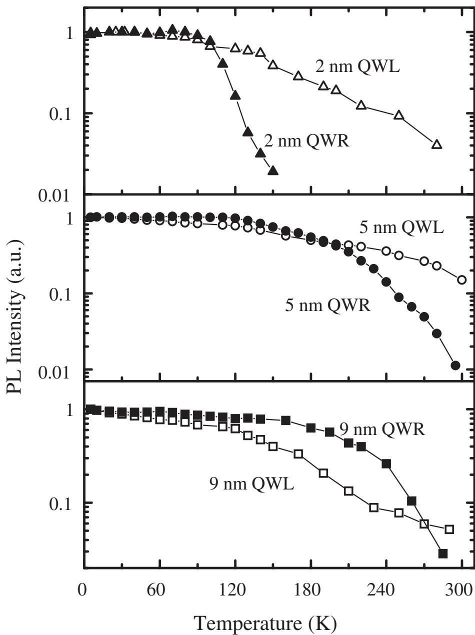


Figure 14. Comparison of the temperature-dependent integrated PL intensities of QWRs and QWLs with different thicknesses.

results show that stronger PL intensity of a QWR compared to that of a QWL structure is achieved at room temperature.

3.3. Piezoelectric Effects in High Quality Quantum Wires

Zinc-blende semiconductors are piezoelectric [38, 39]. Strain-induced piezoelectric effects have been shown to strongly affect the optical properties of quantum structures in strained III–V heterostructures [40–46]. Liu et al. demonstrated strong modification of optical properties by the tiny lattice-mismatch-induced piezoelectric field along the wire in ultrahigh-quality AlGaAs/GaAs V-groove QWR structures [47, 48]. In this study, a 5 nm thick $Al_{0.39}Ga_{0.61}As$/GaAs single QWR was grown using the FME method. To improve the quality of the interface, a short period superlattice buffer layer (AlGaAs (2.8 nm)/GaAs (1.4 nm) × 150) was grown before the growth of the AlGaAs barrier and the QWR structure. QWRs grown by this technique were proved to be of extremely high interface quality; the (001) center facets were composed of 1-D QWR sections with lengths of at least several 100 nm for one section, and the (113)A facets also became much smoother. Figure 15(a) illustrates TEM images of part of the superlattice buffer layer and the single QWR grown on it, (b) shows a schematic diagram of the axis, and (c) shows a schematic diagram of the potential profile along wires. The optical properties of the grown QWRs were characterized by photoluminescence, PL excitation (PLE), and time-resolved PL (TRPL) measurements. Figure 15(d) demonstrates the temperature-dependent PLE spectra and the 5 K PL spectrum. One striking feature shown in Figure 15 is that the PLE intensity of the ground state (e_1–hh_1) transition is almost zero at 5 K over an energy range as large as 20 meV. With increasing temperature, the e_1–hh_1 transition appears gradually, and clear e_1–hh_1 transition peaks can be observed when the temperature is higher than 100 K. The zero PLE intensity at low temperatures strongly suggests complete spatial separation of electron and hole wavefunctions, which might be caused by an internal electric field effect. With the temperature increased, the piezoelectric field is screened, and the PLE peak can be observed clearly.

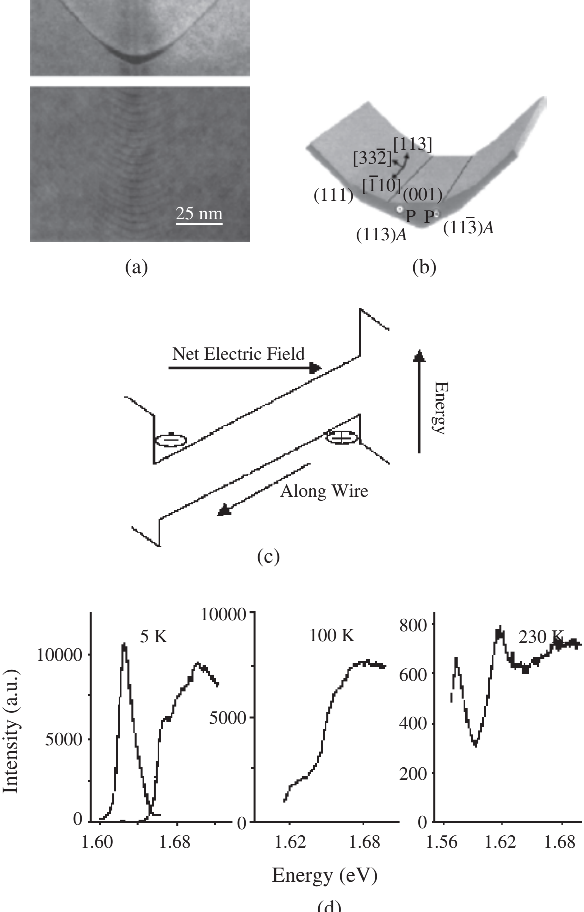


Figure 15. (a) Cross-sectional TEM images of the superlattice buffer and the QWR. (b) Schematic diagram of axis. (c) Schematic of potential profile along wires. (d) Temperature-dependent PLE spectra and PL spectrum at 5 K. Reprinted with permission from [47], X. Q. Liu et al., *Appl. Phys. Lett.* 80, 1894 (2002). © 2002, American Institute of Physics.

3.4. Disordered Quantum Wires Superlattice Structures

In the last decade, much attention has been focused [49–57] on the two-dimensional disordered superlattice (d-SL) proposed by Sasaki et al. [58]. A d-SL is formed by disordered distribution of layer thickness and/or confinement potential, which causes carrier localization along one direction (growth direction). Superior optoelectronic properties resulting from the effects of strong carrier localization have been shown in d-SL as compared with ordered SL (o-SL), such as enhanced photoluminescence, particularly at higher temperatures, and slow PL decay with increasing temperature. Further stronger carrier localization, and thus further stronger PL intensities, are expected to result from a lower dimensional disordered structure, for example, a disordered quantum wire SL (d-QWR-SL) and a disordered quantum dot SL. To realize the lower dimensional disordered structures, accurate control of the size and multilayer potential profile, that is, energy bandgap engineering (EBE) in more than one direction, is required. The V-groove QWR is fabricated by direct epitaxial growth on a V-grooved substrate, during which the EBE can easily be realized in the growth and lateral directions. Liu and co-workers reported the first successful fabrication of the d-QWR-SL structures and the observation of significantly high thermal PL stability [59, 60]. PL properties and the carrier lifetimes in comparison with those of an ordered QWR-SL (o-QWR-SL) and a single QWR structure are systematically investigated. The results demonstrated superthermal PL stability of the d-QWR-SL structures, which probably comes from carrier localization phenomena resulting from artificial disorder effects. The samples were grown on a 4 μm pitch V-grooved substrate by means of MOVPE using the FME technique. The layer structures of the o-QWR-SL sample consist of a 200 nm thick GaAs buffer layer, a 900 nm thick AlGaAs barrier layer, a 32-period GaAs/AlGaAs SL layer, and a 200 nm thick AlGaAs barrier layer. The Al

composition of the AlGaAs barrier is different for different crystalline facets, which are about 0.425 and 0.336 for (111)A sidewall facets and the vertical quantum well (VQWL) formed at the V-groove center, respectively. All layers were grown at 630 °C. The nominal AlGaAs barrier thickness is 12 monolayers (MLs), and that of GaAs wire layer is 6 MLs. The d-QWR-SL structures consist of nominal 3(W_1), 6(W_2), and 9(W_3) MLs for wire layers and 7(B_1), 12(B_2), 17(B_3) MLs for barrier layers. Individual thicknesses appear randomly, but with equal probability. The average thicknesses of the wire and barrier in the d-QWR-SL structures are same as those in the o-QWR-SL. The sequence of the different thickness layers in d-QWR-SL sample is as follows:

$$W_3B_1W_1B_3W_1B_3W_2B_2W_1B_2W_2B_1W_2B_1W_3B_3W_2B_3W_3B_2 \\ \times W_3B_3W_1B_1W_1B_3W_2B_3W_1B_1W_2B_2W_2B_3W_3B_1W_2B_2 \\ \times W_3B_1W_3B_1W_1B_3W_3B_3W_1B_2W_1B_2W_2B_1W_1B_1W_2B_3 \\ \times W_2B_3W_3B_2W_3B_2W_1B_2.$$

The cross section was checked by TEM. Figure 14 shows the TEM image graphs. Figure 16(a) is of the o-QWR-SL, and Figure 16(b) is of the d-QWR-SL. The integrated PL intensity is shown in Figure 17, which demonstrates a much worse thermal stability of the o-QWR-SL than that of the d-QWR-SL. The thermal stabilities of single QWRs with thicknesses of W_1 and W_3 were also confirmed to be worse than that of the d-QWR-SL sample. Therefore, the thermal PL stability of the d-QWR-SL is indeed much better than those of a single QWR with a thick barrier or o-QWR-SL structures.

The carrier lifetime measurements confirmed the above PL results. Figure 16 shows the temperature-dependent lifetimes of the o-QWR-SL and the d-QWR-SL. For comparison, the lifetime of the single QWR is also shown in Figure 18. For the d-QWR-SL sample, the lifetimes of the low (labeled d-QWR-SL) and the high (labeled d-QWR-S_1) energy transitions reach maximums of 3.20 and 2.64 ns at about 170 and 180 K, respectively, and then decrease continuously. These temperatures are consistent with those of

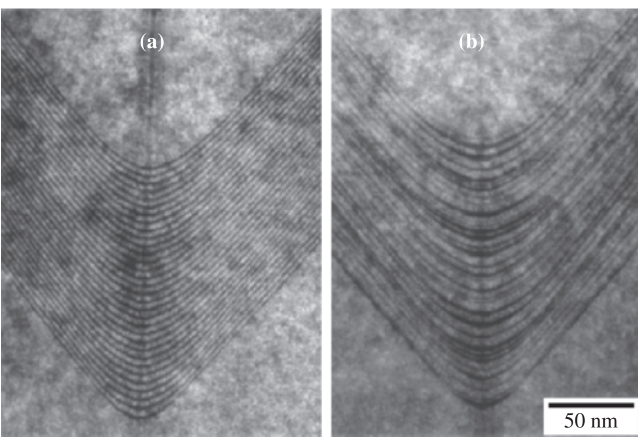


Figure 16. TEM images of cross section of (a) o-QWR-SL, (b) d-QWR-SL. Reprinted with permission from [59], X. Q. Liu et al., *J. Appl. Phys.* 90, 6363 (2001). © 2001, American Institute of Physics.

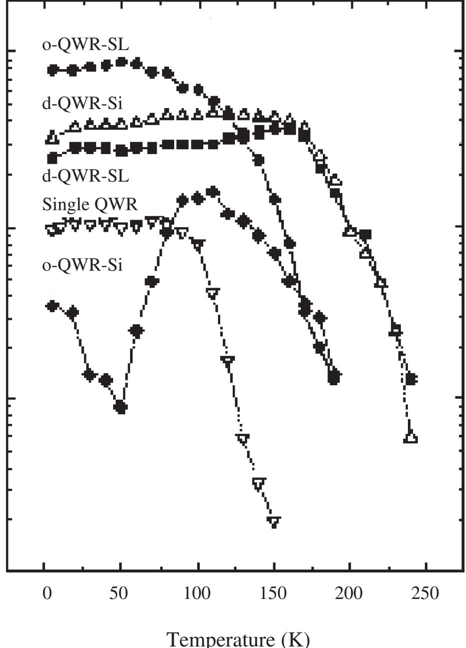


Figure 17. PL integrated intensity of the d-QWR-SL and the o-QWR-SL structures. Reprinted with permission from [59], X. Q. Liu et al., *J. Appl. Phys.* 90, 6363 (2001). © 2001, American Institute of Physics.

the PL intensity maxima of the d-QWR-SL. On the other hand, for the o-QWR-SL sample, the lifetimes of the ground and the excited state transitions reach maxima at about 110 K, which is much lower than that of the d-QWR-SL, and then decrease continuously. In the temperature range from 60 to 110 K, in which the radiative recombination

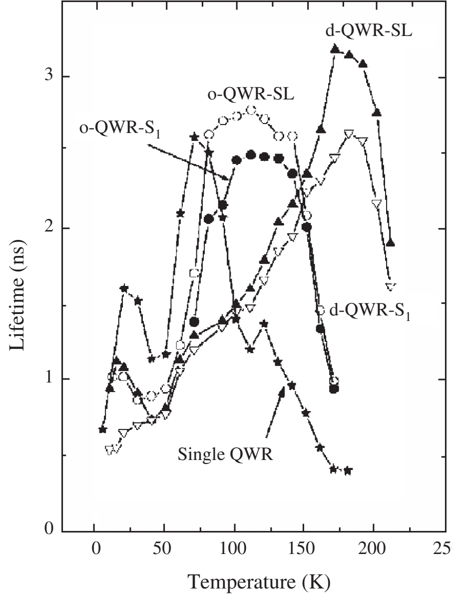


Figure 18. Temperature-dependent carrier lifetimes of the d-QWR-SL and o-QWR-SL structures. Reprinted with permission from [59], X. Q. Liu et al., *J. Appl. Phys.* 90, 6363 (2001). © 2001, American Institute of Physics.

is dominant for both the o-QWR-SL and the d-QWR-SL, the lifetimes of the d-QWR-SL are obviously shorter than those of the o-QWR-SL, while for the single QWR sample, the lifetime reaches a maximum of about 2.6 ns at about 70 K before decreasing, which is a temperature much lower than that of the d-QWR-SL. It is evidence of the contribution of the nonradiative recombination as the wavefunction penetrates into the barriers. These results demonstrate again the superior thermal stability of the PL of the d-QWR-SL.

4. QUANTUM WIRES DEVICES

4.1. Quantum Wires Laser

The interest in QWR laser structures is mainly focused on the predicted peaked density of states due to the lateral quantum confinement experienced by the charge carriers in such low-dimensional systems [1]. An extremely low threshold current is expected compared with QWL laser structures. However, the expectation for improved performance with such lower dimensional structures is difficult to realize because of fabrication difficulties. Kapon and co-workers reported lasing behavior in their pioneering papers [3, 5]; however, the laser emission is from excited states at low temperature. With the improvement of the QWR structures including the improved interfaces and the enhanced lateral confinement, the ground state QWR laser was demonstrated at room temperature [61, 62]. Figure 19 shows typical light output versus current characteristics (a) and stimulated emission spectra at various power levels (b) taken for the 600 nm long uncoated laser. Excitonic lasing behavior in a V-groove QWR was also reported at 10 K by optical pumping arising from a population inversion of localized excitons within the inhomogeneously broadened luminescence line [63].

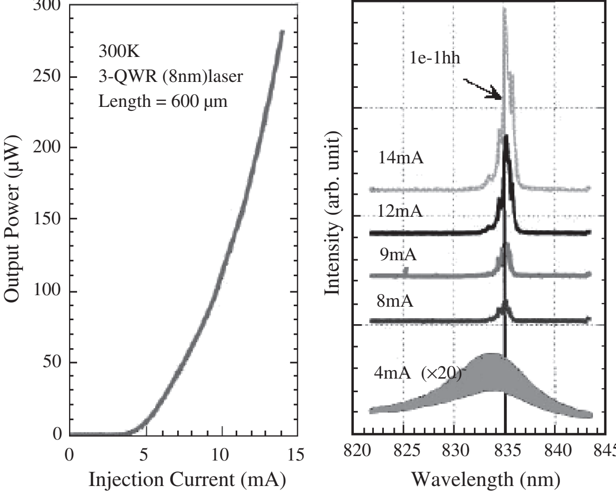


Figure 19. (a) Light output versus current characteristic. (b) Stimulated emission spectra at various power levels taken for a 600 nm long uncoated laser. Reprinted with permission from [61], T. G. Kim et al., *Phys. E* 7, 508 (2000). © 2000, Elsevier.

4.2. Other Kinds of Quantum Wires Devices

The success of QWL infrared photodetectors [64–67] stimulated further interests in lower dimensional infrared quantum structures, such as photodetectors, including QWR and QD photodetectors, in which the photoabsorption will be enhanced, and also the normal incidence absorption can be realized. QD infrared photodetectors have been intensively investigated [68–70]; however, for QWR infrared photodetectors, it is not the case because of the difficulties of fabrication of the QWR structures. Liu and co-workers [71, 72] demonstrated a QWR photodetector device; however, the performances still need to be improved.

Another promising application of QWR structures is the single QWR FET [73–76], which can realize ballistic transportation. Sugaya et al. made an extensive study of the QWR FET, in which the QWR is a trench type grown by using MBE. The novel properties and promising prospects of the QWR FET are attracting great attention. The obvious improvement of the interfaces of QWR structures will speed up the application in electronic devices.

5. SUMMARY

In summary, we made a review of the growth, optical properties, and devices of semiconductor quantum wires. Different types of growth techniques and the structures associated with quantum confinement were reviewed. The optical properties were discussed in detail, followed by several kinds of devices.

GLOSSARY

Detector A kind of device, which can detect optical radiation, or other kind of signals.

Field effect transistor (FET) A kind of electronic device, which is used as electrical amplifier or switch.

Laser Light amplification by stimulated emission of radiation. A kind of optical device, which can emit light of ultrahigh density. The emitting media can be gas or solid.

Localization The carriers (holes or electrons) confined in a structure and the structure dimension is smaller than carrier mean-free length.

Metal organic chemical vapor deposition (MOCVD) A promising method to get high-quality semiconductor quantum structures for electronic devices.

Molecular beam epitaxy (MBE) An advanced method to grow semiconductor quantum structures. It can achieve mono-atomic layer control.

Nanometer One nanometer is 10^{-9} m.

Quantum well A structure that can confine the electrons or holes in a very thin layer. The electrons or holes will be free in two dimensions in the well plane. The well layer is generally around 10^{-9} to 10^{-7} m.

Quantum wire A structure that can confine electrons or holes in a very narrow line. The electrons or holes will be free along the line. The dimension of the line is generally about 10^{-9} to 10^{-7} m.

Superlattice When the confinement barrier is thin, and multiple quantum well stacked together, the carriers states in different quantum wells will couple with each other and form a new kind of quantum structure.

Transmission electron microscope (TEM) A method making use the electron beam to transmit a thin solid sample and can show up to atomic level image.

REFERENCES

1. E. Kapon, *Semicond. Semimet.* 40, 259 (1994).
2. P. S. Zory, Jr., "Quantum Well Lasers." Academic, San Diego, CA, 1993.
3. E. Kapon, M. C. Tamargo, and D. M. Hwang, *Appl. Phys. Lett.* 50, 347 (1987).
4. R. Bhat, E. Kapon, D. M. Hwang, M. A. Koza, and C. P. Yun, *J. Cryst. Growth* 93, 850 (1988).
5. E. Kapon, D. M. Hwang, and R. Bhat, *Phys. Rev. Lett.* 63, 430 (1989).
6. C. Constantin, E. Martinet, F. Lelarge, K. Leifer, A. Rudra, and E. Kapon, *J. Appl. Phys.* 88, 141 (2000).
7. X. L. Wang, M. Ogura, and H. Matsuhata, *J. Cryst. Growth* 195, 586 (1998).
8. A. Gustafsson and E. Kapon, *Phys. Rev. Lett.* 81, 2962 (1998).
9. G. Biasiol, A. Gustafsson, K. Leifer, and E. Kapon, *Phys. Rev. B* 65, 205306 (2001).
10. X. Q. Liu, W. Lu, X. S. Chen, S. C. Shen, H. H. Tan, S. Yuan, C. Jagadish, M. B. Johnston, L. V. Dao, M. Gal, J. Zou, and D. J. H. Cockayne, *J. Appl. Phys.* 87, 1566 (2000).
11. Y. Kim, S. Yuan, R. Leon, C. Jagadish, M. Gal, M. B. Johnston, M. R. Phillips, M. A. S. Kalceff, J. Zou, and D. J. H. Cockayne, *J. Appl. Phys.* 80, 5014 (1996).
12. Y. Fu, M. Willander, W. Lu, X. Q. Liu, S. C. Shen, C. Jagadish, M. Gal, J. Zou, and D. J. H. Cockayne, *Phys. Rev. B* 61, 8306 (2000).
13. X.-L. Wang, M. Ogura, and H. Matsuhata, *Appl. Phys. Lett.* 67, 3629 (1995).
14. X.-L. Wang, M. Ogura, and H. Matsuhata, *Appl. Phys. Lett.* 67, 1506 (1995).
15. X.-L. Wang, V. Voliotis, R. Grousson, and M. Ogura, *J. Cryst. Growth* 213, 19 (2000).
16. X. L. Wang and M. Ogura, *J. Cryst. Growth* 221, 556 (2000).
17. Y. C. Chang, L. L. Chang, and L. Esaki, *Appl. Phys. Lett.* 47, 1324 (1985).
18. L. N. Pfeiffer, K. W. West, H. L. Störmer, J. P. Eisenstein, K. W. Baldwin, D. Gershoni, and J. Spectror, *Appl. Phys. Lett.* 56, 1697 (1990).
19. L. N. Pfeiffer, H. L. Störmer, K. W. Baldwin, K. W. West, A. R. Goni, A. Pinczuk, R. C. Ashoori, M. M. Dignam, and W. Wegscheider, *J. Cryst. Growth* 127, 849 (1993).
20. M. Yoshita, H. Akiyama, T. Someya, and H. Sakaki, *J. Appl. Phys.* 83, 3777 (1998); T. Someya, H. Akiyama, and H. Sakaki, *J. Appl. Phys.* 79, 2522 (1996).
21. J. Hasen, L. N. Pfeiffer, A. Pinczuk, S. He, K. W. West, and B. S. Denniset, *Nature* 390, 54 (1997).
22. H. Akiyama, T. Someya, and H. Sakaki, *Phys. Rev. B* 53, R4229 (1996).
23. H. Akiyama, S. Koshiba, T. Someya, K. Wada, H. Noge, Y. Nakamura, T. Inoshita, A. Shimizu, and H. Sakaki, *Phys. Rev. Lett.* 72, 2123 (1994).
24. R. Nötzel, J. Menniger, M. Ramsteiner, A. Ruiz, H.-P. Schönherr, and K. H. Ploog, *Appl. Phys. Lett.* 68, 1132 (1996).
25. R. Nötzel, M. Ramsteiner, J. Menniger, A. Trampert, H.-P. Schönherr, L. Däweritz, and K. H. Ploog, *J. Appl. Phys.* 80, 4108 (1998).
26. J. Fricke, R. Nötzel, U. Jahn, Z. C. Niu, H.-P. Schönherr, M. Ramsteiner, and K. H. Ploog, *J. Appl. Phys.* 86, 2896 (1999).
27. K. H. Ploog and R. Nötzel, *Phys. E* 11, 78 (2001).
28. P. M. Petroff, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.* 45, 620 (1982).
29. T. Fukui and H. Saito, *Appl. Phys. Lett.* 50, 824 (1987).
30. M. S. Miller, H. Weiman, C. E. Pryor, M. Krishnamurthy, P. M. Petroff, H. Kroemer, and J. L. Merz, *Phys. Rev. Lett.* 68, 3463 (1992).
31. H. Weiman, M. S. Miller, C. E. Pryor, Y. J. Li, P. Bergman, P. M. Petroff, and J. L. Merz, *Phys. Rev. B* 48, 8047 (1992).
32. G. Mayer, B. E. Maile, R. Germann, A. Forchel, P. Grambow, and H. P. Meier, *Appl. Phys. Lett.* 56, 2016 (1990).
33. M. Kohl, D. Heitmann, W. W. Ruhle, P. Grambow, and K. H. Ploog, *Phys. Rev. B* 41, 12338 (1990).
34. M. Notomi, M. Naganuma, T. Nishida, T. Tamamura, H. Iwamura, S. Nojima, and M. Okamoto, *Appl. Phys. Lett.* 58, 720 (1991).
35. A. Crottini, J. L. Staehli, B. Deveaud, X.-L. Wang, and M. Ogura, *Phys. Rev. B* 63,121313 (2001).
36. X. Q. Liu, X. L. Wang, and M. Ogura, *Appl. Phys. Lett.* 79, 1622 (2001).
37. X. Q. Liu, X. L. Wang, and M. Ogura, *J. Appl. Phys.* Lett. (in press).
38. D. L. Smith and C. Mailhiot, *Phys. Rev. Lett.* 58, 1264 (1987).
39. L. D. Caro and L. Tapfer, *Phys. Rev. B* 48, 2298 (1993); L. D. Caro and L. Tapfer, *Phys. Rev. B* 49, 11127 (1994).
40. L. D. Caro and L. Tapfer, *J. Appl. Phys.* 79, 9188 (1996).
41. M. Ilg, K. H. Ploog, and A. Trampert, *Phys. Rev. B* 50, 17111 (1994).
42. R. Nötzel, M. Ramsteiner, Z. Niu, H. Schönherr, L. Däweritz, and K. H. Ploog, *Appl. Phys. Lett.* 70, 1578 (1997).
43. R. Rinaldi, R. Cingolani, L. Decaro, M. Lomascolo, M. DiDio, L. Tapfer, U. Marti, and F. K. Reinhart, *J. Opt. Soc. Am. B* 13, 1031 (1996).
44. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 50, 14187 (1994).
45. C. Constantin, E. Martinet, F. Lelarge, K. Leifer, A. Rudra, and E. Kapon, *J. Appl. Phys.* 88, 141 (2000).
46. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 50, 14187 (1994).
47. X. Q. Liu, X. L. Wang, M. Ogura, T. Guillet, V. Voliotis, and R. Grousson, *Appl. Phys. Lett.* 80, 1894 (2002).
48. X. Q. Liu, X. L. Wang, M. Ogura, T. Guillet, V. Voliotis, and R. Grousson, *Phys. E* 13, 194 (2002).
49. T. Yamamoto, M. Kasu, S. Noda, and A. Sasaki, *J. Appl. Phys.* 68, 5318 (1990).
50. M. Kasu, T. Yamamoto, S. Noda, and A. Sasaki, *Appl. Phys.* 59, 800 (1991).
51. X. L. Wang, A. Wakahara, and A. Sasaki, *Appl. Phys. Lett.* 62, 888 (1992).
52. K. Uno, S. Noda, and A. Sasaki, *J. Appl. Phys.* 77, 4693 (1995).
53. K. Uno. S. Noda, and A. Sasaki, *Jpn. J. Appl. Phys.* 35, 2566 (1996).
54. D. J. Arent, R. G. Alonso, G. S. Horner, D. Levi, M. Bode, A. Mascarenhas, J. M. Olson, X. Yin, M. C. Delong, A. J. Spring Thorpe, A. Majeed, D. J. Mowbray, and M. S. Skolnick, *Phys. Rev. B* 49, 11173 (1994).
55. K. A. Mäder, L. W. Wang, and A. Zunger, *Phys. Rev. Lett.* 74, 2555 (1995).
56. L. W. Wang, A. Zunger, and K. A. Mäder, *Phys. Rev. B* 53, 2010 (1996).
57. X. S. Chen and S. Xiong, *Phys. Rev. B* 48, 5273 (1993).
58. A. Sasaki, M. Kasu, T. Yamamoto, and S. Noda, *Jpn. J. Appl. Phys.* 28, L1249 (1989).
59. X. Q. Liu, A. Sasaki, N. Ohno, X. L. Wang, and M. Ogura, *J. Appl. Phys.* 90, 6363 (2001).
60. A. Sasaki, R. Okanishi, X. Q. Liu, X. L. Wang, and M. Ogura, *Appl. Phys. Lett.* 79, 1870 (2001).
61. T. G. Kim, X. L. Wang, R. Kaji, and M. Ogura, *Phys. E* 7, 508 (2000).
62. T. G. Kim, X.-L. Wang, Y. Suzuki, K. Komori, and M. Ogura, *IEEE J. Select. Topics Quantum Electron.* 6, 511 (2000).

63. L. Sirigu, D. Y. Oberli, L. Degiorgi, A. Rudra, and E. Kapon, *Phys. Rev. B* 61, 10575 (2000).
64. B. F. Levine, K. K. Choi, C. G. Bethea, J. Walker, and R. J. Malik, *Appl. Phys. Lett.* 51, 934 (1987).
65. B. F. Levine, *J. Appl. Phys.* 74, R1 (1993).
66. X. Q. Liu, N. Li, X. S. Chen, W. Lu, W. L. Xu, X. Z. Yuan, N. Li, S. C. Shen, S. Yuan, H. H. Tan, and C. Jagadish, *Jpn. J. Appl. Phys.* 38, 5044 (1999).
67. Y. Fu, S. Bagge, M. Gustavsson, A. Haglund, M. Willander, N. Li, N. Li, W. Lu, X. Q. Liu, X. Z. Yuan, Z. F. Li, H. F. Dou, and S. C. Shen, *Superlatt. Microstruct.* 29, 309 (2001).
68. Z. M. Ye, J. C. Campbell, Z. H. Chen, E. T. Kim, and A. Madhukar, *J. Appl. Phys.* 92, 4141 (2002).
69. S. F. Tang, S. Y. Lin, and S. C. Lee, *IEEE Trans. Electron Devices* 49, 1341 (2002).
70. L. Chu, A. Zrenner, M. Bichler, and G. Abstreiter, *Appl. Phys. Lett.* 79, 2249 (2001).
71. X. Q. Liu, Z. F. Li, N. Li, W. Lu, X. Z. Yuan, S. C. Shen, H. H. Tan, C. Jagadish, and J. Zou, *Jpn. J. Appl. Phys.* 39, 5124 (2000).
72. Y. Fu, M. Willander, X. Q. Liu, W. Lu, S. C. Shen, H. H. Tan, C. Jagadish, J. Zou, and D. J. H. Cockayne, *J. Appl. Phys.* 89, 2351 (2001).
73. T. Sugaya, M. Ogura, Y. Sugiyama, K. Matsumoto, K. Yonei, and K. Y. Jang, *Appl. Phys. Lett.* 78, 2369 (2001).
74. T. Sugaya, T. Takahashi, T. Nakagawa, M. Ogura, and Y. Sugiyama, *Electron. Lett.* 34, 926 (1998).
75. S. J. Kim, T. Sugaya, M. Ogura, Y. Sugiyama, and K. Tomizawa, *Phys. B* 272, 117 (1999).
76. T. Sugaya, S. J. Kim, N. Nakagawa, Y. Sugiyama, and M. Ogura, *J. Vac. Sci. Technol. B* 18, 1680 (2000).

www.aspbs.com/enn

SiGe/Si Heterostructures

C. W. Liu
National Taiwan University, Taipei, Taiwan, ROC

L. J. Chen
National Tsing Hua University, Hsinchu, Taiwan, ROC

CONTENTS

1. Introduction
2. Strain Effect
3. Device Performances
4. Fabrication Technologies
Glossary
References

1. INTRODUCTION

The Si/SiGe:C heterostructures add a rich variety to the monotonic Si world. The misfit between Si and Ge, which is the initial obstacle to fabricating such material systems, becomes an advantage in tailoring the electronic properties of Si-based devices. Si/SiGe heterojunction bipolar transistors have a cutoff frequency of 210 GHz [1] and a maximum oscillation frequency of 285 GHz [2] with carbon incorporation in the base. The figures of merit of metal-oxide-silicon-field-effect transistors (MOSFETs) are also enhanced to some extent with the use of the strained Si on the relaxed SiGe buffers for both N and P channel devices [3, 4]. The reported peak hole mobility is 2700 cm^2/Vs in the strained Ge channel [5], and the reported peak electron mobility is 800 cm^2/Vs in the strained Si channel [6]. Heterojunction devices such as resonant tunneling diodes [7–9] and modulation-doped FETs [10, 11], which were only possible with III–V semiconductors in the past, can now be made from SiGe/Si heterostructures. Table 1 gives a summary of SiGe device performance.

$Si_{1-x}Ge_x$ material is a solid solution of Si and Ge with an intrinsic lattice constant larger than that of Si. A simplified way to look at $Si_{1-x}Ge_x$ is that it is made of "$Si_{1-x}Ge_x$ atoms" in a diamond structure, and the virtual "$Si_{1-x}Ge_x$ atom" inherits the characteristics from both Si and Ge, with the atomic fraction as a weighting factor. In most cases, a linear weighting assumption is good for applications. For example, the relaxed lattice constant of a $Si_{1-x}Ge_x$ crystal is a liner sum of the lattice constants of Si crystal and Ge crystal. The virtual $Si_{1-x}Ge_x$ atom approximation ignores the random nature of $Si_{1-x}Ge_x$, where the traveling carriers in the $Si_{1-x}Ge_x$ crystal visit Ge atoms with the frequency of x and visit Si atoms with the frequency of $1 - x$, but in the virtual $Si_{1-x}Ge_x$ crystal, the carrier sees only one kind of atom, the $Si_{1-x}Ge_x$. This kind of random alloy effect has to be taken into account when randomness is important. For example, the alloy scattering allows nonphonon radiative recombination in $Si_{1-x}Ge_x$, which does not occur in Si and reduces the carrier mobility. For a good approximation, $Si_{1-x}Ge_x$ is a diamond structure occupied by virtual $Si_{1-x}Ge_x$ atoms, and randomness has to be taken into account by alloy scattering in certain circumstances.

2. STRAIN EFFECT

Si and Ge have the same diamond lattice structures and are completely miscible for a full range of composition, but have different lattice constants. The misfit is defined as the lattice constant difference between Si and Ge with respect to the Si lattice constant and is around 4%. For a linear assumption, the misfit between $Si_{1-x}Ge_x$ and Si is $\sim 0.04x$. For example, the misfit between $Si_{0.75}Ge_{0.25}$ and Si is 1%, that is, the 100 Si lattice units have the same length as the 99 $Si_{0.75}Ge_{0.25}$ units. For device applications such as heterojunction bipolar transistors (HBTs), the misfit dislocation at $Si/Si_{1-x}Ge_x$ interface has to be avoided, and the virtual SiGe atoms have to bond with the Si atoms one by one in the underlying Si substrates. The dislocation-free strained $Si_{1-x}Ge_x$ is also described as "pseudomorphic." The $Si_{1-x}Ge_x$ on the Si is under compressive strain, since the lattice constant of Si is smaller compared with $Si_{1-x}Ge_x$. The strain causes changes in the crystal structures and electronic structure of $Si_{1-x}Ge_x$.

2.1. Critical Thickness

Since the Si substrate is much thicker (~0.7 mm for 200-mm wafers), the original cubic $Si_{1-x}Ge_x$ lattice with a lattice constant larger than that of Si has to be distorted into the

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (775–792)

Table 1. The achieved performance of electronic and optoelectronic devices based on SiGe technology.

Device	Figure of merit	Growth technique
Strained NMOSFET	Peak effective electron mobility 800 cm^2/Vs [6]	Relaxed buffer
Strained PMOSFET	Peak effective hole mobility 2700 cm^2/Vs [5]	Relaxed buffer
NPN-HBT	$f_T = 210$ GHz [1], $f_{max} = 285$ GHz [2]	Pseudomorphic
PNP-HBT	$f_T = 59$ GHz at RT, $f_T = 61$ GHz at 85 K [12]	Pseudomorphic
Heterojunction phototransistors	1.47 A/W at 850 nm, bandwidth = 1.25 GHz [13]	Pseudomorphic
NMODFET	$G_m = 460$ ms/mm $f_T = 76$ GHz, $f_{max} = 107$ GHz at RT [14] $G_m = 730$ ms/mm $f_T = 105$ GHz, $f_{max} = 170$ GHz at 50 K	Relaxed buffer
PMODFET	$G_m = 300$ ms/mm $f_T = 70$ GHz, $f_{max} = 135$ GHz at RT [15]	Relaxed buffer
NRTD	P/V ≥ 7.6 at RT [7], 2 at 4.2 K [8]	Relaxed buffer
PRTD	P/V = 1.8 at RT, 2.2 at 4.2 K [9]	Pseudomorphic
Detector (IR)	1.3 μm, 1.5 μm [16]	Quantum dots
Detector (LWIR)	Schottky barrier: $\lambda_c = 10$ μm [17]	Pseudomorphic
LED	1.3 μm, 1.5 μm at RT [18, 19]	Pseudomorphic, quantum dots

Note: The last column indicates the growth techniques used to obtain device-quality material.

tetragonal symmetry with the built-up strain energy. The in-plane lattice constant of $Si_{1-x}Ge_x$ is reduced to that of Si, and the vertical lattice constant increases by 80% of the in-plane lattice constant reduction. $Si_{1-x}Ge_x$ under stress may yield to form dislocations if the compressive strain energy is high enough. Strained $Si_{1-x}Ge_x$ with dislocations grown on Si is shown schematically in Figure 1. The thickness of strained $Si_{1-x}Ge_x$ has to be thin enough to prevent the dislocation formation. The maximum thickness to have a defect-free $Si_{1-x}Ge_x$ epilayer is the so-called critical thickness. To minimize the sum of the strain energy and dislocation energy, the thermal equilibrium critical thickness (h_c) [20] can be derived as

$$x = 0.55 \ln(10h_c)/h_c \quad (1)$$

where h_c is in nanometers. For example, the critical thickness is ~13 nm and ~1.5 nm for strained $Si_{0.8}Ge_{0.2}$ and Ge on Si, respectively. This value is too small to be used in many applications. The meta-stable growth away from thermal equilibrium is desired to increase the critical thickness. The selective epitaxial growth [21] and the low temperature epi-growth are common meta-stable growth methods. The enhancement by a factor of at least 4 can be obtained for x less than ~0.5, but it really depends on the details of the growth conditions. Generally speaking, the smaller area of growth and the lower temperature of growth can have a larger meta-stable critical thickness. Most $Si_{1-x}Ge_x$ reactors have a growth temperature of 550–650 °C. However, metal-stable films will generate the defect if the thermal budget of the further process is too high. This is a challenge in the integration of $Si_{1-x}Ge_x$ into a CMOS process.

In ultra-large scale integrated circuits (ULSIs), the building devices are N-channel or P-channel MOSFETs. Strained channels can improve the device performance of MOSFET. $Si_{1-x}Ge_x$ grown on Si can be relaxed if it is much thicker than the critical thickness and/or is annealed at high temperature. Relaxed $Si_{1-x}Ge_x$ is not good enough for devices with vertical carrier transportation such as HBTs, since both the threading dislocations and misfit dislocations at the

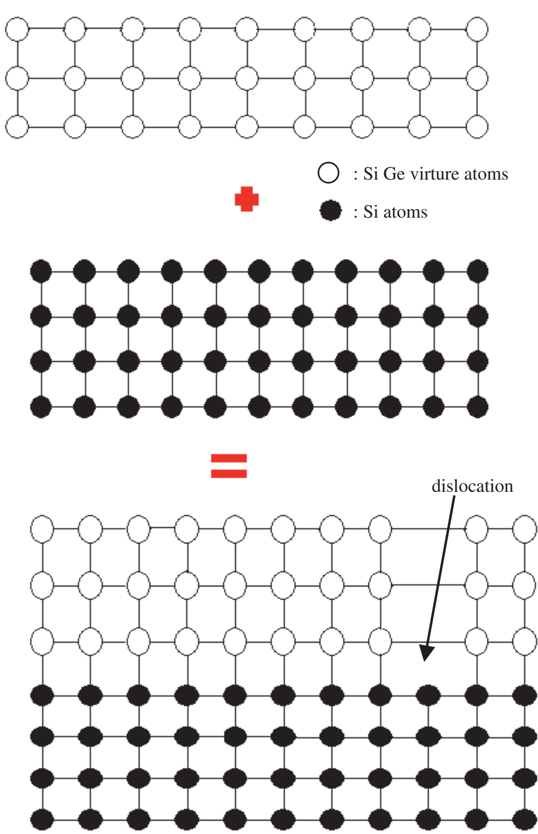


Figure 1. The strained SiGe grown on (100) Si. The misfit dislocation is the missing bond between the Si and SiGe atom.

$Si/Si_{1-x}Ge_x$ interface can act as recombination centers. The Si or Ge grown on relaxed $Si_{1-x}Ge_x$ is under tensile and compressive strain, respectively. The defects in the strained Si or Ge layer are the threading dislocation as shown in Figure 2, and most misfit network exists in the relaxed buffer layers. The threading dislocations in the strained Si and Ge channels have a small effect on in-plane carrier transport, especially for small feature devices, if the density is low enough (below 10^5 cm^{-2}). The strained Si and Ge channels increase the carrier mobility, and thus the device speed is enhanced in the same MOSFET technology node. For the Si or Ge grown on relaxed $Si_{1-x}Ge_x$ layers, the same critical thickness constraint is also applied. Note that the relaxed $Si_{1-x}Ge_x$ can be obtained with graded relaxed buffers [22], silicon on insulators (SOI) [23], and regrowth after smart cut [24]. Equation (1) can be used to estimate the critical thickness of strained Si or Ge, but x is the Ge fraction difference between strained Si (Ge) and relaxed $Si_{1-x}Ge_x$ buffers.

Because of the small lattice constant of diamond (made of carbon), the incorporation of carbon into SiGe on Si can compensate for the compressive strain from the large Ge atoms. A strain-free SiGeC film can be obtained if the ratio of the Ge fraction to the C fraction is ~11 (compensation ratio), obtained from the linear interpolation of the lattice constant between Ge and silicon carbide. Note that the silicon carbide is a compound with the same number of Si and C atoms in the face center cubic sublattices. However, only up to a few percent (~2.5%) of carbon can be incorporated into SiGe or Si [25, 26]. Note that Si and Ge can be mixed at any concentration.

$Si_{1-y}C_y$ and $Si_{1-x-y}Ge_xC_y$ alloys with tensile strain can be fabricated by incorporating carbon into Si and incorporating carbon content above the compensation ratio into $Si_{1-x}Ge_x$, respectively. In principle, the critical thickness concepts can still be applied to alloys with tensile strain. However, relaxation mechanisms in addition to the misfit dislocations, such as silicon carbide precipitation, and stacking fault formation, can also occur for thick and high-temperature annealed $Si_{1-y}C_y$ and $Si_{1-x-y}Ge_xC_y$ films.

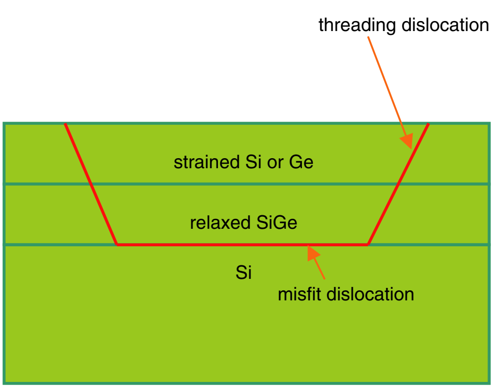


Figure 2. The misfit dislocation and threading dislocation. The misfit dislocation is at the Si/SiGe interface, and threading dislocation ends at the surface of the epilayers.

2.2. Growth Mode, Self-assembled Quantum Dots, and SiO_2 Dots

For $Si_{1-x}Ge_x$ layers thicker than the critical thickness, the relaxation is mainly apportioned by misfit dislocation generation at low Ge fraction (< ~0.4), although the morphology roughening can also lower the elastic energy [27]. For the high Ge fraction, the roughening is the dominant mechanism for relaxation, since the roughening barrier is sufficiently small in energy for most growth conditions. The roughening can be represented by the Stranski-Krastanow growth model. After the two-dimensional layer-by-layer growth reaches some critical thickness, the growth becomes three-dimensional island growth. Note that the critical thickness here can differ from the thermal equilibrium critical thickness (Section 2.1). The two-dimensional layer is also called the "wetting layer," and the islands are called "dots" or "quantum dots," depending on the size. The three-dimensional islands remain dislocation-free, but the top layers of the islands start to relax. This relaxation can stop the build-up of strain energy in the Ge islands. Further growth can cause the islands to coalesce, and misfit dislocations can occur.

Because of the different strain conditions on Ge layers, that is, the strained wetting layer and relaxed islands, the Si layer grown on the Ge dots has a tensile strain, whereas the Si layer on the Ge wetting layer is strain-free. If this Si layer is thinner than the correlation length [28], the strain condition remains, and the further Ge deposited on the Si layer (spacer) can vertically align with the underside of the Ge layers. The Ge dots of individual Ge layers are vertically self-assembled in a row in the same position [29] (Fig. 3). Ge quantum dots can also be fabricated with fine structures such as Ge/Si/Ge composite dots (Fig. 4) to improve uniformity, dot area coverage (Fig. 5), and photoluminescence efficiency. These Ge dot structures are used for photodetectors at wavelengths of 1.3, 1.5, and 10 μm (see Section 3.3.2).

To reduce the thermal budget of $Si_{1-x}Ge_x$-based devices, a low-temperature oxide is often used. Liquid-phase deposition can provide a low-temperature oxide on $Si_{1-x}Ge_x$. It is interesting that silicon dioxide dots can be deposited on the Si cap layers of self-assembled Ge dots with a liquid phase deposition method. The Si capping layer directly above the Ge dots has a tensile strain, whereas the Si cap on the wetting layer is not strained. The tensile strain can enhance the silicon dioxide nucleation and deposition on the Si surface, and SiO_2 dots are formed directly above the Ge dots with the SiO_2 wetting layers between the dots [29] (Fig. 3).

2.3. Electronic Structures

Because of the tetragonal distortion of strained $Si_{1-x}Ge_x$ grown on relaxed (001) $Si_{1-y}Ge_y$ layers, the energy bands of electrons and holes change according to the strain condition. For the compressive strain, the originally sixfold degenerate conduction band valleys split into a low-energy group of four valleys perpendicular to the growth direction and a high-energy group of two valleys along the growth direction. Similarly, the originally degenerate light and heavy hole bands split into the high-energy heavy-hole band and the low-energy light-hole band. For tensile strain such as that of

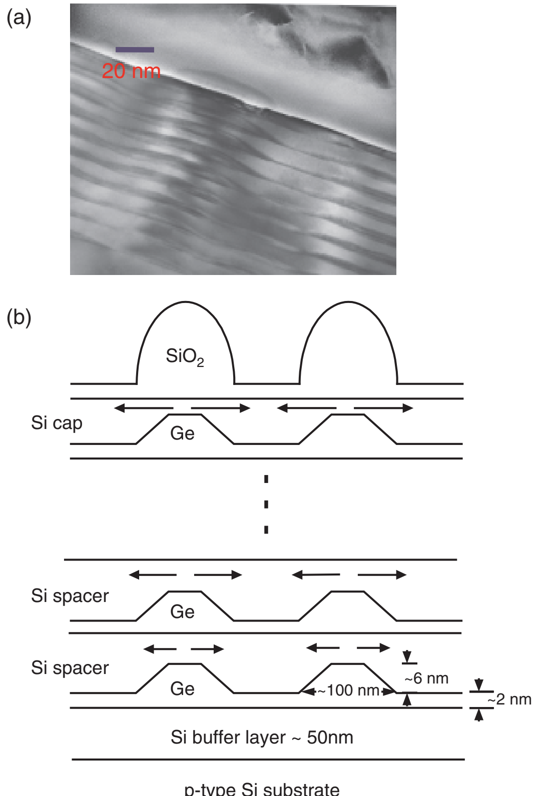


Figure 3. (a) TEM micrograph of multilayer Ge dots. The spacer thickness is ~20 nm. The SiO_2 dots are also formed above the Ge dots. (b) Schematic diagram showing the strain field in Si layers. Adapted with permission from [29], C. W. Liu et al., *Appl. Phys. Lett.* (2003). © 2003, American Institute of Physics.

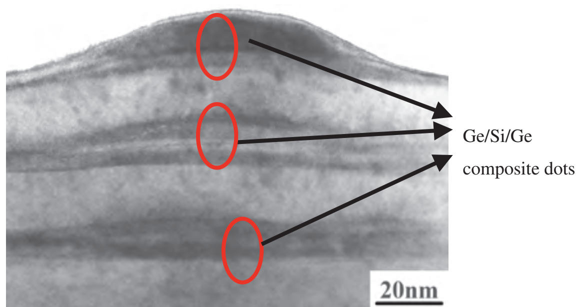


Figure 4. Composite dots with fine structures such as Ge/Si/Ge composite dots.

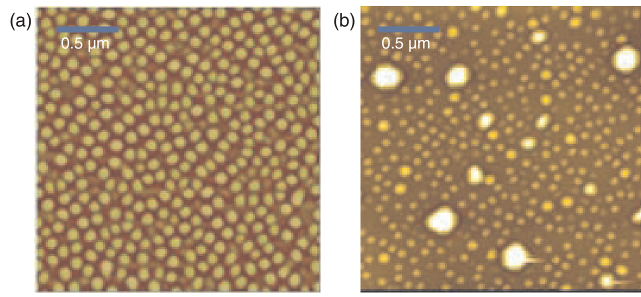


Figure 5. AFM picture of (a) composite dots and (b) conventional dots. The composite dots have a larger coverage area.

strained Si on relaxed $Si_{1-x}Ge_x$, the results are similar, but the energy levels switch between the degenerate groups. The two conduction valleys along the growth direction are lower in energy than the other four valleys in the strained Si. So are the valence bands. The effects are sometimes referred to as "uniaxial splitting."

The strained $Si_{1-x}Ge_x$ on Si has a smaller lattice constant in the two in-plane directions, and larger lattice constants in the growth direction as compared with originally relaxed $Si_{1-x}Ge_x$. The net volume decrease (hydrostatic term) of the tetragonal $Si_{1-x}Ge_x$ unit cell grown on Si leads to a shrinkage of the bandgap, that is, the valence band moves up and the conduction band moves down. Note that the strained Si on relaxed $Si_{1-x}Ge_x$ has the opposite effect, and the net volume increases. The bandgap of strained $Si_{1-x}Ge_x$ on Si is given in Figure 6 with the relaxed bandgap, and the optoelectronic devices for optical communication (1.3 and 1.5 μm wavelength) can be easily implemented by the strained $Si_{1-x}Ge_x$ bandgap. At low Ge concentration, the bandgap reduction is ~7.5 meV per % Ge for strained $Si_{1-x}Ge_x$ grown on Si. The conduction band discontinuity between strained $Si_{1-x}Ge_x$ and Si is often negligible. For strained Si grown on relaxed $Si_{1-x}Ge_x$ buffers, the bandgap reduction of Si is ~4 meV per % Ge. The smaller bandgap reduction of strained Si on relaxed $Si_{1-x}Ge_x$ as compared with the strained $Si_{1-x}Ge_x$ on Si is due to the fact that the tetragonal distorted lattice of strained Si has a net volume expansion to widen the conduction band edge and valence band edge. The conduction band edge of strained Si is lower than that of relaxed $Si_{1-x}Ge_x$ ($\Delta E_c \sim 6$ meV/%), and the electron confinement is possible in the relaxed $Si_{1-x}Ge_x$/strained-Si structures. The modulation doped devices can thus be made (Section 3.2.3). The bandgap of strained $Si_{1-y}C_y$ on Si is also reduced with respect to Si with the experimental value of 68 meV/% C [30].

There are two effects for carbon incorporation into $Si_{1-x}Ge_x$ on the bandgap:

1. Carbon can change the strain condition of the pseudomorphic film, and this effect on band structure can

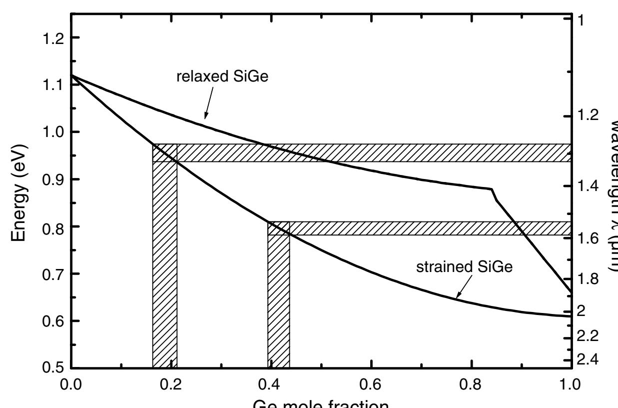


Figure 6. The bandgap of strained $Si_{1-x}Ge_x$ on a (001) Si substrate and relaxed bandgap of $Si_{1-x}Ge_x$ alloys. The shadow areas indicate the optical communication wavelength of 1.3 and 1.5 μm. Reprinted with permission from [16], B.-C. Hsu et al., "International Electron Device Meeting" 2002, pp. 91–94. © 2002, IEEE.

be taken account of in a manner similar to that of Ge incorporation.

2. Carbon also changes the fundamental bandgap [25, 31], and the relaxed $Si_{1-y}C_y$ bandgap decreases with carbon concentration initially because of the local lattice disorder, despite the large bandgap of diamond.

2.4. Mobility

2.4.1. Mobility in Relaxed and Strained $Si_{1-x}Ge_x$ alloys

The theoretical calculation [32] and experimental [33] values of the majority hole mobility in relaxed $Si_{1-x}Ge_x$ alloys are plotted in Figure 7. The decrease in mobility with Ge concentration is due to the alloy scattering, and an alloy scattering potential of 0.7 eV is used in the calculation to fit the measurement data.

Because of the impurity scattering, the mobility depends on the doping concentration. The in-plane and out-of-plane majority hole mobilities in strained $Si_{1-x}Ge_x$ layers for different doping concentrations are shown in Figure 8 [34]. Note that the hole mobilities in n-type material and p-type material are different from the terms of majority hole mobility and minority hole mobility, respectively. So are electron mobilities.

In the calculation, the alloy scattering potential determines the mobility value [37]. The majority hole mobilities in the strained $Si_{1-x}Ge_x$ alloys based on Monte Carlo simulation decrease by a factor of 2–3 if an alloy potential of 1.4 eV is used, as compared with 0.7 eV up to the doping concentration of $10^{19}cm^{-3}$. The hole mobility of strained $Si_{1-x}Ge_x$ calculated by different groups [31, 35, 36] is qualitatively similar. For a doping concentration of $\geq 10^{19}cm^{-3}$, the mobility depends weakly on Ge content, and the alloy scattering is not important. The majority hole mobility is important for the p-channel strained $Si_{1-x}Ge_x$ FET, where the current is made of the majority hole current.

Both in-plane and out-of-plane minority electron mobilities at different doping concentrations are plotted in Figure 9 [37, 38]. The out-of-plane mobility is especially

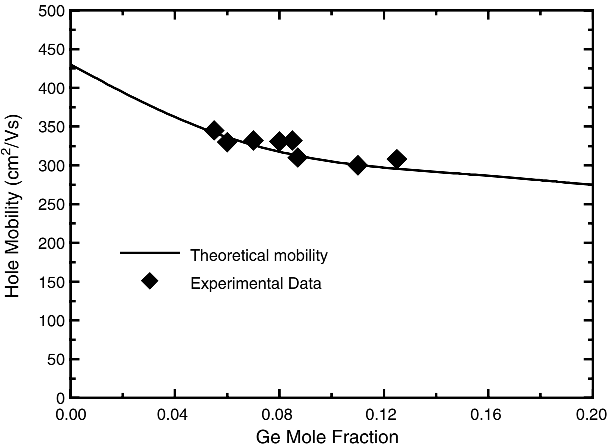


Figure 7. Theoretical values of the hole mobility in relaxed SiGe alloys at 300 K and a doping concentration of $N_A = 10^{16}$ cm^{-3}, with experimental results of Gaworzewski et al. [33].

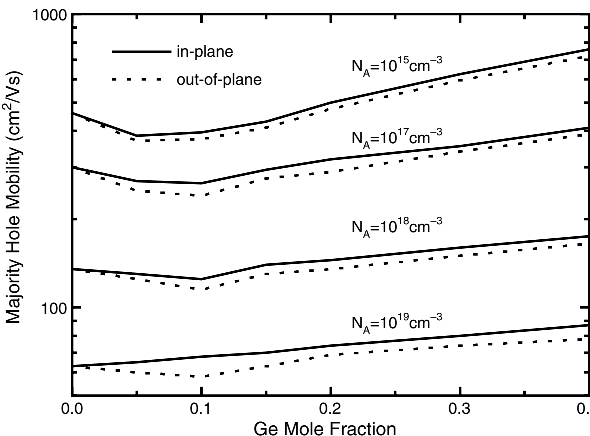


Figure 8. The in-plane and out-of-plane majority hole mobility of SiGe strained alloys for different doping concentrations.

important for npn Si/SiGe heterojunction bipolar transistor applications, where the electron transit time in the base is shortened by the electrical field generated by the graded Ge profiles. In-plane electron mobility decreases with increasing Ge fraction, whereas out-of-plane electron mobility increases with Ge fraction at small Ge content. The out-of-plane mobility is larger than the in-plane mobility.

Klaassen model [39, 40] is commonly used for minority and majority mobility. Klaassen's model takes into account four scattering mechanisms: lattice scattering, ionized impurity scattering, phonon scattering, and carrier-carrier scattering. The last mechanism makes the minority and majority carrier mobilities different.

2.4.2. Mobility in Strained Si

Figure 10 shows the in-plane electron and hole mobilities at 300 K in strained Si, based on [41]. Here the carrier-carrier scattering is neglected, and majority mobility and minority mobilities are the same. For tensile strain (positive strain value), the in-plane electron mobility increases rapidly, and electrons populate the two lower energy Δ valleys along the growth direction. The lower conductivity mass

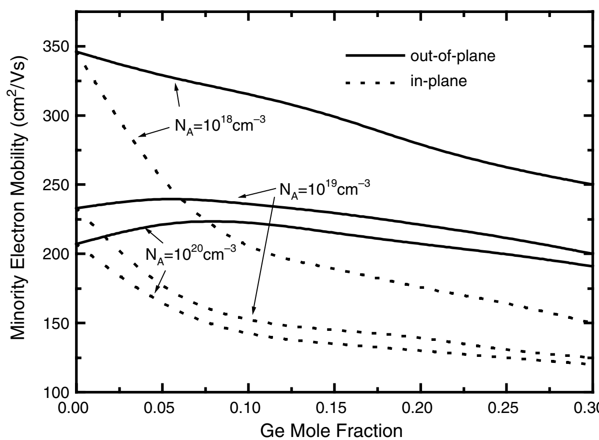


Figure 9. In-plane and out-of-plane minority electron mobilities of SiGe alloys at different doping levels at 300 K.

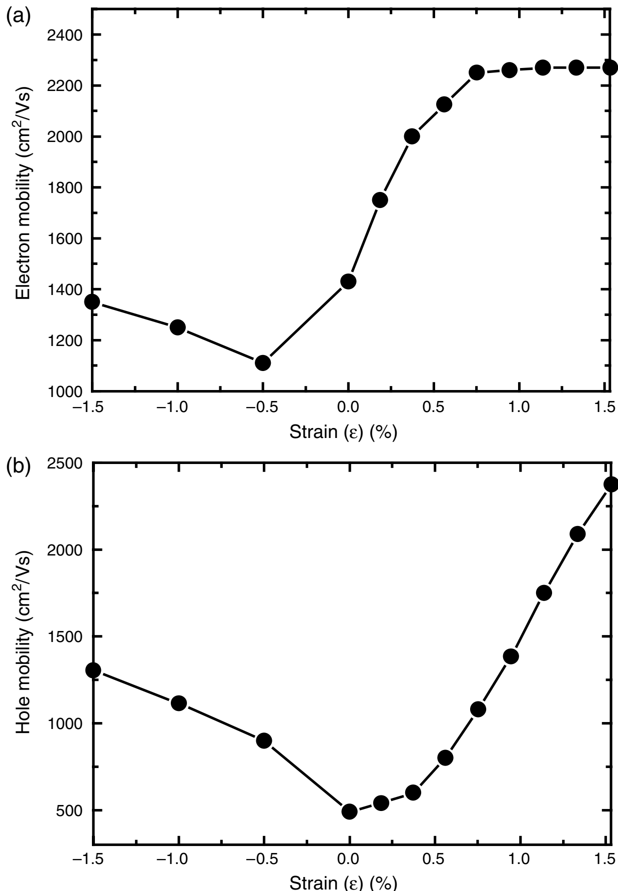


Figure 10. The in-plane (a) electron and (b) hole mobility in strained Si. Reprinted with permission from [70], F. Schafer, *Semicond. Sci. Technol.* 12, 1515 (1997). © 1997, Institute of Physics Publishing.

is the main cause of increased mobility with the help of the small decrease of intervalley scattering (the strong g-type process remains effective). The electron mobility reaches ~2300 cm^2/Vs at a tensile strain of about 0.8% (corresponding to a fully relaxed $Si_{.8}Ge_{.2}$ buffer layer). On the other hand, the out-of-plane mobility in Si with tensile strain drops rapidly with strain, since the conductivity mass is now the large longitudinal mass. However, only the in-plane mobility matters in the field effect device applications.

For the compressively strained Si, the in-plane electron mobility initially decreases with the strain, since electrons mainly populate the four Δ valleys perpendicular to the growth direction with a large effective conductive mass. With the further increase in compressive strain, the separation energy between the twofold degenerate valleys and fourfold degenerate valleys increases, and this reduces the intervalley scattering. The in-plane electron mobility increases again. Note that the out-of-plane electron mobility, controlled by the small transverse mass, remains high.

The hole mobility in strained Si increases for strain of either sign. The principal cause of the enhancement is the split of the degeneracy between the heavy-hole and light-hole bands at the Γ point in the reciprocal space and the reduction of the conductivity mass. For tensile strain, the light hole band has a much higher energy than the heavy-hole band, whereas for compressive strain the splitting of these two bands remains comparable to kT. This results in larger hole mobility in Si with tensile strain.

Tensile strain improves both electron and hole mobilities by reducing the effective transport masses and suppressing intervalley phonon scattering. For the MOSFET applications, the carrier in the inversion layers also suffers the oxide roughness scattering, which decreases the mobility value in the Figure 10. The calculated electron and hole inversion layer mobility enhancements in strained Si are plotted in Figure 11 [42, 43]. The amount of strain in Si is proportional to the Ge content in the relaxed SiGe buffer layer. The electron and hole mobilities are expected to increase with increasing strain and saturate at Ge concentrations of 20% and 40%, respectively, in the buffer layers. The saturation of mobility enhancement is related to the diminishing benefit of strain-induced energy splitting. The suppression of the intervalley scattering by energy splitting increases the mobility, but other scattering mechanisms, such as oxide roughness scattering and other phonon scattering, begin to dominate the low field mobility. At low temperature, there are fewer phonons, and the smaller energy splitting is enough to suppress the intervalley phonon scattering. Therefore, the mobility enhancement saturation is expected at even lower Ge content. Please note that both NMOSFET and PMOSFET can use the strained Si channel to increase the device speed.

2.4.3. Mobility in Strained Ge

Figure 12 shows the in-plane electron and hole mobilities at 300 K in strained Ge [41]. The electron mobility shows a noticeably complicated behavior. For tensile strain, electrons begin to populate the conduction states at the Γ point at strains larger than 1%. Because of the small in-plane effective electron mass in the Γ valley, there is a dramatic enhancement of mobility for strains larger than 1%. For compressive strain, the electrons mainly populate the L valleys at small strain, but the energy separation between

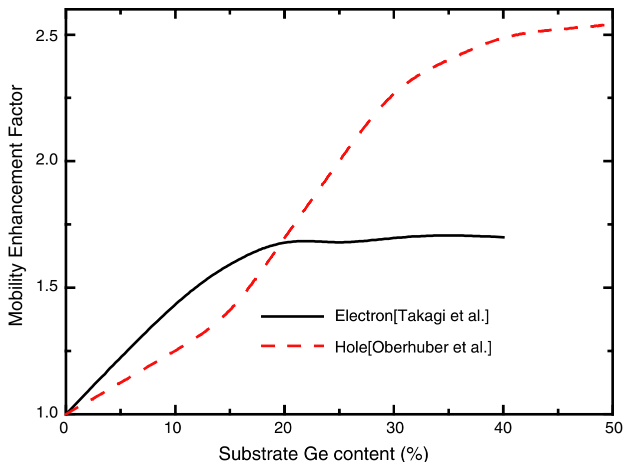


Figure 11. The calculated electron and hole mobility enhancement factor in MOS inversion layers of strained Si.

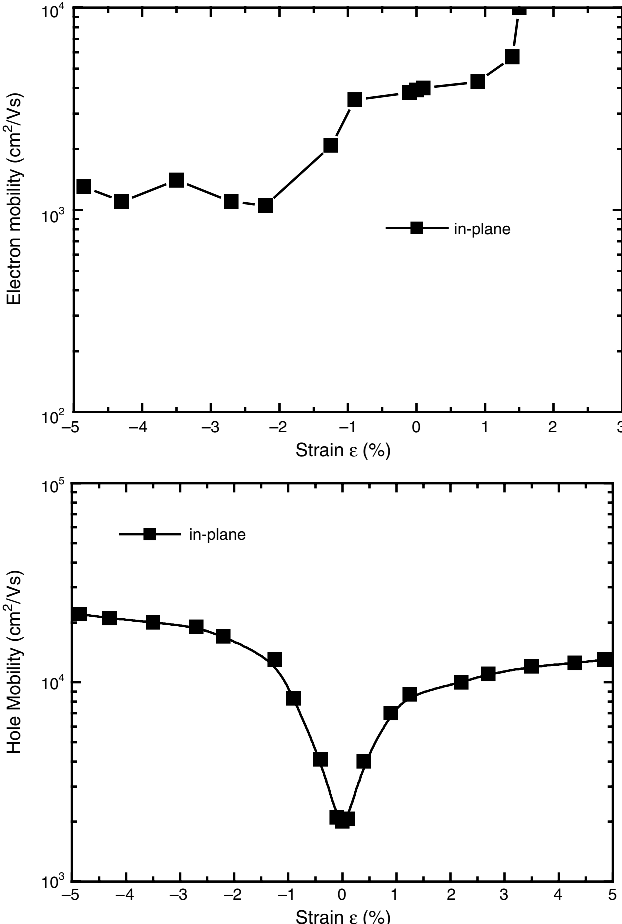


Figure 12. Electron and hole mobility in strained Ge.

L valleys and the four-fold Δ valleys perpendicular to the growth direction shrinks. The resulting larger intervalley scattering causes a lower mobility. At large compressive strain, all electrons stay in the Δ valleys. The mobility now approximates a constant value, nearly equivalent to the Δ valleys' electron mobility in relaxed Ge. In practice, only the compressively strained Ge can be grown. Therefore, there is no advantage to using the Ge channel in the NMOSFET, but it is interesting to note that the Ge with tensile strain can have a direct bandgap.

For in-plane hole mobility, the situation is qualitatively similar to that of strained Si. The compressive Ge has an in-plane hole mobility greater than 2000 cm^2/Vs and is beneficial for the PMOSFET application.

2.5. Relaxed SiGe on an Insulator

To have fully relaxed $Si_{1-x}Ge_x$ grown on Si, the $Si_{1-x}Ge_x$ thickness has to be much larger ($>10x$) than the critical thickness. The growth rate has to be high enough to ensure a reasonable growth time. The growth of films at low temperature takes a long time, but they might also have to be annealed at high temperature after growth to obtain sufficient relaxation.

The misfit and threading dislocations are generated in the relaxed films. All of the problems arise from the thick Si (500–800 μm) substrates. The thin epitaxial $Si_{1-x}Ge_x$ film has to fight with thick substrate to become relaxed with the pay-off of defects. It is unlikely that Si substrates can be thinned down far below 50 μm with mechanical stability. A smart way to effectively grow $Si_{1-x}Ge_x$ on thin Si is to use the silicon-on-insulator (SOI) substrates [44], and the Si can be as thin as 10 nm on the buried oxide (BOX, commonly used insulator). The SOI devices have the extra advantages of high speed, latch-up resistance, and radiation hardening. If the bonding force between Si and BOX can be small compared with Si and $Si_{1-x}Ge_x$, a $Si_{1-x}Ge_x$ thicker than SOI can easily be relaxed without any defect generation, and the thin SOI becomes strained. Several variations have been invented based on the basic idea.

The BOX layer can be created by separation-by-implanted-oxygen (SIMOX) technology after the $Si_{1-x}Ge_x$ layer is grown on conventional bulk Si wafers [45]. The implantation depth can reach Si or $Si_{1-x}Ge_x$. The relaxed films and BOX can be formed after high-temperature (>1100 °C) annealing. To obtain a relaxed $Si_{1-x}Ge_x$ layer with a large Ge fraction, the relaxed films can be further oxidized. The oxidation of $Si_{1-x}Ge_x$ can inject Ge into underlying $Si_{1-x}Ge_x$ films, and the underlying films can have a higher Ge fraction.

Bond-and-etch-back SOI (BESOI) wafers have the advantage of a cleaner silicon/oxide interface at the BOX than SIMOX. Oxidation of the seed and handle wafer, followed by bonding of the two wafers, generates the BESOI. The Si can be thinned down by the combination of CMP (chemical-mechanical polish) and etching to the desired film thickness. The relaxed $Si_{1-x}Ge_x$ can be grown on the BESOI wafers with thin Si (<100 nm) on the BOX.

The hydrogen ions implanted in Si at a sufficient dose (10^{16} to 10^{17} cm^{-2}) can break Si into two pieces after annealing at 400 °C to 600 °C because of the formation of hydrogen molecules (bubbles) in the implanted region. The way to cut Si by hydrogen implantation is called "smart cut." The splitting depth depends on the ion energy, and the splitting temperature depends on the ion dose. The combination of smart cut and wafer bonding can also yield relaxed $Si_{1-x}Ge_x$ with the use of one host wafer with oxide and the other handle wafer with hydrogen-implanted $Si_{1-x}Ge_x$ [46, 47]. Figure 13 shows one of the procedures for the smart-cut and layer transfer process to make strained Si on SiGe-on-insulator (SGOI) material, and Figure 14 is a transmission electron microscopy (TEM) picture of the resulting relaxed $Si_{1-x}Ge_x$ films with strained Si on the top. The as-split SGOI has a surface roughness of 5–12 nm (rms value); therefore a CMP process is needed to smooth the surface of the transferred layer. The relaxation of the $Si_{1-x}Ge_x$ layer is confirmed by X-ray diffraction measurement.

3. DEVICE PERFORMANCES

Advanced BiCMOS technology has fully utilized graded-base SiGe:C heterojunction bipolar transistors, which will dominate the radiofrequency (up to 10 GHz) and optical communication (up to 40 Gbit/s) markets. For mainstream Si technology, to increase the MOSFET speed with the same lithography feature, the strained Si MOSFET will be used in the 90-nm technology node and beyond. $Si_{1-x}Ge_x$ is one of

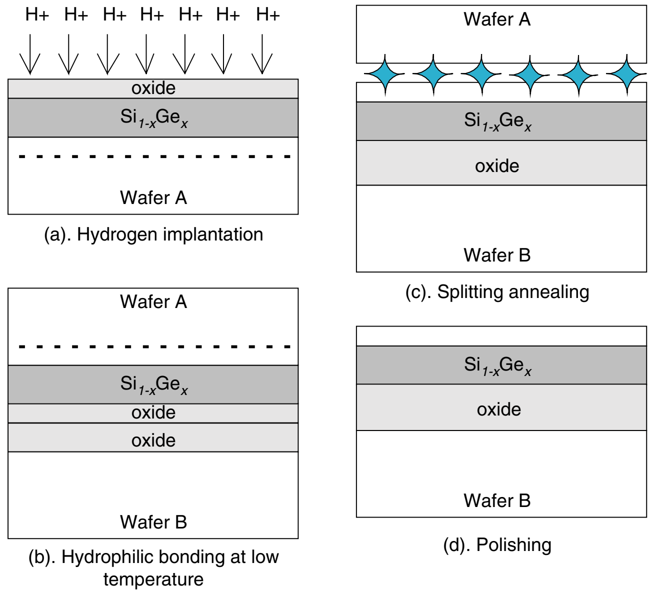


Figure 13. Smart-cut and layer transfer process flow for making strained Si on SiGe-on-insulator material.

the enabling technologies that can make Si prosperous for at least two more decades.

In addition to conventional Si devices, because of the heterostructures of $Si/Si_{1-x}Ge_x$, some III–V device structures, such as modulation doped FET and resonant tunneling diodes, can be made with Si. It is possible that these Si based quantum devices can outperform the original III–V devices.

The bandgap of Si (~1.1 eV) almost makes it impossible to have Si-based devices operating at 1.3-μm, 1.5-μm wavelengths and at the far-infrared region (~10 μm) for opto-electronic and defense applications. However, $Si_{1-x}Ge_x$ with a small bandgap can absorb the light of 1.3 and 1.5 μm, and Ge/Si quantum structures with intraband transition can have far-infrared absorption. For light emission devices, $Si_{1-x}Ge_x$ p-i-n light-emitting diodes are also demonstrated.

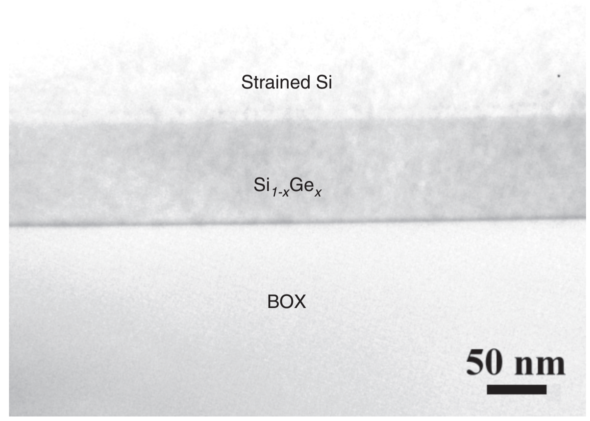


Figure 14. Cross-sectional TEM picture of relaxed SiGe on SOI from the smart-cut method.

3.1. Si/SiGe:C Heterojunction Bipolar Transistors

Since the first Si/SiGe heterojunction bipolar transistor (HBT) was reported in 1988 [48], tremendous progress of Si/SiGe:C HBTs has been achieved, reaching a maximum oscillation frequency of 210 GHz [1] and a cut-off frequency of 285 GHz [2], with circuit applications in wireless and optical communication.

In the NPN bipolar transistor, the collector current is controlled by application of a forward bias V_{BE} to the emitter-base junction [49]. In the forward active region, the base-collector junction is reverse biased by V_{CB}, as shown in Figure 15. More electrons are injected into the base region in HBT than in bipolar junction transistor (BJT) because of the lowering of the potential barrier ΔV_n. Note that the same base doping concentration is assumed in HBT and BJT, and therefore the separation between the Fermi level and the valence band edge is the same for HBT and BJT. This results in an increase in collector current as compared with Si BJT with the same base doping. In other words, to have the same barrier ΔV_n (same collector current), the HBT can have a heavy base doping, which moves the conduction band edge in the base upward. The narrow bandgap of the $Si_{1-x}Ge_x$ base can have a concentration of up to 2×10^{20} cm^{-3} [50], which significantly reduces the base resistance and hence leads to a high maximum oscillation frequency, high Early voltage, and a low noise figure, as compared with Si bipolar junction transistors.

A built-in electric field for electrons in the p-type base region can be obtained with graded $Si_{1-x}Ge_x$ (shown in Fig. 16). The electrical field accelerates the electron motion in the base and thus reduces the base transit time. The cut-off frequency f_t can be enhanced. Many studies [51–56] are focused on the optimal Ge profile of SiGe base HBT′ to minimize the base transit time. The narrow base also reduces the base transit time. However, because of the boron out-diffusion, especially by transient-enhanced diffusion (TED) in the fabrication process, the base width is limited. An ultra-high-speed HBT requires a heavily doped and extremely narrow base. The carbon incorporation in

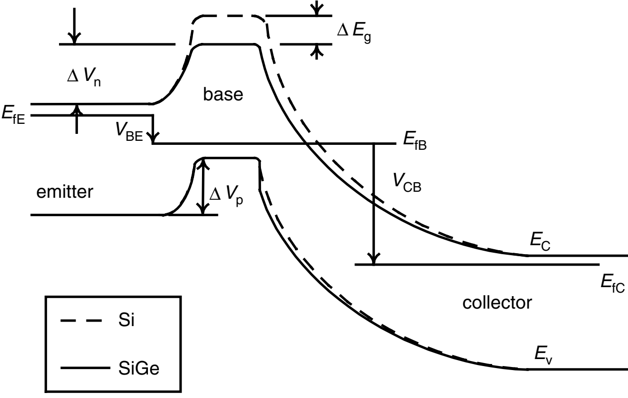


Figure 15. Band diagram of typical SiGe HBT. In the forward active region, the base-emitter junction is forward-biased (V_{BE}) and the base-collector junction reverse-biased (V_{CB}). E_{fC}, E_{fB}, and E_{fC} are the Fermi levels in the emitter, the base, and the collector, respectively.

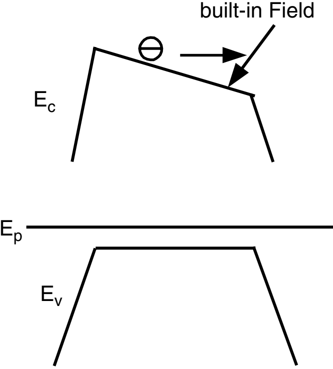


Figure 16. The compositionally graded $Si_{1-x}Ge_x$ can build an electrical field in the base of Si/SiGe HBTs.

the $Si_{1-x}Ge_x$ base can reduce the boron out-diffusion to a narrow base with small base resistance and without the parasitic barrier in the collector. A low concentration of carbon atoms ($\sim< 0.2\% = 10^{20}$ cm^{-3}) is introduced into the base region to effectively suppress the TED of boron [57]. The suppressed boron diffusion found in carbon-doped Si and SiGe is caused by an undersaturation of Si interstitials due to carbon out-diffusion from the base. However, it is reported that the addition of a carbon concentration higher than $\sim$1% can degrade the electron mobility by a factor of 2 [58].

3.2. Field Effect Transistors

3.2.1. N Channel Metal-Oxide-Semiconductor Field Effect Transistors

Very high electron mobilities in strained Si layers suggest a great potential for this material in high-performance N channel metal-oxide-semiconductor field effect transistors NMOSFETs. The low defect density buffers and high quality oxide at low temperature are key technologies for ensuring a mobility advantage in device performance.

The enhancement of effective electron mobility has been reported for long-channel MOSFETs with both surface and buried channels [59]. Figure 17 shows schematic diagrams of these two possible configurations of strained Si channel NMOSFETs [60]. Both structures have relaxed SiGe buffer layers grown on either a graded-Ge layer or SOI. The surface channel device (Fig. 17a) has a single layer of thin strained Si grown on top of a relaxed buffer layer. This layer is oxidized to form gate oxide. The buried strained Si-channel device (Fig. 17b) has a layer of strained Si buried beneath a thin layer of relaxed SiGe. An additional layer of strained Si is necessary to form gate oxide on top of the SiGe, and ideally this additional Si layer should be consumed during oxidation. If it is not, the residual sacrificial Si layer between the gate oxide and the SiGe barrier layer can act as a parallel conducting channel and strongly affects the device performance. Depending on the dopant type in the layers, these structures can also be used for PMOSFETs or NMOSFETs. The long channel ($W = 10\ \mu$m $\times 168\ \mu$m) surface and buried-channel NMOSFET devices fabricated on relaxed $Si_{0.7}Ge_{0.3}$ buffer layers showed well-behaved output characteristics. The effective low-field mobilities for these device structures [60] are shown in Figure 18. For the surface channel strained Si device, the effective mobility is enhanced with a similar dependence on the effective electric field as compared with a bulk-Si control device. The peak mobility value is 1000 cm^2/Vs, which shows 80% enhancement over the Si control device (550 cm^2/Vs). The peak mobility value of the buried channel device is over 1600 cm^2/Vs, which is almost 3 times that of the Si control device, but drops rapidly with increasing effective field. The enhancement of strained Si NMOSFET mobility becomes less significant, if the low Ge content buffer is used [60], since the strained Si mobility increases with increasing strain (more Ge content in the relaxed buffer layer). Note that buried channel devices with deep submicron length suffer from degraded threshold voltage roll-off and subthreshold characteristics.

Recently, IBM [6] has demonstrated current drive enhancements in the strained-Si NMOSFET with sub-100-nm physical gate lengths. Measured effective electron mobility characteristics are shown in Figure 19. For $\sim$1.2% strain (i.e., 28% Ge content in buffer layer), effective electron

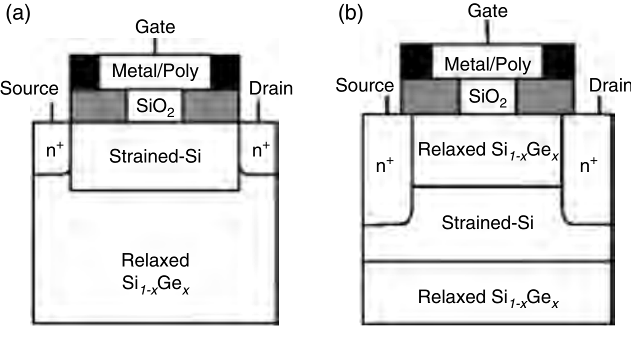


Figure 17. Device structures for strained Si NMOSFETs with (a) a Si surface channel and (b) a Si buried channel. Adapted with permission from [59], J. Welser et al., *IEEE Electron Device Lett.* 15, 100 (1994). © 1994, IEEE.

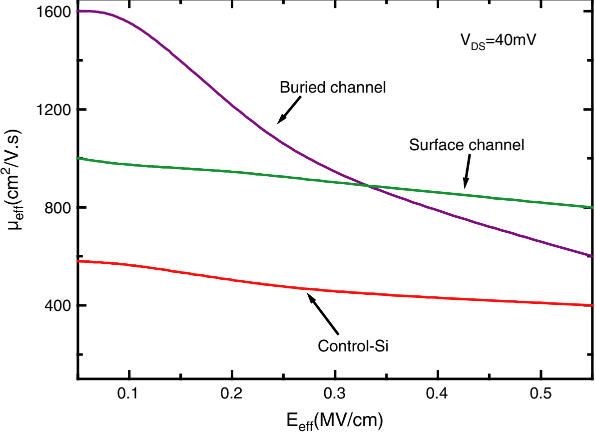


Figure 18. Effective low-field mobility as a function of effective field for different NMOSFETs at $T = 290$ K. The surface channel mobility shows a fairly constant mobility enhancement compared with that of a control-Si device, whereas the buried channel mobility peaks at low fields but decreases rapidly at higher fields. Adapted with permission from [59], J. Welser et al., *IEEE Electron Device Lett.* 15, 100 (1994). © 1994, IEEE.

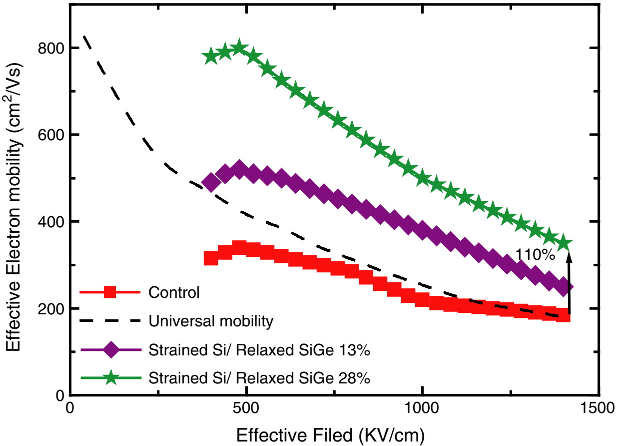


Figure 19. NMOSFET effective mobility versus vertical effective field. Adapted with permission from [6], K. Rim et al., "IEEE Symposium on VLSI Technology Digest, 98," 2002. © 2002, IEEE.

mobility at high field was enhanced by 110% over that of the control Si. Note that to obtain a high current drive in extremely small devices (∼30 nm), the saturation velocity, in addition to the mobility, is also important.

3.2.2. P Channel Metal-Oxide-Semiconductor Field Effect Transistors

Strained-Si P channel metal-oxide-semiconductor field effect transistors (PMOSFETs) have been studied [61–67] through the use of high hole mobility. The tensile strain in silicon grown on relaxed SiGe buffer raises the light-hole band and lowers the heavy-hole band. This reduces an intervalley scattering and increases the low field hole mobility significantly. Recently, IBM [6] has demonstrated current drive enhancements in the strained-Si PMOSFET with sub-100-nm physical gate lengths. Measured hole mobility characteristics are shown in Figure 20. For ∼1.2% strain (i.e., 28% Ge content in the buffer layer), peak hole mobility was enhanced by 45% over that of the control Si. The hole

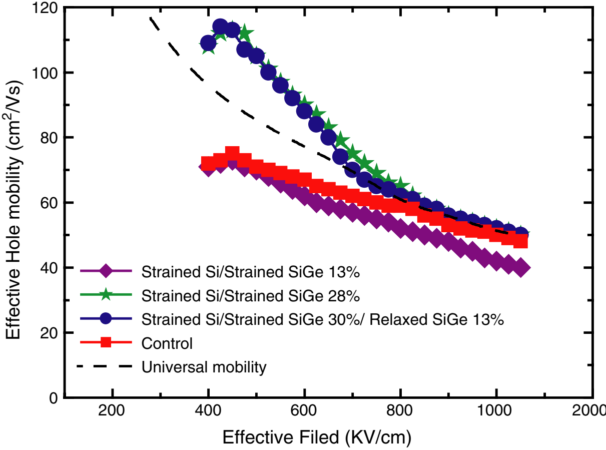


Figure 20. PMOSFET effective mobility versus vertical effective field. Adapted with permission from [6], K. Rim et al., "IEEE Symposium on VLSI Technology Digest, 98," 2002. © 2002, IEEE.

mobility enhancement diminishes at high effective electric field. For lower strain in Si (13% Ge content in buffer layer), hole mobility is slightly degraded compared with the control Si. Since the strained SiGe also has high hole mobility (Fig. 8), it is desirable to use a SiGe channel in the PMOSFETs. With a strained SiGe buried channel below the strained Si in the buried channel device (Fig. 21), the hole mobility has an enhancement comparable to that of the Si surface channel structure with high strain (28% Ge content in the buffer layer). The same strained Si structure suitable for NMOSFETs and PMOSFETs makes it possible to apply strained Si technology in CMOS circuits.

Since the compressively strained Ge has a large hole mobility (Fig. 12), Ge channel PMOSFETs with an effective mobility of 2700 cm^2/Vs have also been reported [5].

3.2.3. Modulation Doped Field Effect Transistors

Carriers can be separated from their parent donor or acceptor atoms for suitable band discontinuity. The impurity scattering can be further reduced if a spacer is inserted between the doped carrier-supply layers and undoped conduction channels. In principle, the modulation doped field effect transistors (MODFET) can have a much higher mobility than the MOSFET because of the lack of impurity scattering. With the band alignment requirement, NMODFETs have to use a strained Si channel on relaxed $Si_{1-x}Ge_x$ [68], whereas PMODFETs have to use strained $Si_{1-x}Ge_x$ on relaxed Si or strained Ge on relaxed $Si_{1-x}Ge_x$ [5, 69]. Figure 22 shows

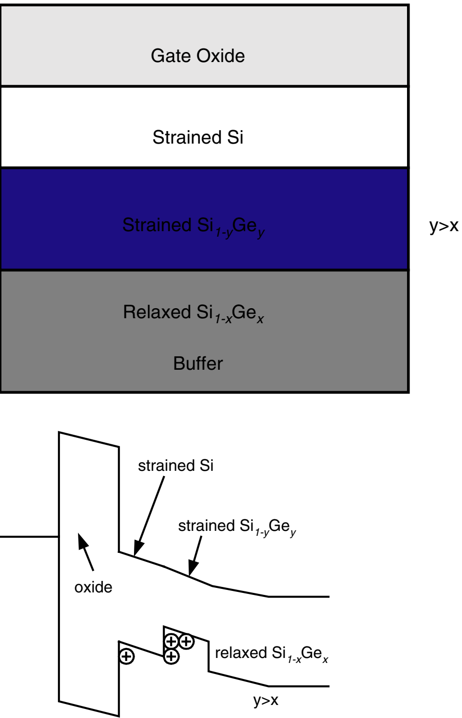


Figure 21. Schematic diagram of a buried strained-SiGe PMOSFET. Most current flow is in the strained SiGe channel. Adapted with permission from [6], K. Rim et al., "IEEE Symposium on VLSI Technology Digest, 98," 2002. © 2002, IEEE.

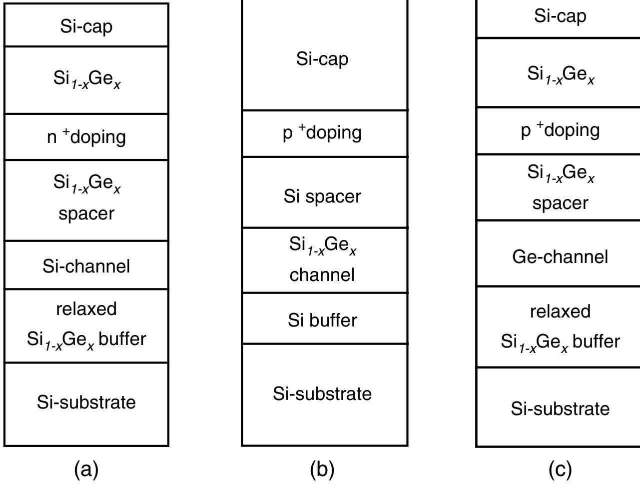


Figure 22. Typical MODFET structures. (a) NMODFETs with Si-channel on relaxed-SiGe buffer. (b) PMODFET with SiGe channel and (c) PMODFET with Ge channel on relaxed-SiGe buffer. Note that the doping layers and spacers can be located below the bottom of the channels (after [70]).

these three possible MODFET structures [70]. As in MOSFETs, the large defect density of these structures can seriously degrade the device performance.

An electron mobility up to 2700 cm^2/Vs at 300 K for a sheet carrier density of 7×10^{12} cm^{-2} in a strained Si channel NMODFET is reported [71], and a hole mobility of 1880 cm^2/Vs at 300 K for a sheet carrier density of 2.1×10^{12} cm^{-2} has been obtained with a pure Ge channel layer on a relaxed $Si_{0.4}Ge_{0.6}$ buffer [71]. All of the available data on effective electron and hole mobilities for MODFET are shown in Figure 23 [71–75]. From a comparison of Figure 19 and 23, it is evident the MODFETs have larger enhancement than MOSFETs.

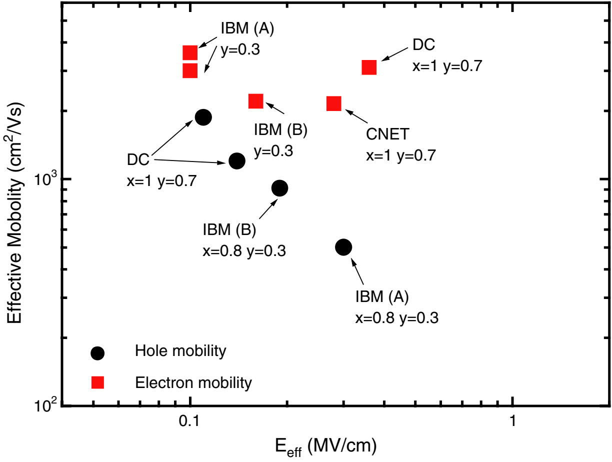


Figure 23. Available experimental data [71–75] at 300 K for effective electron and hole mobility in MODFET. A denotes modulation doping above a strained Si channel, B denotes a doping supply layer below a strained silicon channel, x is the Ge fraction in the channel, and y is the Ge fraction in the buffer layers. DC, DaimlerChrysler Research Center; CNET; France Telecom.

3.3. Optoelectronic Devices

In the last decade, interest in silicon-based optoelectronic devices has grown rapidly [76]. The trend of optoelectronics is to integrate both high-performance optoelectronic and electrical devices on the same chips with silicon processing technology. Among silicon-integrable materials, Ge and SiGe are of great importance to the fabrication of optoelectronic devices because of their optical absorption wavelengths.

Many SiGe/Si devices operated in the 1.3–1.5 μm wavelength range have been reported and demonstrated [77–78]. In this section, we discuss Si/SiGe photodetectors, light-emitting diodes, and heterojunction phototransistors.

3.3.1. Absorption of SiGe

Because of the large absorption length ($\sim$16μm) of Si at 820 nm [79] and the forbidden absorption at 1300 and 1550 nm, Si-based photodetectors have limited detection efficiency and wavelength range. Not only can the incorporation of Ge into Si increase the cutoff wavelength (Fig. 6); it can also enhance the absorption efficiency (small absorption length). The strained Ge could have an absorption length of 0.1 μm or less at a wavelength of 820 nm. Figure 24 shows the absorption length at 820, 1300, and 1550 nm versus Ge mole fraction. The absorption length decreases as the Ge mole fraction increases. For the large Ge fraction, the shadowed areas indicate the uncertainty of the estimation. The incorporation of strained Ge/SiGe into optoelectronic devices makes devices particularly useful over the important fiber optic communications wavelengths.

3.3.2. Photodetectors

The figures of merit for light detection are high efficiency, high responsivity, low noise, low dark currents, and fast response. Previous studies have focused on metal-semiconductor-metal (MSM) diodes and p-i-n diodes.

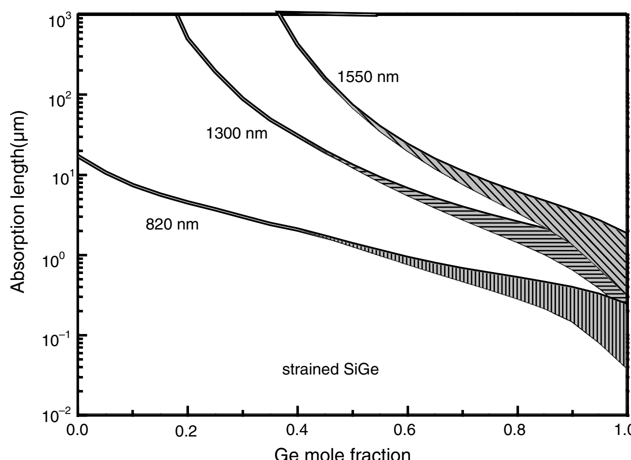


Figure 24. The absorption length at 820, 1300, and 1550 nm versus Ge mole fraction. The absorption length decreases as the Ge mole fraction increases. For the large Ge fraction, the shadowed areas indicate the uncertainty of the estimation. Reprinted with permission from [16], B.-C. Hsu et al., "International Electron Device Meeting," 2002, pp. 91–94. © 2002, IEEE.

In 1992, a metal-semiconductor-metal (MSM) diodes was reported by Splett et al. [80]. The measured responsivity at 1.3 μm is 0.2 A/W over a 1-mm detector length with 500 pA/μm^2 dark current at 5-V bias. In 1996, Huang et al. reported an epitaxial SiGeC/Si photodetector with a response in the 1.3–1.55-μm wavelength range [77]. The active absorption layer of the p-i-n photodiode consists of a pseudomorphic SiGeC alloy grown on a Si substrate with a Ge content of 55% and a thickness of 80 nm. In 2001, a p-i-n Ge on Si photodetector was reported by Masini et al. [78]. The photodiodes exhibit short-circuit responsivities of 0.3 and 0.2 A/W at 1.3 and 1.55 μm, respectively, reverse dark currents of 20 mA/cm^2, and response times of 800 ps.

In 1999, photodetectors and light-emitting diodes for 1.1-μm wavelength have been demonstrated with metal-oxide-silicon (MOS) tunneling structures on both n-type and p-type silicon substrates [81–84]. The ultrathin thickness ($<$3 nm) of thermal oxide is required in these novel photodetectors to provide sufficiently large tunneling probability. As biased in inversion region, the tunneling diode works in the deep depletion region with the soft pinning of oxide voltage, instead of pinning of surface potential, which is very different from the conventional MOS diode with thick oxide. The Ge MOS detector can operate at 1.3 and 1.55 μm with high responsivity [85]. To avoid material degradation such as strain relaxation and Ge out-diffusion, low-temperature liquid phase deposition (LPD) oxide is introduced [86].

A MOS Ge quantum dot photodetector is demonstrated [16]. The photodetector has responsivities of 130, 0.16, and 0.08 mA/W at detection wavelengths of 820 nm, 1300 nm, and 1550 nm, respectively. The dark current is extremely low (0.06 mA/cm^2). Si/Ge quantum dots are prepared by ultra-high-vacuum chemical vapor deposition (UHVCVD) on p-type Si (001) substrates. The structure is shown in Figure 25. With the careful design of intraband transition, the long-wavelength infrared (6–10 μm) can also be detected with the same devices.

3.3.3. Light-Emitting Diodes

Two different Si light emitters have been reported. A narrow-band infrared emitter at 1160 nm was implemented with a pn junction under forward bias with an external quantum efficiency of $\sim 10^{-4}$ [87]. A broad-band (450–850 nm) visible-light emitter was also realized with an avalanche pn diode with an external quantum efficiency of $\sim 10^{-8}$ [87] and $\sim 10^{-6}$ [88]. Recently, the band-edge (1.1 μm) electroluminescence (EL) of a MOS tunneling light-emitting diode has also been reported [83]. Because of the indirect bandgap of Si, additional momentum is required for the light emission process. The phonon provides the additional momentum in bulk Si and bulk SiGe [89]. In the MOS structure, the interface/surface roughness can seriously affect the carrier transport in the inversion layer [90]. This indicates that the interface/surface roughness can scatter the carrier and can change the carrier momentum. Therefore, the rough oxide can enhance the emission efficiency of MOS light emitting diodes [91].

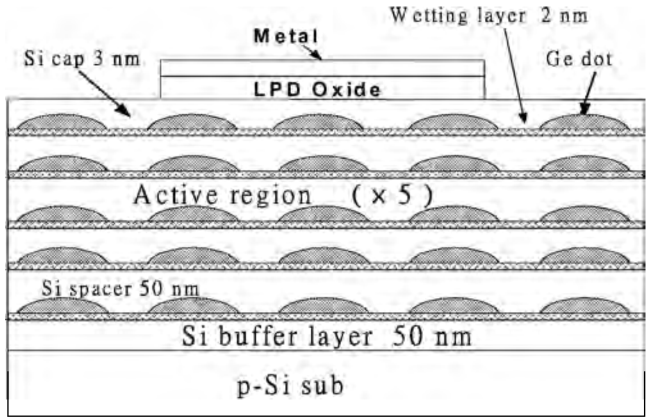


Figure 25. The MOS detector structure with five layers of Ge quantum dots.

The addition of $Si_{1-x}Ge_x$ to p-i-n diodes can increase the emission wavelength. Both 1.3-μm [18] and 1.5-μm [19] emissions with $Si_{1-x}Ge_x$/Si quantum wells and Ge/Si quantum dots, respectively, are demonstrated at room temperature, and erbium-doped $Si_{1-x}Ge_x$ can also have 1.5-μm emission [92]. However, MOS diodes with $Si_{1-x}Ge_x$ or Ge quantum dots are not reported to have long-wavelength emission so far.

3.3.4. Heterojunction Phototransistors

With high gain and low noise, the phototransistor can be used in the front end of an optical receiver. The $Si_{1-x}Ge_x$/Si multiple quantum well heterojunction phototransistor with an ultrahigh responsivity of 1.47 A/W and a bandwidth of $>$1.25 GHz at a wavelength of 850 nm has been reported at an operation voltage as low as $\leq$1.5 V [13]. The responsivity at 1310 nm is 0.15 A/W. The multiple quantum wells in the collectors can enhance the photon absorption and increase the cutoff wavelength. The current gain of the transistors can further amplify the optical current. Si/$Si_{1-x}Ge_x$ quantum well phototransistors with far-infrared (6–20 μm) response with low leakage current and high gain were also proposed [93].

4. FABRICATION TECHNOLOGIES

4.1. Growth of $Si_{1-x}Ge_x$ Films

$Si_{1-x}Ge_x$ can be epitaxially grown by chemical vapor deposition (CVD) and molecular beam epitaxy (MBE). Depending on the growth pressures, CVD has some variations: APCVD (atmospheric pressure), rapid thermal CVD (RTCVD, several torr with lamp heating), low-pressure CVD (LPCVD, several tenths of a torr), and ultrahigh vacuum pressure CVD (UHV/CVD). UHV/CVD [94] has a growth pressure of 10^{-2} to 10^{-3} torr and a base pressure of UHV ($\sim 10^{-9}$ torr). The designation UHV is sometimes confusing and only indicates the base pressure of the reactor chamber, not the growth pressure. The wafer temperature can be obtained by lamp heating or thermal equilibrium furnaces. For lamp heating, the wafer temperature can rise and fall very rapidly (on the order of 10 s), and is often called "rapid thermal CVD" [95]. The chamber used for RTCVD is often made of quartz to avoid contamination during growth. The quartz chamber does not absorb as much emission from the lamps

as Si wafers. A cold-wall RTCVD can be obtained, whereas a quartz tube with hot walls is used in the furnace-type reactor, such as that used for UHV/CVD. Because of the configuration of heating lamps, RTCVD can process only a single wafer at a time, whereas a furnace-type reactor such as that used for UHV/CVD can process a batch of wafers (more than 20 wafers). Silane (SiH_4) and dichlorosilane (DCS) are the common precursors of Si, and germane (GeH_4) is the precursor of Ge. Phosphine (PH_3) is used for n-type doping, and diborane (B_2H_6) is for p-type doping. Both RTCVD and UHV/CVD are used in semiconductor manufacturing laboratories as well as research laboratories.

MBE is a physical deposition system with a stainless chamber. The base pressure and growth pressure are about 10^{-11} and 10^{-9} torr, respectively. The electron beams are used to heat source materials such as Si and Ge, and the Si and Ge atoms are evaporated on Si substrates. The chamber is cold and substrates are heated by a local heater. Because of the difficulty of using large-diameter wafers and some process issues, MBE is used in research laboratories, and HBTs with ideal base currents are grown by MBE [96] after CVD.

4.2. Metal Contacts

As the microelectronics industry moves into the era of deep submicron devices, metallization has become the main performance-limiting factor. Metal silicides have been widely used to reduce the contact resistance of the source/drain of microelectronic devices. The self-aligned silicidation technique has become a crucial part of recent ultra-high-speed silicon device technologies [97]. $TiSi_2$ is currently the most commonly used silicide in the IC industry. In the deep submicron regime, a linewidth dependence of sheet resistance was observed for $TiSi_2$ [98, 99]. $CoSi_2$ has been introduced to replace $TiSi_2$ in sub-quarter-micron technology. However, its use in sub-0.1-micron devices is in doubt unless a raised source/drain scheme becomes feasible. NiSi is the only silicide left with resistivity comparable to that of $TiSi_2$ and $CoSi_2$. On the other hand, intensive efforts have been made to extend IC devices to other substrate frames, such as Si-Ge. Metal germanosilicide/$Si_{1-x}Ge_x$/Si heterostructures are promising for use in devices such as the heterojunction bipolar transistor and infrared detectors with high cutoff wavelengths [100, 101].

Metal germanosilicides may be used for making metallic contacts with $Si_{1-x}Ge_x$ alloys, and knowledge about the formation and stability of thin metal germanosilicide films is essential for such applications. The thermally induced metal/$Si_{1-x}Ge_x$ reaction has previously been studied for many metals. Various degrees of germanium segregation and/or the formation of segregated layered structures were observed in the reactions of these metal/$Si_{1-x}Ge_x$ systems. The difference in heats of formation of silicide and germanide offers the driving force for the segregation of Ge-rich Si-Ge alloy [102]. The phase formation paths were found to depend strongly on the composition of substrates. By extension from silicon to Si-Ge devices, the contact materials considered for Si-Ge devices have been focused on Ti, Co, and Ni contacts.

4.2.1. Co/Si-Ge System

Direct deposition of metal thin films on $Si_{1-x}Ge_x$ in a self-aligned silicide process would be an efficient technique and would take advantage of established technologies. However, the major difficulties in using Co as a contact material for $Si_{1-x}Ge_x$ appear to be the preferential reaction of Co with Si and the high consumption ratio of the Si to Co to form $CoSi_2$. The undesirable characteristics lead to Ge segregation, deterioration of the $Si_{1-x}Ge_x$ layer, degradation of junction integrity, and film agglomeration at low temperatures, resulting in poor thermal stability and high-resistivity contacts [103–105]. In addition, it has been demonstrated that $CoSi_2$ is formed at much higher temperatures on a $Si_{1-x}Ge_x$ layer than that on Si, owing to Ge expelled from the Co-Si-Ge compound, which blocks the Co diffusion paths, slowing down the reaction [106–108]. Successful formation of good quality $CoSi_2$ on epitaxial $Si_{0.7}Ge_{0.3}$ in the presence of a sacrificial a-Si layer has been achieved. $CoSi_2$ was previously found to form at a lower temperature on a-Si with a smoother interface than that on single-crystal Si [109].

For convenience, the two sets of samples Co(15 nm)/$Si_{0.7}Ge_{0.3}$ and Co(15 nm)/a-Si(50 nm)/$Si_{0.7}Ge_{0.3}$ are designated as samples A and B, respectively. Five hundred-nanometer-thick $Si_{0.7}Ge_{0.3}$ and 1-μm-thick strained layers of Si_yGe_{1-y} (y varies from 1 to 0.7) were grown on (001)Si wafers at 550 °C by MBE. A 50-nm-thick sacrificial a-Si layer was deposited on a $Si_{0.7}Ge_{0.3}$ substrate followed by the deposition of a 15-nm-thick Co thin film without breaking the vacuum at room temperature. Figure 26 shows the sheet resistance data of the two sets of samples after different heat treatments. For samples annealed at 600 °C, the sheet resistance of samples B is much lower than that of samples A. The low resistance is attributed to the early formation of low-resistivity $CoSi_2$, which is consistent with the X-ray diffraction data. In contrast, $CoSi_2$ was not detected in samples A annealed at a temperature as high as 800 °C. The further substantial increase in sheet resistance for Co(15 nm)/$Si_{0.7}Ge_{0.3}$ samples after 800 °C annealing,

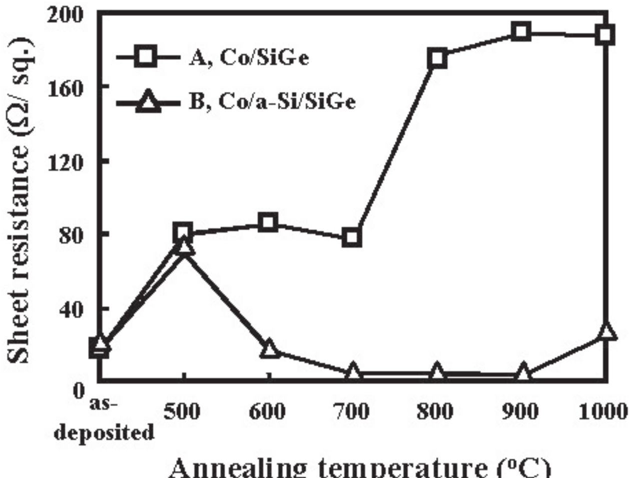


Figure 26. Sheet resistance versus annealing temperature curves for samples A and B after annealing at 500–1000 °C for 30 s by RTA. Reprinted with permission from [109], W. W. Wu et al., *Appl. Phys. Lett.* 81, 820 (2002). © 2002, American Institute of Physics.

as shown in Figure 26, is correlated to the agglomeration of $Co(Si_{1-x}Ge_x)$. Figure 27 is a plane-view TEM micrograph showing the formation of $Co(Si_{1-x}Ge_x)$ islands in a Co(15 nm)/$Si_{0.7}Ge_{0.3}$ sample annealed at 800 °C. On the other hand, for the Co(15 nm)/a-Si(50 nm)/$Si_{0.7}Ge_{0.3}$ samples, the formation of $CoSi_2$ was completed after 700 °C annealing. The sheet resistance maintained the same low level for samples B after annealing at 700–900 °C. In 900 °C annealed samples, the silicide films were found to be continuous with the smooth interface with $Si_{0.7}Ge_{0.3}$, as revealed by plane-view TEM and XTEM images shown in Figure 28. No Ge segregation was detected from the analysis of TEM direct images and electron diffraction patterns. In contrast with samples B, rough interface and Ge segregation in samples A are evident. An example is shown in Figure 29.

Previous study showed that the retardation of the silicidation process on $Si_{1-x}Ge_x$ is related to a higher effective Ge concentration at the reaction front, therefore causing an increase in the interface energy. The expelled and segregated Ge tends to diffuse to the surface and interface of the silicide layer [104, 105]. These Ge atoms can decorate the grain boundaries, resulting in an increase in the grain boundary energy, which makes the silicide film more prone to agglomeration. On the other hand, the sacrificial a-Si layer is shown to improve the interfacial roughness and thermal stability of a $CoSi_2$ film grown on $Si_{0.7}Ge_{0.3}$. The increased uniformity of the silicide/$Si_{0.7}Ge_{0.3}$ interface is speculated to result from small-grained $CoSi_2$. The average grain size is smaller for silicides formed on a-Si than that on single-crystal Si and poly-Si in samples subjected to the same heat treatment [110]. The results indicated that the reaction between Co thin films and a-Si incurs the formation of small-grained $CoSi_2$, which in turn enhances the morphological stability of $CoSi_2$. In addition, the formation of thermally stable $CoSi_2$ leads to a low sheet resistance value.

4.2.2. Ti/Si-Ge System

Titanium disilicide has been widely used for contacts to Si in microelectronics devices. Its use is complicated by the existence of two crystalline phases. The high-resistivity C49-$TiSi_2$ forms first on heating of Ti on Si above 500 °C, and additional heating above 700 °C is needed to transform C49 into the low-resistivity C54-$TiSi_2$. However, in the Ti/$Si_{1-x}Ge_x$ system, Ge segregation was observed near the C54 phase formation temperature [111]. In addition, the resistivity of C54-$Ti(Si_{1-z}Ge_z)_2$ forms on $Si_{1-x}Ge_x$ layer is higher than that of C54-$TiSi_2$. The appearance and agglomeration temperature of low-resistivity C54-$Ti(Si_{1-z}Ge_z)_2$ were both found to decrease with the Ge concentration [111].

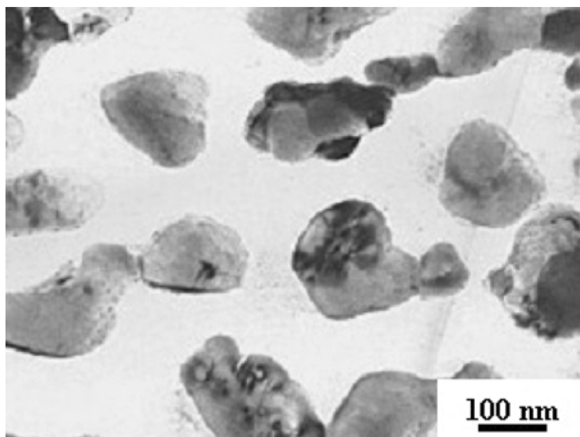


Figure 27. Plane-view TEM image of a Co(15 nm)/$Si_{0.7}Ge_{0.3}$ sample after annealing at 800 °C for 30 s by RTA. Reprinted with permission from [109], W. W. Wu et al., *Appl. Phys. Lett.* 81, 820 (2002). © 2002, American Institute of Physics.

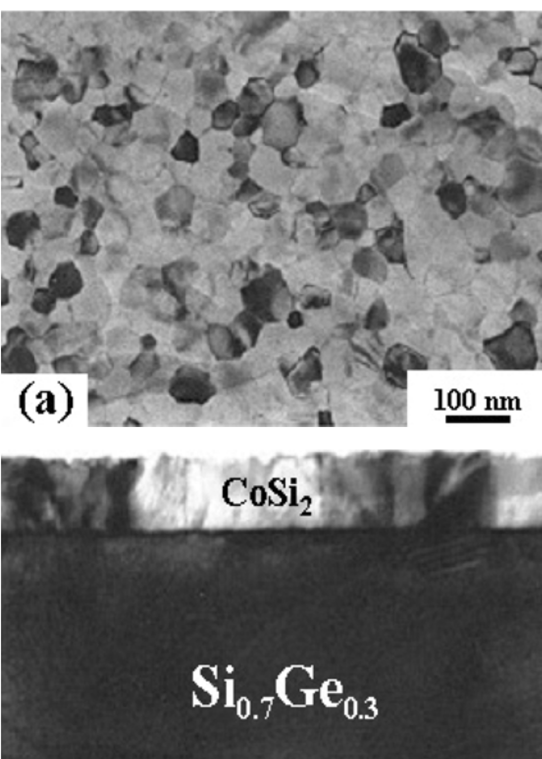


Figure 28. (a) Plane-view and (b) cross-sectional TEM images of Co(15 nm)/a-Si(50 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 900 °C for 30 s by RTA. Reprinted with permission from [109], W. W. Wu et al., *Appl. Phys. Lett.* 81, 820 (2002). © 2002, American Institute of Physics.

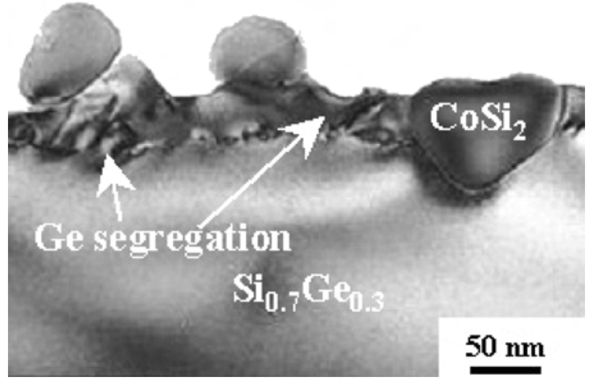


Figure 29. Cross-sectional TEM image of Co(15 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 900 °C for 30 s by RTA.

An amorphous Si was also successfully used to decrease the formation temperature of low-resistivity C54-$TiSi_2$ and alleviate the formation of islands as well as the segregation of Ge. For the Ti/a-Si/$Si_{1-x}Ge_x$ system, the thicknesses of Ti and amorphous Si were selected to be 15 and 30 nm, respectively. Figure 30 shows an XTEM image of a Ti/$Si_{0.7}Ge_{0.3}$ sample annealed at 900 °C, revealing the formation of C54-$Ti(Si_{1-z}Ge_z)_2$ islands and Ge segregation. An amorphous Si was used to alleviate the formation of islands as well as the segregation of Ge. The selection of the thickness ratio of Ti and a-Si was such that the a-Si was completely consumed in forming C54-$TiSi_2$. If the a-Si was too thin, Ti would start to react with the $Si_{1-x}Ge_x$ layer and accompanying problems in forming higher-resistivity $Ti(Si_{1-x}Ge_x)$ compound became troublesome [102]. Ge segregation [111], a rough interface [112], and strain relaxation of the $Si_{1-x}Ge_x$ layer are expected to occur [113].

GIXRD spectra of Ti(15 nm)/a-Si(30 nm)/$Si_{0.7}Ge_{0.3}$ samples revealed that low-resistivity C54-$TiSi_2$ was the dominant silicide phase above 650 °C. The formation temperature of C54-$TiSi_2$ was lowered by about 100 °C in samples with a sacrificial a-Si interlayer compared with that of the Ti/$Si_{0.7}Ge_{0.3}$ system [111]. Figure 31 shows the resistivity curves of the Ti(30)nm/$Si_{0.7}Ge_{0.3}$ and Ti(15 nm)/a-Si(30 nm)/$Si_{0.7}Ge_{0.3}$ samples (samples A and B) after annealing at different temperatures. For samples annealed at 700 °C, the resistivity of samples B is 23 much lower than that of samples A. It resulted from the formation of low-resistivity C54-$TiSi_2$ at low temperature. On the other hand, for samples A, as the annealing temperature was increased to 800 °C, a sharp drop in resistivity occurred. In samples B, the temperature range of low-resistivity C54-$TiSi_2$ was extended to 700–900 °C. From both resistivity and TEM data, it was concluded that the resistivity of C54-$TiSi_2$ in samples B is lower than that of C54-$Ti(Si_{1-z}Ge_z)_2$ in samples A. The magnitude of such differences will be affected by the homogeneity of the microstructures and will depend specifically upon the grain size [110]. On the other hand, with higher solute concentration, the electronic structure or phonon spectrum of the alloy will start to suffer from perturbation, and then the resistivity is expected to increase. For Ti(15 nm)/a-Si(30 nm)/$Si_{0.7}Ge_{0.3}$ samples, the a-Si was completely consumed in forming C54-$TiSi_2$, and the phase possesses lower resistivity than ternary C54-$Ti(Si_{1-z}Ge_z)_2$ phase. In addition, no Ge segregation was detected. The silicide films were also found to be continuous with the smooth interface with $Si_{0.7}Ge_{0.3}$, as shown in Figure 32. As a result, a sacrificial a-Si layer was found to decrease the formation temperature of C54-$TiSi_2$, prevent Ge segregation, lower the resistivity of silicide, and improve the morphological stability of C54-$TiSi_2$ on the $Si_{0.7}Ge_{0.3}$ layer. If the a-Si was too thin, Ti would react with the $Si_{0.7}Ge_{0.3}$ layer significantly and result in a rough interface.

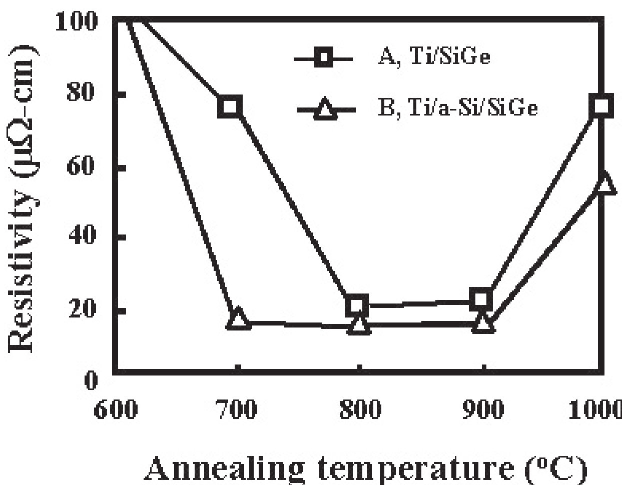


Figure 31. Resistivity versus annealing temperature curves for samples A and B after annealing at 500–1000 °C for 30 s by RTA.

4.2.3. Ni/Si-Ge System

Low-resistivity NiSi is currently one of the most promising silicides to replace $TiSi_2$ for the self-aligned technology of ULSI, owing to its favorable properties low resistivity, low silicon consumption, low processing temperature, and relative insensitivity to the linewidth of the silicide.

For the Ni/$Si_{0.7}Ge_{0.3}$ system, Ge tended to be less reactive with Ni than with of Si. Moreover, Ge was found to segregate to grain boundaries in 600 °C annealed samples, and $Ni(Si_{0.9}Ge_{0.1})$ islands were found to form in 700 °C annealed samples. In the XTEM image observation, the Ge segregation toward the grain boundary was evident. An example is shown in Figure 33. $Ni(Si_{0.9}Ge_{0.1})$ islands were observed to form in 700 °C annealed samples [114].

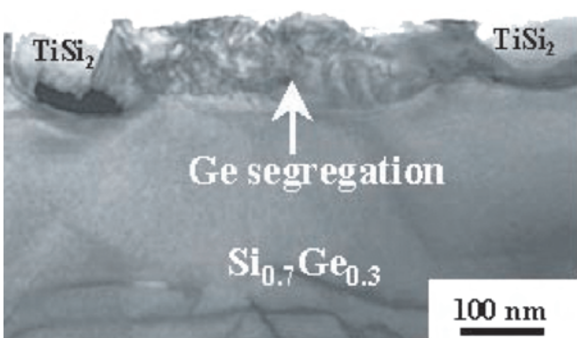


Figure 30. Cross-sectional TEM image of Ti(30 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 900 °C for 30 s by RTA.

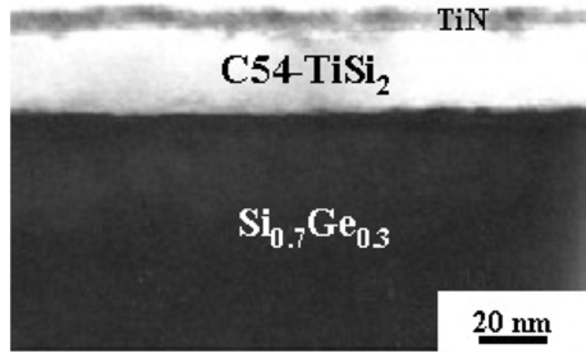


Figure 32. Cross-sectional TEM image of Ti(15 nm)/a-Si(30 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 900 °C for 30 s by RTA.

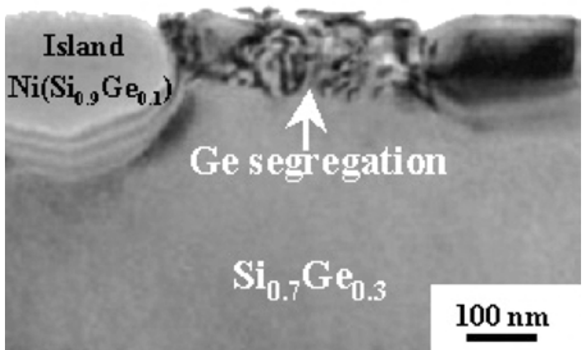


Figure 33. Cross-sectional TEM image of a Ni(30 nm)/$Si_{0.7}Ge_{0.3}$ sample after annealing at 700 °C for 30 s by RTA.

The scheme of an interposing a-Si layer was also attempted to overcome the detrimental effects of Ge segregation as well as island formation in Ni/Si-Ge systems. The thicknesses of Ni and amorphous Si were selected to be 15 and 27 nm, respectively. GIXRD spectra for Ni(15 nm)/a-Si(27 nm)/$Si_{0.7}Ge_{0.3}$ samples annealed at 300–700 °C indicated that low-resistivity NiSi was the only phase present in all samples annealed above 500 °C.

Figure 34 shows the sheet resistance data for Ni(30 nm)/$Si_{0.7}Ge_{0.3}$ and Ni(15 nm)/a-Si(27 nm)/$Si_{0.7}Ge_{0.3}$ samples (samples C and D). After annealing at 400 °C, low-resistivity NiSi phase was formed in both sets of samples. The apparent increase in sheet resistance for Ni(30 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 700 °C results from the formation of the island structure of $Ni(Si_{0.9}Ge_{0.1})$. On the other hand, even after annealing at 800 °C, the sheet resistance remained at the same low level for the Ni(15 nm)/a-Si(27 nm)/$Si_{0.7}Ge_{0.3}$ samples. Furthermore, the sacrificial a-Si layer was also seen to improve the interface roughness and thermal stability of NiSi film grown on $Si_{0.7}Ge_{0.3}$ from XTEM micrographs. In addition, no Ge segregation was observed. An example is shown in Figure 35. In the Ni thin films on the Si system,

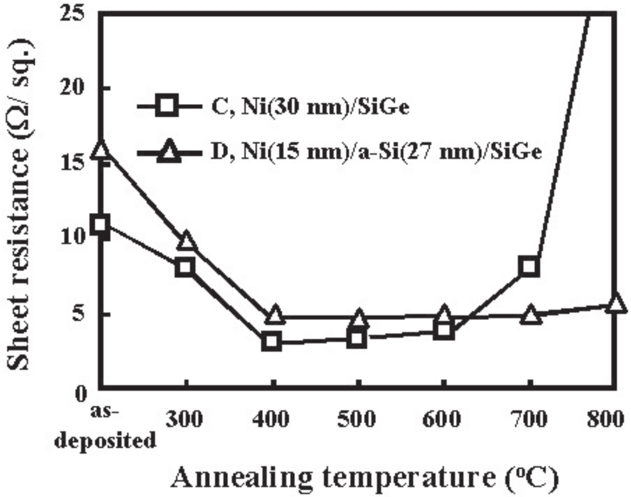


Figure 34. Sheet resistance versus annealing temperature curves for Ni(30 nm)/$Si_{0.7}Ge_{0.3}$ and Ni(15 nm)/a-Si(27 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 300–800 °C for 30 s by RTA.

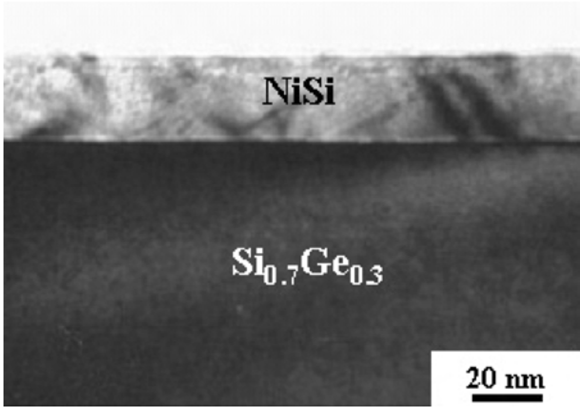


Figure 35. Cross-sectional TEM image of Ni(15 nm)/a-Si(27 nm)/$Si_{0.7}Ge_{0.3}$ samples after annealing at 800 °C for 30 s by RTA.

$NiSi_2$ is normally formed at 750 °C. The extraordinary stability of NiSi on $Si_{0.7}Ge_{0.3}$ with a sacrificial a-Si layer is likely due to the high consumption ratio of the Si to Ni to form $NiSi_2$, which requires the excess Ni atoms to react directly with the $Si_{1-x}Ge_x$ layer. Ge was then expelled from the Ni-Si-Ge compound and slowed down the reaction. The blocking of Ni diffusion paths by Ge atoms may also slow down the reaction and result in a delay of the formation of $NiSi_2$. The mechanism is similar to that of the formation of $CoSi_2$ on a $Si_{1-x}Ge_x$ layer. In contrast, if the a-Si is too thin, as in Ni(15 nm)/a-Si(22 nm)/$Si_{0.7}Ge_{0.3}$ samples, a rough interface is formed. The results indicate that a sacrificial a-Si layer with appropriate thickness works beneficially for Ni contacts on Si-Ge.

GLOSSARY

Chemical-mechanical polish (CMP) A chemical-mechanical method utilized to eliminate the surface roughness of a sample. This method is commonly used in ULSI processes.

Chemical vapor deposition (CVD) A method utilizing chemical gas to deposit film on substrates.

Dislocation A kind of line defect due to the lattice misfit between SiGe and Si. There are two segments. The segment parallel to the Si/SiGe interface is called "misfit dislocation," and the segment propagating to the surface is called "threading dislocation."

Heterojunction bipolar transistors (HBT) The narrow bandgap base in the bipolar transistor to enhance the device performance.

Metal germanosilicide A compound of metal, Si and Ge.

Strained Si-on-SiGe-on-insulator (SGOI) Introducing strained Si to the silicon-on-insulator (SOI) technology, which provides high-performance CMOS circuits due to combination of carrier mobility enhancement in strained Si with the advantage of SOI device/circuit.

Silicidation process The process used to form silicide.

Strained Si FET Strained Si field effect transistors. The tensile strained Si has higher carrier mobility than conventional Si. The new type of device can give larger current drive in VLSI circuits.

Universal mobility The carrier mobility in the field effect transistor depending only on effective transverse electrical field regardless of the doping concentration of the channel.

ACKNOWLEDGMENTS

The help of S. T. Chang (Department of Electrical Engineering, National Taiwan University, Taiwan) and W. W. Wu (Department of Materials Science and Engineering, National Tsing Hua University, Taiwan) in manuscript preparation are gratefully appreciated. The work on the composite quantum dots by Dr. P. S. Chen, ERSO/ITRI, Taiwan, is highly appreciated.

REFERENCES

1. S. J. Jeng, B. Jagannathan, J.-S. Rieh, J. Johnson, K. T. Schonenberg, D. Greenberg, A. Stricker, H. Chen, M. Khater, D. Ahlgren, G. Freeman, K. Stein, and S. Subbanna, *IEEE Electron Device Lett.* 22, 542 (2001).
2. B. Jagannathan, M. Khater, F. Pagette, J.-S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D. R. Greenberg, R. Groves, S. J. Jeng, J. Johnson, E. Mengistu, K. T. Schonenberg, C. M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, and S. Subbanna, *IEEE Electron Device Lett.* 23, 258 (2002).
3. T. Mizuno, S. Takagi, N. Sugiyama, H. Satake, A. Kurobe, and A. Toriumi, *IEEE Electron Device Lett.* 21, 230 (2000).
4. U. Konig, and F. Schaffler, *IEEE Electron Device Lett.* 14, 205 (1993).
5. T. Irisawa, S. Tokumitsu, T. Hattori, K. Nakagawa, S. Koh, and Y. Shitake, *Appl. Phys. Lett.* 81, 847 (2002).
6. K. Rim, J. Chu, H. Chen, K. A. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H.-S. Wong, "IEEE Symposium on VLSI Technology Digest, 98," 2002.
7. Y. Suda and H. Koyama, *Appl. Phys. Lett.* 79, 2273 (2001).
8. Z. Matutinoviac-Krstelj, C. W. Liu, X. Xiao, and J. C. Sturm, *J. Vac. Sci. Technol., B* 11, 1145 (1993).
9. H. C. Liu, D. Landheer, N. Buchanan, and D. C. Houghton, *Appl. Phys. Lett.* 52, 1809 (1988).
10. K. Ismail, B. S. Meyerson, S. Rishton, J. Chu, S. Nelson, and J. Noccra, *IEEE Electron Device Lett.* 13, 229 (1992).
11. J. Welser, J. L. Hoyt, and J. F. Gibbons, *IEEE Electron Device Lett.* 15, 100 (1994).
12. E. F. Crabbe, D. L. Harame, B. S. Meyerson, J. M. C. Stork, J. Y. C. Sun, "Device Research Conference Digest, 50th Annual," 1992, p. 26.
13. Z. Pei, C. S. Liang, L. S. Lai, Y. T. Tseng, Y. M. Hsu, P. S. Chen, S. C. Lu, C. M. Liu, M.-J. Tsai, and C. W. Liu, "International Electron Device Meeting," 2002.
14. F. Aniel, M. Enciso-Aguilar, L. Giguerre, P. Crozat, R. Adde, T. Mack, U. Seiler, Th. Hackbarth, and B. Raynor, "Proceedings of the International Semiconductor Device Research Symposium," 2001, p. 482.
15. U. Konig, "Proceedings of the IEEE international Symposium on High Performance Electron Devices for Microwave and Optoelectronic Application," 2000, pp. 1–7.
16. B.-C. Hsu, S. T. Chang, C.-R. Shie, C.-C. Lai, P. S. Chen, and C. W. Liu, "International Electron Device Meeting," 2002, pp. 91–94.
17. J. R. Jimenez, X. Xiao, J. C. Sturm, and P. W. Pellegrini, *Appl. Phys. Lett.* 67, 506 (1995).
18. Q. Mi, X. Xiao, J. C. Sturm, L. Lenchyshyn, and M. Thewalt, *Appl. Phys. Lett.* 60, 3177 (1992).
19. T. Brunhes, P. Boucaud, S. Sauvage, F. Aniel, J.-M. Lourtioz, C. Hernandez, Y. Campidelli, O. Kermarrec, D. Bensahel, G. Faini, and I. Sagnes, *Apply. Phys. Lett.* 77, 1822 (2000).
20. C. Houghton, C. J. Gibbings, C. G. Tuppen, M. H. Lyons, and A. G. Halliwell, *Appl. Phys. Lett.* 56, 460 (1990).
21. C. W. Liu, J. C. Sturm, P. V. Schwartz and E. A. Fitzgerald, *Mater. Res. Soc. Symp. Proc.* 238, 85 (1992).
22. F. Schaffler, D. Tobben, H.-J. Herzog, G. Abstreiter, and B. Hollander, *Semicond. Sci. Technol.* 7, 260 (1992).
23. T. Mizuno, N. Sugiyama, H. Satake, and S. Takagi, *Symp. VLSI Technol. Dig.* 210 (2000).
24. T. Mizuno, N. Sugiyama, T. Tezuka, T. Numata, and S. Takagi, *Symp. VLSI Technol.* 106 (2002).
25. K. Brunner, K. Eberl, and W. Winter, *Phys. Rev. Lett.* 76, 303 (1996).
26. C.-L Chang, A. St Amour, and J. C. Sturm, *Appl. Phys. Lett.* 70, 1557 (1997).
27. E. Jesson, S. J. Pennycook, J. M. Baribeau, and D. C. Houghton, *Phys. Rev. Lett.* 71, 1744 (1993).
28. O. G. Schmidt and K. Eberl, *Phys. Rev. B* 61, 13721 (2000).
29. C. W. Liu, B.-C. Hsu, K.-F. Chen, M. H. Lee, C.-R. Shie, and Pang-Shiu Chen, *Appl. Phys. Lett.* (2003).
30. O. G. Schmidt and K. Eberl, *Phys. Rev. Lett.* 80, 3396 (1998).
31. S. T. Chang, C. Y. Lin, and C. W. Liu, *J. Appl. Phys.* 92, 3717 (2002).
32. F. M. Bufler and B. Meinerzhagen, *J. Appl. Phys.* 84, 5597 (1998).
33. P. Gaworzewski, K. Tittelbach-Helmrich, U. Penner, and N. Abrosimov, *J. Appl. Phys.* 83, 258 (1998).
34. S. C. Jain, "Germanium-Silicon Strained Layers and Heterostructures," Advances in Electrons and Electron Physics, (P. W. Hawkes, Ed.), Suppl. l24, Academic, Boston.
35. P. J. Briggs, A. B. Walker, and D. C. Herbert, *Semicond. Sci. Technol.* 13, 692 (1998).
36. T. Manku. J. M. McGregor, A. Nathan, D. J. Roulston, J.-P. Noel, and D. C. Houghton, *IEEE Trans. Electron Devices* 40, 1990 (1993).
37. F. M. Bufler, P. Graf, B. Meinerzhagen, B. Adeline, M. M. Rieger, H. Kibbel, and G. Fischer, *IEEE Electron Device Lett.* 18, 264 (1997).
38. T. Manku and N. Nathan, *IEEE Trans. Electron Devices* 39, 2082 (1992).
39. D. B. M. Klaassen, *Solid-State Electron.* 35, 953 (1992).
40. D. M. Richey, J. D. Cressler, and A. J. Joseph, *IEEE Trans. Electron Devices* 44, 431 (1997).
41. M. V. Fischetti and S. E. Laux, *J. Appl. Phys.* 80, 2234 (1996).
42. R. Oberhuber, G. Zandler, and P. Vogl, *Phys. Rev. B* 58, 9941 (1998).
43. S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, *J. Appl. Phys.* 80, 1567 (1996).
44. T. Mizuno, S. Takagi, N. Sugiyama, H. Satake, A. Kurobe, and A. Toriumi, *IEEE Electron Device Lett.* 21, 230 (2000).
45. A. R. Powell, S. S. Iyer, and F. K. LeGoues, *Appl. Phys. Lett.* 64, 1856 (1994).
46. L. J. Huang, J. Chu, S. A. Goma, C. D'Emic, S. J. Koester, D. F. Canaperi, P. M. Mooney, S. A. Cordes, J. L. Speidell, R. M. Anderson, and H. S. P. Wang, *Symp. VLSI Technol.* 57 (2001).
47. M. Bruel, B. Aspar, B. Charlet, C. Malevilie, T. Poumeyrol, A. Soubie, A. J. Auberton-Herve, J. M. Lamure, T. Barge, F. Metral, and S. Trucchi, "Proceedings of the 1995 IEEE International SOI Conference," 1995, p. 178.
48. G. L. Patton, S. S. Iyer, S. L. Delage, S. Tiwari, and J. M. C. Stork, *IEEE Electron Device Lett.* 9, 165 (1988).
49. E. J. Prinz, Base Transport and Vertical Profile Engineering in Si/SiGe/Si Heterojunction Bipolar Transistor, Ph.D. Thesis, Princeton university, 1992.
50. A. Schuppen, U. Erben, A. Gruhle, H. Kibbel, H. Schumacher, and U. Konig, *IEEE Int. Electron Device Meeting Technical Dig.* 743 (1995).

51. J. Song and J. S. Yuan, *IEEE Trans. Electron Devices* 44, 915 (1997).
52. V. S. Patri and M. J. Kumar, *IEEE Trans. Electron Devices* 45, 1725 (1998).
53. V. S. Patri and M. J. Kumar, *IEE Proc. Circuits Devices Syst.* 146, 291 (1999).
54. S. T. Chang, C. W. Liu, and S. C. Lu, "ISDRS 2001," 2001, p. 490.
55. S. S. Winterton, C. J. Peters, and N. G. Tarr, *Solid-State Electron.* 36, 1161 (1993).
56. P. Rinaldi and N. Rinaldi, *Solid-State Electronic.* 41, 59 (1997).
57. L. D. Lanzerotti, J. C. Sturm, E. Stach, R. Hull, T. Buyuklimanli, and C. Magee, *Int. Electron Device Meeting Technical Dig.* 249 (1996).
58. H. J. Osten, D. Knoll, B. Heinemann, H. Rucker, and B. Tillack, *IEEE BCTM Technical Dig.* 109 (1999).
59. J. Welser, J. L. Hoyt, and J. F. Gibbons, *IEEE Electron Device Lett.* 15, 100 (1994).
60. J. Welser, The Application of Strained-Silicon/Relaxed-Silicon Germanium Heterostructures to Metal-Oxide-Semiconductor Field-Effect Transistors, Ph.D. Thesis, Stanford University, 1994.
61. J. Welser, J. L. Hoyt, and J. F. Gibbons, *Int. Electron Device Meeting Technical Dig.* 1000 (1992).
62. T. Mizuno, S. Takagi, N. Sugiyama, J. Koga, T. Tezuka, K. Usuda, T. Hatakeyama, A. Kurobe, and A. Toriumi, *Int. Electron Devices Meeting Technical Dig.* 934 (1999).
63. D. K. Nayak, J. C. S. Woo, J. S. Park, K. L. Wang, and K. P. MacWilliams, *Appl. Phys. Lett.* 62, 2853 (1993).
64. E. A. Fitzgerald, Y.-H. Xie, M. L. Green, D. Brasen, A. R. Kortan, J. Michel, Y.-J. Mii, and B. E. Weir, *Appl. Phys. Lett.* 59, 811 (1991).
65. D. K. Nayak, K. Goto, A. Yutani, J. Murota, and Y. Shiraki, *IEEE Trans. Electron Devices* 43, 1709 (1996).
66. C. K. Maiti, L. K. Bera, S. S. Dey, D. K. Nayak, and N. B. Chakrabarti, *Solid-State Electron.* 41, 1863 (1997).
67. R. Kim, J. Welser, J. L. Hoyt, and J. F. Gibbons, *Int. Electron Device Meeting Technical Dig.* 517 (1995).
68. K. Ismail, B. S. Meyerson, S. Rishton, J. Chu, S. Nelson, and J. Nocera, *IEEE Electron Device Lett.* 13, 229 (1992).
69. U. Konig and F. Schafer, *Electron. Lett.* 29, 486 (1993).
70. F. Schafer, *Semicond. Sci. Technol.* 12, 1515 (1997).
71. T. Hackbarth, G. Hoeck, H.-J. Herzog, and M. Zeuner, *J. Cryst. Growth* 201/202, 734 (1999).
72. K. Ismail, *Int. Electron Devices Meeting Techhnol. Dig.* 509 (1995).
73. E. H. C. Parker and T. E. Whall, *Solid-State Electron.* 43, 1497 (1999).
74. L. Garchery, P. Warren, I. Sagnes, and P. A. Badoz, *Mater. Res. Soc. Symp. Proc.* 379, 321 (1995).
75. S. J. Koester, R. Hammond, J. O. Chu, P. M. Mooney, J. A. Ott, L. Perraud, K. A. Jenkins, C. S. Webster, I. Lagnado, and P. R. de la Houssaye, *IEEE Electron Device Lett.* 22, 92 (2001).
76. R. A. Soref, *Proc. IEEE* 81, 1687 (1993).
77. F. Y. Huang, Shawn G. Thomas, Michael Chu, and Kang L. Wang, *Int. Electron Device Meeting Technical Dig.* 665 (1996).
78. Gianlorenzo Masini, Lorenzo Colace, Gaetano Assanto, Hsin-Chiao Luan, and Lionel C. Kimerling, *IEEE Trans. Electron Devices* 48, 1092 (2001).
79. Min Yang, J. Schaub, D. Rogers, M. Ritter, K. Rim, J. Welser, and Park Byeongju, *IEDM Technol. Dig.* 547 (2001).
80. A. Splett, B. Schuppert, K. Petermann, E. Kasper, H. Kibbel, and H. J. Herzog, *Dig. Conf. Integrated Photonic Res. (OSA Technical Dig. Ser.)* 10, 122 (1992).
81. C. W. Liu, M. H. Lee, C. F. Lin, I. C. Lin, W. T. Liu, and H. H. Lin, *Int. Electron Device Meeting Technical Dig.* 749 (1999).
82. C. W. Liu, W. T. Liu, M. H. Lee, W. S. Kuo, and B. C. Hsu, *IEEE Electron Dev. Lett.* 21, 307 (2000).
83. C. W. Liu, M. H. Lee, M.-J. Chen, C.-F. Lin, and I. C. Lin, *Appl. Phys. Lett.* 76, 1516 (2000).
84. B.-C. Hsu, C. W. Liu, W. T. Liu, and C.-H. Lin, *IEEE Trans. Electron Devices* 48, 1047 (2001).
85. B.-C. Hsu, W.-C. Hua, C.-R. Shie, and C. W. Liu, "Electrochemical Society 201st Spring Meeting," 2002, p. 662.
86. B.-C. Hsu, W.-C. Hua, C.-R. Shie, K.-F. Chen, and C. W. Liu, *Electrochem. Solid State Lett.* (2002).
87. J. Kramer, P. Seitz, E. F. Steigmeier, H. Auderset, and B. Delley, *Sens. Actuators, A* 37–38, 527 (1993).
88. L. W. Snyman, M. du Plessis, E. Seevinck, and H. Aharoni, *IEEE Electron Device Lett.*
89. X. Xiao, C. W. Liu, J. C. Sturm, L. C. Lenchyshyn, and M. L. Thewalt, *Appl. Phys. Lett.* 60, 1720 (1992).
90. A. Pirovano, A. L. Lacaita, G. Ghidini, and G. Tallarida, *IEEE Electron Device Lett.* 21, 34 (2000).
91. Min-Hung Lee, Kuan-Fu Chen, Chang-Chi Lai, Chee Wee Liu, Woei-Wu Pai, Miin-Jang Chen, and Ching-Fuh Lin, *Jpn. J. Appl. Phys. Part 2 Lett.* 41, L326 (2002).
92. Shoou-Jinn Chang, Deepak K. Nayak, and Yasuhiro Shiraki, *J. Appl. Phys.* 83, 1426 (1998).
93. N. Chand, *IEE Electron. Lett.* 29, 1800 (1993).
94. B. S. Meyerson, *Proc. IEEE* 80, 1592 (1992).
95. J. F. Gibbons, C. M. Gronet, and K. E. Williams, *Appl. Phys. Lett.* 47, 721 (1985).
96. E. Kasper, A. Gruhle, and H. Kibbel, *Int. Electron Devices Meeting Technical. Dig.* 71 (1993).
97. S. P. Murarka, *Mater. Res. Soc. Symp. Proc.* 320, 3 (1994).
98. J. B. Lasky, J. S. Snakos, O. J. Cain, and P. J. Geiss, *IEEE Trans. Electron Devices* ED-38, 262 (1991).
99. K. Maex, *Mater. Sci. Eng. R* 11, 53 (1993).
100. H. K. Liou, X. Wu, U. Gennser, V. P. Kesan, S. S. Iyer, K. N. Tu, and E. S. Yang, *Appl. Phys. Lett.* 60, 577 (1992).
101. X. Xiao, J. C. Sturm, S. R. Rarihar, S. A. Lyon, D. Meyerhofer, S. Palfrey, and F. V. Shallcross, *IEEE Electron. Dev. Lett.* 14, 199 (1993).
102. W. J. Qi, B. Z. Li, W. N. Huang, Z. G. Gu, H. Q. Lu, X. J. Zhang, M. Zhang, G. S. Dong, D. C. Miller, and R. G. Aitken, *J. Appl. Phys.* 77, 1086 (1995).
103. B. I. Boyanov, P. T. Goeller, D. E. Sayers, and R. J. Nemanich, *Appl. Phys. Lett.* 71, 3060 (1997).
104. P. T. Goeller, B. I. Boyanov, D. E. Sayers, R. J. Nemanich, A. F. Meyers, and E. B. Steel, *J. Mater. Res.* 14, 4372 (1999).
105. H. J. Huang, K. M. Chen, C. Y. Chang, T. Y. Huang, T. C. Chang, L. P. Chen, and G. W. Huang, *J. Vac. Sci. Technol., A* 18, 1449 (2000).
106. R. A. Donaton, K. Maex, A. Vantomme, G. Langouche, Y. Morciaux, A. St. Amour, and J. C. Sturm, *Appl. Phys. Lett.* 70, 1266 (1997).
107. B. I. Boyanov, P. T. Goeller, D. E. Sayers, and R. J. Nemanich, *J. Appl. Phys.* 84, 4285 (1998).
108. C. Detavernier, T. R. L. Van Meirhaeghe, F. Cardon, and K. Maex, *Thin Solid Films* 384, 243 (2001).
109. W. W. Wu, T. F. Chiang, S. L. Cheng, S. W. Lee, L. J. Chen, Y. H. Peng, and H. H. Cheng, *Appl. Phys. Lett.* 81, 820 (2002).
110. J. F. Chen and L. J. Chen, *Thin Solid Films* 293, 34 (1997).
111. J. B. Lai and L. J. Chen, *J. Appl. Phys.* 86, 1340 (1999).
112. O. Nur, M. Willander, H. H. Radamson, M. R. Sardela, G. V. Hansson, C. S. Peterson, and K. Maex, *Appl. Phys. Lett.* 64, 440 (1994).
113. H. J. Huang, K. M. Chen, C. Y. Chang, T. Y. Huang, L. P. Chen, and G. W. Huang, *J. Appl. Phys.* 88, 1831 (2000).
114. L. J. Chen, J. B. Lai, and C. S. Lee, *Micron* 33, 535 (2002).

www.aspbs.com/enn

Silicon Nanocrystals in SiO_2 Thin Layers

A. G. Nassiopoulou
IMEL, NCSR "Demokritos" Athens, Greece

CONTENTS

1. Introduction
2. Fabrication of Silicon Nanocrystals in SiO_2
3. Luminescence Properties of Silicon Quantum Dots in SiO_2
4. Charging of Silicon Nanocrystals Single Electron Devices
5. Conclusion
Glossary
References

1. INTRODUCTION

Semiconductor nanocrystals, also called quantum dots (QD), show interesting properties, which are very different from those of the corresponding bulk materials. Their electronic structure is determined by the size and the chemical environment of the nanocrystals. The electronic structure is composed of discrete energy levels into the conduction and valence band of the bulk material. By decreasing the size of the nanocrystals, these confined states are shifted to higher values, and the electronic bandgap of the material increases. The atoms occupying the chemical bonds around a nanocrystal play an important role in the determination of the effective bandgap, by introducing interface states below the conduction band minimum or above the valence band maximum of the silicon nanocrystal. So, there is significant difference in the macroscopic properties of the QDs when their chemical environment changes.

Another interesting property of semiconductor QDs is that of the controlled charging, due to the Coulomb blockade effect. When one single charge is injected into a QD, it will occupy a confined state in the nanocrystal. The energy needed to inject a single charge into the QD is the confinement energy. The presence of one charge in the nanocrystal modifies the electrostatic potential within it, so that injection of a second charge needs to overcome the semiconductor charging energy. This means that under the same electric field, there is blockade of the injection of the next charge.

In this paper, we consider the case of silicon nanocrystal layers embedded in an SiO_2 matrix for application in memory devices or in light emission. Silicon dioxide is the dielectric used in current microelectronics technology both as a gate dielectric and as an isolation layer in different process steps. The use of this material as an insulating matrix of the QDs for the mentioned applications, offers important advantages, which concern not only the compatibility with silicon integrated circuit (IC) processing but also the advantage of a high-quality dielectric material with a high electronic quality interface to silicon, showing reduced interface states.

For memory devices, the basic memory structure consists of a sandwich of layers comprising a tunneling SiO_2 layer on bulk crystalline silicon, the silicon nanocrystal layer, a second thicker SiO_2 layer on top, and a metallic cap, used as electrical contact. The optimum nanocrystal size in the case of memory devices, is in the 5–6 nm range. For efficient light emission, on the other hand, much smaller nanocrystals are needed, with sizes necessarily below 2–3 nm. The efficiency of light emission is directly proportional to the number of luminescent sites so that it is necessary to increase, as much as possible, the number of light emitting nanocrystals in the structure. In this respect, periodic multilayers of silicon nanocrystals and SiO_2 are fabricated. If the SiO_2 layer that separates the silicon nanocrystals in these periodic structures is very thin, in the nanometer range, then the structure is called a nanocrystalline silicon superlattice (nc-Si SL).

Fabrication and fundamental properties of silicon nanocrystals in SiO_2, either arranged in one layer in between SiO_2 or in a nc-Si SL will be discussed in detail in this review paper. The fundamental properties of the Si QDs for application in light-emitting and in memory devices will also be discussed.

2. FABRICATION OF SILICON NANOCRYSTALS IN SiO_2

A composite material of silicon nanocrystals, capped with SiO_2, may be fabricated by using different techniques. One simple technique consists of fabricating porous silicon by electrochemical dissolution of bulk crystalline silicon and

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

then oxidizing it to produce silicon dioxide around the nanoscale silicon structures. Fabrication of porous silicon by anodizing p-type bulk crystalline silicon had been used for the first time by Uhlir in 1956 [1]. The size of the obtained silicon nanostructures depends on the anodization conditions used [2, 3]. Further reductions in size may be achieved by oxidation. Freshly prepared porous silicon, after exposure in air, appears contaminated by a native oxide, which may cover both its external and its internal surface. Controlled oxidation may be realized by anodic oxidation in a nonfluoride electrolyte [4, 5], by chemical oxidation [6, 7], or by thermal oxidation at high temperatures [8, 9]. Most of these processes result in the formation of a Si–SiO_2 interface of poor electronic quality, the best results being achieved by thermal oxidation at temperatures between 800 and 900 °C in diluted oxygen.

The composite material obtained by oxidizing porous silicon consists of a network of SiO_2-passivated silicon nanocrystals, interconnected and randomly oriented, with an important dispersion in their sizes. Size dispersion may be reduced by increasing the oxidation time or the porosity of the starting material before oxidation. This material is well suited for light emission by photoluminescence (PL), but its application in electroluminescent devices is limited by the difficulty of injecting carriers into the Si nanocrystals.

Another technique extensively used for the fabrication of silicon nanocrystals in SiO_2 is ion implantation of Si into SiO_2, followed by annealing at high temperature. Different groups used SiO_2 films thermally grown on silicon, which were then implanted either with high [10–13] or with low [14, 15] energy ions. Implantation with high-energy ions is, in general, combined with thick silicon oxide films (in the micrometer range). The material obtained after annealing is composed of silicon nanocrystals of different sizes, randomly oriented into the SiO_2 matrix. The control on the size of the nanocrystals and the nanocrystal packing density are limited. Many groups reported visible PL from this material at room temperature [10–13, 16], and some groups fabricated structures showing visible electroluminescence (EL) [17, 18]. For EL emission, however, as well as for memory applications, to get effective carrier injection into the nanocrystals by tunneling, the silicon oxide layer has to be very thin. Silicon ion implantation at high energy results in a low density of silicon nanocrystals in the oxide.

A reliable method to obtain a two-dimensional (2-D) array of silicon nanocrystals of high packing density in a very thin silicon oxide layer on silicon, suitable for memory applications [20], is to implant it into the thin oxide, very-low-energy ions (1 KeV and below) [14–19]. This technique has been used successfully to fabricate silicon nanocrystal memories with good characteristics [20].

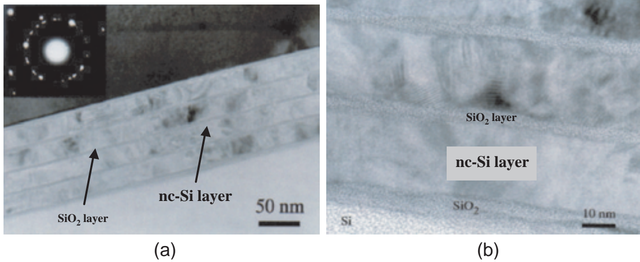
An alternative technique to ion implantation of silicon, by using ions, is ion-beam mixing of SiO_2–Si–SiO_2 layers [21]. The SiO_2 and Si layers in [21] were both amorphous and very thin (60 nm SiO_2/3 nm Si/100 nm SiO_2). They were deposited by ion sputtering of crystalline Si, by using Ar^+ ions for the deposition of amorphous Si (α-Si) and O_2^+ ions in an oxygen ambient for the deposition of α-SiO_2. Ion-beam mixing was performed by 80 KeV argon ions, and it was followed by annealing at 1100 °C.

Other techniques used to fabricate Si nanocrystals in SiO_2 include laser ablation of silicon targets in diluted oxygen gas under vacuum [22], or in an inert He ambient and subsequent exposure of the films to oxygen [23, 24]. Laser breakdown of silane gas in a vacuum chamber also was used [25, 26], followed by oxidation. Electron-stimulated desorption of SiO_2 in an ultrahigh vacuum field-emission transmission electron microscope was used to fabricate size- and position-controlled silicon nanocrystals [27].

SiO_2-capped silicon nanocrystals may be obtained from high-temperature aerosols produced, for example, by pyrolizing disilane [28–30]. A diluted aerosol of surface-oxidized Si nanocrystals is produced, which is bubbled through ethylene glycol to create a nanocrystalline colloid. Size-selective precipitation with tetrahydrofuran is used to reduce size dispersion of the nanocrystals (NCs). This technique is very interesting, since it has produced Si NCs with the highest PL efficiency in Si (quantum efficiency of 50% at low temperatures). However, the technological interest of this technique is limited, due to the very low production rate of the nanocrystals.

Two other promising techniques widely used for the fabrication of a 2-D layer of Si NCs in SiO_2 or a nanocrystalline silicon SL are: (a) direct current (DC) or radiofrequency (rf)-magnetron sputtering [31–37] and (b) low-pressure or plasma-enhanced chemical vapor deposition (LPCVD or PECVD) of silicon on an oxidized silicon wafer [31, 37–48]. Lu et al. [49] and Sullivan et al. [50] deposited ultrathin α-Si layers separated by α-SiO_2, with the aim to crystallize the silicon layer by annealing in a second step. However, crystallization of these thin silicon layers proved to be very difficult and even impossible for the thinnest layers [33]. This has been attributed to the high-strain fields within the layers. Crystallization has been achieved only for the larger layer thicknesses [31, 32], leading to relatively large Si NCs. A more complicated process of solid-phase crystallization of 4- to 20-nm-thick silicon layers, confined between amorphous SiO_2 layers, was used [33] to produce smaller NCs in the SL. It consisted of growing a disordered silicon–α-SiO_2 superlattice on silicon by rf-magnetron sputtering and plasma oxidation, followed by a two-step annealing process: rapid thermal annealing at 900 °C for 60 s, followed by quasi-equilibrium furnace annealing starting from 750 °C up to 1100 °C.

Silicon nanocrystal layers in SiO_2 may be produced in a more controllable way by deposition of a thin silicon layer on a thermally grown silicon oxide, by using LPCVD, followed by high-temperature oxidation [38, 39–41]. Silicon nanocrystals of sizes as small as 1–2 nm may be obtained, arranged in a 2-D layer in between SiO_2. By successive runs of silicon deposition and oxidation, multilayers and superlattices may be fabricated as shown in Figures 1–3. Figure 1 [39] is a TEM image of a multilayer structure with relatively large silicon crystals (10- to 20-nm thick), separated by thin SiO_2 layers. Figure 2 [41], shows a layer of smaller NCs in-between SiO_2 (sizes 2–4 nm), which were produced by oxidizing a much thinner silicon layer (12-nm thick). Figure 2a is a dark field, while Figure 2b is a bright field TEM image. In Figure 3, we see a nanocrystalline silicon superlattice, fabricated by successive runs of the process used in Figure 2. Figure 4a and b shows high-resolution TEM images of silicon nanocrystals in SiO_2, fabricated by the same technique,



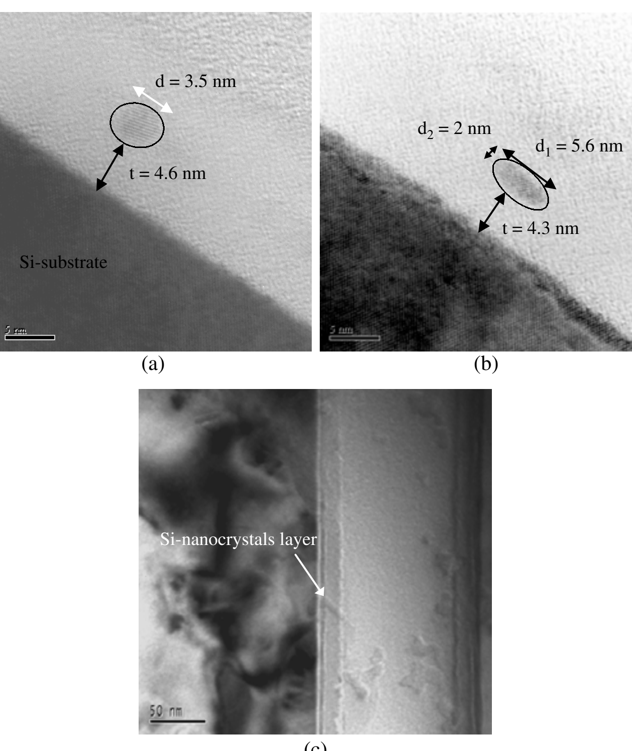
(a) (b)

Figure 1. TEM images of Si–SiO_2 multilayers. The silicon nanocrystal layers are ≅20 nm thick and are separated by thinner SiO_2 layers. They were fabricated by successive runs of LPCVD silicon deposition and high temperature thermal oxidation [39]. The electron diffraction pattern from the nanocrystalline layers also is shown as an inset in (1a), while the image in (1b) was taken with higher magnification. Reprinted with permission from [39], P. Photopoulos et al., *Mater. Sci. Eng., B* 69–70, 345 (2000). © 2000, Elsevier.

while Figure 4c shows a bright field image of the same layer in between a tunneling oxide (bottom) and a thicker oxide on top. This structure is adequate for charge injection from the substrate and charge storage in memory applications.

This process of LPCVD deposition and oxidation is fully compatible with complimentary metal-oxide-semiconductor (CMOS) processing. The nanocrystals are capped with a high-quality silicon dioxide layer, with a reduced density of interface states at both the silicon substrate–SiO_2 interface and the interfaces of silicon nanocrystals with SiO_2. The only drawback of the technique is the complexity of the process in the case of SL fabrication (silicon deposition in the LPCVD system and *ex-situ* furnace oxidation for each superlattice period).

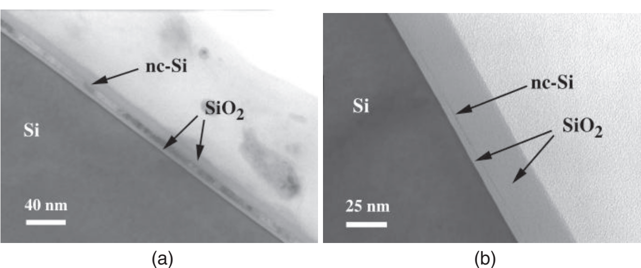
The material obtained by LPCVD deposition and oxidation is appropriate both for light-emission devices (LED) and for memory devices. As mentioned in the introduction, for efficient light emission, very small NCs are needed, of sizes preferably below ≅2 nm. The superlattice structure is used to increase the number of light-emitting sites. For memory devices, slightly larger NCs than those in light emission give optimum results (sizes in the range of 5–6 nm). In general, one single layer of silicon nanocrystals on a thin tunneling silicon oxide is used. In this case, to increase the number of silicon nanocrystals participating in the memory effect for a given device surface area, one has to increase



(a) (b)
(c)

Figure 3. (a)(b) High-resolution TEM images of silicon nanocrystals in a 2-D layer in between SiO_2. The SiO_2 layer between the silicon substrate and the silicon nanocrystal layer is 4.3- to 4.6-nm thick. The diameter of the nanocrystal in (a) is 3.5 nm, while in (b), the nanocrystal is not spherical (d1, d2 equal 5.6 nm and 2 nm, respectively). In (c), the silicon nanocrystal layer in between SiO_2 is shown at lower magnification (A. G. Nassiopoulou, A. Travlos, and V. Ioannou-Sougleridis, unpublished results).

the packing density of the NCs in the layer, while keeping them isolated from each other. In this respect, different processes have been used. One method consists of modifying the surface chemistry of the SiO_2 layer before silicon deposition and in creating a high density of preferential nucleation sites for the growth of the NCs [51]. This is achieved by surface treatment of the oxidized surface in diluted HF to create surface silanol bonds (Si–OH) on SiO_2, which act as



(a) (b)

Figure 2. A 2-D layer of silicon nanocrystals in between SiO_2, produced by growing a thin silicon oxide on silicon, on which a silicon layer, 12-nm thick was deposited by LPCVD at 580 °C, 300 mTorr, and oxidized at 900 °C for 40 min. Two different magnifications are shown in (a) and (b) Reprinted with permission from [40], P. Photopoulos et al., *Appl. Phys. Lett.* 76, 3588 (2000). © 2000, American Institute of Physics.

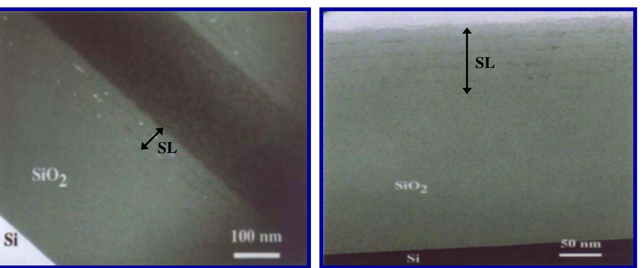


Figure 4. TEM images of a nanocrystalline silicon superlattice with 4 periods, fabricated by successive runs of silicon deposition and oxidation, by using the same conditions as in Figure 2. A dark field TEM image of silicon nanocrystals in shown in (a), while (b) is a bright field TEM image (A. G. Nassiopoulou, A. Travlos, and P. Photopoulos, unpublished results).

nucleation sites for silicon nanocrystal formation by LPCVD [47, 48]. In non-HF treated surfaces, nucleation sites on SiO_2 probably are generated by the thermal dissociation of surface Si–O bonds, their density is, however, at least one order of magnitude lower than that of HF-treated surfaces. Another technique applied to promote lateral self-organization of Si QDs during CVD deposition of Si on a very thin SiO_2 layer was to use a buried array of dislocations, located at the bonding interface between an ultrathin silicon film and the silicon substrate [51]. Self-organized vertical QDs were produced in this way.

More recently, a technique consisting of seeding the SiO_2 surface with a submonolayer of silicon prior to chemical vapor deposition (CVD) was proposed [52]. An increase in Si NC density by one order of magnitude, compared to the nonseeded surface, may be achieved.

3. LUMINESCENCE PROPERTIES OF SILICON QUANTUM DOTS IN SiO_2

3.1. Photoluminescence

The different techniques described previously for the fabrication of SiO_2-capped silicon nanocrystals, result, in general, in a composite material composed of an amorphous SiO_2 matrix in which the Si nanocrystals are embedded. Depending on the technique used, the material is different as to the silicon nanocrystal sizes, size dispersion and their distribution and density in the matrix, the thickness and the quality of the SiO_2 film surrounding each nanocrystal, the strain fields within the material, etc. All these parameters significantly influence the PL properties of the material.

The following luminescence bands were, in general, obtained from silicon nanocrystals embedded in SiO_2 (see an example in Fig. 5):

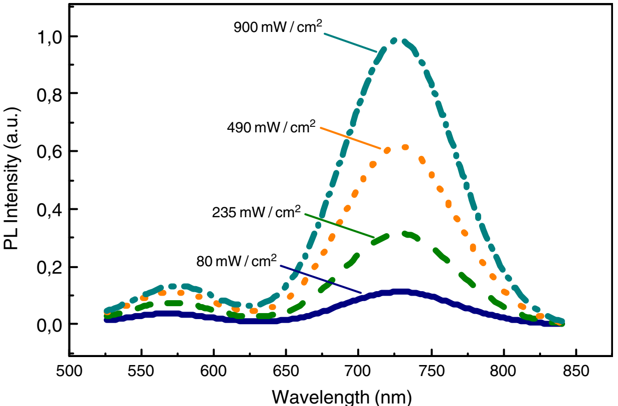


Figure 5. PL spectra from the nanocrystalline silicon superlattice of Figure 4 at different laser excitation intensities, obtained by using the laser excitation line at 457.9 nm. One broad band with a peak at around 720 nm is observed, due to quantum confinement in silicon nanocrystals and a yellow-green band at ≅575 nm, which is related to defects in SiO_2. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

3.1.1. Defect-Related Blue-Green Band

This band is, in general, observed in samples grown by ion implantation of Si ions into SiO_2 or in samples prepared by CVD of silicon and oxidation, when the oxide is grown at relatively low temperatures, and it is not post-annealed at higher temperatures [39, 40, 53]. This is a weak PL band (see Fig. 5), and it has been observed both by PL and by EL. It is rapidly quenched under continuous wavelength (cw) laser illumination or under current stressing [39]. Its intensity increases linearly with increasing laser intensity. It does not saturate at high laser power. Its decay time is in the nanosecond range [40], and it can be resolved from the tail of the red PL band by gated photon counting techniques [39].

Due to the above characteristics, the red–green PL band is attributed to defects in SiO_2 or at the Si–SiO_2 interface, and it does not show any significant technological interest.

3.1.2. Tunable Red PL Band

3.1.2.1. General Characteristics This band is much more intense, and it is stable with time and under laser illumination. For large nanocrystals, the maximum of the emission is at about 900–950 nm, while it is blue shifted down to 680–700 nm by decreasing the size of the nanocrystals. This band is attributed to quantum confinement of electron-hole pairs in the nanocrystals. It has the general characteristics of PL observed in other low-dimensional silicon structures [54–70], starting from porous silicon in 1990 [54, 55] and including both zero-dimensional [54, 55, 59–70] and one-dimensional structures [56–58]. These characteristics are: (i) the tunability of the emission by changing the size of the nanostructure, (ii) the high quantum yield at room temperature, which is orders of magnitude higher than that of crystalline silicon (c-Si), and is due first, to the physical isolation of superimposed electron-hole pairs, and second, to the remarkably efficient suppression of nonradiative recombination, and (iii) the broad and featureless PL spectrum.

Suppression of nonradiative recombination in confined structures is due to the physical isolation of electron-hole pairs. This isolation suppresses Auger recombination at moderate excitation densities because a rare defect existing in the material is confined to just one nanocrystal containing the defect, so that nonradiative quenching of the PL is limited, and the luminescence is enhanced.

Photoluminescence tunability with the nanocrystal size is related to the important bandgap opening when the size is below ≈3–4 nm.

In the case of hydrogen-terminated Si nanocrystals, the nanocrystal bandgap E_g as a function of the nanocrystal diameter is fitted by the equation [71, 72]:

$$E_g(d) = 1.167 + \frac{88.34}{d^{1.37}}, \quad (eV)$$

where the effective diameter d is expressed in angstroms.

Experimental results from hydrogen-terminated silicon nanocrystals [73, 74] showed that PL emission follows the bandgap opening. In an experiment under controlled conditions of hydrogenation of porous silicon films, it was demonstrated that when the nanocrystal size is decreased, PL is tuned from the near infrared (IR) to the ultraviolet (UV) wavelength.

The size dependence of PL from silicon nanocrystals capped with SiO_2 was first demonstrated experimentally by Wilson et al. [70], by using size-selective precipitation and size-exclusion chromatography to separate the silicon nanocrystals from larger crystallites and aggregates. The optical spectra and photophysics of the nanocrystals provided direct evidence for quantum confinement. Quantum yields as high as 5.6% at room temperature and 50% at temperatures below 50 K were measured, which approach the yields obtained from direct bandgap semiconductors. However, their data indicated that although these nanocrystals had high luminescence yields, they behaved fundamentally as indirect bandgap materials with low oscillator strength. The measured high-emission quantum yields were attributed to the suppression of nonradiative recombination and the extremely efficient nanocrystal oxide passivation. The average lifetime of PL from their nanocrystals at 630 nm was found to be 50 μsec at room temperature and about 2.5 msec at 20 K. These values are typical for indirect gap semiconductors [73–75].

Photoluminescence tunability by size tunability of silicon nanocrystals in SiO_2 also was demonstrated in [41], where arrays of silicon nanocrystals of different sizes, arranged in a layer in-between SiO_2, were obtained by depositing a thin LPCVD silicon layer on a thermally grown silicon oxide on bulk silicon, followed by oxidation at a high temperature. The size of silicon nanocrystals was controlled by controlling the oxidation time. Silicon nanocrystal superlattices were obtained by successive cycles of silicon deposition and oxidation. By varying the size of nanocrystals from 3.5 to 1.5 nm, the PL peak intensity was shifted from 800 to 680 nm.

Tunable PL from silicon nanocrystals in Si–SiO_2 superlattices also was reported in [44]. The Si–SiO_2 superlattice was produced by depositing substoichiometric silicon oxide by PECVD, followed by high-temperature annealing. The nanocrystal size distribution was, in this case, tuned either by varying the annealing temperature or by changing the excess silicon content in the deposited layer. Tuning of the PL signal from 910 nm for a nanocrystal diameter of 4.6 nm to $\cong$800 nm for a nanocrystal diameter of 2.2 nm was reported.

Compared to hydrogen-terminated silicon nanocrystals, where the PL emission was tuned experimentally from the near IR to the UV, in the case of SiO_2–capped silicon nanocrystals, the emitted wavelengths do not go below 650–680 nm [41]. Below these wavelengths, there is no PL. The existence of this limit suggests that the luminescence is not related to free exciton recombination involving only silicon nanocrystal confined states in the conduction and valence band of the nanocrystal, but some localized states are involved in the recombination mechanism.

The existence of localized states also has been considered in a theoretical study for porous silicon by Allan et al. [75] to account for the large energy difference observed experimentally between absorption and luminescence. Porous silicon is a material composed of a network of silicon nanocrystals or nanowires terminated by hydrogen or silicon oxide or –OH, depending on whether the material is freshly prepared, aged in air, or oxidized. By using total energy calculations, they demonstrated the existence of self-trapped excitons at some surface bonds of silicon nanocrystals. These bonds are, for example, dimer bonds passivated by hydrogen atoms or by silicon oxide. Stretching of the covalent Si–Si bond stabilizes the self-trapped exciton. They attributed part of the observed luminescence to such surface states, while optical absorption is characteristic of quantum confinement, and it follows the bandgap opening when the nanocrystal size is decreased.

Figures 6, 7 and 8 illustrate the results of their calculations [75]. Figure 6 is a schematic configuration coordinate diagram, showing the energies of the ground state G, the normal excitonic state E, and the self-trapped exciton state as a function of the configuration coordinate Q, which corresponds to the stretching of the covalent bond. For a small Q, the ground and first exited states are delocalized over the crystallite and show a normal parabolic behavior. For Q larger than a critical value Q_c, the system localizes the electron-hole pair on one particular bond. Such a self-trapped exciton is likely to be favored at surfaces of nanocrystals where the elastic response of the environment is likely to be weaker than in the bulk. Calculations showed that in the case of Si–H surface bonds, it is always possible to trap an exciton when these bonds are sufficiently stretched. The stable situation corresponds to the broken bond that leads to hydrogen desorption. In the case of stretching of the Si–Si bond of a surface dimer, the stable situation for the excited state corresponds to the surface silicon atoms almost returned to their original lattice sites (see Fig. 7 from [75]).

Calculations by Allan et al. [75] were extended to the case of dimers at the Si–SiO_2 interface. It has been shown that the exciton can be trapped on the dimer. The Si–Si dimer distance was found to be 3.24 Å in the self-trapped state and the optical gap equal to 1.51 eV.

Later, a theoretical study of the electronic states of SiO_2-capped Si QDs has been published by the same group [73], combined with experimental results on the PL from oxidized

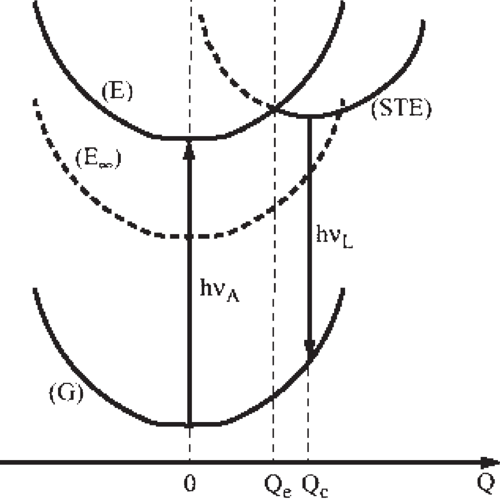


Figure 6. Schematic configuration coordinate diagram showing the energies of the ground state (E) and the self-trapped exciton state (STE) in silicon nanocrystals, The curve (E_∞) corresponds to a very large crystallite with no blue shift, showing that the STE state might not exist for large crystallites. Reprinted with permission from [75], G. Allan et al., *Phys. Rev. Lett.* 76, 2961 (1996). © 1996, American Physical Society.

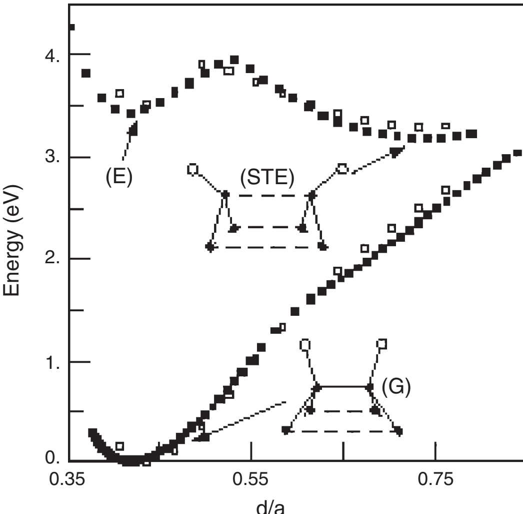


Figure 7. Total energy (full symbols: tight binding; empty symbols: LDA) of a spherical crystallite with 29 silicon atoms in the ground state and in the excitonic state as a function of the dimer interatomic distance d ($\alpha = 0.54$ nm). The arrows indicate the energy minima. Schematic side views of the cluster surface dimer in the ground state (G) and in the self-trapped state (STE) are also shown (o = hydrogen, • = silicon). Reprinted with permission from [75], G. Allan et al. *Phys. Rev. Lett.* 76, 2961 (1996). © 1996, American Physical Society.

porous silicon, obtained by a group from the University of Rochester [73]. The results of this study showed that the recombination mechanism in oxidized nanocrystals is different from that in hydrogen-passivated nanocrystals. It is proposed that in oxide-passivated nanocrystals, the carriers are trapped in an Si=O double bond, which is formed when a silicon nanocrystal is oxidized to stabilize the Si–SiO_2 interface. This bond does not require either large deformation energy or an excess element. It also terminates two dangling bonds. Depending on the size of the nanocrystals, the Si=O surface state is stable or unstable. In large nanocrystals, (size $\geq$3 nm), the bandgap is not wide enough to stabilize the Si=O state, and recombination occurs via free excitons, as in the case of hydrogen-terminated silicon nanocrystals. In the range of sizes between 1.5 and 3 nm, a p-type localized state on the silicon atom of the Si=O bond traps the electron, so recombination involves a free hole and a trapped electron. For sizes below $\approx$1.5 nm, both a p state localized on the Si atom and a p state localized on the oxygen atom are found, which localize, respectively, the hole and the electron. In this case, recombination is via trapped excitons. The calculated electronic states as a function of silicon nanocrystal size are illustrated in Figure 9 (from [75]).

The above simple model explains the red shift in the PL observed in oxidized silicon nanocrystals compared to hydrogen-terminated ones.

A detailed experimental investigation of PL from silicon nanocrystals in nc-Si–SiO_2 multilayers and superlattices, fabricated by successive runs of silicon deposition and high-temperature thermal oxidation, is found in [39–41] and [76]. Limited tunability of PL emission with silicon nanocrystal size was observed, and a clear red shift of PL, compared to the wavelength corresponding to the calculated nanocrystal bandgap. This discrepancy was more pronounced in the case of smaller nanocrystals. This result is in agreement with the theoretical calculations previously described. An example of PL spectra obtained at 70 and 300 K is given in Figure 10 (from [76]).

Another interesting result in [40] is the superlinear increase of PL intensity by increasing the number of periods in the superlattice. This is illustrated in Figure 11, where the PL signal from SLs with 1 to 5 periods is shown in (a), while,

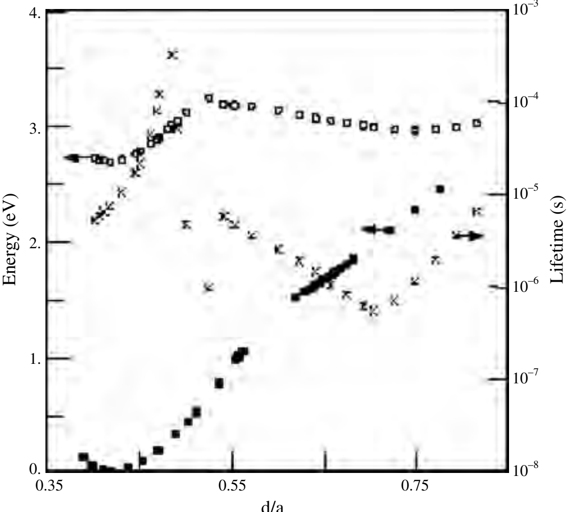


Figure 8. Energy of a spherical crystallite with 123 silicon atoms (diameter = 1.67 nm) in the ground state (■) and in the excitonic state (o) as a function of the interatomic distance d ($\alpha = 0.54$ nm). Crosses are the radiative lifetimes of the excitonic state. Reprinted with permission from [75], G. Allan et al., *Phys. Rev. Lett.* 76, 2961 (1996). © 1996, American Physical Society.

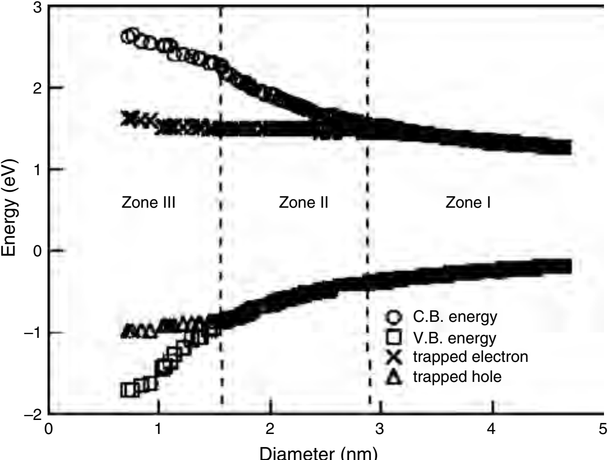


Figure 9. Electronic states in Si nanocrystals as a function of cluster diameter and surface passivation. When the silicon cluster is passivated by hydrogen, recombination is via free excitons for all sizes (symbols: “oo” for conduction band and “□□” for valence band). If the silicon cluster is passivated by oxygen, a stabilized electronic state may be formed on the Si=O bond. The trapped electron state is a p state, localized on the Si atom of the Si=O bond, and the hole state is a p state localized on the oxygen atom. Reprinted with permission from [73], M. V. Wolkin et al., *Phys. Rev. Lett.* 82, 197 (1999). © 1999, American Physical Society.

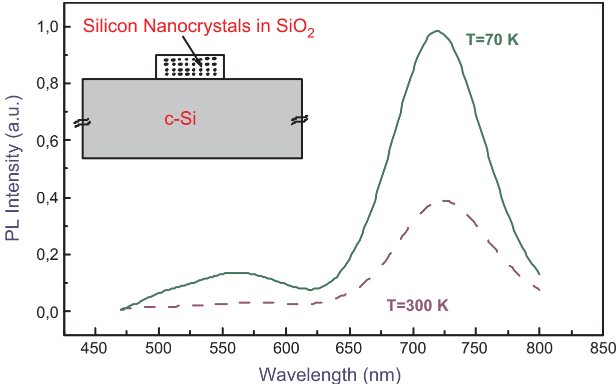


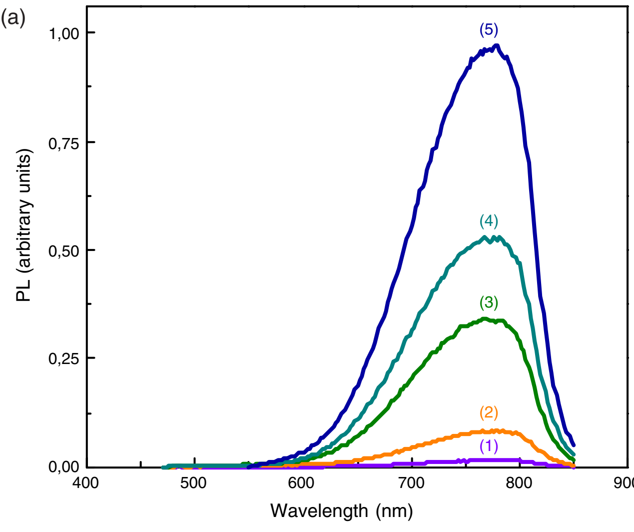
Figure 10. PL spectra from silicon nanocrystals in nanocrystalline Si–SiO_2 superlattice at two different temperatures. The size of silicon nanocrystals was ≅1.5 nm. A laser line at 457.9 nm, and an excitation intensity of 30 mW/cm^2 are used. Inset: schematic representation of the nanocrystalline silicon superlattice. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

in (b), we see the PL peak intensity variation as a function of the number of periods. This interesting result may be attributed either to annihilation of defects during processing for the fabrication of the next SL period or to a decrease of silicon nanocrystal sizes and narrowing of their size dispersion after prolonged processing and thermal treatment. Smaller, defect free, silicon nanocrystals are more efficient light emitters. The hypothesis of decreasing the sizes of NCs in the first layers during formation of the next bilayers is evidenced in Figure 3.

3.1.2.2. Recombination Dynamics The recombination dynamics in silicon nanocrystals embedded in SiO_2 depend strongly on the characteristics of the sample used. This is because the radiative lifetime depends on silicon nanocrystal size, size dispersion, shape [77], and separation distance between the nanocrystals. As a result, different authors gave different experimental results for the radiative lifetimes of silicon nanocrystals. From Si nanocrystals produced by laser breakdown of silane gas [25], which were composed of a crystalline silicon core covered by a 1.6-nm-thick amorphous SiO_2 surface layer, the red PL decay profiles were described by a stretched exponential function, as follows:

$$I(t) = I_0\left(\frac{\tau}{t}\right)^{1-\beta} \exp\left[\left(-\frac{t}{\tau}\right)\right]^{\beta},$$

where τ is an effective decay time, β is a constant between 0 and 1, and I_0 also is constant. This stretched exponential decay usually is observed in the PL decay of disordered systems. It has been systematically observed in the case of porous silicon [78–80], and it was attributed either to size, shape, or crystallographic orientation [81] dispersion of the nanostructures. In the case of silicon nanocrystals in SiO_2 fabricated by ion implantation of silicon [62, 63], strongly nonexponential PL decay was obtained, which was associated with a nanocrystal–nanocrystal interaction. Multiexponential PL decay also was obtained from size-selected, surface-oxidized nanocrystals fabricated by size-selective precipitation and size-exclusion chromatography



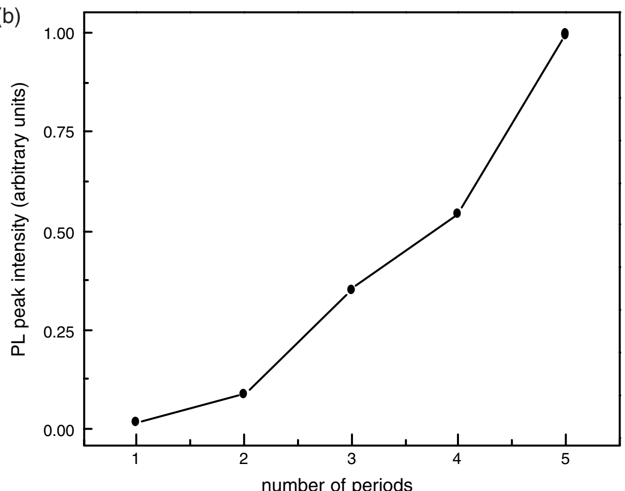


Figure 11. (a) PL spectra from nanocrystalline SiO_2 superlattices with one to five periods. (b) PL peak intensity as a function of the number of periods. A superlinear increase in PL intensity with the increase of the number of periods is obtained. Reprinted with permission from [40], P. Photopoulos et al., *Appl. Phys. Lett.* 76, 3588 (2000). © 2000, American Institute of Physics.

[82]. Since in that case, the size distribution was limited, the multiexponential PL behavior was attributed to the distribution of shapes and decay dynamics of nanocrystals for a given emission wavelength. The 630-nm average lifetime was found to be 50 nm at 293 K and about 2.5 nm at 20 K. On the other hand, single [44] or double [76] exponential PL was found in the case of well-separated silicon nanocrystals in Si–SiO_2 superlattices. This behavior was attributed to better size control and larger nanocrystal separation, which limited the nanocrystal–nanocrystal interaction. For example, in [44], it was found that for a nanocrystal diameter equal to 2.2 nm, the PL band was peaked at 790 nm and the radiative lifetime was in the range of 0.8 ms at 17 K and 0.25 ms at 250 K. The radiative rate, calculated by dividing the PL intensity by the decay time, was found to increase monotonically by a factor of 4 when the temperature was increased from 17 to 300 K. This increase has been explained by considering the model of Calcott et al. [83]for porous silicon.

According to this model, the exchange electron-hole interaction splits the excitonic levels by an energy Δ. The lowest level is a triplet state and the upper level is a singlet state, the triplet state having a radiative decay rate much smaller than that of the singlet. Once excited, carrier population in this state will be distributed according to a thermal equilibrium law. The temperature dependence of the radiative decay is, thus, deduced, and it is found that by increasing the temperature, the relative population of the singlet state, compared to the triplet state, will increase. As a result, since the radiative rate of the singlet state is much higher than that of the triplet state, the total radiative rate will increase with the increase of the temperature.

Slightly different behavior was found in the case of silicon nanocrystals of sizes below $\approx$1.5 nm, arranged in multilayers in between SiO_2, and studied in detail in [76]. The PL peak intensity was found to increase by increasing the excitation intensity, as illustrated in Figure 12. This increase was linear at low excitation intensity, and it saturated at higher intensities, both under cw and under chopped excitation. The PL saturation was observed both at room and at low temperatures (Fig. 13). This saturation is typical of systems with a finite number of light-emitting centers. In PL from silicon nanocrystals due to exciton recombination, PL saturation is attributed to the switching on of the fast nonradiative Auger recombination channel, which is due, at high excitation rates, to an additional carrier added to the first electron-hole pair within the nanocrystal [83–85].

The temperature dependence of PL intensity is different when we operate at low excitation than in the case of high excitation. As it was found experimentally in [76], at low

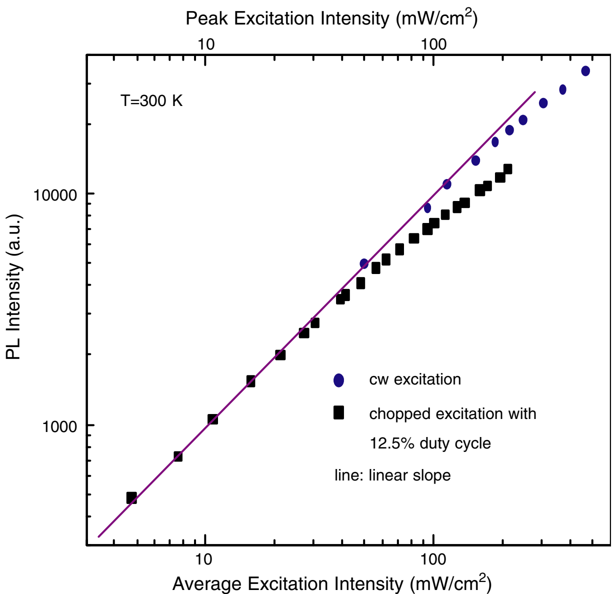


Figure 12. PL intensity from silicon nanocrystals in a nanocrystalline silicon superlattice as a function of the excitation intensity under continuous wave (cw) and chopped excitation with 12.5% duty cycle. The line is the linear slope at low excitation intensity. At high excitation intensity, there is saturation of PL. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

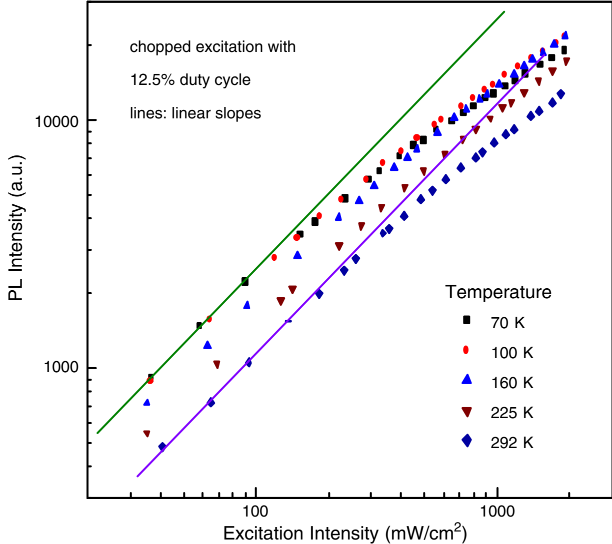


Figure 13. Same as in Figure 12 for different temperatures. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

excitation, PL intensity was increased by decreasing the temperature, and it saturated at temperatures below ~180 K. Under high excitation, PL intensity increased as the temperature was decreased down to 180 K, and it then decreased again at lower temperatures (Fig. 14).

The thermally activated behavior of PL emission is better understood if we consider PL intensity in combination with PL relaxation times. It was found that the PL decay time does not follow the stretched exponential law observed in the case of porous silicon, but it is fitted by two exponentials with relaxation times at room temperature equal, respectively, to 25–30 μsec (τ_F) and 80–100 μsec (τ_S). Typical PL transients at different temperatures are shown in Figure 15. By examining separately each term of the double exponential function of PL, it was found that while the temperature dependence of the two decay times was similar

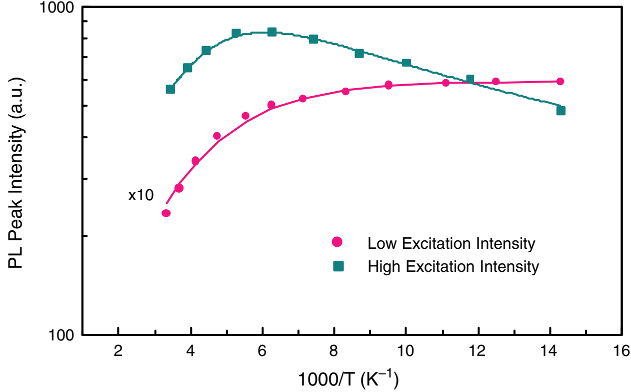


Figure 14. PL peak intensity from silicon nanocrystals of an average size of ≅1.5 nm from a nanocrystalline $Si–SiO_2$ superlattice as a function of the inverse temperature under low (open circles) and high (dark squares) laser power excitation. The solid lines are fittings. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

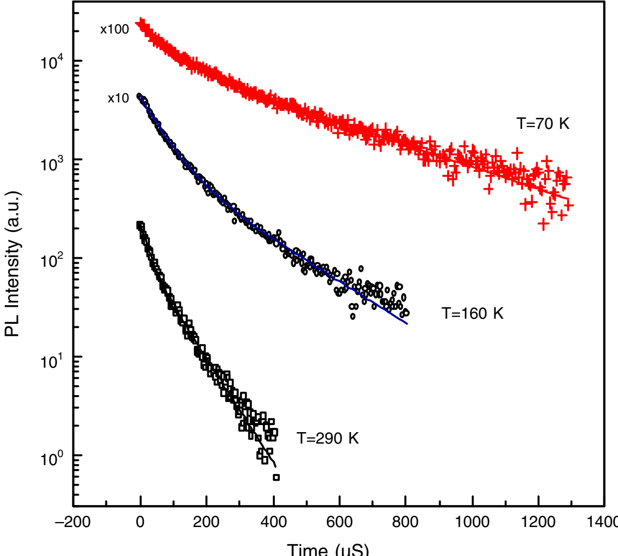


Figure 15. Typical PL decay times from silicon nanocrystals, ≅1.5 nm in size, at different temperatures. Lines are fitted by using a double exponential function. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

(Fig. 16), the PL amplitude of each component showed different temperature dependence. The PL amplitude of the slower component, I_{PLS}, with decay time τ_S was independent of temperature, while that of the faster component I_{PLF}, with decay time τ_F, showed a maximum at around ~160 K. From room temperature down to 160 K the variation of I_{PLF} followed that of the decay time τ_F, while at lower temperatures, down to 70 K, I_{PLF} decreased (Fig. 17).

The increase of I_{PLF} from 70 to 160 K suggests that its excitation is thermally activated with small activation energy.

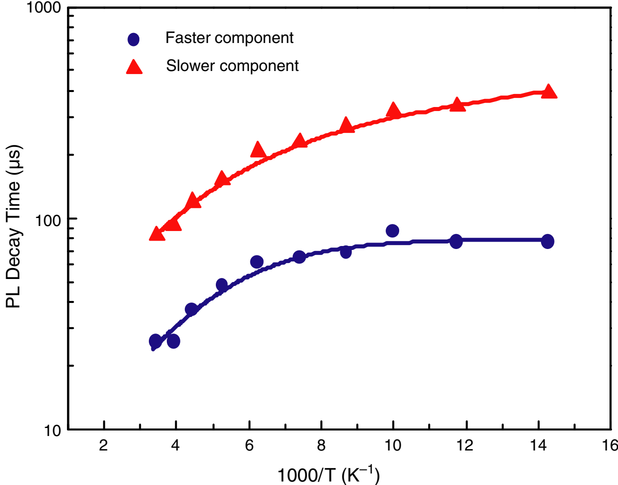


Figure 16. Variation of the two different components of PL decay times from ≅1.5 nm silicon nanocrystals in a nanocrystalline Si–SiO_2 superlattice as a function of the inverse of temperature. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

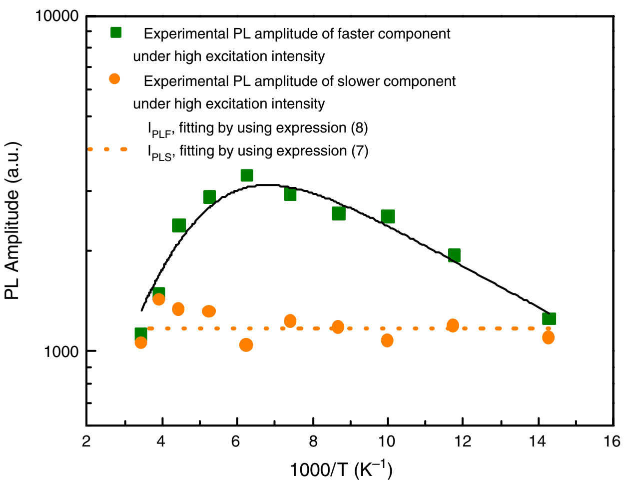


Figure 17. PL amplitude as a function of the inverse temperature for the two PL components obtained from 1.5-nm silicon nanocrystals. Reprinted with permission from [76], B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001). © 2001, American Institute of Physics.

However, its decrease, by increasing the temperature in the range 160–300 K, which directly follows the decrease of the corresponding decay time, suggests that for this recombination channel, in this temperature range, Auger recombination is absent. This component was attributed to self-trapped excitons, related to Si–Si dimers at the surface or to the Si–SiO_2 interface of the very small silicon nanocrystals under investigation (size ≈1.5 nm). It also was suggested that the formation of self-trapped excitons on Si–Si dimers was in competition with the recombination mechanism of other active centers in the system, resulting in an apparent absence of Auger recombination in the case of this PL component. The slower PL component was ascribed to radiative recombination involving localized interface states in Si═O bonds [73, 76, 84].

3.1.2.3. Nature of Absorbing and Emitting States in Si Nanocrystals Bulk crystalline silicon is an indirect bandgap material. As a result, optical transitions are allowed only if phonons are absorbed or emitted to conserve the crystal momentum (phonon-assisted (PA) process). The relevant phonon modes include transversal optical (TO) phonons (E_{TO} = 56 meV), longitudinal optical (LO) phonons (E_{LO} = 56 meV), and transversal acoustic (TA)

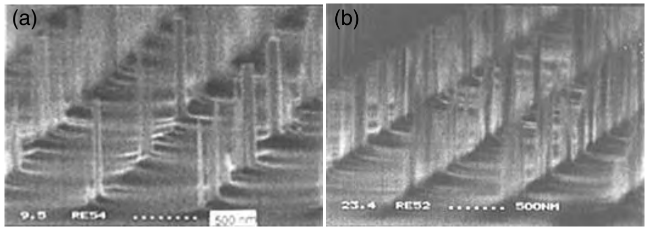


Figure 18. Vertical silicon pillars on a silicon substrate fabricated by using lithography and highly anisotropic silicon etching. The etching process was based on SF_6 and CHF_3 gases (S. Grigoropoulos and A. G. Nassiopoulou, unpublished results).

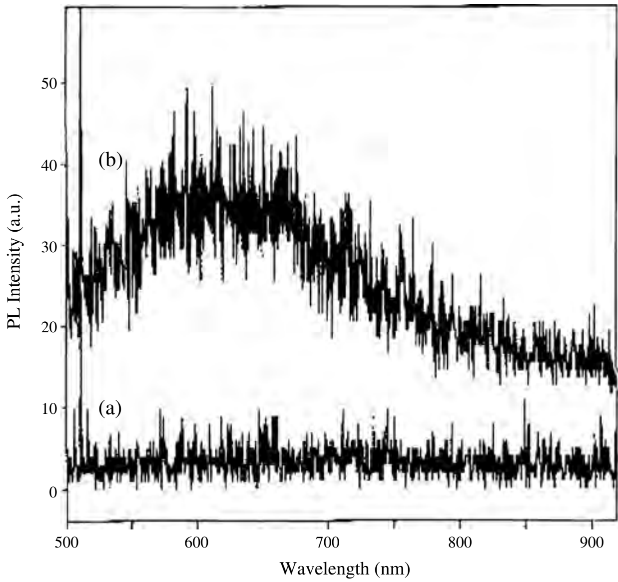


Figure 19. (a) PL spectrum from silicon nanowires, fabricated by using lithography and silicon etching. After etching, different steps of high temperature oxidation at 900 °C and oxide removal were performed to reduce the diameter of the nanowires below ≅4 nm. (b) Zero spectrum from the silicon wafer in an area outside the pillars. Reprinted with permission from [103], A. G. Nassiopoulou et al., *Appl. Phys. Lett.* 66, 1114 (1995). © 1995, American Institute of Physics.

phonons ($E_{TA} = 56$ meV). In a nanocrystal, the situation is quite different. First, due to spatial confinement, both the absorbing and luminescing states are shifted to higher energies, thus the density of electronic states is changed, and second, this spatial confinement results in delocalization of carriers in K space, thus allowing no-phonon (NP) optical transitions to occur and significantly enhancing the oscillator strength of these transitions [86, 87]. A third result is that due to the better overlap of electron and hole envelope wavefunctions, one can expect a strong enhancement of the exchange interaction, inducing a splitting of the exciton levels [88].

The clear evidence that emitting states are driven to higher energies by confinement is coming from PL measurements. Photoluminescence tunability is obtained by changing the silicon nanocrystal size. Due to size distribution of the nanocrystals, a broad and featureless spectrum is obtained. A more detailed understanding of the nature of the absorbing and luminescing states in silicon nanocrystals is obtained from resonant PL emission (PLE) studies [89, 90], where the excitation energy approaches the detection energy. Under these conditions, only a small subset of nanocrystals is probed. It has been observed that the PL spectrum is no longer featureless, but it exhibits a set of onsets, whose energy separation corresponds to the energy of TO- and TA-momentum-conserving phonons of bulk silicon [89, 90]. Moreover, it has been found that phonon-related spectral features of the PLE signal are replicated in energy. This behavior is explained as follows: Spatial confinement partially breaks down the K-conservation rule. This allows NP transitions to be possible, and, therefore, in both absorption and emission, NP and PA processes coexist. At each particular energy, two groups of nanocrystals are contributing, one having an energy gap $E_{gap} = E_{det}$ and the other with a bandgap $E_{gap} = E_{det} + E_{TA}$ and $E_{gap} = E_{det} + E_{TO}$. The same is valid for the absorption, and this is why the related spectral features in PLE are replicated in energy [90].

Theoretically, it is predicted that the probability of NP transitions is increased by decreasing the size of the nanocrystals. Therefore, the optical properties of Si nanocrystals are governed by a competition between indirect and quasi-direct recombination channels. By using extremely energy-selective optical spectroscopy, Kovalev et al. [90] estimated the relative strengths of the NP transitions and phonon-assisted processes. It was found that in the vicinity of the crystalline silicon bandgap, in the weak confinement regime, TO-phonon-assisted processes dominate, while at 1.8 eV, NP processes begin to take over. The ratio between TO- and TA-phonon-assisted processes was estimated to be of the order of 10, independent of the confinement energy.

The above results were obtained from hydrogen-terminated silicon nanocrystals [90] in porous silicon. The authors claimed that SiO_2-capped silicon nanocrystals behave in a similar way, but, in this case, the strength of NP processes is significantly enhanced at the same emission energies. However, their overall conclusion was that even nanometer-size silicon crystallites do not become totally direct semiconductors. The more direct nature of the recombination leads to an increase of the oscillator strength, resulting in a shortening of lifetime of the optically active exciton state by two orders of magnitude over the whole spectral range.

3.1.3. Absorption Cross Sections

To estimate the quantum efficiency of PL, it is important to know the absorption cross sections for different excitation wavelengths. A detailed study has been made for porous silicon by Kovalev et al. [91] and for silicon nanocrystals embedded in SiO_2 by Priolo et al. [92] and Garcia et al. [93]. In [92] and [93], it was found that in SiO_2-capped silicon nanocrystals, the value of the absorption cross section per nanocrystal varies in the range of 10^{-16} to 10^{-15} cm^2 for excitation energies in the visible (457-, 488-, and 514-nm lines) and UV ranges (325 nm). The absorption cross section per nanocrystal increases as the nanocrystal size decreases for all excitation wavelengths, implying that the variation of the oscillator strength dominates over the reduction of the density of states.

Different results were obtained from porous silicon by Kovalev et al. [91], who found that absorption cross section depends both on the excitation and detection energy difference and it varies over nearly 5 orders of magnitude, from 10^{-19} cm^2 under low-energy resonant excitation to 10^{-14} cm^2 under excitation at higher energies above the bulk silicon direct bandgap. At resonant excitation, it was found that the value of the absorption cross section is a very strong function of the difference of the excitation and detection energy difference $\hbar\omega_{ex} - \hbar\omega_{det}$, this energy dependence reflecting the increase in the density of electronic states with increasing energy above the "gap."

3.1.3.1. Optical Gain in Silicon Nanocrystals For the practical use of silicon nanostructures for the fabrication of silicon-based optoelectronics, it is important to reach quantum efficiencies at least similar to those obtained from III–V

semiconductors. In an effort to demonstrate that this would be possible, experiments were carried out to investigate the possibility of obtaining stimulated emission from silicon nanocrystals. By using samples fabricated by ion implantation of silicon into SiO_2, Pavesi et al. [94] claimed, in the year 2000, that they observed optical gain from silicon nanocrystals. Other works then followed [95–97], reporting stimulated emission in silicon nanocrystals, in severe competition with fast nonradiative Auger processes. To explain the gain mechanism, it was suggested that interface states associated with oxygen atoms were involved in light emission. More recently, it was reported that either positive or negative gain could be observed [98], depending on the sample and on whether Auger fast processes or stimulated emission predominate. An effective four-level model was proposed to explain the obtained results.

However, although the above results on optical gain and light amplification seem to be promising, silicon LEDs still have a factor of 10 lower efficiencies than those required in optoelectronics. One important limiting factor seems to be the low density of light-emitting centers in the composite material, composed of an insulating matrix in which the nanocrystals are embedded.

3.2. Electroluminescence

Electroluminescence was observed both from silicon nanocrystals in SiO_2 [17, 18, 39, 43, 99] and from silicon nanowires [100–102] passivated by silicon dioxide. Compared to PL, the main issue in EL is carrier injection into the nanostructures.

3.2.1. Electroluminescence from SiO_2-Passivated Silicon Nanowires

Light-emitting silicon nanowires were fabricated by Nassiopoulou et al. [100, 102] and Grigoropoulos et al. [101] in 1995. These nanowires were in the form of nanopillars vertically standing on a silicon substrate, and they were fabricated by using lithography, combined with a highly anisotropic silicon-etching process [103], and followed by high-temperature thermal oxidation and oxide removal. The diameter of the wires was the same all along their length, and the aspect ratio of height to diameter after removal of the surface oxide was equal to 120. An example of these nanowires, arranged in line arrays on a silicon substrate, is shown in Figure 18. Figure 19 shows an example of PL spectrum from these nanowires. The PL intensity was limited because of the small number of light-emitting nanowires probed (their packing density was low). The characteristics of PL were similar to those of the red PL band from silicon nanocrystals, and they were attributed to quantum confinement.

To obtain EL from these wires, a device structure was fabricated as follows: The oxidized silicon nanowires were embedded in a nonconductive transparent polymer (polymethylmethacrylate), deposited by spinning on the silicon wafer containing the nanowires, which provided a matrix for planarization of the whole structure and extra electrical isolation of the nanowires, since both the wires and the silicon surface on which they were lying were covered by a thin SiO_2 layer. After planarization, the surface was slightly etched in an oxygen plasma to remove the resist from the top of the nanopillars. The oxide from the nanopillar top was removed by dipping the sample in diluted HF for a few seconds. A thin gold film, 15-nm thick, was then evaporated on top, which was used as gate metal. An ohmic contact was formed on the back side of the wafer, which was used as second contact. The process flow is schematically illustrated in Figure 20. Figure 21 shows EL spectra at two different polarization voltages. These results demonstrated clearly that light emission from silicon nanowires at room temperature is possible, both under optical and under electrical excitation, when their diameter is below few nanometers.

3.2.2. Electroluminescence from Silicon Nanocrystals in SiO_2

Samples fabricated by different techniques were used to investigate EL from silicon nanocrystals embedded in SiO_2.

Samples obtained by ion implantation of silicon into an oxide thin film on a silicon substrate showed, in general, EL at relatively high voltages. The main difficulty is carrier injection into the nanocrystals through the insulating oxide,

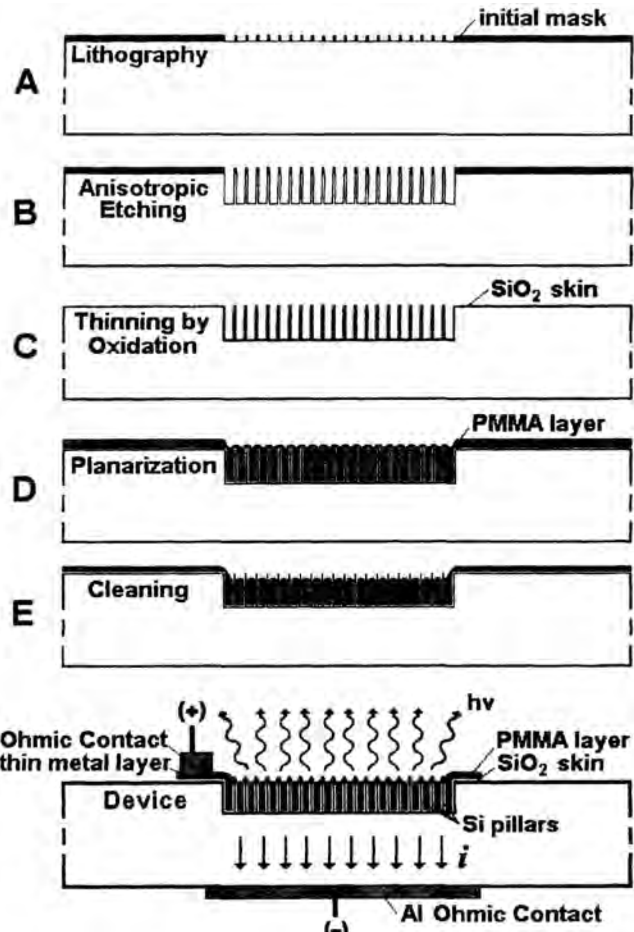


Figure 20. Process flow for the fabrication of a light-emitting structure based on silicon nanowires. Reprinted with permission from [103], A. G. Nassiopoulou et al., *Appl. Phys. Lett.* 66, 1114 (1995). © 1995, American Institute of Physics.

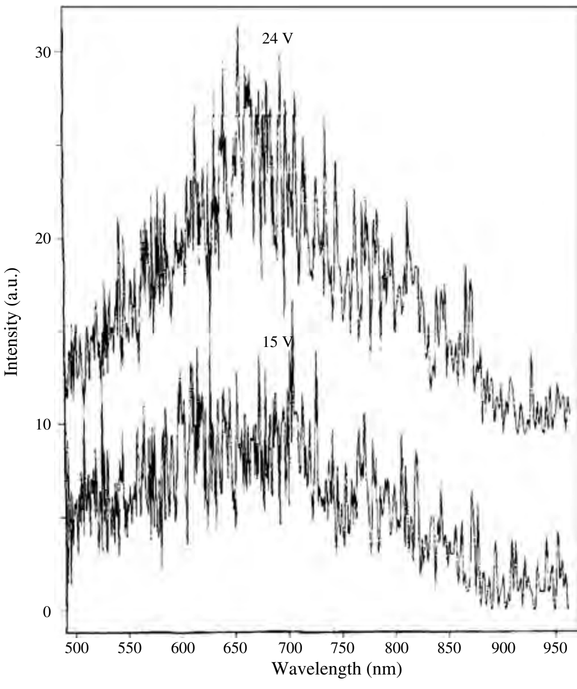


Figure 21. EL spectra from silicon nanowires, by using the structure of Figure 20 at two different applied voltages. Reprinted with permission from [103], A. G. Nassiopoulou et al., *Appl. Phys. Lett.* 66, 1114 (1995). © 1995, American Institute of Physics.

which leads to a low quantum efficiency. To improve carrier injection, two solutions were proposed, either (a) to grow a very high concentration of silicon nanocrystals into the oxide (close to the percolation threshold), implanted at high voltages [104]; or (b) to use very thin SiO_2 layers, implanted at low voltages, to get tunneling transport through the oxide. Song et al. [17] used 34-nm-thick SiO_2films on p-type silicon implanted with Si^+ ions of 25 KeV and annealed in a controlled nitrogen atmosphere at temperatures up to 1100 °C. Electroluminescence was observed at forward bias, with an onset at 8 V. The EL spectrum was broad and centered at 580 nm, with two weaker shoulders at around 470 and 730 nm. Song et al. [17] suggested that part of the 730 nm band was due to silicon nanocrystals, while some contribution was due to defects in SiO_2. The bands at 470 nm and 580 nm also were attributed to defects in the oxide. Valenta et al. [18] fabricated light-emitting diodes based on Si nanocrystals, by using thermal silicon oxides of thickness between 12 and 100 nm and covered by an amorphous silicon layer, 210-nm thick, which were implanted with Si^+ ions at 150 KeV. The EL band from radiative recombination in silicon nanocrystals was peaked at 800 nm. An external quantum efficiency of 3×10^{-5} was obtained.

More promising for EL emission were samples containing 2-D arrays of nanocrystals in between SiO_2 [80], fabricated by LPCVD deposition of silicon and high-temperature oxidation. The structure used for EL is shown in Figure 22. The substrate was p type, and a thin transparent aluminum film, ≈15-nm thick, was deposited on top of the SiO_2 layer containing the nanocrystals, and it was patterned to be used as gate metal. A variable voltage was applied between the gate and an ohmic contact on the backside of the wafer.

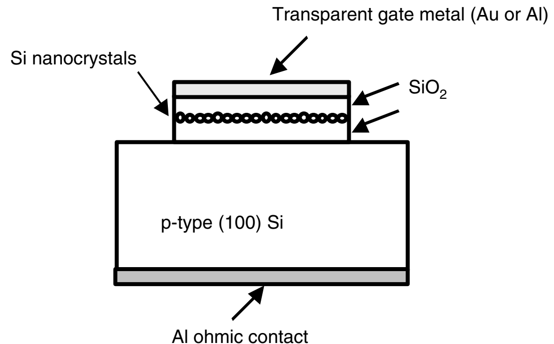


Figure 22. MOS structure with silicon nanocrystals in the SiO_2 layer, used to observe EL from silicon nanocrystals.

Figure 23 shows typical EL spectra at different temperatures, obtained with an applied voltage of 10 V, while Figure 24 shows EL spectra at different voltages. Figure 25 shows the evolution of EL intensity measured at two different emitted wavelengths as a function of the applied voltage, normalized by the electrical current through the structure. At low voltages, the normalized EL intensity increased linearly by increasing the voltage, while it saturated at higher voltages. Moreover, the normalized EL intensity at the two different emitted wavelengths (680 and 770 nm), showed that at low voltages, the intensity of the signal at 770 nm was

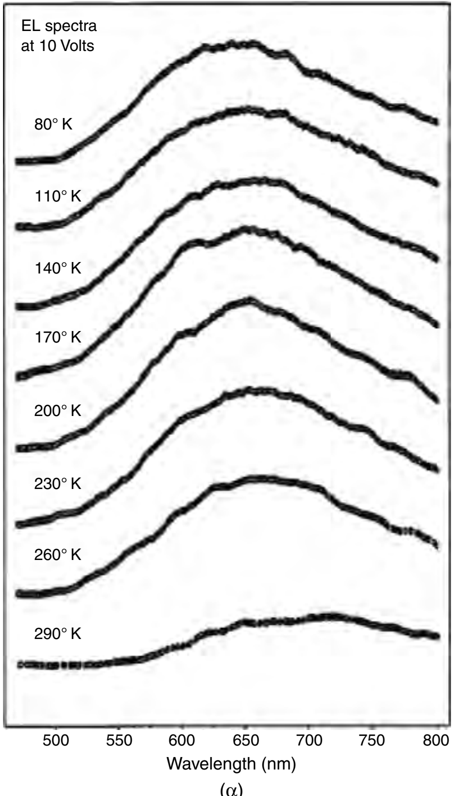


Figure 23. EL spectra from silicon nanocrystals in SiO_2 at different temperatures. Reprinted with permission from [99], P. Photopoulos and A. G. Nassiopoulou, *Appl. Phys. Lett.* 77, 1816 (2000). © 2000, American Institute of Physics.

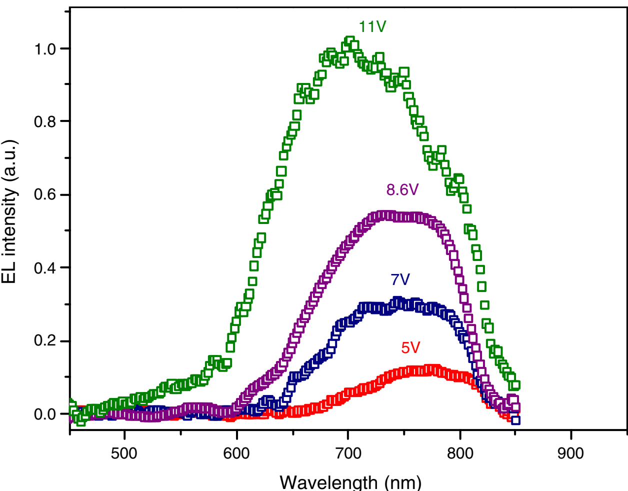


Figure 24. EL spectra from silicon nanocrystals at different applied voltages (P. Photopoulos and A. G.Nassiopoulou, unpublished results).

higher than that at 680 nm, while this was inverted at higher voltages. The saturation of EL at a high applied electric field occurs because part of the light-emitting nanocrystals are quenched, due to two different main effects: (a) the higher carrier injection when the applied voltage is increased and (b) the effect of the electric field itself (quantum-confined Stark effect, see [105]). The first effect also occurs in PL at high excitation power. It has been observed experimentally both in the case of nanocrystalline silicon SLs [105, 106] and in the case of porous silicon [83–85]. It is attributed to the fact that at high excitation power, it is more likely that more than one electron-hole pair is generated or injected into the nanocrystal [84, 85]. The presense of an extra carrier or a second electron-hole pair into the nanocrystal results in the fast opening of a nonradiative Auger recombination channel, which quenches PL and EL. This effect is expected to be more pronounced in larger nanocrystals, and this is why the PL signal at 680 nm (due to smaller nanocrystals) saturates at a higher voltage than that at 770 nm (due to larger nanocrystals). This is illustrated schematically in Figure 26.

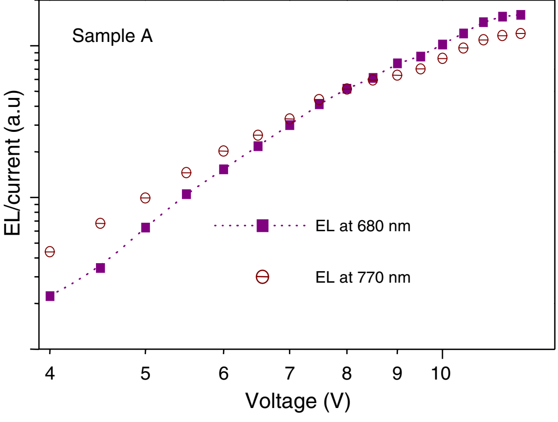


Figure 25. Variation of EL intensity from silicon nanocrystals, normalized by the current through the structure, measured at two different emitted wavelengths. Reprinted with permission from [99], P. Photopoulos and A. G. Nassiopoulou, *Appl. Phys. Lett.* 77, 1816 (2000). © 2000, American Institute of Physics.

The second effect, which is at the origin of EL saturation but is not present in the case of PL, is the effect of the electric field on the electron and hole of an electron-hole pair in the nanocrystal. This effect is known as the quantum-confined Stark effect, and it has been studied in detail in the case of quantum wells of III–V compounds [107–110]. For the case of $Si–SiO_2$ superlattices, a detailed study has been made by Quisse and Nassiopoulou [105] by means of a variational calculation. Due to the good dielectric properties of SiO_2, an electric field as high as 6–7 MV/cm can be applied to the $Si–SiO_2$ superlattice without causing a breakdown to the structure. This high electric field results in an increase of the separation between the electron and the hole of the electron-hole pair and, thus, exciton lifetime is increased. In [105], a cubic quantum box of size L with infinite barrier heights was considered and the possibility of carrier escape from the box was neglected. It was found that for a nanocrystal of a given size, as the electric field increases, the electron and the hole wavefunctions are gradually separated from one another, so that the radiative lifetime increases. Figure 27 schematically illustrates this effect. On the other hand, the transition energy decreases. At moderate electric field, the energy shift is proportional to the square of the electric field, as predicted by the second-order perturbation theory [105]. Also, to a good approximation (a few percentages), the dependence of the phonon-assisted lifetime on the applied electric field at room temperature is empirically given by:

$$\tau = \tau_0(1 + \gamma L^{5.25} E^2)$$

where $\gamma \approx 2.1 \times 10^{-8}$, if E is expressed in MV per centimeter, L is in angstroms, and τ_0 is the lifetime in the absence of an external electric field. The overall effect of the electric field on a nanocrystal of a given size is thus manifested by: (a) a red shift of the luminescence and (b) a quenching of the radiative recombination rate. This effect is size dependent, and it is more important for larger nanocrystals, as illustrated in Figures 26–28. Figure 28 shows the variation of the gap of cubic silicon nanocrystals versus the cube size and the electric field applied along the (100) crystallographic direction. Figures 29 and 30 show, respectively, the

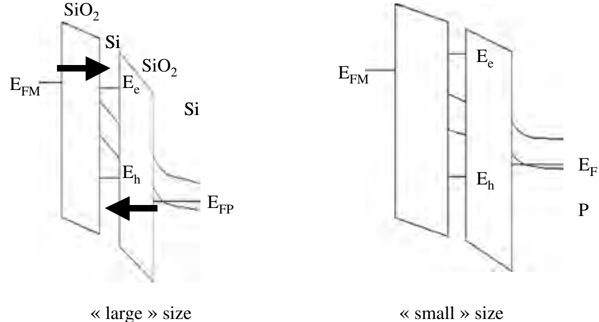


Figure 26. Effect of the applied electric field on energy levels in the case of a large and a smaller silicon nanocrystal, surrounded by two tunneling SiO_2 layers. See the position of the confined states E_h and E_e with respect to the Fermi level of the metal and the semiconductor in the two cases.

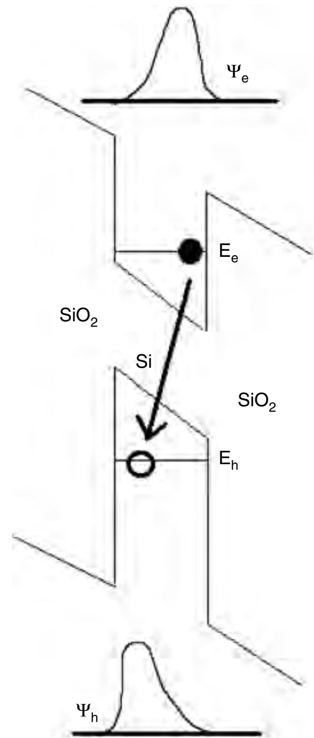


Figure 27. Effect of an electric field on the electron and hole wavefunctions of an exciton in a silicon nanocrystal surrounded by two thin silicon oxide layers.

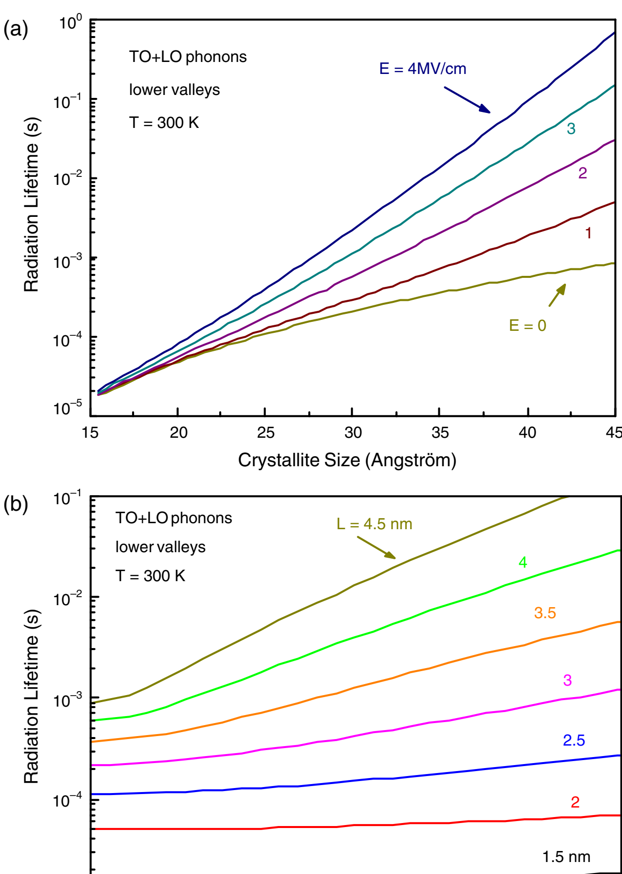


Figure 28. Variation of the radiation lifetime of silicon nanocrystals as a function of the crystallite size for different values of the electric field (a) and as a function of the electric field for different values of crystallite size (b) (T. Ouisse and A. G. Nassiopoulou, unpublished results)

direct and TO + LO phonon-assisted radiative recombination lifetime as a function of nanocrystal size and electric field along a (100) direction (lower valleys), at $T = 300$ K. As it is seen, the effect of the electric field is more substantial in larger nanocrystals. By increasing the electric field in these nanocrystals, the radiative recombination lifetime increases substantially.

3.3. Effect of a High Electric Field on PL and EL

3.3.1. Effect on PL

The effect of a high applied electric field on PL from silicon nanocrystals in SiO_2 has been systematically investigated in [111]. The structures used were MOS-type (metal–oxide–semiconductor) structures on silicon with the silicon dioxide layer containing 2-D arrays of silicon nanocrystals in SiO_2. The thickness of SiO_2 was chosen so as to minimize carrier injection into the oxide under the influence of the vertically applied electric field. The PL signal was excited by laser pumping at 457.9 nm. Two different samples were investigated, one with a mean nanocrystal size in the range of 1–2 nm and the other with a mean nanocrystal size in the 2–3 nm range. The PL was measured under the application of an increasing electric field at constant laser pumping. By current-voltage (I-V) and capacitance-voltage (C-V) measurements, it was found that there was no current flow through the structure if the electric field was kept below 2.5 MV/cm for the sample with the larger nanocrystals

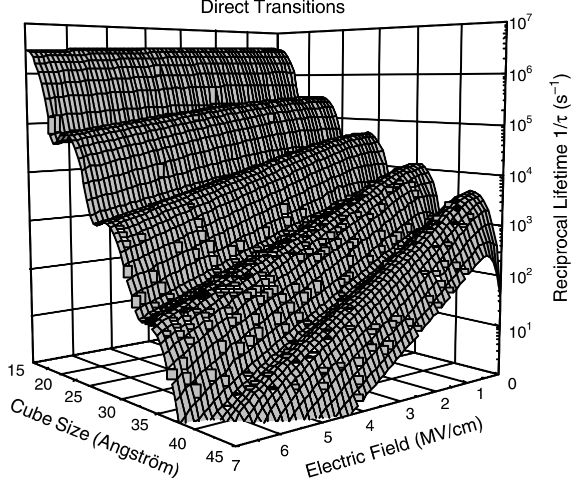


Figure 29. Variation of direct radiative recombination lifetime as a function of Si crystallite size and electric field, along a (100) direction (lower valleys). Reprinted with permission from [105], T. Ouisse and A. G. Nassiopoulou, *Europhys. Lett.* 51, 168 (2000). © 2000, EDP Sciences.

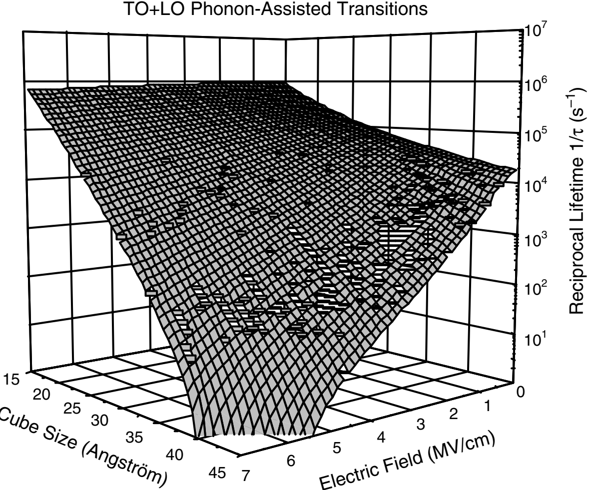


Figure 30. Variation of TO + LO phonon-assisted radiative recombination lifetime as a function of crystallite size and electric field along a (100) direction (lower valleys). Reprinted with permission from [105], T. Ouisse and A. G. Nassiopoulou, *Europhys. Lett.* 51, 168 (2000). © 2000, EDP Sciences.

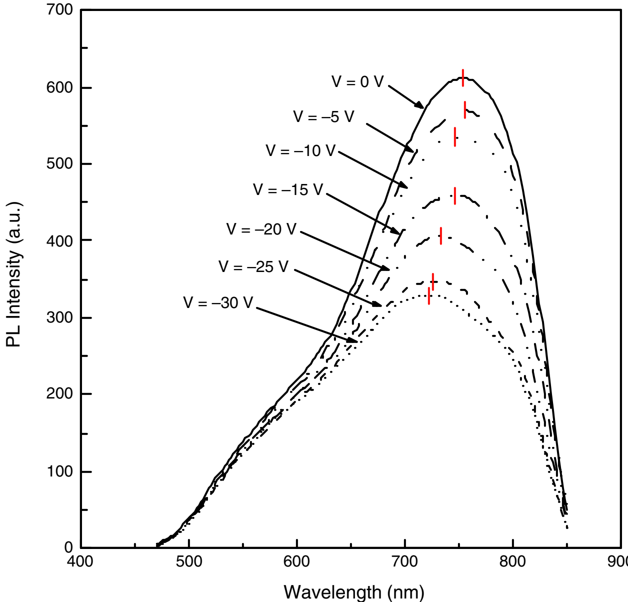


Figure 31. Effect of an increasing electric field on PL emission from silicon nanocrystals, of a diameter of 2–3 nm, in between SiO_2. The sample structure used was similar to that of Figure 22, with a bottom SiO_2 layer, 30-nm thick and a top SiO_2 layer, 18-nm thick. Reprinted with permission from [111], Ioannou-Sougleridis et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier Science.

and below 4 MV/cm for the sample with the smaller nanocrystals. The observed overall effect of the electric field on PL was as follows:

(i) In the sample with the larger nanocrystals, it was found that by applying an increasing electric field, the PL peak position was first slightly red shifted, and it was then blue shifted at a higher field, while PL intensity was continuously quenched by increasing the electric field. This is illustrated in Figures 31, 32, and 33.
(ii) In the case of the sample with the smaller nanocrystals, it was found that PL peak position was almost constant under the influence of an increasing electric field (slightly red shifted), up to an electric field of 2.3 MV/cm, while it showed a constantly increasing blue shift by increasing the electric field above this value. The PL intensity was also constant up to ≈2.3 MV/cm, while it was rapidly quenched above this value. This is illustrated in Figures 32–34.

The observed shift in PL peak position and the PL quenching under the influence of an external electric field is explained in terms of the quantum-confined Stark effect, described previously. In the case of very small nanocrystals, PL was insensitive to the electric field, up to a certain field, while larger nanocrystals showed an important PL quenching under the influence of the same electric field, in agreement with theoretical calculations [105]. The fact that the observed shift in PL peak wavelength did not follow a constant red shift by increasing the electric field was attributed to the size distribution of silicon nanocrystals. The broad PL signal measured, contains the contribution from nanocrystals with different sizes, which show different behavior under the influence of the electric field. At a given electric field, PL from larger nanocrystals is quenched, while that from smaller nanocrystals is not affected by the field. The macroscopic effect observed is so manifested by a blue shift in PL position, accompanied by an important PL quenching. The obtained results were consistent with this explanation.

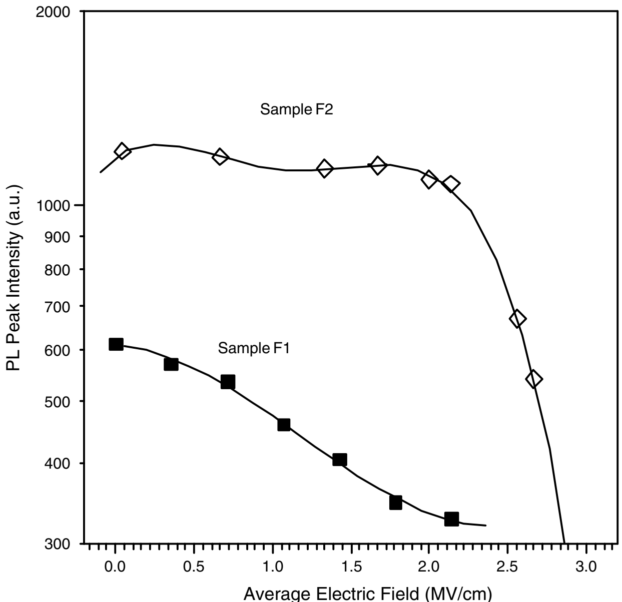
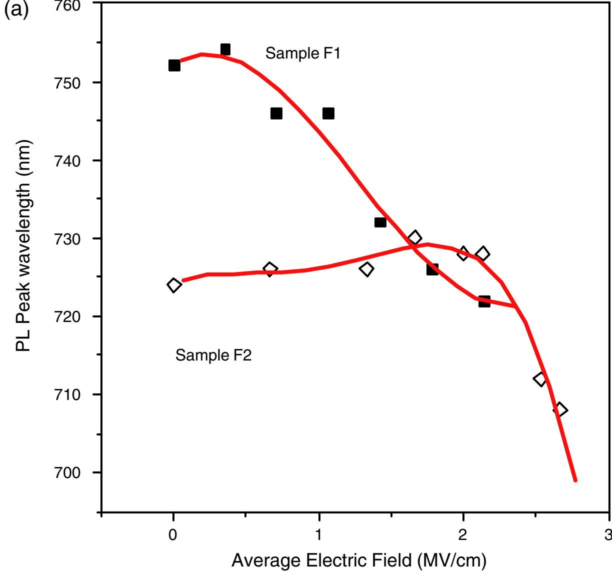


Figure 32. PL peak intensity from silicon nanocrystals in SiO_2 as a function of an applied electric field for two different sizes of silicon nanocrystals (2–3 nm [F1] and 1–2 nm [F2]). The influence of the electric field is more important in the case of larger nanocrystals. Reprinted with permission from [111], Ioannou-Sougleridis et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier Science.



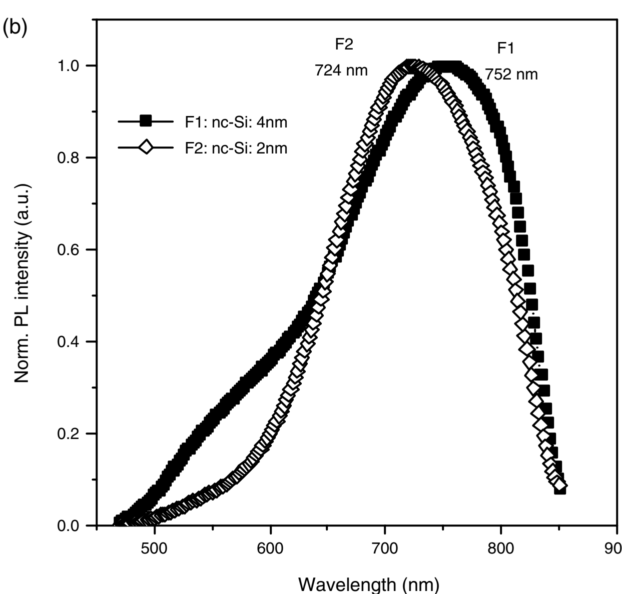


Figure 33. (a) PL peak wavelength as a function of an applied electric field for larger (2–3 nm, F1) and smaller (1–2 nm, F2) silicon nanocrystals in SiO_2. Reprinted with permission from [111], Ioannou-Sougleridis et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier; (b) PL spectra from samples F1 and F2 at zero electric field.

3.3.2. Effect on EL

In the case of EL, there is no laser pumping, so carrier injection is necessary, to create electron-hole pairs into the nanocrystals. In the presence of carrier injection, the electric field has different effects on the EL signal:

(i) By increasing the electric field, carrier injection is increased, so we expect an increase in the EL signal proportional to the current. This is valid for low carrier injection, while at high carrier injection, Auger quenching of the EL signal occurs, which leads to saturation of the EL intensity.

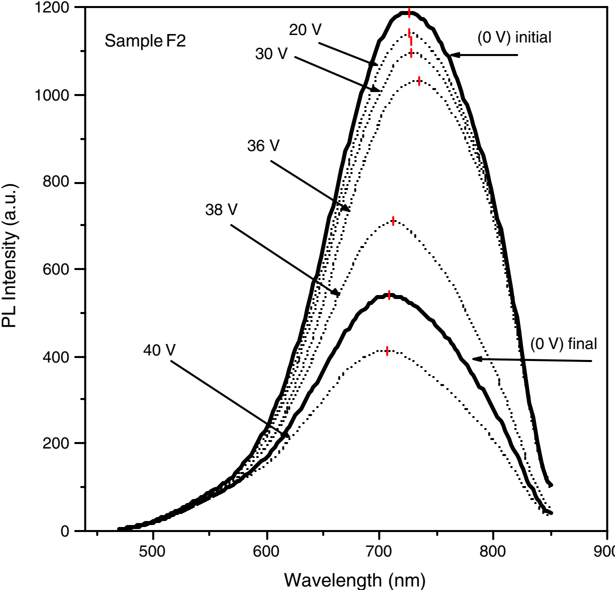


Figure 34. PL spectra from silicon nanocrystals, 1- to 2-nm thick, at different applied electric fields from zero to 40 V. The PL spectrum at zero voltage, obtained after the application of 40 V for some time, shows PL quenching and a slight blue shift compared with the initial spectrum before the application of the electric field. Reprinted with permission from [111], Ioannou-Sougleridis et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier Science.

(ii) The quantum-confined Stark effect plays an important role in EL quenching at high electric fields. The effect is more pronounced in larger nanocrystals.

(iii) Carrier injection into the nanocrystals is also a size-dependent effect. Smaller nanocrystals have a significantly smaller cross section for carrier injection.

The effect of an increasing electric field on EL was investigated in [99]. The EL intensity as a function of the applied electric field has been investigated at different wavelengths, and an example has been given in Figure 24 for two detected EL wavelength. The evolution of the EL position as a function of the applied voltage was given in Ref. [99] for nanocrystals of different sizes. The overall behavior was a blue shift in EL wavelength by increasing the applied field. This was observed in all samples. observed in all samples.

3.4. EL Efficiency

In the case of PL, it was demonstrated that the use of silicon nanocrystal–SiO_2 multilayers, instead of a single 2-D silicon nanocrystal layer in SiO_2, leads to a significant increase in PL emission, due to the increase of light-emitting centers. As mentioned, it was even found that a superlinear increase in PL emission is obtained when the number of multilayers is increased [40]. In the case of EL, however, the increase of recombination centers by increasing the number of silicon nanocrystal layers does not necessarily imply an increase in EL emission. This is because the main limiting factor in EL is carrier injection into the nanocrystals. By using vertical carrier injection, which was the configuration used in the literature [99, 104, 111] in light emitting devices, Quisse et al. [112] simulated the tunneling and recombination current in

$Si–SiO_2$ superlattices, to find the optimum conditions for EL. A p-type silicon substrate was considered for hole injection and an aluminum gate for electron injection. The basic assumptions used were that (i) the Si wells were thin enough to get an almost ideal 2-D system and (ii) quantization was represented by a square well approximation. They first simulated the one-well configuration, and it was found that in the absence of recombination, the hole-electron density product in the well is maximum for equal top and bottom oxide thickness ($t_{ox1} = t_{ox2}$). This is illustrated in Figure 35, where the density of electrons and holes in the well is plotted versus the top oxide thickness for a bottom oxide of thickness $t_{ox2} = 2.6$ nm, a well thickness $t = 1.5$ nm, a gate voltage $Vg = -2$ V, lower than the breakdown voltage of the Si well, and a temperature of 300 K. The optimum for a high density of both electrons and holes in the well is at ($t_{ox1} = t_{ox2}$). As it now concerns the optimum thickness for t_{ox1} and t_{ox2}, it was found that the thinner the insulating layers are, the higher are the currents and, therefore, the recombination rate.

In the presence of recombination, however, accumulation of holes is only possible if the top oxide thickness, t_{ox1}, is below a critical value t_{crit}. For accumulation of electrons, on the other hand, the necessary condition is $t_{ox1} = t_{crit}$. This means that it is impossible to get simultaneously both high-electron and high-hole density in the well in the presence of recombination. Another interesting result in [112] concerns the material of the injecting electrodes. The injection of carriers for radiative recombination in the well is difficult when the injecting material has the same or a similar work function as that of the confined layer, in the present case of silicon. Theoretically, to get acceptable electron injection currents in the direction of confinement, one has to use injection electrodes with a work function much lower than the electron affinity of Si. Similarly, one should inject holes from a material with a Fermi level as close as possible to the top of the Si valence band. For example, platinum could be used as a gate metal (chemical potential lower than the top of the Si valence band) and an n-type silicon substrate for electron injection. Compared with the $Al–SiO_2–Si–SiO_2$–p-Si structure, an improvement of a factor of 3 may be achieved.

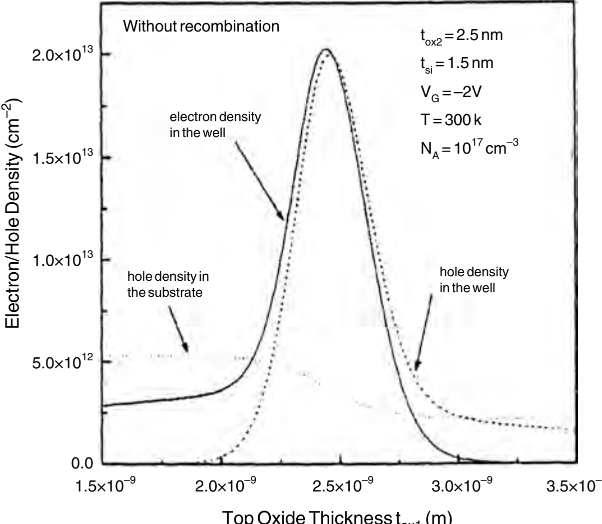


Figure 35. Variation of the electron and hole concentrations in an $Al–SiO_2–Si–SiO_2$–Si structure versus top oxide thickness $t_{ox2} = 2.5$ nm and voltage, without any recombination process in the single well. Reprinted with permission from [112], T. Quisse et al., *J. Phys. D: Appl. Phys.* 33, 2691 (2000). © 2000, Institute of Physics.

Coming back to the $Al–(Si–SiO_2)SL$–p-Si structure, a systematic investigation has been made in [112] by varying the number of wells and other parameters, for example, the voltage and temperature. The obtained results showed that almost all the recombination takes place in one well, the closest to the p-type substrate. There are no holes in all the other wells, and even the first well is almost empty, since both electron and hole concentrations are low. As in the case of one well, all the injected holes recombine with electrons, and they do not reach the metal electrode. In contrast, a non-negligible part of electrons tunnels into the silicon substrate.

In general, in contrast to what happens in PL, in EL, we do not expect any gain from the use of a multiple-well configuration.

Although the study in [112] concerns silicon wells in SiO_2, the obtained results could give a rough idea of what is expected in the case of 2-D arrays of nanocrystals in a nc-$Si–SiO_2$ superlattice. The use of a superlattice, instead of a single layer of 2-D arrays of nanocrystals does not offer any advantage in EL compared to PL. The obtained EL efficiency in the literature was, in general, low, below 10^{-5} [104], and this mainly is attributed to the intrinsic difficulty to inject carriers into the nanocrystals. The optimum case is expected to be a structure with a 2-D array of nanocrystals in a layer at a tunneling distance from the silicon substrate.

4. CHARGING OF SILICON NANOCRYSTALS SINGLE ELECTRON DEVICES

Silicon nanocrystals in SiO_2 constitute a very good candidate for use in single electron devices, due to material and process compatibility with the well-established IC technology (113–118). The IC compatible processes used involve either ion implantation of silicon into SiO_2 followed by annealing or LPCVD or PECVD deposition of silicon on a very thin SiO_2 layer, followed by high-temperature oxidation.

The active part in a single electron memory is a MOS capacitor, in which the oxide layer contains a 2-D layer of silicon nanocrystals at a tunneling distance from the silicon substrate, as shown in Figure 36. The top oxide is relatively thicker than the tunneling oxide, to prevent carrier injection from the gate electrode. The silicon nanocrystal memory uses, in general, two more electrodes (source and drain), and it operates as a three-terminal device. The MOS structure may be used to evaluate the quality of the nanocrystals. Single electron charging of the nanocrystals is evidenced by applying a vertical electric field between the gate metal and an ohmic contact on the backside of the silicon substrate. Capacitance-voltage (C-V) and current-voltage (I-V) characteristics may be used in this respect. The C-V curves show an hysteresis effect, due to charging of the silicon nanocrystals, while I-V curves show N-shaped behavior [119–121]. To correlate the observed hysteresis with charge trapping in the nanocrystals, one has to minimize all other causes

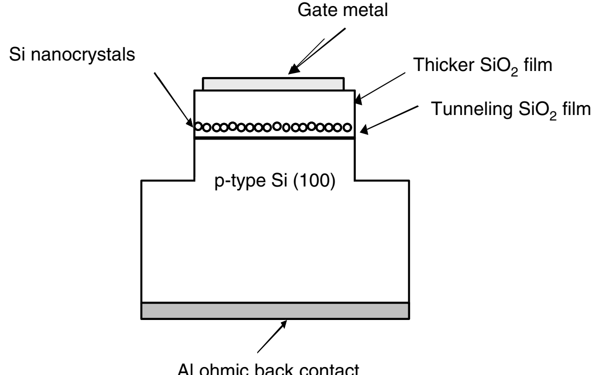


Figure 36. Basic structure used in a silicon nanocrystal memory. The bottom oxide is a tunneling oxide and the top oxide is thicker.

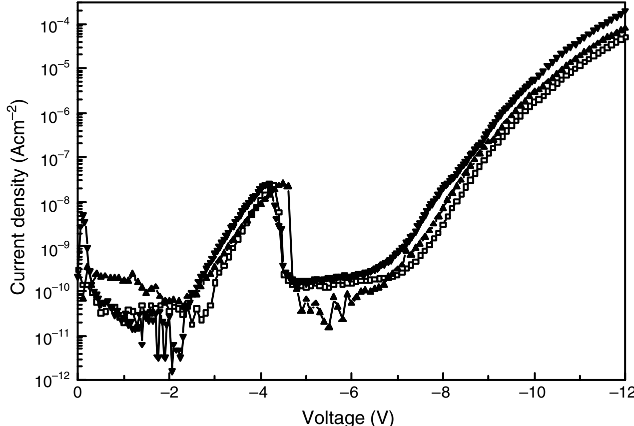


Figure 38. Current-voltage (I-V) curves from a MOS structure with silicon nanocrystals as in Figure 36. The I-V curves from three different samples are shown. These curves exhibit N-shaped behavior, attributed to charging of silicon nanocrystals. Reprinted with permission from [119], D. Kouvatsos et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier.

of hysteresis (defects within the oxide and at the different Si–SiO_2 interfaces, including the interface of silicon dioxide with the nanocrystals). The hysteresis ΔV due to the nanocrystals is observed when the voltage is swept from deep inversion to deep accumulation and back. An example is given in Figure 37 [119], where the substrate was p type and trapping of holes was observed (counterclockwise hysteresis). The measured hysteresis ΔV depends on nanocrystal size, being smaller for smaller nanocrystals. An example of a N-shaped I-V curve from the same sample is shown in Figure 38 for three different structures examined. The current shows a peak at a voltage of -4 to -5 V. This peak corresponds to hole charging of the nanocrystals. There is direct correlation between this peak in the current and the onset of C-V curves. This is illustrated in Figure 39, which shows C-V curves corresponding to successive voltage sweeps, starting from $+1$ V and going to negative values, with successively higher negative end voltage. The curves corresponding to sweeps from 1 to -2 V and back, and from 1 to -4 V coincide. The curve corresponding to a sweep from -4 to 1 V is considerably shifted with respect to the 1 to -4 V sweep. The sweeps with negative end voltage higher than -4 V are further shifted to more negative values. A capacitance drop also is observed at around -4 V, which also correlates with the peak in the current. However, the observed shift ΔV may be related not only to nanocrystal charging but also to charges at the interfaces between silicon nanocrystals and silicon oxide. By appropriate annealing, these interface states may be minimized to separate the nanocrystal charging effect [120, 121].

Silicon nanocrystal memory devices published so far are based on the threshold voltage shift, induced by the presence of nanocrystals in the gate dielectric of a field effect transistor [118, 120]. Compared to standard floating gate Flash electrically erasable programmable ready only memory (EEPROM), the difference is that in the case of silicon nanocrystal memories, the floating gate is not composed of a continuous nanocrystalline silicon layer, but it is composed of discrete nanocrystal dots, which are not electrically

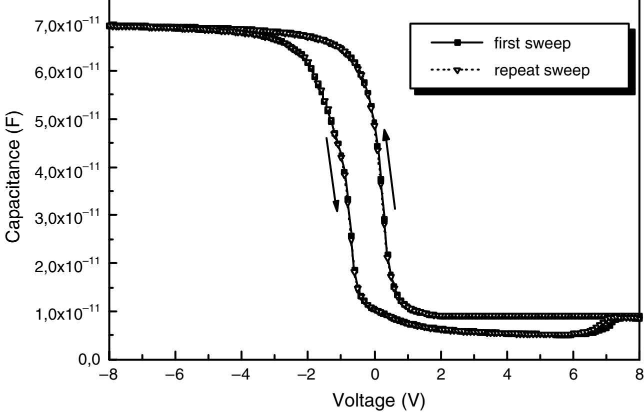


Figure 37. Double C-V curve from a MOS structure with silicon nanocrystals as in Figure 36. An hysteresis is observed, attributed mainly to silicon nanocrystal charging. Reprinted with permission from [119], D. Kouvatsos et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier.

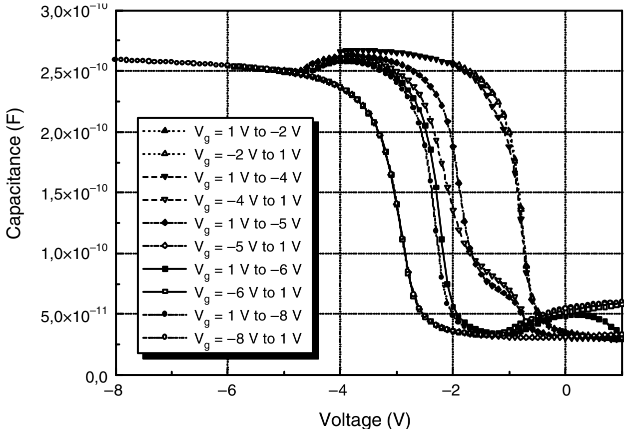


Figure 39. Multiple double C-V sweeps of MOS structure with silicon nanocrystals in SiO_2, biased from 1 V to increasing voltages into accumulation. Solid or open symbols are used for forward or reverse sweeps, respectively. The dip in the curves starting at $\cong$3 V is attributed to the silicon nanocrystal layer. Reprinted with permission from [119], D. Kouvatsos et al., *Mater. Sci. Eng., B* (to appear, 2003). © 2003, Elsevier.

continuous. In this way, a discrete defect in one of the nanocrystals does not lead to total charge leakage through the structure, and the memory device may still be functional. Moreover, there are other advantages of silicon nanocrystal memories compared to floating gate flash EEPROMs, such as the reduced degradation of the devices, faster write–erase times, lower operation voltage and lower power consumption.

To scale down the memory-device size, while increasing the stored charge, one has to maximize the density of silicon nanocrystals in the 2-D layer in-between SiO_2. In the case of LPCVD deposition of silicon on SiO_2, it was found that the silane (SiH_4) pressure, as well as the chemistry of the SiO_2 surface, influence significantly the silicon nanocrystal density [51], which decreases almost linearly with the increase in silane pressure. On the other hand, a chemical treatment of the SiO_2 surface before silicon deposition by 0.05% HF in deionized water, which creates surface silanol bonds (Si–OH), favorizes silicon dot nucleation and growth [51]. A QD density increase from 4×10^{11} cm^{-2} to 1.2×10^{12} cm^{-2} was measured by increasing the silanol density from 0.4 to 1.4 OH/nm^2.

5. CONCLUSION

Silicon nanocrystals in SiO_2 thin layers on silicon constitute a composite material with many interesting properties, the most important being the bright and stable PL at room temperature in the visible range and the controlled charging of the nanocrystals. These properties are size dependent and are significantly influenced by the chemical composition of the interface between silicon nanocrystals and silicon dioxide. They lead to very important applications, including light-emitting and memory devices and sensors.

GLOSSARY

Coulomb blockade effect The effect of blocking the injection of a second charge into a semiconductor under a certain electric field, due to modification of the electrostatic potential within it by the presence of the first injected charge. Injection of a second charge needs to overcome the semiconductor charging energy.

Ion implantation A technique used to incorporate different atoms into a solid at a certain depth from the surface. The atoms are ejected from different sources, they are ionized and accelerated by an electromagnetic potential under vacuum so as to be injected at high energy into the sample. Their depth distribution depends on their mass and energy.

Low-pressure chemical vapor deposition A technique used to deposit thin dielectrics or semiconductors on a solid surface from gas reactions under vacuum at high temperature.

Luminescence Intrinsic property of a material consisting in light emission at a certain wavelength under excitation by photons (photoluminescence) or under voltage excitation (electroluminescence).

Light-emitting devices Devices that emit light under electrical excitation.

Memory devices Devices used to store information. In conventional semiconductor memories, the information is stored in the form of charge.

Nanocrystalline silicon superlattice A structure consisting of ultrathin Si nanocrystal layers separated by nanometer-thick silicon dioxide.

Nanowires or quantum wires Wires with diameters below ~10 nm.

Optical gain Amplification (positive optical gain) or reduction (negative optical gain) of light from a source after interaction with a material.

Resonant photoluminescence emission Photoluminescence of the material with an excitation energy that approaches the detection energy.

Semiconductor quantum dots Semiconductor nanocrystals of diameter below ~10 nm.

Single electron devices Devices based on semiconductor nanocrystals, in which there is a possibility of controlling the injection of a single electron charge into each nanocrystal by controlling the applied electric field. Single electron devices are based on the Coulomb blockade effect.

REFERENCES

1. Uhlir, *Bell Syst. Techn. J.*, 35, 333–347 (1956).
2. P. Allonge, in "Properties of Porous Silicon" (L. T. Canham Ed.), emies Datareviews Series No 18, 3 (1997).
3. A. Halimaoui, in "Properties of Porous Silicon" (L. T. Canham Ed.), emies Datareviews Series No 18, 12 (1997).
4. A. Halimaoui, *NATO ASI Series E. (Netherlands)* 244, 11 (1993).
5. A. Bsiesy, Y. F. Nikolau, A. Ermolieff, F. Muller, and F. Gaspard, *Thin Solid Films* 255, 43 (1995).
6. A. Nakajima, T. Itakura, S. Watanabe, and N. Nakayama, *Appl. Phys. Lett.* 61, 46 (1992).
7. G. Mauckner, T. Walter, T. Baier, K. Thonke, and R. Sauer, *Mater. Res. Soc. Symp. Proc.* 283, 109, (1993).
8. V. Petrova-Koch, T. Muschik, A. Kux, B. K. Meyer, F. Koch, and V. Lehmann, *Appl. Phys. Lett.* 61, 943 (1992).
9. K. D. Hirschman, L. Tsybeskov, S. P. Duttagupta, and P. M. Fauchet, *Nature* (London) 384, 338, (1996).
10. P. Mutti, G. Ghislotti, S. Bertoni, L. Bonoldi, G. F. Cerofolini, L. Mede, E. Grelli, and M. Guzzi, *Appl. Phys. Lett.* 66, 851 (1995).
11. G. A. Kachurin, I. E. Tyschenko, K. S. Zhuravlev, N. A. Pazdnikov, V. A. Volodin, A. K. Gutakouskii, A. F. Leier, W. Skorupa, and R. A. Yankov, *Semiconductors* 31, 626 (1977).
12. H. Z. Song and X. M. Bao, *Phys. Rev. B* 55, 6988 (1997).
13. Soumyendu Guha, *J. Appl. Phys.* 84, 5210 (1998).
14. P. Normand, K. Beltsios, E. Kapetanakis, D. Tsoukalas, A. Travlos, J. Stoemenos, J. Van Den Berg, S. Zhang, C. Vieu, H. Launois, J. Gautier, F. Jourdan, and L. Palem, *Nucl. Instrum. Methods Phys. Res., Sect. B* 178, 74 (2001).
15. T. S. Iwayama, T. Hama, D. E. Hole, and I. W. Boyd, *Surf. Coat. Technol.* 158–159, 712 (2002).
16. T. Shimizu-Iwayama, S. Nakao, and K. Saitoh, *Appl. Phys. Lett.* 65, 1814 (1994).
17. H.-Z. Song, X.-M. Bao, N.-S. Li, and J.-Y. Zhang, *J. Appl. Phys.* 82, 4028 (1997).
18. J. Valenta, N. Lalic, and J. Linnros, *Opt. Mater.* 17, 45 (2001).
19. E. Kapetanakis, P. Normand, D. Tsoukalas, K. Beltsios, J. Stoemenos, S. Zhang, and J. van den Berg, *Appl. Phys. Lett.* 77, 3450 (2000).
20. E. Kapetanakis, P. Normand, D. Tsoukalas, and K. Beltsios, *Appl. Phys. Lett.* 80, 2794 (2002).

21. K. H. Chae, J. H. Son, G. S. Chang, H. B. Kim, J. Y. Jeong, S. Im, J. H. Song, K. J. Kim, H. K. Kim, and C. N. Whang, *Nanostruct. Mater.* 11, 1239–1243 (1999).
22. T. Makimura, Y. Kunii, N. Ono, and K. Murakami, *Appl. Surf. Sci.* 127–129, 288 (1998).
23. A. V. Kabaslim, M. Charbonneau-Lefort, M. Meunier, and R. Leonelli, *Appl. Surf. Sci.* 168, 328 (2000).
24. T. Yoshida, Y. Yamada, and T. Orii, *J. Appl. Phys.* 83, 5427 (1998).
25. Y. Kanemitsu, *Phys. Rev. B.* 49, 16845 (1994).
26. M. Rückschloss, B. Landkammer, and S. Veptek, *Appl. Phys. Lett.* 63, 1474 (1993).
27. M. Takeguchi, M. Tanaka, and K. Furuya, *Appl. Surf. Sci.* 146, 257 (1999).
28. W. L. Wilson, P. F. Szajowski, and L. E. Brus, Science (Wahington, D.C.) 262, 1242 (1993).
29. L. E. Brus, P. F. Szajowski, W. L. Wilson, T. D. Harris, S. Schuppler, and P. H. Citrin, *J. Am. Chem. Soc.* 117, 2915 (1995).
30. K. A. Littau, P. F. Szajowski, A. J. Muller, R. F. Kortan, and L. E. Brus, *J. Phys. Chem.* 97, 1224 (1993).
31. L. Tsybeskov, K. D. Hirschman, S. P. Duttagupta, M. Zacharias, P. M. Fauchet, J. P. McCaffrey, and D. J. Lockwood, *Appl. Phys. Lett.* 72, 43 (1998).
32. M. Zacharias, J. Bläsing, K. Hirschman, L. Tsybeskov, and P. M. Fauchet, *J. Non-Cryst. Solids* 266–269, 640 (2000).
33. D. J. Lockwood, G. F. Grom, L. Tsybeskov, P. M. Fauchet, H. J. Labbé, J. P. McCaffrey, and B. White, Jr., *Physica E* 11, 99 (2001).
34. M. Zacharias, L. Tsybeskov, K. D. Hirschman, P. M. Fauchet, J. Bläsing, P. Hohlert, and P. Veit, *J. Non-Cryst. Solids* 227–230, 1132 (1998).
35. D. J. Lockwood, G. F. Grom, P. M. Fauchet, and L. Tsybeskov, *J. Cryst. Growth* 237–239, 1898 (2002).
36. P. Mishra and K. P. Jain, *Mater. Sci. Eng., B* 95, 201 (2002).
37. L. Tsybeskov, K. D. Hirschman, S. P. Duttagupta, P. M. Fauchet, M. Zacharias, J. P. McCaffrey, and D. J. Lockwood, *Phys. Status Solidi A* 165, 69 (1998).
38. A. G. Nassiopoulos, P. Photopoulos, V. Ioannou-Sougleridis, S. Grigoropoulos, and D. Papadimitriou, *Mater. Res. Soc. Proc.* 452, 663 (1997).
39. P. Photopoulos, A. G. Nassiopoulou, D. N. Kouvatsos, and A. Travlos, *Mater. Sci. Eng., B* 69–70, 345 (2000).
40. P. Photopoulos, A. G. Nassiopoulou, D. N. Kouvatsos, and A. Travlos, *Appl. Phys. Lett.* 76, 3588 (2000).
41. A. G. Nassiopoulou, V. Ioannou-Sougleridis, P. Photopoulos, A. Travlos, V. Tsakiri, and D. Papadimitriou, *Phys. Status Solidi A* 165, 79 (1998).
42. L. Tsybeskov, K. D. Hirschman, S. P. Duttagupta, M. Zacharias, P. M. Fauchet, J. P. McCaffrey, and D. J. Lockwood, *Appl. Phys. Lett.* 72, 43 (1998).
43. A. G. Nassiopoulou, T. Ouisse, and P. Photopoulos, in "Frontiers of Nano-Optoelectronics Systems," L. Pavesi and E. Buzaneva (Eds.). Kluwer, 137, 2000.
44. V. Vinciguerra, G. Franzo, F. Priolo, F. Iacona, and C. Spinella, *J. Appl. Phys.* 87, 8165 (2000).
45. L. Tsybeskov, G. F. Grom, P. M. Fauchet, J. P. McCaffrey, J. M. Baribeau, G. I. Sproule, and D. J. Lockwood, *Appl. Phys. Lett.* 75, 2265 (1999).
46. F. Iacona, G. Franzo, and C. Spinella, *J. Appl. Phys.* 87, 1295 (2000).
47. M. Otobe, T. Kanai, T. Ifuku, H. Yajima, and S. Oda, *J. Non-Cryst. Solids* 198–200, 875 (1996).
48. K. Nakagawa, M. Fukuda, S. Miyazaki, and M. Hirose, *Mat. Res. Soc. Proceed.* 452, 243 (1997).
49. Z. H. Lu, D. J. Lockwood, and J. M. Baribeau, *Nature* (London) 378, 258 (1995).
50. B. T. Sullivan, D. J. Lockwood, H. J. Labbe, and Z. H. Lu, *Appl. Phys. Lett.* 69, 3149 (1996).
51. T. Baron, F. Mazen, C. Busseret, A. Souifi, P. Mur, F. Fournel, M. N. Séméria, H. Moriceau, B. Aspard, P. Gentile, and N. Magnea, *Micr. Engin.* 61–62, 511 (2002).
52. W. T. Leach, J. Zhu, and J. G. Ekerdt, *J. Cryst. Growth* 240, 415 (2002).
53. J. Y. Jeong, S. Im, M. S. Oh, H. B Kim, K. H. Chae, C. N. Whong, and J. H. Song, *J. Lumin.* 80, 285 (1999).
54. L. T. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
55. A. G. Cullis, L. T. Canham, and P. D. J. Calcott, *J. Appl. Phys.* 82, 909 (1997).
56. A. G. Nassiopoulou, S. Grigoropoulos, E. Gogolidis, and D. Papadimitriou, *Appl. Phys. Lett.* 66, 1114 (1995).
57. D. Papadimitriou and A. G. Nassiopoulou, *J. Appl. Phys.* 84, 1059 (1998).
58. D. Papadimitriou, Y. S. Raptis, and A. G. Nassiopoulou, *Phys. Rev. B* 58, 14 089 (1998).
59. F. Arnaud d'Avitaya, L. Vervoort, F. Bassani, S. Ossicini, and F. Bernadini, *Europhys. Lett.* 31, 25 (1995).
60. F. Bassani, L. Vervoort, I. Mihalcescu, J. C. Vial, and F. Arnaud d'Avitaya, *J. Appl. Phys.* 79, 4066 (1996).
61. A. G. Nassiopoulou, V. Tsakiri, V. Ioannou-Sougleridis, P. Photopoulos, S. Menard, F. Bassani, and F. Arnaud d'Avitaya, *J. Lumin.* 80, 81 (1999).
62. A. G. Nassiopoulou, V. Tsakiri, V. Ioannou-Sougleridis, P. Photopoulos, S. Menard, F. Bassani, and F.Arnaud d'Avitaya, *J. Lumin.* 80, 81 (1999).
63. V. Ioannou-Sougleridis, V. Tsakiri, A. G. Nassiopoulou, P. Photopoulos, F. Bassani, and F. Arnaud d'Avitaya, *Phys. Status Solidi A* 165, 97 (1998).
64. V. Ioannou-Sougleridis, T. Ouisse, A. G. Nassiopoulou, F. Bassani, and F. Arenaud d'Avitaya, *J. Appl. Phys.* 89, 610 (2001).
65. D. Papadimitriou, A. G. Nassipoulou, F. Bassani, and F. Arnaud d'Avitaya, *Mater. Sci. Eng., B* 69–70, 546 (2000).
66. V. Ioannou-Sougleridis, V. Tsakiri, A. G. Nassiopoulou, F. Bassani, S. Menard, and F. Arnaud d'Avitaya, *Mater. Sci. Eng., B* 69–70, 309 (2000).
67. P. Photopoulos, A. G. Nassiopoulou, D. N. Kouvatsos, and A. Travlos, *Appl. Phys. Lett.* 76, 3588 (2000).
68. J. Linnros, A. Galeckas, N. Lalic, and V. Grivickas, *Thin Solid Films* 297, 197 (1997).
69. J. Linnros, N. Lalic, A. Galeckas, and V. Grivickas, *J. Appl. Phys.* 86, 6128 (1999).
70. William L. Wilson, P. F. Szajowski, and L. E. Brus, *Science* (Washington, D.C.), 262, 1242 (1993).
71. C. Delerue, in "Properties of Porous Silicon" (L. T. Canham, Ed.), Vol. 212. IEE INSPEC, The Institution of Electrical Engineers, London, 1997.
72. A. Zunger and S. B. Zhang, *Appl. Surf. Sci.* 102, 350 (1996).
73. M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allan, and C. Delerue, *Phys. Rev. Lett.* 82, 197 (1999).
74. H. Mizuno, H. Koyama, and N. Koshida, *Appl. Phys. Lett.* 69, 3779 (1996).
75. G. Allan, C. Delerue, and M. Lannoo, *Phys. Rev. Lett.* 76, 2961 (1996).
76. B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001).
77. L. Brus, *J. Phys. Chem.* 98, 3575 (1994).
78. N. Ookudo, H. Ono, Y. Ochiai, Y. Mochizuchi, and S. Matsui, *Appl. Phys. Lett.* 61, 940 (1992).
79. H. Koyama, T. Ozaki, and N. Koshida, *Phys. Rev. B* 52, 562 (1995).
80. Y. Kanemitsu and S. Okamoto, *Phys. Rev. B* 56, R15561 (1997).
81. X. Zianni and A. G. Nassiopoulou, *Phys. Rev. B* 65, 326 (2002).
82. L. E. Brus, P. F. Szajowski, W. L. Wilson, T. D. Harris, S. Schuppler, and P. H. Citrin, *J. Am. Chem. Soc.* 117, 2915 (1995).
83. I. Mihalcescu, J. C. Vial, A. Bsiesy, F. Muller, R. Romestain, E. Martin, C. Delerue, M. Lannoo, and G. Allan, *Phys. Rev. B* 51, 17605 (1995).

84. C. Delerue, M. Lannoo, G. Allan, E. Martin, I. Mihalcescu, J. C. Vial, R. Romenstain, F. Muller, and A. Bsiesy, *Phys. Rev. Lett.* 75, 2228 (1995).
85. L. Efros, M. Rosen, B. Averboukh, D. Kovalev, M. Ben-Chorin, and F. Koch, *Phys. Rev. B.* 56, 3875 (1997).
86. M. S. Hybertsen, *Phys. Rev. Lett.* 72, 1514 (1994).
87. G. Vijaya Prakash, N. Daldosso, E. Degoli, F. Iacona, M. Cazzanelli, Z. Gaburro, G. Pucker, P. Dalba, F. Rocca, E. Ceretta Moreira, G. Franzo, D. Pacifici, F. Priolo, C. Arcangeli, A. B. Filonov, S. Ossicini, and L. Pavesi, *J. Nanosci. Nanotechnol.* 1, 159 (2001).
88. D. J. Norris, A. L. Efros, M. Efros, M. Rosen, and M. G. Bawende, *Phys. Rev. B.* 53, 16347 (1995).
89. D. Kovalev, H. Heckler, M. Ben-Chorin, G. Polisski, M. Schwartzkopft, and F. Koch, *Phys. Rev. Lett.* 81, 2803 (1998).
90. D. Kovalev, H. Heckler, G. Polisski, J. Diener, and F. Koch, *Opt. Mater.* 17, 35 (2001).
91. D. Kovalev, J. Diener, H. Heckler, G. Polisski, N. Künzner, and F. Koch, *Phys. Rev. B* 61, 4485 (2000).
92. F. Priolo, G. Franzo, D. Pacifici, and V. Vincinguerra, *J. Appl. Phys.* 89, 264 (2001).
93. C. Garcia, B. Garrido, P. Pellegrino, R. Ferre, J. A. Moreno, L. Pavesi, M. Cazzanelli, and J. R. Morande, *Physica E* 16, 429 (2003).
94. L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzo, and F. Priodo, *Nature* (London) 408, 440 (2000).
95. L. Khriachtchev, M. Rasanen, S. Novikov, and J. Sinkkonrn, *Appl. Phys. Lett.* 79, 1249 (2001).
96. M. Nayfeh, S. Rao, N. Barry, J. Therrien, G. Belomoin, A. Smith, and S. Chaieb, *Appl. Phys. Lett.* 80, 121 (2002).
97. K. Luterova, J. Pelant, I. Mikulskas, R. Tomasiunaset, D. Muller, J.-J. Grob, J.-L. Rehspringer, and B. Honeslage, *J. Appl. Phys.* 91, 2896 (2002).
98. L. Dal Negro, M. Cazzanelli, N. Daldosso, A. Gaburro, L. Pavesi, F. Priolo, D. Pacifici, G. Franzo, and F. Iacona, *Physica E* 16, 297 (2003).
99. P. Photopoulos and A. G. Nassiopoulou, *Appl. Phys. Lett.* 77, 1816 (2000).
100. A. G. Nassiopoulou, S. Grigoropoulos, and D. Papadimitriou, *Appl. Phys. Lett.* 69, 2267 (1996).
101. S. Grigoropoulos, A. G. Nassiopoulou, A. Travlos, D. Papadimitriou, S. Kennou, and S. Ladas, *Appl. Surf. Sci.* 102, 377 (1996).
102. A. G. Nassiopoulou, S. Grigoropoulos, and D. Papadimitriou, *Thin Solid Films* 297, 176 (1997).
103. A. G. Nassiopoulou, S. Grigoropoulos, E. Goggolidis, and D. Papadimitriou, *Appl. Phys. Lett.* 66, 1114 (1995).
104. N. Lalic and J. Linnros, *J. Lumin.* 80, 263 (1999).
105. T. Quisse and A. G. Nassiopoulou, *Europhys. Lett.* 51, 168 (2000).
106. T. Ouisse, A. G. Nassiopoulou, and D. N. Kouvatsos, "Proceedings of MMN 2000" (A. G. Nassiopoulou and X. Zianni, Eds.), 65, World Scientific, 2000.
107. G. Bastard, E. E. Mendez, L. L. Chang, and L. Esaki, *Phys. Rev. B* 28, 3241 (1983).
108. E. J. Austin and M. Jaros, *Phys. Rev. B* 31, 5569 (1985).
109. E. E. Mendez, G. Bastard, L. L. Chang, L. Esaki, H. Morkoc, and R. Fischer, *Phys. Rev. B* 26, 7101 (1982).
110. G. Bastard and J. Brum, *IEEE J. Quantum Electron.* QE 22, 1625 (1986).
111. V. Ioannou-Sougleridis, B. Kamenev, D. N. Kouvatsos, and A. G. Nassiopoulou, *Mater. Scien. Eng., B* (2003) (to appear).
112. T. Ouisse, V. Ioannou-Sougleridis, D. Kouvatsos, and A. G. Nassiopoulou, *J. Phys. D: Appl. Phys.* 33, 2691 (2000).
113. S. H Choi and R. G Elliman, *Appl. Phys. Lett.* 75, 968 (1999).
114. Q. Ye, R. Tsu, and E. H. Nicollian, *Phys. Rev. Lett.* 76, 539 (1996).
115. S. Tiwari, F. Rama, H. Hanafi, A. Hartstein, E. F. Grabbe, and K. Chan, *Appl. Phys. Lett.* 68 (1996).
116. S. Tiwari, F. Rama, K. Chan, L. Shi, and H. Hanafi, *Appl. Phys. Lett.* 69, 1232 (1996).
117. A. T. Tilke, F. C. Simmel, R. H. Blick, H. Lornz, and J. P. Kotthaus, *Prog. Quant. Electron.* 25, 97 (2001).
118. H. Hanafi, S. Tiwari, and I. Khan, *IEEE Trans. Electron Devices* 43, 1553 (1996).
119. D. N. Kouvatsos, V. Ioannou-Sougleridis, and A. G. Nassiopoulou, *Mater. Sci. Eng., B* (2003) (to appear).
120. T. Maeda, E. Suzuki, I. Sakata, M. Yamanaka, and K. Ishii, *Nanotechnology.* 10, 127 (1999).
121. D. N. Kouvatsos, V. Ioannou-Sougleridis, and A. G. Nassiopoulou, *Appl. Phys. Lett.* (2003) (to appear).

www.aspbs.com/enn

Silicon Nanowires

Klaus Sattler
University of Hawaii at Manoa, Honolulu, Hawaii, USA

CONTENTS

1. Introduction
2. Bottom-Up Fabrication
3. Top-Down Fabrication
4. Analytic Techniques
5. Morphology, Surface, Doping
6. Electronic and Optical Properties
7. Transport Properties
8. Applications
9. Summary
Glossary
References

1. INTRODUCTION

In the past decade there have been major advances in materials growth techniques as well as electron-beam lithography and reactive ion etching. Scanning probe microscopy manipulation became more advanced and able to form nanoscale structures. These developments made it possible to produce and study quasi-one-dimensional structures from metals and semiconductors.

The development of patterning and self-assembly procedure now allows fabrication of high-quality nanowires. Many techniques have been used to grow and fabricate nanowires in large quantities, from materials such as Si [1], Ge [2], GaAs [3], GaN [4], GaSe [5], Al_2O_3 [6], SiC [7, 8], Si_3N_4 [9], and various metals [10]. The nanowires not only allow the study of new physical phenomena characteristic for materials with reduced dimensions, but may also lead to future nanoscale device applications. In order to apply nanowires, one needs to study their physical and chemical properties and to understand the growth mechanism.

Semiconductor whiskers such as Si, Ge, GaN, etc. have been studied over 30 years. In recent years, however, the development of new techniques allowed much better control of size and structure of low-dimensional materials and the fabrication of nanowires. Among the nanowires from different materials, silicon nanowires (SiNWs) have attracted much attention because silicon is the most important semiconductor material as of now. Silicon nanowires differ significantly from bulk silicon and may find applications in the new emerging field of nanoelectronics.

A variety of techniques for synthesizing silicon nanowires has been developed, including lithography and etching techniques [11, 12], scanning tunneling microscopy [13], thermal evaporation [14], and laser ablation [3]. Yet it is still a challenge to control the morphology and diameter of silicon nanowires.

One of the most challenging tasks in recent years has been the synthesis of SiNWs on large scales. The well-known metal-catalyzed vapor-liquid-solid (VLS) growth model [10] was introduced by Wagner and Ellis for the unidirectional growth of silicon whiskers. In recent years, this technique has also been used for growth of silicon nanowires [3, 15, 16]. Large-scale synthesis of SiNWs has been achieved using the VLS process [17]. The preparation of bulk quantities of SiNWs makes it possible to study these materials in detail by various analytic techniques.

Bottom-up techniques deal with the growth of SiNWs from individual atoms. This can be performed with different approaches, for example by simple evaporation [18, 19], laser ablation [20], chemical vapor deposition [21], vapor-liquid-solid (VLS) growth [22], or epitaxial growth [23]. Some techniques, such as the template-directed vapor-liquid-solid growth [24] or magnetron sputter deposition [19], have been used to form aligned SiNWs.

SiNWs can also be fabricated by top-down approaches. These include electron-beam lithography with etching [25], pattern transfer [26], SiO_2 reduction under electron-beam radiation [27], or stress-limited oxidation [28].

In the area of scanning probe microscope (SPM) lithography, a variety of techniques has been developed: (1) local oxidation of silicon and metals [29–31], (2) superfine e-beam lithography using the tip of a STM as a low-energy electron source [32, 33], (3) direct movement of atoms and molecules [34], and direct mechanical modification of structures [35, 36].

SiNWs with different morphologies could be fabricated including amorphous and crystalline configurations. In particular, much attention has been given to single-crystalline SiNWs which are defect-free. Control of wire structure,

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

thickness, and surface layers has been achieved by various synthetic methods and post-treatments.

Noncrystalline but ordered structures, such as silicon nanotubes [37], superatom-stacking [38], and fullerene-structured SiNWs [19], have been proposed.

The surface of SiNWs has been studied in pristine form [19], hydrogen-passivated [39], silicon oxide sheathed [40], and SiC-coated [40].

SiNWs have been studied which were lightly doped with foreign atoms (such as boron [41] or lithium [42]) or had nanocrystallites such as copper silicide incorporated [43].

With most methods the SiNWs are grown or fabricated on solid supports. A few studies report on the fabrication of freestanding silicon nanowires [5, 44]. Other techniques were used to generate SiNWs incorporated in silicon oxide [45], in suspended form [46], as metal-silicon networks [47], or as composites with carbon nanotubes [48]. Silicon nanowires are widely considered to constitute the microstructure of porous silicon. Therefore, assemblies of SiNWs in parallel or other configurations have been investigated [49–52].

Various analytic techniques for structural investigation of SiNWs have been applied. Size, atomic structure, and morphology have been determined by transmission electron microscopy (TEM) [20], X-ray diffraction [14, 42], STM [41], and AFM [53]. Raman spectroscopy of silicon nanowires has been discussed with respect to bulk silicon [54, 55]. Field-emission has been studied under various conditions [56].

Electronic and thermal properties of silicon nanowires have been studied such as electron density of states [57], and phonon heat conduction [58]. In particular, the quantum confinement of charge carriers, excitons, and phonons is of interest. Charge transport is characterized by anomalous nonlinear behavior such as quantized conductance [59], negative resistance [60], hysteresis characteristics [60], and single-electron effects [61]. Optical properties, such as nonlinear behavior [62], photoluminescence [55], optical anisotropy [63], excitonic effects [63], and radiative lifetimes [64], have been studied. The results show that silicon nanowires have optical and thermal properties that differ significantly from bulk when their diameter is a few nanometers. This is mainly due to their one-dimensional electronic structure, which hardly resembles the bulk crystalline structure of silicon. The energy band structure of silicon nanowires depends on the reduction in dimension, morphology and atomic structure, and the surface conditions.

Theory of silicon nanowires was performed using the effective mass approximation (EMA) [65], first-principles techniques such as pseudopotential [65] or local density functional [39], empirical tight-binding [66], or semiempirical techniques [19]. Much attention has been given to understanding the electronic and optical properties related to the wire structures. Wires with crystalline or other network structures were considered and surfaces were either pristine or passivated. The bandgap, electronic density of states, and band edge electronic states were determined within first-principles calculations [49, 50, 57, 63, 65, 67–70] and the tight-binding method [71, 72].

The efforts in theory of silicon nanowires had intensified after in 1990 Canham [52] proposed that the observed efficient, visible photoluminescence in porous silicon is due to an array of silicon nanowires. This led to extensive theoretical studies focusing on the origin and mechanism of the luminescence and in particular on the quantum confinement in SiNWs. The two most important explanations of the photoluminescence are the quantum confinement (to Si nanoparticles and/or Si nanowires) and chemical luminescence models. There has been enormous work on porous silicon and other Si-based low-dimensional structures, in order to understand their unusual optical behaviors.

Most of the theoretical work has been done on nanowires cut out from bulk diamond-structure Si. Dangling bonds at the surface of the wires were usually passivated with hydrogen. Bandgaps and interband oscillator strengths were calculated [73]. Structural relaxation greatly improves the oscillator strength, which suggests that free nanowires are stronger light emitters than those embedded in porous silicon or other matrixes.

The strong interest in understanding porous silicon and other low-dimensional silicon structures is due to the possibility of developing Si-based optoelectronic integrated circuits (OEIC). In the past decade, many studies on porous-silicon-related nanotechnology have been stimulated by its potential applications in silicon-based optoelectronic devices.

SiNWs contain a quasi-one-dimensional electron gas which gives them novel properties for silicon-based electronic devices, such as the field-effect transistors [74], single-electron memory devices [75], silicon quantum wire lasers [76], metal-semiconductor junctions [77], or thermoelectric quantum wire arrays [78].

2. BOTTOM-UP FABRICATION

2.1. Vapor-Liquid-Solid Growth

For more than 30 years, the growth of silicon whiskers has been studied by using the vapor-liquid-solid (VLS) method. For the VLS method, small metal clusters are used to catalyze the whisker growth. The metal particle is either supported on a surface or produced in the gas phase. The diameters of the whiskers obtained from VLS are determined by the size of the metal particles. Variations of VLS techniques are solution-liquid-solid [79] and vapor-solid-solid [80] methods. All these techniques require the formation of nanoscale catalyst particles. Recently, the VLS method has been applied to grow silicon nanowires [3].

Silicon nanowires 15–35 nm in diameter were grown using SiH_4 chemical vapor deposition via Au and Zn particle-nucleated vapor-liquid-solid growth at 440 °C [81]. The Au-nucleated wires (Au-SiNWs) were grown by first evaporating 1 nm of Au onto a SiO_2 wafer. If the wafer then is annealed to 450 °C, Au particles form which will act as catalytic sites for the SiNW growth. The morphology and width of the wires critically depend on temperature and pressure used in the process [22]. At low pressure and high temperature, relatively thick nanowires grow, and thinner ones grow at higher pressure and lower temperature. The nanowires can be thinned further by oxidation. Surface oxidation can yield silicon cores as small as 5 nm.

Zn particles can also be used as catalysts but are fabricated with a different method [81]. For synthesis of Zn-catalyzed

silicon nanowires (Zn-SiNWs), a $ZnCl_2$/ethanol solution was deposited on a p-doped silicon substrate after removing the native oxide layer with HF. The wires then were grown using chemical vapor deposition from 5% SiH_4/He gas at 100 Torr. Wires in the diameter range 14–35 nm and length of 1–10 μm were grown. The wires are crystalline in the core with a 1-nm oxide coating.

The required synthesis temperatures for VLS growth of SiNWs range from 450–500 °C for gold and zinc to 950–1150 °C for transition-metal catalysts. By using low-melting-point metals low-temperature VLS growth of SiNWs can be achieved. Six-nanometer-wide SiNWs with uniform diameters were synthesized using gallium as a molten solvent, at temperatures less than 400 °C in hydrogen plasma [82]. Liquid-metal gallium has low solubility for silicon. Gallium forms a eutectic with silicon near room temperature.

2.2. Chemical Vapor Deposition

Ti-containing islands have been grown on silicon substrates by chemical vapor deposition using the decomposition of TiO_4 [83, 84]. When these islands are exposed to SiH_4 or SiH_2Cl_2, the metal catalyzes the decomposition of the silicon precursor and long SiNWs can grow. Exposing the wires to an ion beam after deposition promotes alignment of nanowires.

Quite straight SiNWs were grown using catalytic synthesis over Fe containing silica gel substrates by thermal chemical vapor deposition where silane is decomposed [85]. The fact that the nanowires are straight indicates that there are few defects and little growth stress. The nanowires are of uniform diameter of about 20 nm with a length of the order of 1 μm. A nanoparticle with a diameter of ~50 nm was found at the end of each nanowire. The center part of the nanowires is well crystallized and there is a thin amorphous surface layer, as revealed by high-resolution electron microscopy (HREM) and electron diffraction. The Raman spectrum of the wires has a prominent peak at 511 cm^{-1} with a shoulder at 490 cm^{-1}. The peak has a larger line width (about 15 cm^{-1}) and is down-shifted (about 9 cm^{-1}) as compared to that of bulk crystalline silicon. This is considered to be caused by the quantum confinement effects since similar behavior has also been found in porous silicon [86] and freestanding nanowires [87].

2.3. Laser Ablation

Laser ablation has become an important method for the synthesis of silicon nanowires [3, 88, 89]. Depending on the applied conditions, nanowires of different form and structural quality are produced. Of particular interest are long, uniform-sized and single-crystal wires, which can be fabricated in bulk quantities [90].

The internal structure of nanowires is closely related to the growth kinetics. SiNWs fabricated by laser ablation can have a high density of defects [20] and show various morphologies [91] which are closely related to the insertion of twins or the special twisting along the growth direction during the growth process. The nanowires can show morphological characteristics with straight, curved, kink, braided, and coiled shapes.

The combination of laser ablation cluster formation and vapor-liquid-solid (VLS) growth was used to synthesize SiNWs [3]. Laser ablation of a Si target with small Fe content was used to prepare metal catalyst clusters that determine the width of the silicon nanowires. The method produces bulk quantities of single-crystal silicon nanowires with diameters of 6 to 20 nm and lengths ranging from 1 to 30 μm.

It has been found that the addition of metal to the starting material has negligible effect on the yield of the SiNWs produced by laser ablation. Rather, the addition of SiO_2 powder to the Si target significantly increased the yield [17]. This suggests that silicon monoxide (SiO), which is produced by the reaction between Si and SiO_2, plays an important role in the growth of silicon nanowires (oxide-assisted growth model).

Silicon nanowires were synthesized by laser ablation of mixed SiC and SiO_2 powders at 1400 °C [92]. The SiNWs were about 14 nm in diameter and coexisted with a small amount of SiC nanoparticles. Such silicon nanowires have the diamond-type crystal structure and the nanoparticles are cubic SiC.

Silicon nanowires with high yield, uniform diameter distribution, and high purity were grown by excimer laser ablation at high temperature [88]. The target was made by mixing silicon powder with nano-sized Ni and Co powder (about 5 wt.% each). The mixed powder was hot-pressed at 150 °C to form a plate target. The target was ablated at a temperature of 1200 °C by using an excimer laser with a wavelength of 248 nm at pressure of about 500 Torr Ar. The ablated Si atoms self-organize into nanowires as a dark yellow colored deposit on a quartz tube wall [88]. For the ablation of targets containing a Co-Ni catalyst, most silicon nanowires have a curve-shaped morphology and some are largely coiled. More than 99% of the deposit is as silicon nanowires. Most of the silicon nanowires have diameters of 13 ± 3 nm. Their length ranges from a few tens of micrometers to hundreds of micrometers. Silicon nanowires produced from Co-Ni catalyst-free targets are different. Most of the nanowires have a diameter, which is not uniform at all, ranging from 15 nm to 60 nm.

Using laser ablation at high temperature, silicon nanowires were fabricated with very small diameters from 3 to 43 nm up to a few hundred microns long [16]. Twins and stacking faults have been observed in the silicon core of the nanowires.

By ablation of a silicon monoxide target with a pulsed KrF excimer laser at 1200 °C in an argon atmosphere, high-purity silicon nanowires were fabricated in bulk quantities [93]. The yield and linear growth rate can reach 30 mg/h and 500 μm/h, respectively.

Along with SiNWs, chains of silicon nanoparticles have also been formed. These were synthesized by laser ablation of a target containing metals over a temperature range 910–1120 °C [27].

2.4. Thermal Evaporation

Thermal evaporation of SiO powders was found to produce high-purity silicon nanowires in bulk quantity [94]. Growth of the SiNWs was studied in a closed system with different ambient conditions [95]. A silicon nanowire growth rate of

10 nm/h could be routinely obtained. With increasing ambient pressure, the yield of SiNWs decreased and the diameter of the SiNWs increased.

Thermal evaporation and laser ablation of SiO_2 powders were studied [17], by both methods separately and by combining the methods. The growth rate of silicon nanowires was higher with the assistance of laser ablation than with thermal evaporation alone. The solid source was pure silicon powder mixed with 70% of SiO_2 powder. After 12 hours of thermal evaporation, a silicon nanowire product (sponge-like, dark red in color) formed on the inside of the quartz tube. The method leads to two types of products, silicon nanowires with uniform and smooth surfaces (the major component) and silicon nanoparticles. Most silicon nanowires were very long, 10 microns and more, and randomly oriented. Most silicon nanoparticles appeared in form of chain [17].

Elevated temperature synthesis has been used to fabricate SiO_2-passivated (sheathed) crystalline SiNWs [80]. The nanowires were grown by heating a crucible with a Si-SiO_2 mixture in flowing argon gas. Condensation of nanowires as dark brown deposits was found in a narrow region on the wall of the tube enclosing the crucible. The central crystalline silicon core was $\sim$30 nm in diameter and the outer SiO_2 sheathing was 15 nm in thickness [80].

Thermal evaporation of a powder mixture of Si and SiO_2 [15, 40] was found to be an effective method for the growth of silicon nanowires on a silicon wafer [17]. They consist of a core of crystalline silicon clothed by a shell of silicon oxide with a thickness up to half of the wire radius [88]. Very weak photoluminescence at 630 nm (2 eV) was obtained for the as-grown sample. The as-grown SiNW sample was transferred to an ion-beam deposition chamber and a mixture of methane, hydrogen, and argon (mixture ratio 1:50:173) was used to deposit a thin film of SiC onto the sample. TEM analysis showed that the silicon dioxide layer was replaced by a cubic silicon carbide layer coating the silicon nanowires.

The synthesis of SiNWs on iron patterned silicon substrates has been performed by thermal evaporation of pure silicon powder [18]. This results in straight crystalline SiNWs with diameters of 10–60 nm. The growth of the nanowires can be controlled by positioning of iron squares on the silicon substrate using current semiconductor technology. For nanowire formation, p-type silicon(100) substrates were patterned with 5-nm-thick iron film by electron-beam evaporation through a shadow mask, which contains square-shaped openings of micrometer size. The substrates were then annealed in argon at 300 °C for 3 h. Thermal evaporation of silicon powder from a furnace was then used to grow the SiNWs.

Si nanowires have been grown by vapor deposition on various substrates including sapphire, Al_2O_3, and Si(111). Step-flow growth of silicon nanowires by gas-source molecular beam epitaxy on sapphire was used to fabricate parallel silicon nanowires [96]. The wires were aligned along the substrate steps and formed uniformly with 50 nm width and 1 nm height. Si nanowires were grown on an Al_2O_3 substrate at ambient pressure by using $SiCl_4$ as Si source [97]. Highly oriented amorphous SiNWs have been grown on Si(111) by a solid-liquid-solid (SLS) mechanism [98]. The length and diameter of the nanowires are almost uniform, 1 μm and 25 nm, respectively.

2.5. Magnetron Sputter Deposition

Bundles of parallel silicon nanowires were produced by using a magnetron sputter source for the deposition of pure silicon vapor onto highly oriented pyrolytic graphite (HOPG) [19]. The experiments were done in UHV at 10^{-10} torr and no catalyst particles were used. Therefore the nanowires grew from quenching the pure silicon vapor on the cold substrate. Since HOPG is atomically flat and chemically inert, the wires are formed by quasi-free growth. Uniform diameter for the wires in a bundle is found (Fig. 1).

2.6. STM-Induced Growth

In the setup of an STM, silicon nanowires were grown by applying a voltage at constant current between a heated silicon substrate (700 °C) and a gold tip [13]. Silicon atoms were deposited onto the apex of the gold tip by field evaporation. The applied sample voltage was 5 V and the tunneling current was 10 nA which yields a tip-sample distance of 0.6 nm. The voltage was applied for 15 minutes during which the deposition of silicon on the gold tip was monitored by *in-situ* scanning electron microscopy (SEM). The diameter and length of the wire were 20–150 nm and 3 μm, respectively. Growth of the silicon nanowire occurs due to atom transfer from the substrate to the tip. The maximum growth rate (for the maximum diameter of the wire of 150 nm) was 6.0 nm/s. The deposition rate was 2.1×10^7 atoms/s.

2.7. Template-Directed Growth

SiNWs have been synthesized in nanoporous alumina membranes using a combination of Au electrodeposition and vapor-liquid-solid growth at 500 °C using SiH_4 as the silicon source [24]. The diameters of the SiNWs were 200 ± 54 nm,

Figure 1. Several bundles of parallel silicon nanowires produced by magnetron sputter deposition of silicon on highly oriented pyrolytic graphite (HOPG) in UHV. Reprinted with permission from [19], B. Marsen and K. Sattler, *Phys. Rev. B* 60, 11593 (1999). © 1999, American Physical Society.

which was close to the pore size distribution of the membranes. The nanowires consist of a crystalline Si core, oriented in the (100) or (211) growth direction, with a thin (<3 nm) native oxide coating. The two ends of the nanowires are terminated by segments of gold, which result from the VLS mechanism of growth. This technique could also be applied for the synthesis of smaller diameter nanowires by using anodized aluminum oxide templates with pore sizes in the range of 10–100 nm [99].

3. TOP-DOWN FABRICATION

3.1. Lithography and Etching

Electron-beam or optical lithography combined with reactive ion etching are top-down methods for silicon nanostructure formation. These are conventional techniques employed in the microelectronic industry. SiNWs fabricated with these techniques are typically broader than 100 nm. However, by using sequences of sophisticated lithographic, etching, and oxidation techniques, much narrower SiNWs could be produced (Fig. 2). Crystalline SiNWs with diameters down to 10 nm and smooth surfaces have been fabricated by advanced procedures of electron-beam (EB) lithography and wet etching [100]. Amorphous SiNWs have been generated by optical lithography and reactive ion etching of a 9-nm-thick amorphous silicon film on a silicon crystal surface [58].

SiNWs were fabricated using an inorganic SiO_2 electron-beam (EB) resist process [25]. In this procedure, a conventional organic resist is substituted by an inorganic SiO_2 resist. The conventional EB lithography process using an organic resist shows relatively low resolution of around 30 nm. In the inorganic EB resist process, a resolution of 15 nm can be achieved. This is because the inorganic SiO_2 resist is hardly affected by the secondary electrons produced in the primary EB scattering. The technique was used to generate 15-nm-wide silicon nanowires and to fabricate a silicon nanowire memory transistor [25].

Linear windows 15 nm wide in a 0.3-nm-thick silicon oxide layer have been formed by focused electron-beam

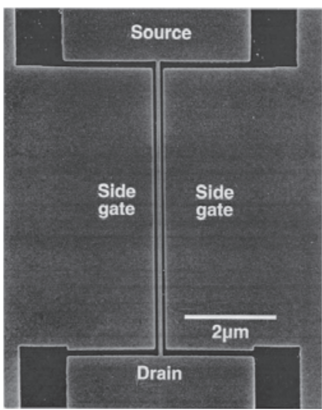


Figure 2. A silicon nanowire device fabricated with high-resolution electron-beam lithography. The wire is 70 nm wide over the 6-μm length of the side gates with a side gate separation of 80 nm. Reprinted with permission from [181], R. A. Smith and H. Ahmed, *J. Appl. Phys.* 81, 2699 (1997). © 1997, American Institute of Physics.

(EB) irradiation and successive HF-chemical etching [101]. Si-selective epitaxial growth (Si-SEG) with Si_2H_6 gas in the linear window of the mask has then been applied to fabricate silicon nanowires.

Silicon nanowires with SiO_2 boundaries were fabricated using a process based on SiGe/Si heterostructure [102]. First a Si/SiGe/Si heterostructure is grown on a silicon substrate with very low-pressure chemical vapor deposition. This is followed by lithography and reactive ion etching to form trench structures. Then SiGe is selectively etched over silicon and finally thermal oxidation is applied to obtain the silicon nanowires. A nanowire diameter of about 40 nm is obtained. The diameter can be controlled by the parameters used for selective chemical etching and thermal oxidation.

Orientation-dependent etching with subsequent oxidation has been applied to obtain silicon nanowires [12]. Silicon nanowires 2 nm wide were fabricated through a combination of a partly shifted resist pattern and orientation-dependent etching [53]. A $\{110\}$ silicon substrate is etched with KOH using a line pattern mask, which is shifted in the $\langle 111 \rangle$ direction at the line center position. The technique takes advantage of the fact that during silicon etching with KOH, the etching rate of the $\{111\}$ plane is more than 100 times lower than that of the other planes. Due to this selective etching property and to the shift technique, line patterns smaller than the beam diameter of lithographic tools can be formed.

SiNWs were generated by application of vertical ultrafine SiO_2 wall masks using oxidation of a polycrystalline silicon (poly-Si) layer [103]. For generation of a wall mask, electron cyclotron resonance (ECR) plasma etching and wet etching of the poly-Si layer after the sidewall oxidation has been applied. The wall masks are 10 nm wide and 90 nm high, and Si nanowires 10 nm wide and 18 nm thick have been obtained.

3.2. Selective Thermal Desorption

Silicon nanowires were fabricated using electron-beam-induced selective thermal desorption (EB-STD) of SiO_2 in silicon substrates [104, 105]. Electron-beam interference fringes have been used to form a 10-nm periodic surface structure of Si and SiO_2 by STD [104].

3.3. Pattern Transfer

Pattern transfer is an effective formation method for SiNWs. A process has been used for preparing high densities (up to 500 μm^{-2}) of well-passivated sub-5-nm diameter silicon nanowires [26]. In this formation process, a fine-grained natural mask is deposited. Subsequently, the pattern is transferred into the silicon substrate by reactive ion etching. The process was applied on silicon-on-insulator material, where the buried oxide layer acts as an etch stop. This isolates the nanowires from the bulk and from each other. The nanowire diameter was then brought below 5 nm by self-limiting oxidation. Subsequently, hydrogen-rich amorphous silicon nitride, prepared by plasma enhanced chemical vapor deposition, was deposited to act as a source of hydrogen for passivation of surface dangling bonds.

A resist film can be patterned in various ways in order to produce nanoscale structures on a substrate. Dots or

lines of small size or diameter can be written into the resist leaving openings to the substrate material. Applying such procedures, silicon nanowires have been fabricated using electron cyclotron resonance plasma deposition of a SiO_2 film through the openings of a patterned resist film [106].

3.4. Self-Limiting Oxidation

Self-limiting oxidation can be used to drastically reduce the active diameter of silicon nanowires. An initially broad diameter distribution can become much narrower and is significantly shifted to smaller diameters if the SiNWs are oxidized at elevated temperatures [26] (Figs. 3, 4).

Planar single-crystal silicon nanowires down to 8 nm in diameter were generated using self-limiting oxidation [28]. The SiNWs were formed on silicon-on-insulator (SOI) substrates with electron-beam lithography followed by a metal liftoff process and a silicon plasma etch. Oxidation was then used to shrink the wires to a sub-10-nm diameter. The stress generated by the expansion of the oxide during oxidation eventually stops the reaction, leaving a narrow silicon core. Extremely thinned SiNWs with 3–5 nm width were obtained by utilizing SiO_2 side-wall masks and self-limiting oxidation [107].

3.5. Electron-Beam-Induced Reduction of SiO_2

A thin film of SiO_2 can be directly reduced to Si under electron-beam irradiation [27]. If SiO_2 is irradiated with a high-intensity 100-keV electron beam with a dose of $\geq 3 \times 10^9$ Cm^{-2} of nanometer scale, silicon nanowires can be formed as small as 2 nm in diameter. No resist or chemical development is required.

3.6. Oxygen Implantation

Silicon nanowires embedded in SiO_2 have been fabricated using *in-situ* multiple low-energy oxygen implantation in combination with silicon molecular beam epitaxy [23].

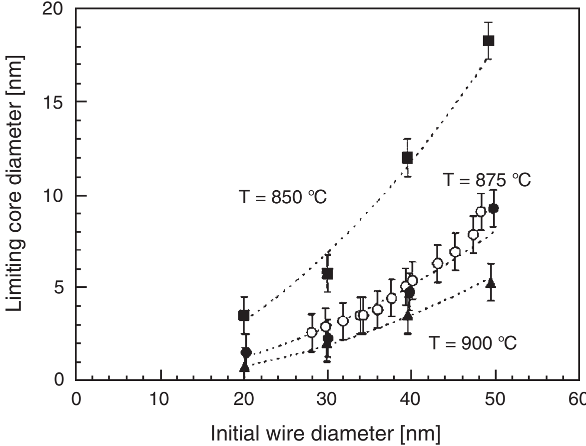


Figure 3. Core diameter of silicon nanowires as a function of the initial diameter after oxidation at different temperatures. Reprinted with permission from [26], M. Gotza et al., *J. Vac. Sci. Technol. B* 16, 582 (1998). © 1998, American Vacuum Society.

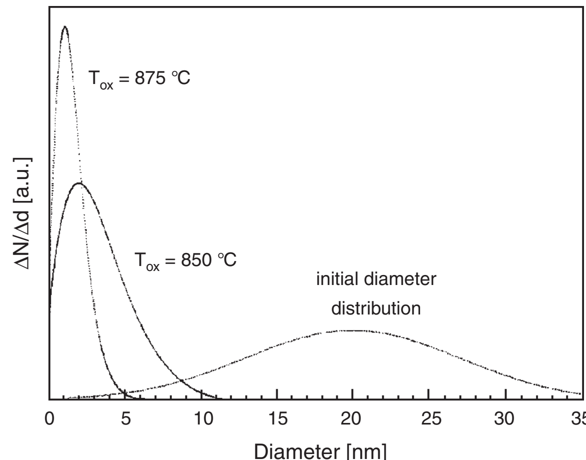


Figure 4. Initial silicon nanowire diameter distribution and calculated final diameter distribution of silicon oxide covered cores. Reprinted with permission from [26], M. Gotza et al., *J. Vac. Sci. Technol. B* 16, 582 (1998). © 1998, American Vacuum Society.

A patterned silicon substrate was used for the formation of the nanowires. The starting substrate was a p-type on-axis Si(100). V-grooves along [011] with well-developed (111) facets were produced by electron-beam lithography and KOH etching at a 1-μm interval. Silicon nanowires can be generated by oxygen implantation at the bottom of the V-groove. Using silicon molecular beam epitaxy (MBE), SiNWs can be formed at the hilltop. Thickness and width of V-groove and hilltop-type nanowires are about 10 nm and 10 to 50 nm, respectively. In view of optoelectronic applications for these techniques, parallel SiNWs were formed and their use as Bragg reflector was studied.

3.7. AFM and STM Lithography

SiO_2 nanoscale structured layers on a silicon surface could be basic constituents for ultra-large-scale integrated (ULSI) nanocircuits. Silicon oxide pattern and SiNWs have been performed on the silicon surface by using STM and AFM [108–111]. Silicon nanowires were also fabricated by using mechanical atomic force lithography [112].

4. ANALYTIC TECHNIQUES

4.1. Microscopy and Electron Diffraction

There are various analytic techniques used to investigate the properties of silicon nanowires. Some of these are listed and discussed in the following paragraphs.

The most often used technique for a first study of silicon nanowires is high-resolution electron microscopy (HREM) [20, 88, 112–115], often combined with selected area electron diffraction (SAED) [42, 114]. The techniques are applied to determine the diameter distribution and morphology of the nanowires. The atomic structure of SiNWs, such as single-crystalline and amorphous, has been determined. Some growth techniques produce crystalline SiNWs, however, with a high density of defects, such as micro-twins and stacking faults [20]. These can well be studied with transmission electron microscopy (TEM). A profile projector has

been used to enhance the magnification for examination of the TEM micrographs [113]. A thin amorphous silicon oxide layer is usually visible on the surface of the SiNWs, determined by electron diffraction. In some studies, field-emission scanning electron microscopy (FESEM) [115] has been used for structural characterization.

A technique has been developed to improve imaging of SiNWs with the plane-view transmission electron microscope [107]. The nanowires are covered with a Si_3N_4 holding layer in order to fix them and to give high contrast in the TEM image of the crystalline fine structure. The lattice image of very fine (3- to 5-nm-wide) SiNWs could be studied.

The X-ray diffraction (XRD) spectrum of crystalline silicon nanowires [88] contains several peaks, which are clearly distinguishable. They all can be indexed to the diamond structure of bulk silicon, both by their peak position and by their relative intensity. The lattice parameter of the silicon nanowires is close to the value obtained for bulk silicon.

A scanning reflection microscope (SREM) has been used in combination with a scanning tunneling microscope (STM) for SiNW studies [105]. By applying the combined SREM/STM methods one can obtain the information on surface geometric and crystallographic properties at the same position on a sample.

A high-resolution electron microscope with energy dispersive X-ray fluorescence (EDAX) was used for morphology and composition analysis [88]. In the EDAX spectrum of pure SiNWs only one peak is visible, which corresponds to silicon.

Extended X-ray absorption fine-structure (EXAFS) studies are used for structural investigation, in particular to study short-range versus long-range order. EXAFS measurements showed a progressive degradation in long-range order going from bulk silicon to nanowires to porous silicon [116]. For silicon nanowires, various types of defects are observed, but EXAFS spectra show that the degree of disorder in SiNWs is usually relatively small and the nanowires are essentially crystalline.

Microprobe reflection high-energy electron diffraction (μRHEED) [105] and microprobe Auger electron spectroscopy (μAES) have been used to study the surface atomic structure and composition of silicon nanowires and oxide masks [101, 105].

Scanning tunneling microscopy (STM) has been applied both for structural analysis [41] and for the fabrication [13] of SiNWs. Si nanowires from a liquid suspension were deposited on highly oriented pyrolytic graphite (HOPG) and analyzed by STM [46]. Scanning tunneling spectroscopy (STS) has been applied for SiNW electronic structure studies [41]. STM and STS measurements have been performed on boron-doped and undoped silicon nanowires [41]. Regular nanoscale domains were observed on the silicon nanowire surface, which were attributed to boron-induced surface reconstruction. STS measurements in silicon nanowires showed an enhancement of electrical conductivity by boron doping [41]. An STM was used for nanofabrication by taking the gap between a gold-tip and a silicon substrate as active region to grow silicon nanowires [13].

The width of silicon nanowires was measured using atomic force microscopy (AFM) [53, 117]. SiNWs deposited on the substrate from a liquid suspension were studied. Different AFM modes can be applied and tuned to the specific applications. Tapping mode and contact mode AFM were used to image SiNWs on HOPG and mica [46] where the coupling to the substrate is weak.

4.2. Raman Spectroscopy

Raman scattering is very sensitive to the lattice microstructure and the crystal symmetry of microcrystalline materials. The Raman spectra of silicon nanowires, single-crystalline silicon (c-Si), and completely oxidized silicon nanowires were recorded and compared [16]. The first-order Raman peak of c-Si at 521 cm^{-1} is symmetrical with a full width at half maximum (FWHM) of 5 cm^{-1}. The Raman spectrum for fully oxidized silicon nanowires shows a very broad asymmetric peak. A redshift of the phonon frequency was observed in Raman spectra of SiNWs and assigned to the compressive stress from the oxide skin at the wire surface [55]. The Raman peaks are found to shift and broaden with decreasing diameter [54, 118] (Fig. 5). Similar results were obtained with Raman spectra that were taken from porous silicon layers with different confinement size [119]. Theoretical studies resulted in similar conclusions for Raman spectra of SiNWs [120].

Raman spectra of SiNWs were obtained for several phonon modes at room temperature [118]. In addition to the fundamental phonon mode, overtone and combination modes were observed. Phonon confinement is used to explain the results. The confinement effect becomes more obvious for wire width less than 22 nm. Silicon nanowires with average diameter of 15 nm and length from a few micrometers to millimeters were studied. Compared to bulk silicon, the silicon nanowires show redshift of all Raman peaks and a broadening.

Raman scattering has also been performed on silicon nanoparticle chains and showed that phonons are confined in the silicon nanospheres [121].

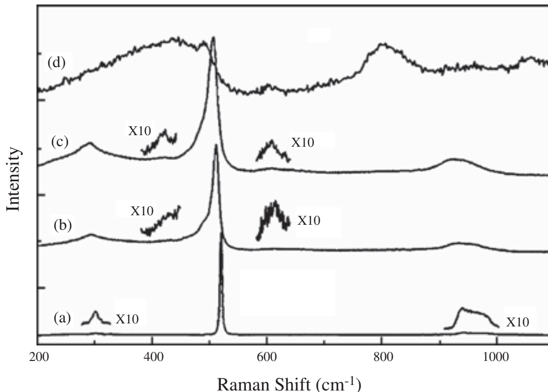


Figure 5. Experimental Raman spectra for (a) crystalline silicon, (b) SiNW with 20 nm diameter, (c) SiNW with 13 nm diameter, (d) nanowires of SiO_2. Reprinted with permission from [54], B. Li et al., *Phys. Rev. B* 59, 1645 (1999). © 1999, American Physical Society.

4.3. Electron Spectroscopy

X-ray photoelectron spectroscopy (PES) [113] and electron energy dispersive spectroscopy (EDS) [113] were used to perform chemical analysis on SiNWs. These techniques together with transmission electron microscopy and Raman spectroscopy showed that the wires consist of a crystalline silicon core with an amorphous silicon oxide surface.

Ultraviolet photoelectron spectroscopy (UPS) can be used to study both the crystalline structure and the chemical composition of a material. Silicon nanowires are essentially crystalline silicon encapsulated by silicon oxide [116] when some of the common fabrication methods are applied, a fact which can be obtained by the measurement of Si 2p and valence-band photoelectron spectra.

The bonding state and chemical composition of SiNWs were determined by parallel electron-loss spectroscopy (PEELS). It has been used to measure the concentration of Li ions in Li-doped SiNWs [42].

4.4. Optical Spectroscopy

UV-photoluminescence (UV-PL) spectroscopy was used to study the optical properties of nanowires and their electronic structures. Photoluminescence of porous silicon with varying confinement size was experimentally studied and explained by considering silicon nanowire and nanoparticle properties [119]. PL spectra have been taken, for example, for SiNWs fabricated on crystalline silicon (100). The wires emitted photoluminescence in the range of 500–600 nm [55]. Direct evidence of quantum confinement from size dependence of the PL of SiNWs was obtained [122]. The temperature dependence of the visible PL was determined over a wide range. For SiNWs on sapphire substrates the PL was observed at 9 and 300 K with a broad emission peak at 750 nm. The PL intensity at 9 K is over 20 times larger than that of 300 K. The PL peak barely shifts between 9 and 300 K [96].

4.5. Electrical Conductivity

In most studies of electrical conductivity, current-versus-voltage curves are taken, often at different temperature. Other techniques have been applied as well. Scanning Maxwell-stress microscopy (SMM) was used to investigate anomalous electrical conductance in a 100-nm-wide silicon wire [60]. SMM is capable of obtaining topographic and surface potential images simultaneously and gives the relationship between surface charging and electrical characteristics [123, 124].

5. MORPHOLOGY, SURFACE, DOPING

5.1. Form and Size

Most growth techniques yield nanowires with cylindrical shape, with a crystalline silicon core and an amorphous silicon oxide surface (Fig. 6). Theoretical expressions were derived for the calculation of energy eigenvalues in cylindrical and elliptic nanowires with finite potential barriers [125]. SiNWs with other cross sections, such as rectangular, have been obtained with top-down fabrication methods.

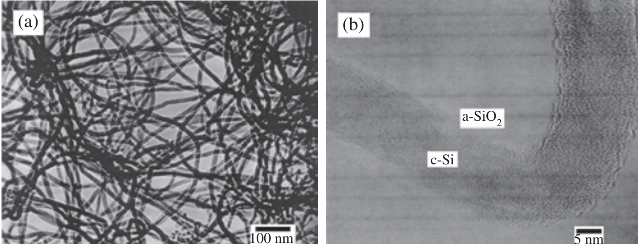


Figure 6. (a) TEM images of silicon nanowires with smooth surface. (b) Closer view TEM image showing crystalline core with amorphous oxide surface layer. Reprinted with permission from [139], Y. H. Tang et al., *J. Appl. Phys.* 85, 7981 (1999). © 1999, American Institute of Physics.

Such SiNWs have been used as model systems in several theoretical studies [63, 65, 126]. Pentagon-shaped silicon nanowires with widths around 300 nm were obtained by using Si/SiGe epitaxy technique, reactive ion etching, and subsequent selective chemical etching [127]. SiNWs with rectangular, triangular, and trapezoidal cross sections were studied using the effective mass approximation [128]. It was shown that the order of energy levels depends strongly on both the cross-sectional shape and the lattice orientation in wire direction.

For many experiments and applications it is desirable to have very thin and very long SiNWs. By using a sequence of procedures, the fabrication of such SiNWs can be achieved.

A combination of electron-beam lithography, NF_3 reactive ion etching, and dry thermal oxidation was used to fabricate 2-nm-wide silicon nanowires with an aspect ratio of more than 100 to 1 [129] (Fig. 7). The small width of the nanowires was realized by applying the self-limiting oxidation process. Fabrication of silicon nanowires as long as 10 μm has been reported [15].

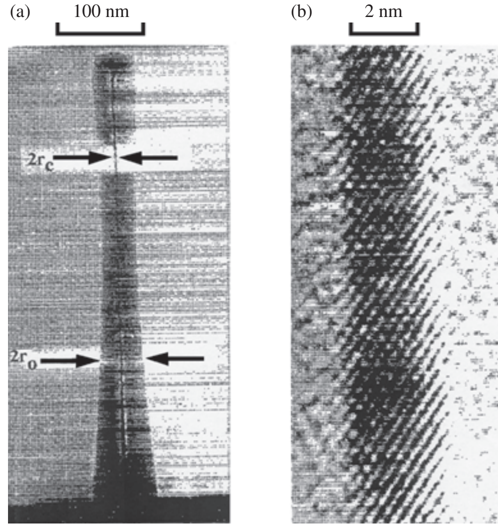


Figure 7. (a) TEM image of an oxidized silicon nanowire where the inner core has reached its limiting dimensions. (b) A high-resolution lattice image of the 2-nm-wide inner silicon core. Reprinted with permission from [129], H. I. Liu et al., *Appl. Phys. Lett.* 64, 1383 (1994). © 1994, American Institute of Physics.

An important property of SiNWs is the uniformity of their diameter and the smoothness of the wire surface. Many growth and fabrication techniques lead to SiNWs with well-defined and uniform diameters. However, under some experimental conditions, SiNWs with periodic instabilities of the diameter were found [130].

SiNWs were fabricated in different environments such as quasi-free [44], dispersed in liquid [46], parallel aligned in bundles [19], horizontally placed on substrates [18], perpendicular to a substrate [131], or inside a solid host (SiO_2 [45] or diamond nanocrystal [131]). An air-bridge-structured SiNW was made by micromachining a silicon-on-insulator (SOI) substrate [44]. Nanowires 20–100 nm in diameter and 300–600 nm in length were isolated from the substrate by an air bridge with a gap of about 300 nm. SiNWs have been dispersed in liquid using a mild etching treatment [46]. Subsequently, an electrophoresis mounting process was used to mount SiNWs onto electrodes of a prototype device. SiNWs were also fabricated perpendicular to the base of a substrate, analogous to the bristles of a brush [131]. The nanowire diameters were ~60 nm, the lengths were 500 nm, and the spaces between the wires were 130 nm.

Silicon nanoparticle chains can be formed together with silicon nanowires by using combined laser ablation and thermal evaporation of Si powder mixed with SiO_2 powder [17]. Using this method, bulk quantities of silicon nanoparticle chains have been produced [132] (Figs. 8, 9). The nanoparticle chains consist of equally spaced crystalline spherical silicon particles connected by silicon oxide bars. The transition from silicon nanowires to silicon nanosphere chains was determined by the annealing temperature, ambient pressure, initial silicon nanowire diameter, and the oxide-state of the outer layers of the silicon nanowires.

5.2. Width Reduction and Control

Various techniques for width reduction and control have been applied, for both the bottom-up and top-down techniques of synthesis. The growth techniques usually produce SiNWs with not exactly the same diameter but rather a distribution of diameters. It is desirable to have this distribution as narrow as possible. Also, one would like to have methods available to vary the size distributions.

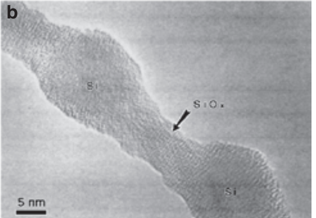


Figure 8. (a) Si nanoparticle chains converted from silicon nanowires by annealing. (b) The microstructure of the silicon nanoparticle chain. Reprinted with permission from [132], H. Y. Peng et al., *J. Appl. Phys.* 89, 727 (2001). © 2001, American Institute of Physics.

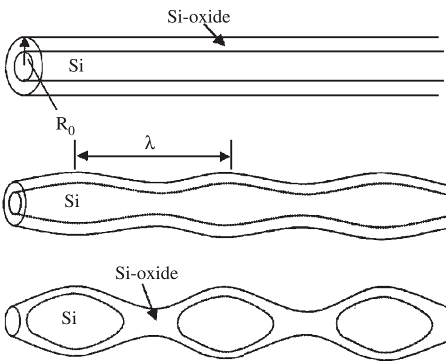


Figure 9. Schematic drawing showing the formation of a nanoparticle chain from a cylindrical nanowire. Reprinted with permission from [132], H. Y. Peng et al., *J. Appl. Phys.* 89, 727 (2001). © 2001, American Institute of Physics.

Si nanowires with different diameters have been synthesized by laser ablation in different ambient gases [113]. SiNWs with the diameter distribution peaks at ~13.2 and ~9.5 nm have been obtained respectively in He and Ar (5% H2). SiNWs produced in N_2 had the smallest peak diameter at 6 nm.

For SiNWs produced by thermal evaporation in ambient gas, the diameter can be controlled by the ambient pressure. In one of these studies, the ambient pressure was changed between 150 and 600 torr and an increase of the average wire diameter from 12 to 23 nm was observed [133]. The mean diameter of the SiNWs was found to be proportional to the 0.4th power of ambient pressure.

Silicon nanowires with different diameters and morphologies were synthesized by laser ablation of a target containing metals over a temperature range 910–1120 °C [27]. SiNWs with larger diameters were formed in lower-temperature zone while silicon nanowires and silicon nanoparticle chains of smaller diameters were formed in higher-temperature zone. The study shows that the morphology and diameter of silicon nanowires synthesized by laser ablation not only correlate with the growth temperature of SiNWs, but also with the nature of the catalyst.

Diameter-controlled vapor-liquid-solid (VLS) synthesis of single-crystal silicon nanowires was reported [115]. Gold nanoclusters were used as catalysts. Starting with 5-, 10-, 20-, and 30-nm-sized gold clusters, silicon nanowires with mean diameters of 6, 12, 20, and 31 nm, respectively, have been grown. The nanowires have single-crystal cores sheathed with 1–3 nm of amorphous oxide and the cores remain highly crystalline for diameters as small as 2 nm.

The width of silicon nanowires grown by thermal evaporation of SiO at 1350 °C in a closed system depends on the applied ambient pressure [95]. With increasing ambient pressure, the diameter of the SiNWs increases.

In experiments where the nanowires are grown on a substrate, the type and temperature of the substrate may be important. It has been found [134] that the substrate temperature played a dominant role in controlling the diameter and morphology of SiNWs produced by evaporation.

Thickness of silicon nanowires can well be controlled by top-down techniques such as lithography with etching, using mask transfer, etc. For example, SiNWs of controlled width were obtained when fabricated by plasma deposition of a SiO_2 film through a mask with subsequent reactive ion etching [106].

5.3. Atomic Structures

Most preparation techniques for SiNWs lead to the formation of crystalline structures (Fig. 10). The degree of crystallinity and the lattice orientation have been determined for such SiNWs [114]. Defects inside such silicon nanowires could be significantly reduced by annealing the nanowires at 1100 °C for 6 h [135]. Stacking faults and twins were annihilated upon annealing. Raman spectra confirmed that the bulk specimen became almost defect-free by the heat treatment. Polycrystalline silicon nanowires were obtained by electron-beam lithography and thermal oxidation in standard polycrystalline silicon material [136].

Amorphous silicon nanowires have been grown using AuPd nanoparticles as catalysts [137]. The growth process appears to be the vapor-liquid-solid (VLS) mechanism. The controlled growth of oriented amorphous silicon nanowires on Si(111) was also reported [98].

Most applied synthetic techniques produce slightly curved or straight SiNWs. However, under some experimental conditions, uncommon shapes and structures have been obtained.

Silicon nanowires synthesized by laser ablation can exhibit straight, curved, kink, braided, and coiled shapes [91]. The formation of various morphologies of silicon nanowires is closely related to the insertion of twins or special twisting along the growth direction during the growth process. Flower-like and octopus-like forms of silicon nanowires were reported by heating a SiO_2 plate at 1600 °C under argon [138]. Spring-shaped, fishbone-shaped, and necklace-shaped silicon nanowires were observed after synthesis by laser ablation of silicon powder targets at 1200 °C [139].

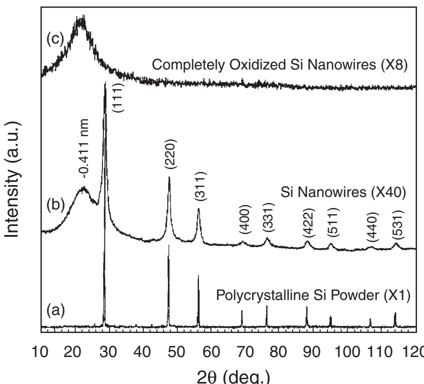


Figure 10. X-ray diffraction spectra for silicon nanowires with thin oxide surface layer, compared with polycrystalline silicon powder and completely oxidized silicon nanowires. Reprinted with permission from [113], Y. F. Zhang et al., *Appl. Phys. Lett.* 75, 1842 (1999). © 1999, American Institute of Physics.

The stability, morphology, and electronic structures of "hypothetical" silicon nanotubes were calculated with first-principles techniques based on density functional theory [37]. The band structure calculations show that, similar to carbon nanotubes, depending on their chiralities, silicon nanotubes may have metallic (armchair) or semiconductor (zigzag and mixed) properties (Fig. 11).

The electronic properties of silicon nanowires with "super-atom stacking structure" were investigated using a generalized tight-binding molecular dynamics scheme [38]. The nanowires are stable when the core consists of fourfold coordinated atoms and is surrounded by a threefold coordinated outer surface (Fig. 12).

Silicon nanowires were grown on highly oriented pyrolytic graphite in ultrahigh vacuum at 10^{-10} torr from pure silicon vapor [19]. No catalyst particles are used where the nanowires could start to grow. It was postulated that a fullerene-based structure with a Si_{24}-unit is the seed for nanowire growth under UHV conditions and that the wire formation occurs by adding further Si_{24} units in one direction. Such fullerene-structured silicon nanowires (Fig. 13) are calculated to be stable.

Highly pure SiNWs are required for many experimental studies. Among the various fabrication techniques, those without metal participation usually produce SiNWs with less impurities. High-purity silicon nanowires in bulk quantity, for example, have been grown from SiO in a high-temperature tube furnace [94]. Also, high-purity SiNWs of uniform diameters around 15 nm were obtained by sublimating a hot-pressed silicon powder target at 1200 °C in a flowing carrier gas [14]. A review of the synthesis of highly pure, ultralong, and uniform-sized semiconductor nanowires in bulk quantity by the methods of laser ablation and thermal evaporation of semiconductor powders mixed with metal or oxide catalysts has been given [140].

5.4. Surface Reactivity and Passivation

Silicon nanowires are usually made of a crystalline core of pure silicon with a surface layer of amorphous silicon oxide. Other types of coatings have also been studied. For example, a reaction of SiNWs with methane and hydrogen has been performed to produce a thin coating layer of cubic silicon carbide (beta-SiC) [40].

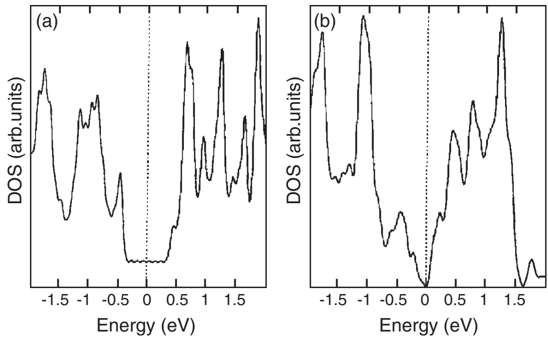


Figure 11. The calculated density of states for (a) metallic armchair-structured silicon nanotube, and (b) semiconducting zigzag tube. Reprinted with permission from [37], S. B. Fagan et al., *Phys. Rev. B* 61, 9994 (2000). © 2000, American Physical Society.

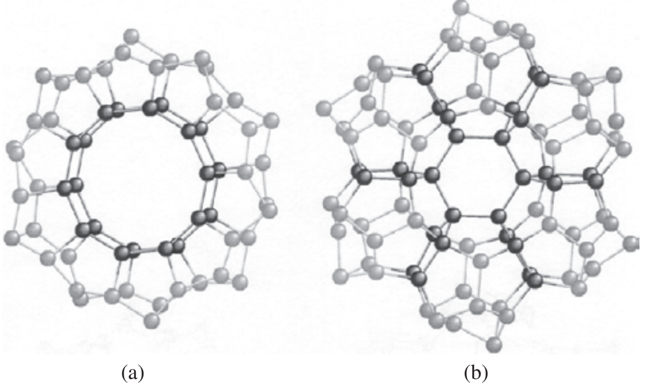


Figure 12. Two superatom cluster units proposed for silicon nanowires. Reprinted with permission from [38], M. Menon and E. Richter, *Phys. Rev. Lett.* 83, 792 (1999). © 1999, American Physical Society.

In theoretical studies the nanowires are often coated with hydrogen in order to saturate the silicon dangling bonds. Unpassivated and H-passivated SiNWs were compared [141]. The role of H-passivation has been explored by considering unpassivated, partially passivated, and fully passivated NWs. When one H atom is removed from a ring of 14 surface H atoms, one half-filled surface state is in the bandgap leading to metallic character. The removal of two hydrogen atoms leads to two dangling bond states (one occupied and one empty), separated by a small gap of ~0.2 eV. If all H atoms are removed, the unpassivated wire is metallic since a continuum of surface states fills up the bandgap. This result shows that good passivation is required for silicon nanowires to be efficient light emitters. The presence of even low concentration of H vacancies leads to localized states in the bandgap, which drastically change the optical wire properties. A similar result with respect to the bandgap has been obtained with scanning tunneling spectroscopy studies of unpassivated silicon clusters [142]. The pristine clusters, produced in ultrahigh vacuum, have none of their surface states passivated and can have bandgaps close to zero because the dangling bond states fill up the gap.

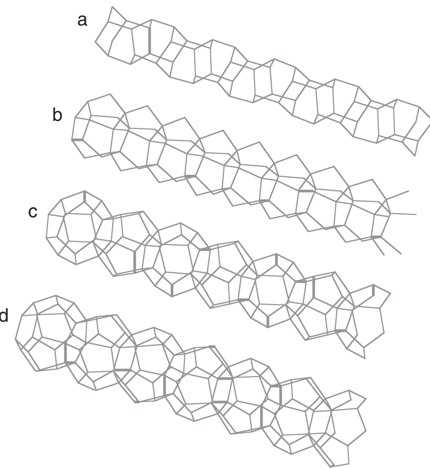


Figure 13. Four models for possible nanowire core structures. (a) Si_{12} cage polymer. (b) Si_{15} cage polymer. (c) Si_{20} cage polymer. (d) Si_{24} cage polymer. Reprinted with permission from [19], B. Marsen and K. Sattler, *Phys. Rev. B* 60, 11593 (1999). © 1999, American Physical Society.

The chemical reactivity of H-passivated SiNWs with silver and copper ions has been studied [143]. Freestanding SiNWs with diameters of ~20 nm and millimeters in length were synthesized by laser ablation. Such wires had a polycrystalline silicon core and were coated by a thin layer of silicon oxide. The surface oxide layer was then removed and replaced by hydrogen after immersion of the SiNWs in HF solution. Then the etched SiNWs were exposed to silver (I) nitrate or copper (II) sulfate solutions. The formation of metal particles and other metal nanostructures at the surface of the SiNWs was observed as a result of the reductive deposition of the metal ions.

5.5. Doping

The transport properties of silicon are very sensitive to the concentration of foreign atoms. Controlled doping of SiNWs is a prerequisite for the realization of many applications as SiNW-based devices. Controlled Li-doping of SiNWs by electrochemical insertion method with different doping levels was reported [42]. Doping levels have to be low enough since the crystalline structure of the silicon nanowires is destroyed with increasing doping concentration. STS measurements in SiNWs showed an enhancement of electrical conductivity by boron doping. These nanowires were produced by oxide-assisted growth [144]. SiNWs with different bandgaps can be obtained by controlling the doping level [42]. Boron-doped silicon nanoparticle chains with 15 nm outer diameter were produced in bulk quantity by laser ablation of SiO powder mixed with B_2O_3 powder [145]. The nanoparticles had perfect lattices with an 11-nm crystalline core and a 2-nm amorphous surface oxide layer. Field-emission studies showed that the turn-on field of silicon nanoparticle chains was 6 V/μm, which was much lower than that of undoped SiNWs (9 V/μm).

5.6. Alignment

Parallel alignment or specific orientation with a substrate structure is desirable for the realization of many applications. Parallel SiNWs can be grown by deposition of silicon vapor on HOPG in UHV [19].

Alignment of SiNWs can be achieved by post-growth treatment such as exposure to an ion beam. Ti-containing islands, for example, have been grown on silicon substrates by chemical vapor deposition using the decomposition of TiO_4 [83]. When these islands are exposed to SiH_4 or SiH_2Cl_2, the metal catalyzes the decomposition of the silicon precursor and long SiNWs can grow. Exposing the wires to an ion beam after deposition promotes alignment of nanowires [83].

Interacting nanowires were considered in theoretical investigations in order to model the structure of porous silicon. One of the studies [146] deals with the effect of interwire interaction on the electronic structure and optical properties through two first-principles techniques: linear

muffin tin orbitals method in the atomic sphere approximation (LMTO-ASA) and norm-conserving pseudopotential. Silicon wires with rectangular cross section (5×4 and 7×4) with their axes along the [001] direction and the atomic structure of silicon bulk were used as model systems. Interwire interaction leads to the presence of localized interface states which strongly lower the bandgap energy, for example from 3.27 to 1.55 eV for two NWs considered [146]. As expected, the energy separation between the silicon core crystalline states of the two wires (i.e., the mobility gap) remains unchanged. However, localized interwire bonded states that emerge from the valence band appear and reduce the bandgap. The stability of the wire structures with respect to wire-wire interaction was also investigated and some reorganization of the structure was found. The Si-Si bond lengths in the wire remain practically unchanged, while there is a small lowering of the Si-H bond distance. When the wires are forced to get closer the bandgap strongly is reduced.

5.7. Composite Structures

Composites of carbon nanotubes and silicon nanowires in longitudinal [77] and transverse [48] have been synthesized. In the "transverse" process, first the silicon nanowires were produced by excimer pulsed laser ablation of a target made of a mixture of Si and SiO_2 powders under 9×10^4 Pa Ar atmosphere at 1200 °C. Then, by using hot filament chemical vapor deposition (HFCVD), carbon was evaporated on the SiNWs. Multiwalled carbon nanotubes were found to form on the SiNWs with good adherence.

Deposition of carbon on silicon nanowires can lead to another type of composite. Diamond crystallites were grown from the vapor phase onto arrays of silicon nanowires [131] (Fig. 14).

Two-dimensional aluminum-silicon alloy nanowire networks were fabricated on glass and silicon substrates by de-alloying an aluminum-silicon thin film through selective chemical etching [47]. The network consists of 3- to 6-nm diameter wires with lengths of 50–200 nm, and a wire density of 10^{10} wires/cm^2. Current-voltage measurements show that the nanowire network is metallic.

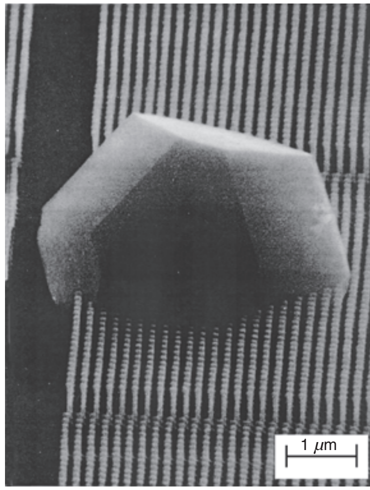


Figure 14. Scanning electron micrograph of a diamond crystallite grown on a silicon nanowire substrate, taken at 45°. The silicon nanowire obstructions have not disturbed the cubo-octahedral faceting of the diamond crystallites. Reprinted with permission from [131], P. A. Dennig et al., *Appl. Phys. Lett.* 67, 909 (1995). © 1995, American Institute of Physics.

Another type of composite was reported, which is composed of copper silicide nanocrystals in silicon nanowires [43]. Copper atoms have a high solubility in the silicon nanowires and precipitate in the nanowires.

6. ELECTRONIC AND OPTICAL PROPERTIES

6.1. Bandgap

Photoluminescence of porous silicon layers has been related to direct interband transitions in nanoparticles and nanowires of silicon [119]. Porous silicon samples with varying confinement size have been fabricated and a blueshift of the photoluminescence peak was observed for smaller confined regions, which is related to an opening of the energy bandgap with reduction of size. Fabricated silicon nanowires have shown unusual photoluminescence and Raman spectra [79, 118, 130] as well, implying a strong quantum size effect, which relaxes the k-selection rule. This leads to direct optical transitions in the SiNWs, which are not allowed for Si bulk crystals. Evidence of quantum confinement was given from the size dependence of the photoluminescence of silicon nanowires [122]. Following such observations, energy levels and charge density distributions have been calculated for silicon nanowires [141].

The bandgap up-shift of the wires with respect to bulk silicon was theoretically determined using effective mass theory (EMT) over a wide range of diameters [65]. It was found that EMT is valid for wires wider than $\sim$2.3 nm. The bandgaps are direct and at the zone center. From another study it was concluded that EMT describes the electronic structure of silicon nanowires well down to diameters of about 1 nm [147].

Nanowires have been considered in EMT studies where the wires either have a free surface or are incorporated in a solid host. The analysis of the electron states within the EMT is straightforward when the barrier potential is very large. It permits the infinite barrier approximation as in structures with free surfaces. However, the barrier potential is finite when the wire is surrounded by another semiconductor. An analytic solution of the equation for the envelope function is not possible for such wires. The wave function has to be expanded in terms of two-dimensional orthogonal functions and the differential equation is reduced to a matrix eigenvalue equation. Solutions to these equations were given [125] and the energy levels for nanowires with finite potential barriers were calculated.

First-principles pseudopotential calculations of band-edge electronic states of SiNWs have been performed [148]. The valence- and conduction-band-edge energies and effective masses of hydrogen-terminated silicon nanowires were calculated and the results were compared with EMT. It was found that the first-principles result for the ordering of states at the valence band maximum is different from the prediction of EMT. The EMT gives accurate values for the

bandgap up-shift for wire widths greater than about 3.3 nm, but it overestimates the bandgap up-shift by nearly 100% for wires of 2.0 nm width.

Using a full-potential linear-muffin-tin-orbital molecular-dynamics method, the geometric and electronic structure of very thin and short silicon nanowires were calculated [87]. Silicon nanowires consisting of tri-capped trigonal prism Si_9 sub-units and uncapped trigonal prisms were studied. These structures are found to be the thinnest stable silicon nanowires and have very small gaps of only a few tens of an electron volt.

The pressure dependence of bandgaps of 6 × 6 atoms (1.15 nm) and 10 × 10 atoms (1.92 nm) wide SiNWs has been calculated using a plane-wave basis and carefully fitted empirical pseudopotentials [149]. The pressure coefficient becomes more negative as the wire diameter increases, eventually approaching a value for the bulk of a = −1.90 meV/kbar. The pressure dependence of the photoluminescence of porous silicon has been measured by several groups [150–153]. It was found that as the pressure increases above ~25 kbar, the photoluminescence peak shifts to lower energies (redshift) with an average pressure coefficient of a~−3 meV/kbar. This value is more negative than the value for the indirect gap of crystalline Si (−1.4 meV/kbar) or that of amorphous silicon (−2.0 meV/kbar). At lower applied pressures, a blueshift is found for the shift of the photoluminescence peak [151, 153]. The crossover from blue- to redshift has been interpreted as a pressure-induced direct-to-indirect transition in the SiNWs of the sample [153]. Calculations of the coefficients for different applied pressures to silicon nanowires suggest that the redshift is an intrinsic property of the silicon wires but that the blueshift is rather a pressure-induced chemical reaction that takes place when alcohols are used as medium in the pressure cell [149].

6.2. Optical Activity

Optical activity and bandgaps have been calculated for SiNWs using an empirical tight-binding model [66]. Due to quantum confinement, the bandgap increases significantly for small wire diameters. The same technique applied to silicon nanowires [126] showed bandgaps equal to 3.29, 2.5, and 2.3 eV for square-shaped wires with diameters 0.77, 1.15, and 1.54 nm (4 × 4 to 8 × 8 atoms), respectively. These nanowires are optically active.

Ab initio electronic structure studies were reported for square-shaped silicon nanowires with 3 × 3, 4 × 4, and 5 × 5 structures and excitonic effects on the optical properties were calculated [63]. The SiNWs have a direct bandgap in the visible energy range. The allowed bandgaps are due to anisotropic effective masses and band mixing. The bandgaps are enlarged by 0.9, 1.3, and 1.9 eV with decreasing silicon wire widths from 1.53, 1.15, to 0.77 nm, respectively. A large optical anisotropy is found.

Blue-shifted electron states were calculated for 1-nm-wide (5 × 5 with $Si_{25}H_{20}$ per unit cell and 6 × 5 with $Si_{30}H_{22}$ per unit cell) H-passivated silicon nanowires using a first-principles approach, based on the self-consistent local-density approximation (LDA) [57] (Figs. 15, 16). LDA calculations [57, 65] as well as the effective mass theory [154] give strong size- and symmetry-dependent radiative times for optical interband transitions.

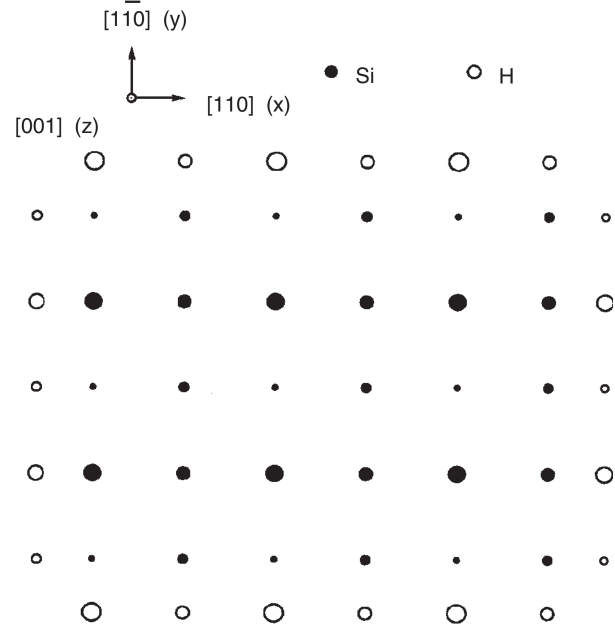


Figure 15. A unit cell of a 6 × 5 silicon nanowire. The size of the circles increases with height. Reprinted with permission from [57], M. S. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993). © 1993, American Physical Society.

6.3. Photoluminescence

Photoluminescence (PL) spectra have been taken for silicon nanowires fabricated with various methods. SiNWs supported on crystalline silicon (100) emit photoluminescence in the range of 500–600 nm [55]. The PL efficiency and

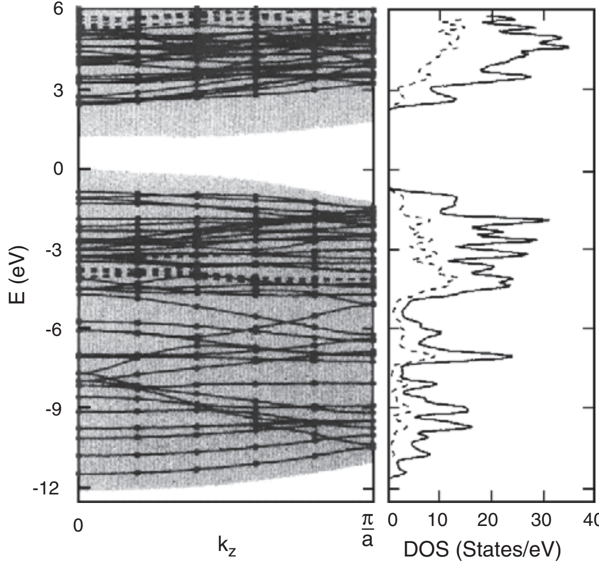


Figure 16. The band structure of a 5 × 5 silicon nanowire together with the projected bulk energy bands of silicon plotted against the wave vector along the wire and the associated density of states. The dashed lines indicate bands with more than 50% Si-H character. Reprinted with permission from [57], M. S. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993). © 1993, American Physical Society.

energy position depend on the wavelength and polarization of the initial laser beam.

SiNWs produced in Ar (5% H_2) and N_2 atmospheres exhibit PL and spectral blueshift with diameter reduction, which was attributed to quantum confinement in the crystalline nanowires [113].

Silicon nanowires 50 nm wide formed by decorating atomic steps of an otherwise atomically smooth sapphire surface were found to exhibit visible PL. A broad PL emission peaked at 750 nm was found with an intensity at 9 K being over 20 times larger than at 300 K [155]. This is similar to what was observed for porous silicon and other nanostructured silicon [156, 157].

Strong multiple-peak emission in the visible spectral range was found for as-grown and partially oxidized silicon nanowires [122] (Fig. 17). The emission ranges from dark red to blue regions. The red light emission was ascribed to a quantum confinement effect originating from the crystalline core of the silicon nanowires.

Multiple-peak photoluminescence was also observed for silicon nanowires grown with $FeSi_2$ catalyst particles [130]. The nanowires were treated at high temperature with oxygen to obtain different diameters for the silicon core.

There have been problems in the past with degradation and low photoluminescence efficiency from silicon nanostructures. The degradation is usually accompanied with a blueshift of the PL peak. It is probably due to the chemical instability of the silicon surface [40]. When exposed to the atmosphere, the thickness of the silicon oxide layer at the surface increases, which leads to increased scattering and absorption of the incident light and to degradation of the PL intensity. The oxidation process also reduces the size of the core of nanowires or nanocrystals, which results in a blueshift. Various passivation methods have been applied to silicon nanostructures such as nitridation [158], and coating with a diamond-like carbon film [159] or a SiC film [40]. Silicon nanowires with SiC coating show stable photoluminescence with high efficiency.

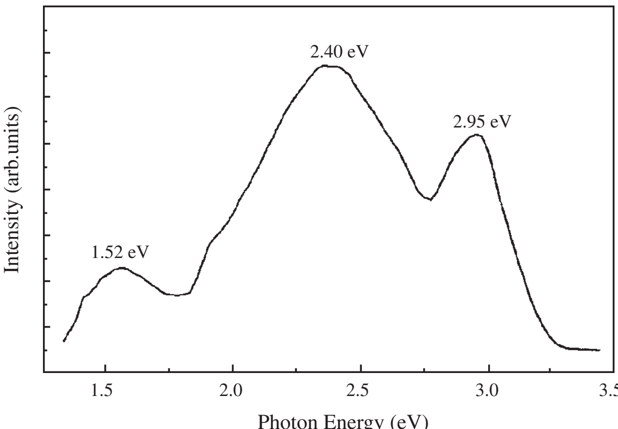


Figure 17. Photoluminescence spectra of silicon nanowires. Multiple emissions corresponding to red, green, and blue light are shown peaking at 1.54 eV (804 nm), 2.40 eV (515 nm), and 2.95 eV (429 nm), respectively. Reprinted with permission from [122], D. P. Yu et al., *Phys. Rev. B* 59, R2498 (1999). © 1999, American Physical Society.

The properties of small-width square hydrogenated silicon nanowires have been studied theoretically by a number of research groups with emphasis on energy level distribution, dispersion configuration, and optical transition matrix elements [63, 65, 67, 160]. A typical result is that SiNWs with diameters 1.0–2.5 nm are optically active in the visible range [146]. The energy gap becomes wider with decreasing wire diameter. Fully hydrogen-passivated 1-nm-wide SiNWs, for example, have a bandgap of about 2.7 eV.

The question has been addressed if the diamond structure is stable for extremely narrow silicon nanowires. It was found that ~1-nm-wide, H-terminated (5×5 and 5×6) silicon nanowires, fully relaxed within the local-density approximation, indeed can have the diamond structure of bulk silicon [57].

Bandgaps, effective masses, and optical matrix elements were studied with first-principles pseudopotential calculations for H-terminated silicon nanowires with thicknesses up to 1.6 nm [64]. For each wire structure the bandgap was found to be direct and at the zone center. Also, an upshift of the bandgap with decreasing wire thickness is a result of the calculations. For the narrowest possible hydrogenated silicon nanowire, a polysilane molecule, a direct bandgap at the zone center of 4.69 eV is calculated. A radiative lifetime of 560 μs for zero-phonon transitions is determined by calculating the zone-center interband optical matrix elements for a 1.6-nm-wide wire. This is close to what is measured for porous silicon. Porous silicon and silicon nanowires have the character of a direct bandgap semiconductor. The radiative transitions occur without phonon participation because the momentum rule is broken by quantum confinement.

The empirical pseudopotential method was used to study the electronic structure and the optical properties of [001] silicon quantum wires [141]. The wires have square cross sections ranging from 4×4 to 14×14 atoms (7.7×7.7 to 26.9×26.9 Å, respectively).

The influence of special orientation on the optical response of hydrogenated crystalline silicon nanowires with small width (0.55 nm) was calculated [69]. (111)-oriented silicon nanowires were studied and the results were compared to (001)-oriented wires [67]. It was found that both the (111)- and (100)-oriented nanowires exhibit a direct gap at $k = 0$. Therefore, if nanowires are present in porous silicon, they are optically active independent of their crystallographic orientation.

The spontaneous emission rate for direct and phonon-assisted transitions was calculated [161]. It was found that light emission from the wires has strong directional character if the wires are directed along the main crystallographic directions or along [110].

The band structure and luminescence properties of pristine and H-passivated SiNWs with three different diameters were compared [39]. The hydrogen-passivated SiNWs have wide bandgaps. For the nonhydrogenated nanowires the bandgap is very small because of the nonbonding nature of the dangling bonds, with the system remaining an insulator but with a minimum direct gap. For instance, calculations for Si_3H_4- and Si_3-based SiNWs give bandgaps of 3.2 eV and 0.3 eV, respectively. Similar results were obtained [66] with an empirical tight-binding approach. Mid-gap states act

as recombination centers for electron-hole pairs and therefore quench the photoluminescence. It explains the observed reduction of the PL intensity after desorption of hydrogen [162, 163].

The first step in a photoluminescence experiment is the optical absorption of ultraviolet photons. Nonorthogonal tight-binding calculations were used to determine the positional dependence of optical absorption of square-shaped 9×9-atom-sized SiNWs [164]. It is found that the optical absorption below 3.4 eV tends to occur at the inner silicon atoms of the nanowires.

7. TRANSPORT PROPERTIES

7.1. Electrical Conductance

The transport properties of silicon nanowires depend very much on the purity of the samples. Also, the wire diameter is important since it determines the sub-band electronic structures and the boundary scattering. It has been shown experimentally that electrical conductivity of SiNWs 15–35 nm in diameter can be varied by four orders of magnitude by doping and thermal treatment [81].

In electrical conductance measurements, air-bridge-structured silicon wires of 100 nm width showed anomalous behavior such as negative resistance and hysteresis at room temperature [44, 60] (Fig. 18). Charge accumulation in surface states is considered to be responsible for the nonlinear electric current behavior. Electrons are trapped in surface states and discharge very slowly, which causes the hysteresis behavior.

Narrow nanowires are expected to show quantized conductance at low temperatures. If the diameter of a nanowire is less than the Broglie wavelength of moving electrons, the band structure becomes divided into sub-bands and the moving electrons experience one-dimensional confinement. The Fermi level increases with the applied gate voltage and the channel conductance shows step-like increase every time when the Fermi level coincides with sub-band edges.

Single-electron effects have been observed at 4.2 K in heavily doped polycrystalline SiNWs with 20 nm $\times$ 30 nm active cross section [136]. The wires were fabricated by electron-beam lithography and thermal oxidation. A double period was seen and explained by grain boundary scattering and Coulomb blockade effects.

Silicon nanowires fabricated by orientation-dependent etching [12] show conductance steps with little fluctuations on plateaus at a temperature of 45 K. Conductance steps up to temperatures as high as 100 K have been observed for very narrow silicon nanowires in the sub-10-nm regime [53].

Step-like conductance was studied for silicon nanowires [61] with different wire widths and applied temperatures (between 25 and 160 K). The step-like conductance remains up to a temperature of about 100 K. The width of the wires was controlled from 20 to 70 nm.

Silicon nanowires have been fabricated using electron cyclotron resonance plasma deposition of a SiO_2 film through the openings of a patterned resist film. The nanowires show quantized conductance up to a temperature of 200 K. (Fig. 19).

A theoretical study [165] used computer simulation of electronic states in SiNWs to determine the conditions for quantum transport at room temperature. The goal is to get insight on the limits of device scalability for silicon. Depending on the configuration of nanowire, gate, and drain in

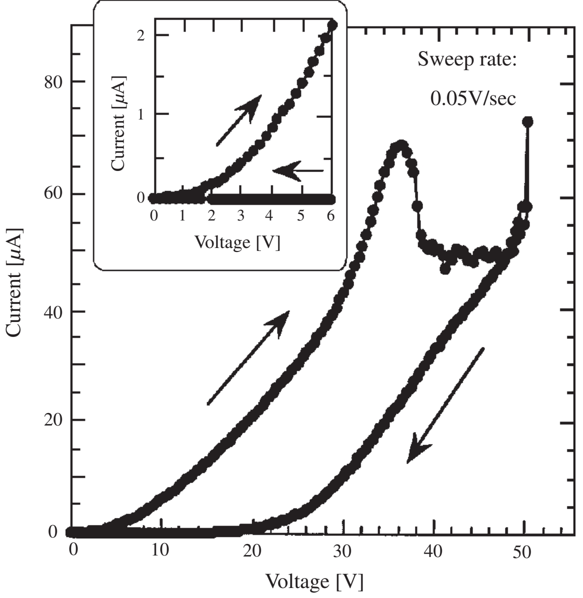


Figure 18. Current-voltage characteristics of a silicon nanowire, showing hysteresis and a region of negative resistance. Inset shows magnification of the I–V curve near zero voltage. Reprinted with permission from [60], H. Fujii et al., *Appl. Phys. Lett.* 78, 2560 (2001). © 2001, American Institute of Physics.

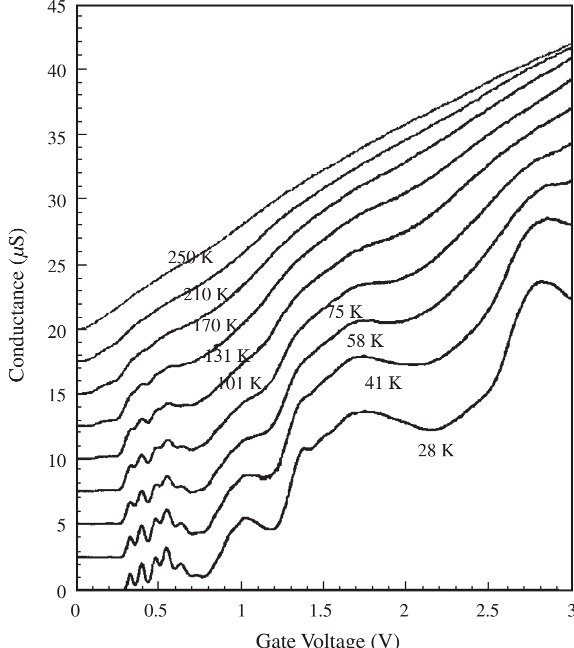


Figure 19. Conductance characteristics of a silicon nanowire at various temperatures. Reprinted with permission from [106], H. Namatsu et al., *J. Vac. Sci. Technol. B* 13, 2166 (1995). © 1995, American Vacuum Society.

a device, the electrical transport properties were studied over a wide parameter space. This led to the prediction of conditions for quantum transport to be observable at room temperature.

For an amorphous silicon nanowire, single-electron effects with nonlinear current-voltage behavior at low temperature have been found [166]. A periodic step-like pattern was observed in the current-voltage curves with two periods. This was explained by two effects: hopping conductivity between trapping sites in amorphous regions or at grain boundaries and Coulomb blockade. The silicon nanowire for the I–V studies was 150 nm long, 9 nm thick, and 50 nm wide and at a temperature of 2 K. Short-period oscillations occur with a period of ~1.4 mV and are explained by single-electron scattering processes. The larger observed period of 10 mV is assigned to Coulomb blockade.

Nanowire-electrode contacts have been studied in experiments of charge carrier transport [81]. Al and Ti/Au contacts were investigated. SiNWs 15–35 nm in diameter were produced using SiH_4 chemical vapor deposition via Au or Zn particle-nucleated vapor-liquid-solid growth at 440 °C. The wires as produced were essentially intrinsic. Thermal treatment of the fabricated devices resulted in better electrical contacts, but also in the diffusion of dopant atoms into the nanowire and an increased nanowire conductance by 10^4.

Passivated silicon nanowires connected to aluminum electrodes were studied by large-scale local-density-functional simulations [167]. Short (~0.6 nm) wires have zero bandgaps due to metal-induced gap states (MIGS), which results in a finite electrical conductance. For longer wires (>2.5 nm) Schottky barriers develop and the conductance spectra exhibit size-dependent oscillations due to interference resonances from scattering of the ballistic electrons from the contacts.

7.2. Thermal Conductance

Heat conduction in one-dimensional structures such as nanowires has been studied only recently [168, 169]. There have been several approaches proposed for calculating the lattice thermal conductivity in nanowires [170–172]. Quantized thermal conductance has been predicted theoretically for nanowires at low temperature [169].

Various effects can influence the phonon heat conduction in a semiconductor nanowire with dimensions comparable to the phonon mean free path. Phonon relaxation mechanisms to be considered are phonon Umklapp scattering, mass-difference scattering, boundary scattering, and phonon-electron scattering. In a recent theoretical study [58], a model was used based on the solution of Boltzmann's equation, which takes into account (1) modification of the acoustic phonon dispersion due to confinement, and (2) change in the nonequilibrium phonon distribution due to partially diffuse boundary scattering. Phonon confinement and boundary scattering lead to significant decrease of the lattice thermal conductivity. In pure or lightly doped semiconductors, phonon-electron scattering has little influence on the lattice thermal conductivity. It becomes, however, important for high doping concentrations.

Similar results were obtained using molecular-dynamics (MD) simulations. The thermal conductivity of SiNWs with square cross sections are found to be about two orders of magnitude smaller than those of bulk silicon crystals in a wide range of temperature (200–500 K) [170]. The thermal conductivity is reduced due to boundary and internal phonon scattering. For nanowire cross sections from 2.58 to 28.62 nm^2, the thermal conductivity becomes length independent when the wire is longer than 8.56 nm.

A strong reduction of the thermal conductivity with decreasing size is a general property of nanowires. It has also been predicted (within the Boltzmann transport equation approach) for gold, sodium, and CdTe nanowires [173].

8. APPLICATIONS

8.1. Optoelectronic Devices

Silicon nanowires coated with SiC show stable photoluminescence at room temperature. Silicon carbide is an insulator with a wide bandgap, and high breakthrough electric field. It is durable at high temperature with high thermal conductivity and chemical stability. It has high tensile strength and high Young's modulus, which makes it an excellent coating material. Silicon nanowires coated with SiC show strong and stable photoluminescence [40], which makes them ideal components for integration in optoelectronic devices.

8.2. Quantum-Wire Transistor

The size of silicon devices has continuously been shrinking in order to obtain high integration and faster performance of electronic circuits. This scaling trend has put much attention on nanometer-scale devices with silicon nanowires as structural units.

Recent advances in formation methods allowed the fabrication of silicon quantum-wire transistors. A number of sophisticated successive fabrication steps are used for nanowire, source, and drain configurations.

One of these techniques uses silicon-on-insulator (SOI) technology, electron-beam lithography, anisotropic dry etching, and thermal oxidation [174]. The quantum wires have a width of 65 nm and are fully embedded in silicon dioxide. A Coulomb staircase, that is, step-like conductance versus gate voltage, was observed at temperatures below 4.2 K. This is typical for one-dimensional quantum conductance.

A silicon single-electron transistor (SET), showing an inverted voltage gain as high as 3.7, has been fabricated [175]. The device has staircase conductance oscillations at 4.2 K. The SET is formed in highly doped and oxidized SiNWs of less than 40 nm $\times$ 50 nm cross section and 1.5 μm length.

A combination of photolithography and side-wall patterning was used to fabricate a SET based on a 30-nm-wide SiNW, which can be operated at 77 K [176]. The device showed clear single-electron tunneling phenomena by an electrostatically well-defined single island and two tunnel junctions. The procedures are compatible with conventional metal-oxide-semiconductor process technology.

Detection of single electron and single holes was demonstrated at room temperature in silicon nanowire transistors [177]. Photogenerated carriers are stored in a quantum dot electrically formed in a silicon wire by a front gate. The

stored charges affect the current of the other type of carriers that flow along the silicon nanowire [177]. The silicon nanowire was cylindrical with a diameter of 20 nm. The device was a silicon-wire metal-oxide-semiconductor field-effect transistor (MOSFET) fabricated on a silicon-on-insulator wafer by electron-beam lithography.

An experimental setup was proposed that potentially allows a single-electron silicon quantum dot transistor to operate at room temperature [178]. The emitter and collector of the device consist of silicon nanowires and the base contains a single silicon dot buried in silicon dioxide. Split gates are added to better confine charge carriers perpendicular to the transport directions in the emitter and collector regions.

8.3. Single-Electron Memory

Single-electron memory cells with control over individual electrons would consume extremely low power and could be used in future integrated circuits [25, 75, 179]. A single-electron memory cell can be realized by using the Coulomb blockade effect. Important components of such a device are a silicon nanowire as a channel, a silicon nanodot as a storage node, and a silicon nanogate as a control gate. In addition, a single-electron read-out device needs to be attached. To realize single-electron or few-electron memory devices, narrow SiNWs need to be generated.

A compact single-electron memory cell has been fabricated in silicon [75]. The memory cell consists of a multiple tunnel junction (MTJ) and a memory node, which occupies an area of 0.5 μm^2. A second, connected device with another MTJ forms an electrometer for detecting the state of the memory node. Each MTJ has a side gate to trim its operating point. The amount of charge on the memory node is exactly defined by a control voltage. Memory operation is demonstrated by two distinct levels in the memory node voltage (Fig. 20). The device was tested at 4.2 K and showed memory operation with a >100-mV gap between "0" and "1" levels.

Applying the technology of an inorganic electron-beam resist process, a 15-nm-wide silicon nanowire was fabricated, together with a 10-nm-diam silicon nanodot, and used for a single-electron memory cell [25]. The observed memory behavior is caused by an electron charging effect.

8.4. Metal-Semiconductor Junction

The catalytic growth of metal-semiconductor junctions between carbon nanotubes and silicon nanowires was reported and shown that the junctions exhibit reproducible rectifying behavior [77]. To fabricate carbon nanotube–silicon nanowire (NT/SiNW) junctions, SiNWs were grown from the ends of the NT tips by using silane. Then, the electrical properties of individual NT/SiNW junctions were measured. While little current flow was measured for negative voltage, the current increased sharply above +0.8 V. This rectifying behavior is characteristic for a metal-semiconductor Schottky diode.

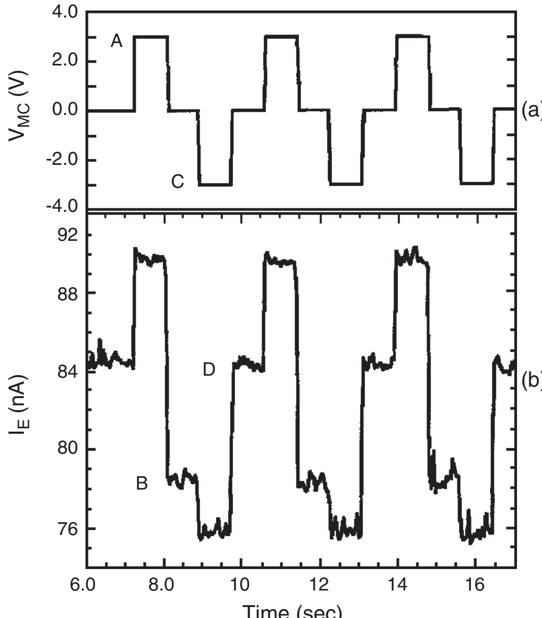


Figure 20. Read-write characteristic of single-electron memory: (a) drive voltage pulse wave form, (b) electrometer current output pulses monitoring the memory node. Reprinted with permission from [75], N. J. Stone and H. Ahmed, *Appl. Phys. Lett.* 73, 2134 (1998). © 1998, American Institute of Physics.

8.5. Proposed Applications

Nanoscale resonating devices could be of interest for sensing applications. The fabrication of such resonators, with low-mass and high-frequency response, could lead to a new generation of sensors. In order to test this idea, a harp-like array of 1- to 8-μm-long, 45-nm-wide silicon nanowires has been fabricated and the mechanical resonance has been measured [180] (Fig. 21). Resonance frequencies up to 380 MHz were detected. The Q-factor was studied as a function of wire width in order to understand the loss mechanism in nanosize structures. The Q-factor is found to decrease with

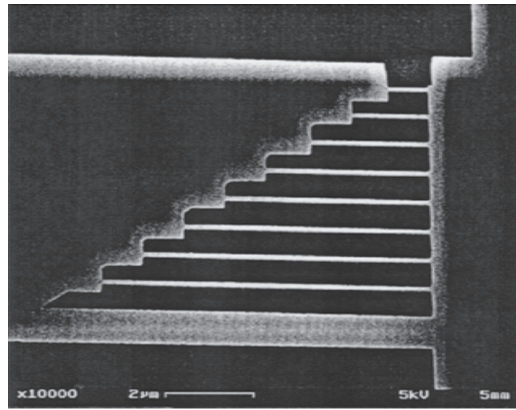


Figure 21. Harp structure made of silicon nanowires. The length of the wires varies between 1 and 8 μm. The center-to-center spacing of the wires is 630 nm. Reprinted with permission from [180], D. W. Carr et al., *Appl. Phys. Lett.* 75, 920 (1999). © 1999, American Institute of Physics.

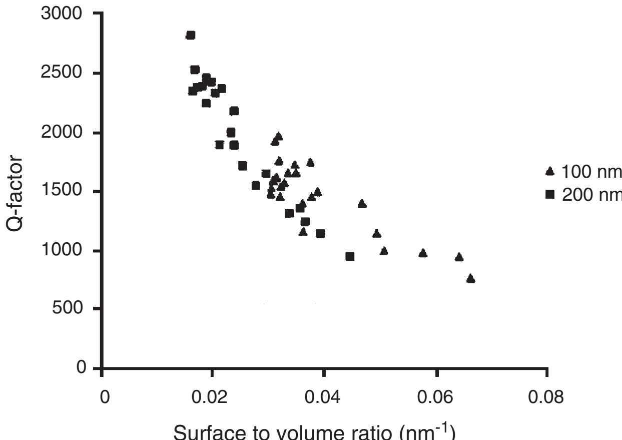


Figure 22. Plot of the Q-factor as a function of the surface-to-volume ratio for the silicon nanowire harp. Reprinted with permission from [180], D. W. Carr et al., *Appl. Phys. Lett.* 75, 920 (1999). © 1999, American Institute of Physics.

smaller wire dimensions, which is expected due to surface-related losses (Fig. 22).

Electron field emission was studied for 20-nm-wide SiNWs placed on a Si wafer with anode-cathode distances of 10–90 μm [56]. Smooth and consistent current-voltage curves were obtained. The turn-on fields of SiNW emitters with nominal diameter ~10 nm, ~20 nm, and ~30 nm are 4.5, 13, and 23 V/μm (Fig. 23). This is comparable to other types of emitters such as carbon nanotubes, diamond, amorphous carbon, and amorphous silicon nitride. It follows that silicon nanowire emitters could be used in future optical displays based on field emission.

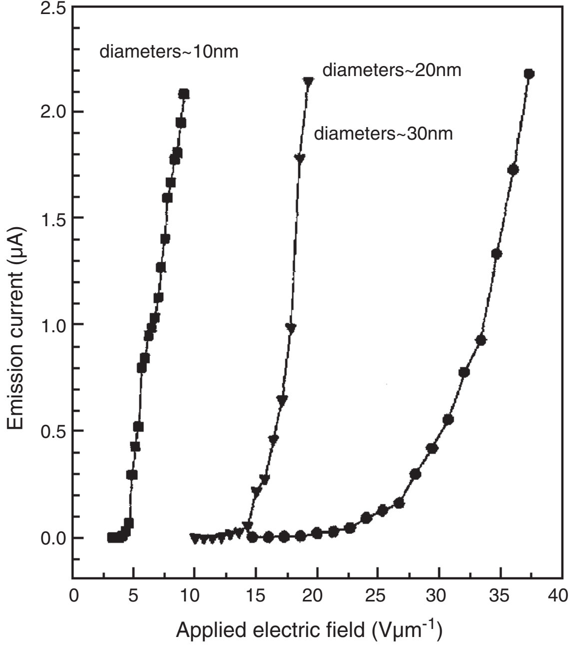


Figure 23. Current-voltage characteristics for field emission of 10-nm-, 20-nm-, and 30-nm-wide silicon nanowires with an anode-sample separation kept at 70 μm. Reprinted with permission from [56], F. C. K. Au et al., *Appl. Phys. Lett.* 75, 1700 (1999). © 1999, American Institute of Physics.

9. SUMMARY

Silicon nanowires are a novel form of nanostructures with exceptional properties. A great deal of work has been devoted to find fabrication techniques to obtain thin, pure, and homogeneous SiNWs in bulk quantities. They have been produced by bottom-up and top-down fabrication techniques. Both approaches usually lead to crystalline SiNWs with an amorphous oxide surface. The crystalline core often has lattice imperfections and defects. Polycrystalline and amorphous SiNWs have also been fabricated. Other morphologies such as graphene and fullerene-like structures have been proposed.

Most bottom-up techniques use catalytic metal clusters for initiating the SiNW growth. This may lead to metal impurities in the nanowires, which can affect the transport properties. Methods have been developed where pure SiNWs are produced. SiNWs with diameters between 2 and several hundred nanometers and several millimeters long have been produced. The grown wires usually have cylindrical shape; the top-down fabricated nanowires have rectangular shape. Top-down fabrication is usually performed by lithography and etching, with many variations in these processes and various types of post-treatments.

SiNWs were usually first studied by electron microscopy and electron diffraction, which give size, length, shape, and atomic structure. Other standard techniques, such as Raman spectroscopy, electron spectroscopy, optical spectroscopy, and electrical conductivity, were subsequently applied. SiNWs are in many respects different from bulk silicon, due to their unique electronic structure, large surface area, optical activity, wide bandgaps, and confinement of electrons, excitons, and phonons. Photoluminescence is observed for SiNWs in the visible range while it is not allowed for bulk silicon. Narrow SiNWs show quantized conductance at low temperatures. Thermal transport through SiNWs is strongly reduced compared to the bulk.

As device dimensions in microelectronics approach the nanometer scale, conventional circuits will reach their operational limits. Low-dimensional electronic materials such as silicon nanowires could be among the solutions to this problem. SiNWs may find applications in optoelectronic devices, in single-electron transistors, and in single- or few-electron memory units. The wide range of novel properties may lead to many other applications for silicon nanowires.

GLOSSARY

Atomic force microscope (AFM) An instrument able to image surfaces to atomic accuracy by mechanically probing their surface contours.

Bottom-up fabrication Building larger objects from smaller building blocks.

Nanoelectronics Electronics on the nanometer scale.

Nanomanipulation The process of manipulating matter at an atomic or molecular scale in order to produce small structures.

Nanoscale 1–100 nanometer size range of structures.

REFERENCES

1. J. R. Heath and F. K. LeGoues, *Chem. Phys. Lett.* 208, 263 (1993).
2. T. J. Trentler, K. M. Hickman, S. C. Goel, A. M. Viano, P. C. Gibbons, and W. E. Buhro, *Science* 270, 1791 (1995).
3. A. M. Morales and C. M. Lieber, *Science* 279, 208 (1998).
4. D. N. Davydov, J. Haruyama, D. Routkevitch, B. W. Statt, D. Ellis, M. Moskovits, and J. M. Xu, *Phys. Rev. B* 57, 13550 (1998).
5. Z. G. Bai, D. P. Yu, J. J. Wang, Y. H. Zou, W. Qian, J. S. Fu, S. Q. Feng, J. Xu, and L. P. You, *Mater. Sci. Eng. B* 72, 117 (2000).
6. C. C. Tang, S. S. Fan, P. Li, L. Chapelle, and H. Y. Dang, *J. Cryst. Growth* 224, 117 (2001).
7. H. Dai, E. W. Wang, Y. Z. Lu, S. Fan, and C. M. Lieber, *Nature* 375, 769 (1995).
8. Y. Zhang, K. Suenaga, C. Colliex, and S. Iijima, *Nature* 281, 973 (1998).
9. W. Q. Han, S. S. Fan, Q. Q. Li, and Y. D. Hu, *Science* 277, 1287 (1997).
10. R. S. Wagner and W. C. Ellis, *Appl. Phys. Lett.* 4, 89 (1964).
11. H. I. Liu, N. I. Maluf, and R. F. W. Pease, *J. Vac. Sci. Technol. B* 10, 2846 (1992).
12. H. Namatsu, S. Horiguchi, M. Nagase, and K. Kurihara, *J. Vac. Sci. Technol. B* 15, 1688 (1997).
13. T. Ono, H. Saitoh, and M. Esashi, *Appl. Phys. Lett.* 70, 1852 (1997).
14. D. P. Yu, Z. G. Bai, Y. Ding, Q. L. Hang, H. Z. Zhang, J. J. Wang, Y. H. Zou, W. Qian, G. C. Xiong, H. T. Zhou, and S. Q. Feng, *Appl. Phys. Lett.* 72, 3458 (1998).
15. N. Wang, Y. H. Tang, Y. F. Zhang, C. S. Lee, I. Bello, and S. T. Lee, *Chem. Phys. Lett.* 299, 237 (1999).
16. Y. F. Zhang, Y. H. Tang, N. Wang, D. P. Yu, C. S. Lee, I. Bello, and S. T. Lee, *Appl. Phys. Lett.* 72, 1835 (1998).
17. N. Wang, Y. H. Tang, Y. F. Zhang, C. S. Lee, and S. T. Lee, *Phys. Rev. B* 58, 16024 (1998).
18. Q. Gu, H. Dang, J. Cao, J. Zhao, and S. Fan, *Appl. Phys. Lett.* 76, 3020 (2000).
19. B. Marsen and K. Sattler, *Phys. Rev. B* 60, 11593 (1999).
20. G. W. Zhou, Z. Zhang, Z. G. Bai, S. Q. Feng, and D. P. Yu, *Appl. Phys. Lett.* 73, 677 (1998).
21. N. M. Hwang, W. S. Cheong, D. Y. Yoon, and D.-Y. Kim, *J. Cryst. Growth* 218, 33 (2000).
22. J. Westwater, D. P. Gosain, and S. Usui, *Phys. Stat. Sol. A* 165, 37 (1998).
23. Y. Ishikawa, N. Shibata, and S. Fukatsu, *Thin Solid Films* 321, 234 (1998).
24. K.-K. Lew, C. Reuther, A. H. Carim, J. M. Tedwing, and B. R. Martin, *J. Vac. Sci. Technol. B* 20, 389 (2002).
25. T. Tsutsumi, K. Ishii, H. Hiroshima, S. Hazra, M. Yamanaka, I. Sakata, H. Taguchi, E. Suzuki, and K. Tomizawa, *J. Vac. Sci. Technol. B* 18, 2640 (2000).
26. M. Gotza, M. Dutoit, and M. Ilegems, *J. Vac. Sci. Technol. B* 16, 582 (1998).
27. G. S. Chen, C. B. Boothroyd, and C. J. Humphreys, *Appl. Phys. Lett.* 62, 1949 (1993).
28. J. Kedzierski, J. Bokor, and C. Kisielowski, *J. Vac. Sci. Technol. B* 15, 2825 (1997).
29. S. Hu, A. Hamidi, S. Altmeyer, T. Koester, B. Spangenberg, and H. Kurz, *J. Vac. Sci. Technol. B* 16, 2822 (1998).
30. P. Avouris, T. Hertel, and R. Martel, *Appl. Phys. Lett.* 71, 285 (1997).
31. J. Servat, P. Gorostiza, and F. Sanz, *J. Vac. Sci. Technol. A* 14, 1208 (1996).
32. A. Majumdar, P. I. Oden, J. P. Carrejo, L. A. Nagahara, J. J. Graham, and J. Alexander, *Appl. Phys. Lett.* 61, 2293 (1992).
33. S. C. Minne, P. Flueckinger, H. T. Soh, and C. F. Quate, *J. Vac. Sci. Technol. B* 13, 1380 (1995).
34. L. Bartels, G. Meyer, and K. H. Rieder, *Phys. Rev. Lett.* 79, 697 (1997).
35. V. Bouchiat and D. Esteve, *Appl. Phys. Lett.* 69, 3098 (1996).
36. R. Magno and B. B. R., *Appl. Phys. Lett.* 70, 1855 (1997).
37. S. B. Fagan, R. J. Baierle, R. Mota, A. J. R. da Silva, and A. Fazzio, *Phys. Rev. B* 61, 9994 (2000).
38. M. Menon and E. Richter, *Phys. Rev. Lett.* 83, 792 (1999).
39. J. W. Mintmire, *J. Vac. Sci. Technol. A* 11, 1733 (1993).
40. X. T. Zhou, R. Q. Zhang, H. Y. Peng, N. G. Shang, N. Wang, I. Bello, C. S. Lee, and S. T. Lee, *Chem. Phys. Lett.* 215 (2000).
41. D. D. D. Ma, C. S. Lee, and S. T. Lee, *Appl. Phys. Lett.* 79, 2468 (2001).
42. G. W. Zhou, H. Li, H. P. Sun, D. P. Yu, Y. Q. Wang, X. J. Huang, L. Q. Chen, and Z. Zhang, *Appl. Phys. Lett.* 75, 2447 (1999).
43. J. Qi and Y. Masumoto, *Mater. Res. Bull.* 36, 1407 (2001).
44. H. Fujii, S. Kanemaru, T. Matsukawa, and J. Itoh, *Appl. Phys. Lett.* 75, 3986 (1999).
45. Y. Ishikawa, N. Shibata, and S. Fukatsu, *Nucl. Instr. Meth. B* 147, 304 (1999).
46. J. Z. He, J. B. Xu, M. S. Xu, Z. Xie, I. H. Wilson, X. L. Ma, Q. Li, N. Wang, L. S. Hung, C. S. Lee, and S. T. Lee, *Appl. Phys. Lett.* 80, 1812 (2002).
47. M. Paulose, C. A. Grimes, O. K. Varghese, and E. C. Dickey, *Appl. Phys. Lett.* 81, 153 (2002).
48. Y. F. Zhang, Y. H. Tang, Y. Zhang, C. S. Lee, I. Bello, and S. T. Lee, *Chem. Phys. Lett.* 330, 48 (2000).
49. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Appl. Phys. Lett.* 63, 3455 (1993).
50. S. Ossicini, *Phys. Stat. Sol. A* 170, 377 (1998).
51. M. L. Ciurea, V. Iancu, V. S. Teodorescu, L. C. Nistor, and M. G. Blanchin, *J. Electrochem. Soc.* 146, 3516 (1999).
52. L. T. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
53. H. Namatsu, K. Kurihara, and T. Makino, *Appl. Phys. Lett.* 70, 619 (1997).
54. B. Li, D. Yu, and S.-L. Zhang, *Phys. Rev. B* 59, 1645 (1999).
55. D. Papadimitriou and A. G. Nassiopoulos, *J. Appl. Phys.* 84, 1059 (1998).
56. F. C. K. Au, K. W. Wong, Y. H. Tang, Y. F. Zhang, I. Bello, and S. T. Lee, *Appl. Phys. Lett.* 75, 1700 (1999).
57. M. S. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993).
58. J. Zou and A. Balandin, *J. Appl. Phys.* 89, 2932 (2001).
59. Y. S. Tang, G. Jin, J. H. Davies, J. G. Williamson, and C. D. W. Wilkinson, *Phys. Rev. B* 45, 13799 (1992).
60. H. Fujii, T. Matsukawa, S. Kanemaru, and J. Itoh, *Appl. Phys. Lett.* 78, 2560 (2001).
61. Y. Nakajima, Y. Takahashi, S. Horigushi, K. Iwadate, H. Namatsu, and K. Kurihara, *Appl. Phys. Lett.* 65, 2833 (1994).
62. V. Dneprovskii, A. Eev, N. Gushina, D. Okorokov, V. Panov, V. Karavanskii, A. Maslov, V. Sokolov, and E. Dovidenko, *Phys. Stat. Sol. b* 199, 297 (1995).
63. T. Ohno, K. Shiraishi, and T. Ogawa, *Phys. Rev. Lett.* 1992, 2400 (1992).
64. R. J. Needs, A. J. Read, K. J. Nash, S. Bhattarcharjee, A. Qteish, L. T. Canham, and P. D. J. Calcott, *Physica A* 207, 411 (1994).
65. A. J. Read, R. J. Needs, K. J. Nash, L. T. Canham, P. D. J. Calcott, and A. Qtiesh, *Phys. Rev. Lett.* 69, 1232 (1992).
66. G. D. Sanders and Y.-C. Chang, *Appl. Phys. Lett.* 60, 2525 (1992).
67. F. Buda, J. Kohanoff, and M. Parrinello, *Phys. Rev. Lett.* 69, 1272 (1992).
68. C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 48, 11024 (1993).
69. A. M. Saitta, F. Buda, G. Fiumara, and P. V. Giaquinta, *Phys. Rev. B* 53, 1446 (1996).
70. J.-B. Xia and K. W. Cheah, *Phys. Rev. B* 55, 15688 (1997).
71. J. Taguena-Martinez, Y. G. Rubo, M. Cruz, M. R. Beltran, and C. Wang, *Appl. Surf. Sci.* 142, 564 (1999).
72. Y. M. Niquet, C. Delerue, G. Allan, and M. Lannnoo, *Phys. Rev. B* 62, 5109 (2000).

73. J. Koga, K. Nishio, H. Ohtani, F. Yonezawa, and T. Yamaguchi, *J. Non-Cryst. Solids* 293–295, 630 (2001).
74. J. R. Gao, C. de Graaf, J. Caro, S. Radelaar, M. Offenberg, V. Lauer, J. Singleton, T. J. B. M. Janssen, and J. A. A. J. Perenboom, *Phys. Rev. B* 41, 12315 (1990).
75. N. J. Stone and H. Ahmed, *Appl. Phys. Lett.* 73, 2134 (1998).
76. T. G. Kim, X.-L. Wang, Y. Suzuki, K. Komori, and M. Ogura, *IEEE J. Sel. Top. Quantum Electron.* 6, 511 (2000).
77. J. Hu, M. Ouyang, Y. P., and C. M. Lieber, *Nature* 399, 49 (1999).
78. X. Sun, Z. Zhang, and M. S. Dresselhaus, *Appl. Phys. Lett.* 74, 4005 (1999).
79. J. D. Holmes, K. P. Johnston, R. C. Doty, and B. A. Kprgel, *Science* 287, 1471 (2000).
80. J. L. Gole, J. D. Stout, W. L. Rauch, and Z. L. Wang, *Appl. Phys. Lett.* 78, 2346 (2000).
81. S.-W. Chung, J.-Y. Yu, and J. R. Heath, *Appl. Phys. Lett.* 76, 2068 (2000).
82. M. K. Sunkara, S. Sharma, R. Miranda, G. Lian, and E. C. Dickey, *Appl. Phys. Lett.* 79, 1546 (2001).
83. T. I. Kamins, R. S. Williams, T. Hesjedal, and J. S. Harris, *Physica E* 13, 995 (2002).
84. T. I. Kamins, R. S. Williams, Y. Chen, Y.-L. Chang, and Y. A. Chang, *Appl. Phys. Lett.* 76, 562 (2000).
85. Z. Q. Liu, S. S. Xie, W. Y. Zhou, L. F. Sun, Y. B. Li, D. S. Tang, X. P. Zou, C. Y. Wang, and G. Wang, *J. Cryst. Growth* 224, 230 (2001).
86. Z. Sui, P. P. Leong, I. P. Heman, G. S. Higashi, and H. Temkin, *Appl. Phys. Lett.* 60, 2086 (1992).
87. B. X. Li, P. L. Cao, R. Q. Zhang, and S. T. Lee, *Phys. Rev. B* 65, 125305 (2002).
88. D. P. Yu, C. S. Lee, B. I., X. S. Sun, Y. H. Tang, G. W. Zhou, Z. G. Bai, Z. Zhang, and S. Q. Feng, *Solid State Commun.* 105, 403 (1998).
89. Y. F. Zhang, Y. H. Tang, N. Wang, C. S. Lee, I. Bello, and S. T. Lee, *J. Cryst. Growth* 197, 136 (1999).
90. S. T. Lee, Y. F. Wang, Y. F. Zhang, and Y. H. Tang, *MRS Bull.*, August, 36 (1999).
91. G. Zhou, Z. Zhang, and Y. D., *J. Crystal Growth* 197, 129 (1999).
92. Y. H. Tang, Y. F. Zhang, H. Y. Peng, N. Wang, C. S. Lee, and S. T. Lee, *Chem. Phys. Lett.* 314, 16 (1999).
93. Y. H. Tang, Y. F. Zhang, N. Wang, W. S. Shi, C. S. Lee, I. Bello, and S. T. Lee, *J. Vac. Sci. Technol. B* 19, 317 (2001).
94. Y. F. Zhang, Y. H. Tang, C. Lam, N. Wang, C. S. Lee, I. Bello, and S. T. Lee, *J. Cryst. Growth* 212, 115 (2000).
95. X. H. Fan, L. Xu, C. P. Li, Z. Y. F., C. S. Lee, and S. T. Lee, *Chem. Phys. Lett.* 334, 229 (2001).
96. S. Yanagiya, S. Kamimura, and H. Koinuma, *Appl. Phys. Lett.* 71, 1409 (1997).
97. Y. Zhang, Q. Zhang, N. Wang, Y. Yan, H. Zhou, and J. Zhu, *J. Cryst. Growth* 226, 185 (2001).
98. D. P. Yu, Y. J. Xing, Q. L. Hang, H. F. Yan, J. Xu, Z. H. Xi, and S. Q. Feng, *Physica E* 9, 305 (2001).
99. D. Al-Mawlawi, C. Z. Liu, and M. Moskovits, *J. Mater. Res.* 9, 1014 (1994).
100. A. Forchel, P. Ils, K. H. Wang, and O. Schilling, *Microelectron. Eng.* 32, 317 (1996).
101. N. Miyata, H. Watanabe, and M. Ichikawa, *J. Vac. Sci. Technol. B* 17, 978 (1999).
102. J. L. Liu, Y. Shi, F. Wang, Y. Lu, R. Zhang, P. Han, S. L. Gu, and Y. D. Zheng, *Appl. Phys. Lett.* 68, 352 (1996).
103. T. Tsutsumi, K. Tomizawa, K. Ishii, S. Kanemaru, T. Maeda, and E. Suzuki, *J. Vac. Sci. Technol. B* 17, 77 (1999).
104. S. Fujita, S. Maruno, H. Watanabe, and M. Ichikawa, *J. Vac. Sci. Technol. B* 16, 2817 (1998).
105. S. Maruno, S. Fujita, H. Watanabe, and M. Ichikawa, *J. Appl. Phys.* 82, 639 (1997).
106. H. Namatsu, Y. Takahashi, M. Nagase, and K. Murase, *J. Vac. Sci. Technol. B* 13, 2166 (1995).
107. T. Tsutsumi, E. Suzuki, K. Ishii, S. Kanemaru, T. Maeda, and K. Tomizawa, *J. Vac. Sci. Technol. B* 17, 1897 (1999).
108. V. Bouchiat, M. Faucher, T. Fournier, B. Pannetier, C. Thirion, W. Wernsdorfer, N. Clement, D. Tonnea, H. Dallaporta, S. Safarov, J. C. Villegier, D. Fraboulet, D. Mariolle, and J. Gautier, *Microelectron. Eng.* 61–62, 517 (2002).
109. J. W. Lyding, T. C. Shen, J. S. Hubacek, J. R. Tucker, and G. C. Abeln, *Appl. Phys. Lett.* 64, 2010 (1994).
110. T. C. Shen, C. Wang, J. W. Lyding, and J. R. Tucker, *Appl. Phys. Lett.* 66, 976 (1995).
111. H. C. Day and D. R. Allee, *Appl. Phys. Lett.* 62, 2691 (1993).
112. I. V. Blonskyy, M. S. Brodyn, A. Y. Vakhnin, V. M. Kadan, and A. K. Kadeshchik, *Phys. Lett. A* 279, 391 (2001).
113. Y. F. Zhang, Y. H. Tsang, H. Y. Peng, N. Wang, C. S. Lee, I. Bello, and S. T. Lee, *Appl. Phys. Lett.* 75, 1842 (1999).
114. N. Wang, Y. H. Tang, Y. F. Zhang, D. P. Yu, C. S. Lee, I. Bello, and S. T. Lee, *Chem. Phys. Lett.* 283, 368 (1998).
115. Y. Cui, L. J. Lauhon, M. S. Gudiksen, and J. Wang, *Appl. Phys. Lett.* 78, 2214 (2001).
116. Y. F. Zhang, L. S. Liao, W. H. Chan, S. T. Lee, R. Sammynaiken, and T. K. Sham, *Phys. Rev. B* 61, 8298 (2000).
117. M. Nagase, S. Horiguchi, A. Fujiwara, Y. Ono, K. Yamazaki, H. Namatsu, and Y. Takahashi, *Appl. Surf. Sci.* 190, 144 (2002).
118. R.-P. Wang, G.-W. Zhou, Y.-L. Liu, S.-H. Pan, H.-Z. Zhang, D.-P. Yu, and Z. Zhang, *Phys. Rev. B* 61, 16827 (2000).
119. E. F. Steigmeier, B. Delley, and H. Auderset, *Phys. Scrip.* T45, 305 (1992).
120. M.-Y. Shen, *Phys. Lett. A* 180, 295 (1993).
121. H. Kohno, T. Iwasaki, Y. Mita, M. Kobayashi, S. Endo, and S. Takeda, *Physica B* 308–310, 1097 (2001).
122. D. P. Yu, Z. G. Bai, and S. Q. Geng, *Phys. Rev. B* 59, R2498 (1999).
123. J. Itoh, Y. Nazuka, S. Kanemaru, T. Inoue, H. Yokoyama, and K. Shimizu, *J. Vac. Sci. Technol. B* 14, 2105 (1996).
124. H. Yokoyama and T. Inoue, *Thin Solid Films* 242, 33 (1994).
125. B. R. Nag and S. Gangopadhyay, *Phys. Stat. Sol. a* 179, 463 (1993).
126. H. M. Polatoglou, *J. Lumin.* 57, 117 (1993).
127. J. L. Liu, Y. Lu, Y. Shi, S. L. Gu, R. L. Jiang, F. Wang, and Y. D. Zheng, *Appl. Phys. A* 66, 539 (1998).
128. S. Horiguchi, Y. Nakajima, Y. Takahashi, and M. Tabe, *Jpn. J. Appl. Phys.* 34, 5489 (1995).
129. H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson, and R. F. W. Pease, *Appl. Phys. Lett.* 64, 1383 (1994).
130. S. Q. Feng, D. P. Yu, H. Z. Zhang, Z. G. Bai, and Y. Ding, *J. Cryst. Growth* 209, 513 (2000).
131. P. A. Dennig, H. I. Liu, D. A. Stevenson, and R. F. W. Pease, *Appl. Phys. Lett.* 67, 909 (1995).
132. H. Y. Peng, N. Wang, W. S. Shi, Y. F. Zhang, C. S. Lee, and S. T. Lee, *J. Appl. Phys.* 89, 727 (2001).
133. H. Z. Zhang, D. P. Yu, Y. Ding, Z. G. Bai, Q. L. Hang, and S. Q. Feng, *Appl. Phys. Lett.* 73, 3396 (1998).
134. A. G. Cullis and L. T. Canham, *Nature* 353, 335 (1991).
135. Y. H. Tang, Y. F. Zheng, C. S. Lee, and S. T. Lee, *Chem. Phys. Lett.* 328, 346 (2000).
136. A. Irvine, Z. A. K. Durrani, and H. Ahmed, *Appl. Phys. Lett.* 73, 1113 (1998).
137. Z. Q. Liu, Z. W. Pan, L. F. Sun, D. S. Tang, W. Y. Zhou, G. Wang, L. X. Qian, and S. S. Xie, *J. Phys. Chem. Solids* 61, 1171 (2000).
138. Y. Q. Zhu, W. K. Hsu, N. Grobert, M. Teronnes, T. H., H. W. Kroto, D. R. M. Walton, and B. Q. Wei, *Chem. Phys. Lett.* 322, 312 (2000).
139. Y. H. Tang, Y. F. Zhang, N. Wang, C. S. Lee, X. D. Han, I. Bello, and S. T. Lee, *J. Appl. Phys.* 85, 7981 (1999).
140. S. T. Lee, N. Wang, and C. S. Lee, *Mater. Sci. Eng. A* 286, 16 (2000).
141. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Phys. Rev. B* 50, 14405 (1994).

142. B. Marsen, M. Lonfat, P. Scheier, and K. Sattler, *Phys. Rev. B* 62, 6892 (2000).
143. X. H. Sun, H. Y. Peng, Y. H. Tang, W. S. Shi, N. B. Wong, C. S. Lee, S. T. Lee, and T. K. Sham, *J. Appl. Phys.* 89, 6396 (2001).
144. S. T. Lee, Y. F. Zhang, N. Wang, Y. H. Tang, I. Bello, C. S. Lee, and Y. W. Chung, *J. Mater. Res.* 14, 4503 (1999).
145. Y. H. Tang, X. H. Sun, F. C. K. Au, L. S. Liao, H. Y. Peng, C. S. Lee, S. T. Lee, and T. K. Sham, *Appl. Phys. Lett.* 79, 1673 (2001).
146. S. Ossicini, C. M. Bertoni, and O. Bisi, *Thin Solid Films* 297, 154 (1997).
147. S. Horiguchi, *Physica B* 227, 336 (1996).
148. R. J. Needs, S. Bhattacharjee, K. J. Nash, A. Qteish, A. J. Read, and L. T. Canham, *Phys. Rev. B* 50, 14223 (1994).
149. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Appl. Phys. Lett.* 64, 3545 (1994).
150. A. K. Sood, K. Jayaram, and D. V. S. Mathu, *J. Appl. Phys.* 72, 4963 (1992).
151. X.-S. Zhao, P. D. Persans, J. Schoeder, and Y.-J. Wu, *Mater. Res. Soc. Symp.* 283, 127 (1993).
152. N. Ookubo, Y. Matsuda, and N. Kuroda, *Appl. Phys. Lett.* 63, 346 (1993).
153. J. M. Ryan, P. R. Wamsley, and K. L. Bray, *Appl. Phys. Lett.* 63, 2260 (1993).
154. M. Yamamoto, R. Hayashi, K. Tsunemoto, K. Kohno, and Y. Osaka, *Jpn. J. Appl. Phys.* 30, 136 (1991).
155. T. I. Kamins, R. S. Williams, D. P. Basile, T. Hesjedal, and J. S. Harris, *J. Appl. Phys.* 89, 1008 (2001).
156. Z. Y. Xu, M. Gal, and M. Gross, *Appl. Phys. Lett.* 60, 1375 (1992).
157. Y. Takahashi, T. Furuta, Y. Ono, T. Ishiyama, and M. Tabe, *Jpn. J. Appl. Phys.* 34, 950 (1995).
158. A. Daami, G. Bremond, L. Stalmans, and J. Poortmans, *J. Lumin.* 80, 169 (1998).
159. Y. P. Piryatisnkii, V. A. Semenovich, N. I. Klyui, and A. G. Rozhin, *Tech. Phys.* 43, 423 (1998).
160. G. D. Sanders and Y.-C. Chang, *Phys. Rev. B: Condens. Mat.* 45, 9202 (1992).
161. X. Zianni and A. G. Nassiopoulou, *Phys. Rev. B* 65, 035326-1 (2002).
162. S. M. Prokes, O. J. Glembocki, V. M. Bermudez, R. Kaplan, L. E. Friesdorf, and P. C. Searson, *Phys. Rev. B* 45, 13788 (1992).
163. C. Tsai, K.-H. Li, J. Sarathy, S. Shih, J. C. Campbell, B. K. Hance, and J. M. White, *Appl. Phys. Lett.* 59, 2814 (1991).
164. K. Nishio, J. Koga, H. Ohtani, T. Yamaguchi, and F. Yonezawa, *J. Non-Cryst. Solids* 293-295, 705 (2001).
165. A. Trellakis and U. Ravaioli, *Comput. Methods Appl. Mech. Engrg.* 181, 437 (2000).
166. V. Ng, H. Ahmed, and T. Shimada, *Appl. Phys. Lett.* 73, 972 (1998).
167. U. Landman, R. N. Barnett, A. G. Scherbakov, and P. Avouris, *Phys. Rev. Lett.* 85, 1958 (2000).
168. L. G. C. Rego and G. Kirczenow, *Phys. Rev. Lett.* 81, 232 (1998).
169. T. S. Tighe, J. W. Morlock, and M. L. Roukes, *Appl. Phys. Lett.* 70, 2687 (1997).
170. S. G. Volz and G. Chen, *Appl. Phys. Lett.* 75, 2056 (1999).
171. A. Balandin and K. L. Wang, *Phys. Rev. B* 58, 1544 (1998).
172. S. G. Walkauskas, D. A. Broido, K. Kempa, and T. L. Reinecke, *J. Appl. Phys.* 85, 2579 (1999).
173. X. Lue, W. Z. Shen, and J. H. Chu, *J. Appl. Phys.* 91, 1542 (2002).
174. M. Je, S. Han, I. Kim, and H. Shin, *Solid-State Electron.* 44, 2207 (2000).
175. R. A. Smith and H. Ahmed, *Appl. Phys. Lett.* 71, 3838 (1997).
176. D. H. Kim, S.-K. Sung, J. S. Sim, K. R. Kim, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang, and D. Ahn, *Appl. Phys. Lett.* 79, 3812 (2001).
177. A. Fujiwara, K. Yamazaki, and Y. Takahashi, *Appl. Phys. Lett.* 80, 4567 (2002).
178. J. K. Vincent, V. Narayan, and M. Willander, *Phys. Rev. B* 65, 125309 (2002).
179. N. J. Stone and H. Ahmed, *Microelectron. Eng.* 41/42, 511 (1998).
180. D. W. Carr, S. Ecoy, L. Sekaric, H. G. Craighead, and J. M. Parpia, *Appl. Phys. Lett.* 75, 920 (1999).
181. R. A. Smith and H. Ahmed, *J. Appl. Phys.* 81, 2699 (1997).

www.aspbs.com/enn

Silicon Quantum Dots

Anri Nakajima

Hiroshima University, Higashi-Hiroshima, Japan

CONTENTS

1. Introduction
2. Fabrication of Si Quantum Dots
3. Optical Properties and Applications
4. Electrical Properties and Applications
5. Conclusions
Glossary
References

1. INTRODUCTION

The current decade has seen great progress in the use of Si nanostructures for single-electron and quantum devices. There are two fields of research into the use of Si nanostructures. One field is their application in light-emitting devices used with large-scale-integrated (LSI) circuits, and the other is their application in ultimate low-power systems or new functional electron devices, such as single-electron or quantum-electronic devices.

With regard to light-emitting devices, considerable activity followed the first observation, in 1990, of intense, visible photoluminescence from porous Si at room temperature [1]. As well as the practical aspects, the mechanism underlying the photoluminescence itself is interesting because bulk Si has an indirect bandgap, which results in only weak photoluminescence in the infrared region. It goes without saying that these basic studies will be of great help in establishing various device applications for Si nanostructures.

Single-electron devices might be of use as elements of future solid-state circuits that require high integration and low power consumption. Single-electron charging effects are in themselves nothing new and have long been studied [2–5]. However, renewed interest has led to much theoretical and experimental work since the first proposal of their application to single-electron transistors (SET) in the mid-1980s [6]. Early work concentrated on proof of the basic operation and elucidation of the physical properties of SETs [7–9]. This involved experiments at temperatures at which the thermal energy was less than the single-electron charging energy of a Coulomb island. Accordingly, when observations were carrieds out at very low temperatures, it was unnecessary for the Coulomb island to be very small. In this early work, single-electron devices were mainly fabricated using metals or compound semiconductors, and the experiments were performed at very low temperatures. Owing to the relatively large device size used at these low temperatures, device fabrication could be achieved using conventional electron-beam (EB) lithography, the metal lift-off process, metal evaporation, and the shadow mask technique, which had already been used to produce metal gate structures for electron confinement in two-dimensional structures present in compound semiconductors.

In contrast, Si is the preferred material for single-electron and quantum-electronic devices from a practical point of view because we can use mature technologies for LSI circuits during the fabrication process. This is a definite advantage in developing commercial devices. One further requirement for the use of single-electron and quantum-electronic devices in future LSI devices is the ability to operate at room temperature. To achieve this using SETs, we need a nanoscale Coulomb island which reduces the total capacitance of the island so that the single-electron charging energy exceeds the thermal energy at room temperature. A similar structure is needed in quantum-electronic devices for electron confinement; the structure must be small enough for subband splitting to exceed the thermal energy at room temperature.

Research into the practical uses of Si single-electron and quantum-electronic devices began in the early 1990s [10, 11] when the fabrication of nanostructures in Si became possible owing to the continuous reduction of device size in LSI circuits for higher integration and performance. Room-temperature operation of Si single-electron memories and SETs was first reported in 1993 [10] and 1994 [11], respectively.

This article gives an overview of Si quantum dots along with their fabrication, microstructure, and optical and electrical characteristics. In this article, quantum dots include Coulomb islands, which are key elements in single-electron devices. In Section 2, natural and self-assembled formation as well as artificial fabrication of Si quantum dots are described. In Section 3, the optical properties and applications of Si quantum dots are reviewed. Their electrical properties and applications are reviewed in Section 4.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (837–857)

Finally, there is a summary of the remaining problems and some suggestions for the direction of future research (Section 5).

2. FABRICATION OF Si QUANTUM DOTS

With the aim of creating quantum devices or single-electron devices, many processes have already been proposed for the fabrication of Si quantum dots. These fabrication techniques can be divided into two broad categories: (1) self-organizing fabrication processes, and (2) artificial fabrication, with the most typical and direct method being EB lithography.

2.1. Natural and Self-Assembled Formation

While EB lithography has many advantages in nanostructure fabrication, its low rate of productivity is a serious drawback to the development of commercial devices. In contrast, the self-organizing fabrication process for nanostructures is a relatively simple and inexpensive method that is also highly productive.

In 1988, Furukawa and Miyasato first reported a quantum size effect in naturally formed Si nanocrystals surrounded by hydrogen atoms with diameters of 2–3 nm. They used a magnetron rf sputtering technique in hydrogen gas onto a low-temperature substrate [12]. Using optical absorption measurements, they observed a wider optical bandgap than that of bulk Si. Then in 1990, Takagi et al. fabricated Si nanocrystals, ranging in size from 2.8 to 5 nm, embedded in a SiO_2 matrix using microwave plasma decomposition of SiH_4 and H_2, or SiH_4, H_2, and Ar mixture [13]. They observed visible photoluminescence, which could be ascribed to the quantum confinement effect of electron-hole pairs.

Around the same time, visible photoluminescence in porous Si at room temperature was first reported [1]. After that discovery, most of the techniques for fabricating Si nanocrystals in the early 1990s were developed with the aim of studying their photo- or electroluminescence properties. In these optical studies, increasing luminescence intensity by increasing the number of Si nanocrystals responsible for it, size uniformity to produce a sharp luminescence line, and size controllability to change the position of the luminescence peak, are more important than achieving precise positional controllability of the nanocrystals. To achieve this, various techniques have been proposed, such as sputtering, gas evaporation, plasma-enhanced chemical vapor deposition (PECVD), and ion implantation followed by thermal annealing.

In the sputtering methods, Si nanocrystals were fabricated by means of an rf sputtering technique using hydrogen or argon plasma. The target materials are Si, SiO_2, or a mixture of Si and SiO_2. Si nanocrystals surrounded by hydrogen atoms were fabricated using a pure Si target with hydrogen plasma [12, 14]. Si nanocrystals embedded in a SiO_2 matrix were prepared by rf co-sputtering using argon plasma followed by N_2 [15] or vacuum [16] annealing at high temperatures. For targets in the co-sputtering, pieces of Si wafer were placed around or on pure silica. Si nanocrystals embedded in a SiO_2 matrix were also fabricated by sputtering from only a pure SiO_2 target using argon plasma followed by N_2 annealing above 1000 °C [17].

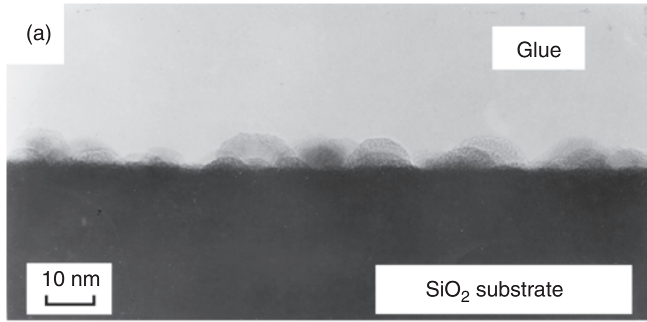
In the gas evaporation methods, high-purity Si powder was evaporated from a ceramic boat heated to high temperatures (about 2000 °C) in an inert gas, or an oxygen-containing inert-gas atmosphere under low pressure [18, 19]. Adding oxygen to the atmosphere during evaporation means that the surface of the deposited Si nanocrystals can be oxidized and the size of the Si core can be reduced [18]. The oxidized Si nanocrystals shows above-bandgap photoluminescence peak. The amount of the increase in the bandgap is 0.3 eV, which can be attributed to a quantum size effect. Also, Si nanocrystals and amorphous Si nanoparticles embedded in a SiO_2 matrix were prepared by evaporation of SiO followed by high-temperature annealing at 1303 K and 973 K, respectively [20]. Photoluminescence has been observed from both amorphous and crystalline nanoparticles and interpreted in terms of band-to-band recombination in nanoparticles with an average size of more than 2.5 nm and carrier recombination through defect states in smaller nanoparticles [20].

In the PECVD methods, Si nanocrystals surrounded by hydrogen atoms were produced using the decomposition of SiH_4 and H_2, or a SiH_4, H_2, and Ar gas mixture [13, 21, 22], or a SiH_4 and Ar gas mixture [23]. Si nanocrystals embedded in a SiO_2 matrix were produced using a SiH_4 and N_2 gas mixture with N_2O followed by thermal annealing in N_2 [24].

Thermal SiO_2 and bulk-fused silica were implanted with Si and Si nanocrystals formed after subsequent thermal annealing in an inert-gas ambient [25, 26]. To produce Si nanocrystals with a narrow size distribution throughout a thick SiO_2 film, Si implantation was carried out at multiple energies with varying doses [26].

Other typical fabrication methods for the synthesis of Si nanocrystals are gas-phase nucleation following pyrolysis of Si_2H_6, which uses an aerosol apparatus [27], and pulsed-laser-induced decomposition of SiH_4 in a flow reactor [28]. Si nanocrystals embedded in a SiO_2 and Si_3N_4 matrix were prepared using an approach based on the deposition of Si-rich Si-O and Si-N films by the ultrahigh vacuum chemical vapor deposition reactions of $O(SiH_3)_2$ and $N(SiH_3)_3$, respectively, followed by high-temperature thermal annealing [29].

However, the above techniques for producing Si nanocrystals are not always suitable for electron-device applications due to the inability to precisely control the position of the nanocrystals. The first clear demonstration of the practical application of a Si single-electron device was a single-electron memory with floating gates reported by Yano et al. [10]. They utilized nanoscale grains in an ultrathin poly-Si film as ultrasmall floating gates and ultranarrow channels in 1993 (three years after the discovery of visible photoluminescence in porous Si) [1]. Improvements in the methods of fabricating Si quantum dots for electron devices started from this point. Tiwari et al. fabricated a single-electron memory with floating dots utilizing Si nanocrystals produced by a specially designed ultralow-pressure CVD on a thin tunnel oxide [30]. Then, Nakajima et al. [31, 32] applied a conventional low-pressure CVD (LPCVD) technique using pyrolysis of SiH_4 to form Si quantum dots on a SiO_2 substrate. Figure 1 shows the transmission electron microscopy (TEM) micrograph of the fabricated Si nanocrystals. Owing to its wide



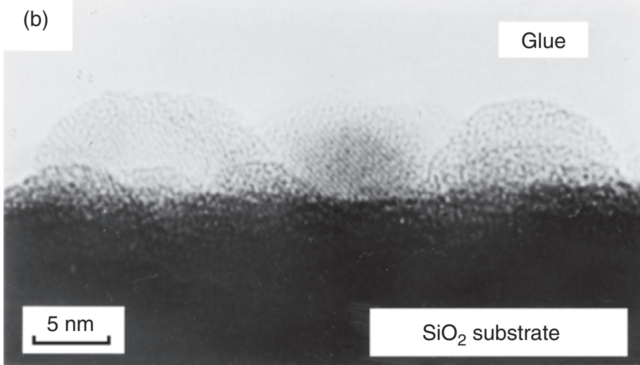


Figure 1. Cross-sectional TEM micrograph of a layer of Si nanocrystals for (a) an LPCVD deposition time of 60 seconds, and (b) a higher resolution micrograph. Reprinted with permission from [31] and [32], A. Nakajima et al., *Jpn. J. Appl. Phys.* 35, L189 (1996); *J. Appl. Phys.* 80, 4006 (1996). © 1996, Institute of Pure and Applied Physics; American Institute of Physics.

use in current ULSI technology and its convenience, Si quantum dots formed utilizing this technique are now extensively used not only for single-electron memories with floating dots [33] but also for Coulomb islands in Si SETs [34]. This technique is based on the early stages of Si film growth using conventional LPCVD on a SiO_2 film. It should be noted that Brownian motion of Si nanocrystals with diameters about 10 nm on SiO_2 was observed during the deposition. Though Brownian migration on KCl of Al or Au nanocrystals with diameters of less than 5 nm had previously been reported [35, 36], this was the first report regarding Brownian migration of such large Si nanocrystals on SiO_2.

Another approach to engineering the spatial position, size, and density of Si nanocrystals is the preparation of amorphous Si/SiO_2 superlattices [37–40]. Si nanocrystals were formed as a result of thermal recrystallization using high-temperature annealing [37]. The crystallization temperature increases rapidly with decreasing thickness of the amorphous Si layer. The size of the Si nanocrystals formed is limited by the thickness of the amorphous Si layer. A short period of steam oxidation was found to improve the surface passivation of Si nanocrystals. Amorphous Si/SiO_2 superlattices were grown using the PECVD method with alternating decomposition of SiH_4 and SiH_4/nitrous oxide [38], or by magnetron sputtering of amorphous Si followed by plasma oxidation [37, 38].

A severe drawback of this method is that most of the nanocrystals are touching. This clearly limits the luminescence efficiency due to the nonradiative processes that occur at the grain boundaries. An interesting development is the phase separation and thermal crystallization technique for SiO/SiO_2 superlattices used by Zacharias et al. [39]. Amorphous SiO/SiO_2 superlattices were prepared by reactive evaporation of SiO powders in an oxygen atmosphere. The SiO layers, ranging in thickness from 1 to 3 nm, were separated by SiO_2 layers of 2–3 nm. After deposition, the samples were annealed at 1100 °C for 1 hour under a N_2 atmosphere. The high-temperature annealing of such initially amorphous SiO film results in phase separation described by

$$SiO \rightarrow 0.5\ SiO_2 + 0.5\ Si$$

and in Si nanocrystals in a SiO_2 matrix. The phase separation of the SiO automatically ensures that the nucleated Si nanocrystals are separated from each other by a SiO_2 shell. Gourbilleau et al. also proposed a method for achieving good phase separation in a Si nanocrystal layer [40]. Si nanocrystals embedded in a SiO_2 matrix were formed by thermal annealing of a Si-rich layer/SiO_2 superlattice produced by magnetron sputtering. Instead of alternating Si target sputtering and plasma oxidation, the SiO_2 targets were sequentially sputtered by pure argon plasma for the deposition of a SiO_2 layer, and by a mixture of argon and hydrogen plasma to produce a Si-rich layer. The latter run made use of the ability of hydrogen to reduce the amount of oxygen originating from the SiO_2 target.

In a different approach, Si nanocrystals/amorphous Si superlattices were fabricated by Tong et al. [41] using the PECVD method. The layers of Si nanocrystals, surrounded by hydrogen atoms and amorphous Si, were prepared by changing some of the deposition parameters: (1) the gas ratio ($SiH_4/SiH_4 + H_2$), and (2) the application of a negative bias voltage to the sample substrate.

2.2. Artificial Fabrication

As mentioned above, the greatest advantage of using natural and self-assembling formation for Si quantum dots is high productivity. Although excellent uniformity of size and depth position was achieved using some of the formation methods described above, there was inadequate control over the number and lateral position of the dots. It is also often difficult to alter the size and position of the dots using natural and self-assembling formation methods. The most typical and direct method of overcoming these problems is to use EB lithography followed by various etching processes.

Using electrostatic confinement with a dual-gate device, Alsmeier et al. [42] fabricated periodic arrays of electron dots on Si, which were widely tunable in diameter and electron number. MOS devices with dual gates were fabricated. A metal (NiCr) bottom gate with a mesh structure was formed on a thermal oxide grown on *p*-type Si substrate. The mesh had circular openings of about 150 nm in diameter with a periodicity of 400 nm fabricated using EB lithography. PECVD was used to deposit the upper layer of SiO_2 on the mesh with a continuous thin metal (NiCr) gate on top. Positive and negative voltages were applied to the top and bottom gates, respectively. In the presence of bandgap radiation, electron dots were induced at the Si/SiO_2 interface underneath the bottom gate openings via the top gate voltage. The bottom gate voltage served to isolate the dots and

the top gate voltage determined the inversion layer electron density.

Nakajima et al. fabricated nanoscale Si dot arrays using EB lithography [43]. Following lithography, anisotropical dry etching and isotropical wet etching in $NH_4OH/H_2O_2/H_2O$ produces completely isolated crystalline Si dot arrays 10 nm wide and 10 nm high (Fig. 2). Since wet etching in $NH_4OH/H_2O_2/H_2O$ is a low-temperature process, problems such as a change in the doping profile can be ameliorated. After wet etching in $NH_4OH/H_2O_2/H_2O$, the expansion in volume caused by the growth of an oxide layer on the Si surface is much less than the dot size. This means that the stress induced by the expansion is low, enabling the integration and miniaturization of devices. The large reduction in device size is possible because the wet-etching process does not have any self-limiting characteristics in contrast to the oxidation of low-dimensional Si structures [44]. The extent of damage that occurs during dry etching can be greatly reduced by using wet etching. A further advantage of wet etching is that, since the technique is commonly used in the cleaning of Si wafers (RCA process), Si quantum dots can be fabricated using conventional Si LSI fabrication methods.

A self-limiting process combined with conventional EB lithography was proposed by Fukuda et al. to fabricate thin pillars consisting of multilayers of Si polycrystalline quantum dots with Si nitride tunnel barriers [44]. Vertically stacked Si dots with dimensions on the order of 10 nm in all three directions, connected by thin Si nitride layers (~2 nm thick), were fabricated using the saturation characteristics of the poly-Si core diameter during lateral pillar oxidation. During the fabrication, the multilayer structures were formed first by repeated deposition of amorphous Si film (~10 nm thickness) followed by thermal nitridation. Then the pillars were patterned using EB lithography followed by dry etching. The patterned pillars were oxidized in a dry oxygen ambient at 850~900 °C, yielding self-limiting oxidation of the Si pillars.

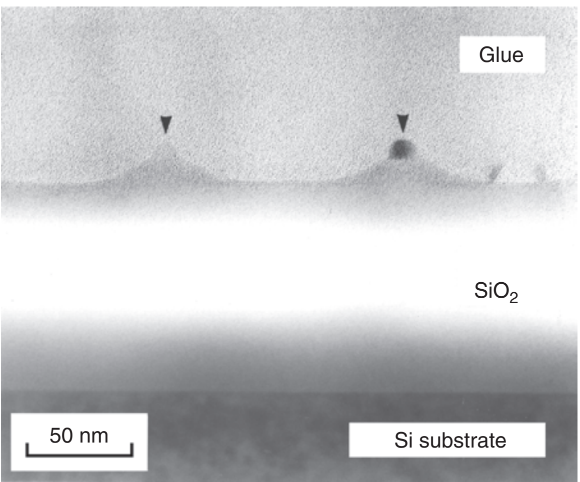


Figure 2. Cross-sectional TEM micrograph of Si dots 10 nm high and 10 nm in diameter (indicated by arrows) after the isotropic wet etching. The difference in contrast between the two dots is attributed to the difference in the crystal orientation of each poly-Si dot. Reprinted with permission from [43], A. Nakajima et al., *Jpn. J. Appl. Phys.* 33, L1796 (1994). © 1994, Institute of Pure and Applied Physics.

3. OPTICAL PROPERTIES AND APPLICATIONS

3.1. Quantum Size Effect

The quantum size effects of Si nanocrystals have been extensively studied using various optical measurements, such as measurement of optical absorption or photoluminescence.

Si nanocrystals surrounded by either hydrogen [12] or oxygen atoms [31, 32] indicate an increase in the optical bandgap due to the three-dimensional quantum confinement effect. Nakajima et al. measured the optical absorption of Si nanocrystals formed on a SiO_2 wafer [31, 32]. They showed that Si nanocrystals with a diameter of about 10 nm essentially maintain the properties of an indirect bandgap semiconductor, with the Si bandgap increasing due to the electron-hole quantum confinement effect. Figure 3 shows $(\alpha h\nu)^{1/2}$ as a function of the photon energy, $h\nu$, where α is the absorption coefficient. For each sample of nanocrystals with varying average sizes, $(\alpha h\nu)^{1/2}$ changes linearly with the photon energy, $h\nu$, near the absorption edge, which indicates that an indirect transition in the absorption is likely to occur for all samples. The bandgap energy of 1.7 eV obtained for a 60-second deposition time (average lateral diameter of 9.1 nm and average height of 4.0 nm including the outer surface oxide layer) showed reasonable agreement with the theoretical calculations of Takagahara and Takeda based on envelope function analysis [45]. A similar analysis also performed for the absorption coefficient of Si nanocrystals embedded in SiO_2 [24] and porous Si [46] indicated that Si nanocrystals maintain an indirect bandgap structure with an increase in the bandgap due to the electron-hole quantum confinement. This result is also consistent with that obtained from site-selective excitation spectroscopy for porous Si described in the Section 3.2.

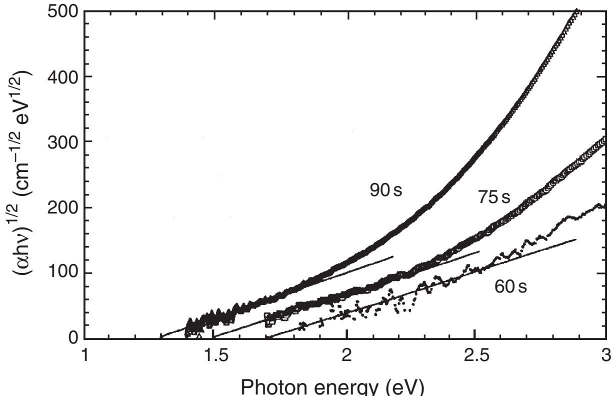


Figure 3. Square root of absorption coefficient (α) times photon energy ($h\nu$) vs. photon energy ($h\nu$) for Si dots with different deposition times. The sample with a 60-second deposition time consisted of isolated Si dots with an average diameter of 9.1 nm and an average height of 4.0 nm. At deposition times exceeding 60 seconds, not only does the diameter of each Si dot become larger, but also the dots tend to coalesce due to Brownian migration, leading to a rapid increase in the Si dot size over time. Reprinted with permission from [32], A. Nakajima et al., *J. Appl. Phys.* 80, 4006 (1996). © 1996, American Institute of Physics.

As described earlier, there seems to be a consensus at present that the photogeneration of electron-hole pairs mainly occurs in the Si core region of Si nanocrystals with an increased bandgap due to the quantum confinement effect. However, in regard to the radiative recombination process, there is still controversy as to whether the band-to-band radiative recombination of electron-hole pairs confined within nanocrystals is essential or whether the effect of the nanocrystal surface cannot be ignored.

Wilson et al. showed that the peak of the photoluminescence spectrum shifts towards a higher energy as the average size of the Si nanocrystals decreases [47]. They formed surface-oxidized Si nanocrystals using homogeneous gas-phase nucleation following pyrolysis of dilute Si_2H_6 in He. They used size-selective precipitation and size-exclusion chromatography to separate the nanocrystals into different groups with larger or smaller average particle size. There was no Stokes shift between the photoluminescence peak energy and the bandgap of the Si nanocrystals estimated from the photoluminescence excitation spectra.

Patrone et al. reported a photoluminescence study of Si nanocrystals produced using a conventional pulsed-laser ablation technique [48]. The Si nanocrystals were produced under an inert-gas ambient and the photoluminescence measurements were carried out *in-situ*. By selecting appropriate experimental parameters, they obtained significant condensation of the Si nanocrystals with reduced size dispersion. The photoluminescence band shifted from near-infrared to near-ultraviolet as the mean size of the nanocrystals varied from $\sim$4 to $\sim$1 nm. The size-dependent photoluminescence was explained by a discrete size model; that is, the spectral position of the photoluminescence band is essentially determined by the volume of nanocrystals with a complete outer atomic layer. The bandgaps of nanocrystals with discrete size differences, which were determined by the quantum confinement effect, corresponded to the photoluminescence peaks.

The above two groups insisted that photoluminescence occurs due to the band-to-band radiative recombination of electron-hole pairs confined within the nanocrystals and the bandgap is determined by the carrier quantum confinement effect.

Fernandez et al. found that the bandgap estimated by the photoluminescence excitation spectra and the photoluminescence peak energy shifted to a higher energy as the average size of the Si nanocrystals embedded in SiO_2 decreased [49, 50]. The Si nanocrystals were synthesized using ion implantation of Si^+ in a thick SiO_2 layer grown on a Si substrate followed by thermal annealing in N_2. The bandgap shifts measured are in good agreement with the model based on quantum confinement. Constant Stokes shifts of 0.26 eV, which is almost twice the energy of Si-O vibration, were observed between the photoluminescence peak energy and the bandgap. This finding suggests that absorption occurs between the quantum confinement states in the Si core and radiative recombination occurs at the Si/SiO_2 interface with the assistance of local Si-O vibration.

Dinh et al. described the optical properties of oxygen- or hydrogen-passivated Si nanocrystals [51]. These nanocrystals had an average core diameter of 2–7 nm and were synthesized by thermal vaporization of Si in an Ar buffer gas, followed by exposure to oxygen or atomic hydrogen. The visible component of the photoluminescence spectra showed blueshifts, which broadened as the size of the Si nanocrystals was reduced. There were differences in the photoluminescence spectra for the hydrogen- and oxygen- passivated nanocrystals. The authors explained the results using a model involving absorption in quantum-confined Si cores and emission due to the transition between surface and/or interface states. Namely, upon absorption of photons, electrons and holes were generated in the conduction and valence bands of the dots. In the emission process, the carriers were trapped and subsequently recombined through surface-related traps with energy levels extending into the energy gaps.

Iacona et al. also reported that the photoluminescence peak showed blueshifts as the size of the Si nanocrystals decreased with a quantitative discrepancy between the peak energies and the theoretically obtained values for the bandgap [52]. Si nanocrystals were formed by high-temperature annealing of substoichiometric Si oxide thin films prepared using the PECVD method. It was speculated that the photoluminescence results were due to absorption occurring as a result of the quantum confinement effect of the carriers in the Si nanocrystal cores, and that recombination occurred on the intermediate states, introducing energy levels inside the bandgap and corresponding to the Si/SiO_2 interfacial states of the Si nanocrystals.

Kanemitsu et al. described the size-independent peak energy of photoluminescence from oxidized Si nanocrystals [53]. The Si nanocrystals were produced by laser breakdown of SiH_4 gas followed by oxidation at room temperature in a clean air box. The average diameter of the Si nanocrystals formed varied from 7 to 13 nm with a constant SiO_2 surface thickness of 1.6 nm achieved by selecting appropriate experimental parameters. The peak position of the photoluminescence was 1.65 eV for all the Si nanocrystals, independent of their size. They also observed that photoluminescence intensity increased with a decrease in the average diameter of the Si nanocrystals in the size region. Photoluminescence intensity was observed to increase at temperatures below $\sim$150 K. The authors used a three-region model to explain these experimental findings. In the model, Si nanocrystals consist of three regions: (a) a Si core with a diameter of less than 5–7 nm, (b) a surface SiO_2 layer with a thickness of about 1.6 nm, and (c) an interfacial layer between (a) and (b). The interfacial layer contains nonstoichiometric amounts of oxygen, which leads to a much lower bandgap (calculated to be about 1.7 eV) than the outer surface SiO_2 layer. The bandgap of the Si core region is theoretically calculated to be higher than that of the interfacial layer when the core diameter is less than 5–7 nm. In this case, the exciton is confined in the interfacial spherical-shell region. This localized exciton increases the oscillator strength. The origin of the photoluminescence was therefore considered to be due to the localized exciton. In a three-region model, photogeneration of excitons mainly occurs within the Si core region. Some of the excitons in the core transfer to the interfacial layer by a thermally activated diffusion process. Then, strong photoluminescence is caused by the radiative recombination of excitons confined in the interfacial layer.

Kanemitsu [54] studied the dynamics of photoluminescence decay and photoluminescence under resonant excitation for these oxidized Si nanocrystals. The

photoluminescence behavior is characterized by a stretched exponential function. The radiative decay rate shows thermally activated temperature dependence with large activation energy (71 meV) at high temperatures, suggesting that the photoluminescence is produced by multilevel light-emitting states in a single nanocrystal. The cascade of carriers from higher to lower states means that the lifetime of the higher states is shorter and an exponential slope is then observed. At low temperatures, however, a long photoluminescence lifetime is mainly determined by the tunneling process. Clear phonon-related steplike structures are not observed in the photoluminescence spectra under resonant excitation of oxidized Si nanocrystals, in contrast to the case of porous Si described later. The large activation energy for radiative recombination and indistinct phonon-related structures indicate that the potential fluctuation in the surface states in oxidized Si nanocrystals is much larger than that in porous Si. The disordered potential of oxidized surface states causes a shallow tail or localized states. Thus, the photoluminescence properties in oxidized Si nanocrystals can be understood in terms of the localization of carriers or excitons in a disordered potential of the surface states.

The photoluminescence decay dynamics and site-selective excitation spectroscopy of surface-oxidized Si nanocrystals fabricated with a SiH_4 plasma cell reported by Kanemitsu et al. suggests that there are different luminescence bands in Si nanocrystals [55]. The photoluminescence decay dynamics show different dependence on monitored photoluminescence wavelengths at photon energies below and above ~1.65 eV for Si nanocrystals with an average core diameter of 4 nm. In site-selective excitation spectroscopy, the phonon-related structures are only observed under excitation levels below ~1.65 eV. Since a larger excitation energy selects smaller Si nanocrystals, the photoluminescence band under laser excitations above ~1.65 eV corresponds to the small Si nanocrystals where the effect of the surface becomes more important. Kanemitsu et al. speculated that the unclear phonon structures under laser excitations above ~1.65 eV are due to the localization of excitons near the disordered interface between the crystalline Si core region and the SiO_2 surface in these small Si nanocrystals. In large Si nanocrystals, on the other hand, excitons are delocalized in the core state, leading to clear phonon-assisted photoluminescence.

Linnros et al. studied the effect of interactions between Si nanocrystals on photoluminescence properties [56]. The decrease in the curvature of the photoluminescence decay curve with the increasing density of Si nanocrystals embedded in SiO_2 suggests a partially interconnected system of nanocrystals where excitons may migrate to and be trapped in large nanocrystals for a high density of nanocrystals.

3.2. Porous Si

Since intense visible photoluminescence was first observed at room temperature in 1990 [1], the mechanism underlying photoluminescence from porous Si has been studied intensively both from a purely scientific standpoint and from the point of view of its practical applications [57–59].

Porous Si samples are typically prepared by electrochemical anodization of the surface of a Si wafer. The microstructure of the porous Si very much depends on the fabrication conditions such as the doping type and resistivity of the Si wafer, current density, concentration of hydrofluoric acid (HF), and presence or absence of illumination. Cullis and Canham showed the existence of columnar Si wires undulating in width in porous Si that emits red light [60]. Nakajima et al. observed similar results for porous Si emitting red light [61]. Microstructures of lightly doped porous Si are shown in Figures 4 and 5 [61]. The existence of threadlike structures (a few nanometers wide) that undulated in width (Fig. 4) was confirmed by TEM. The porous Si also contained numerous spherical Si nanocrystals with diameters of a few nanometers (Fig. 5). This confirmed the existence of Si nanocrystals and undulating crystalline-Si wires with diameters of a few nanometers.

The quantum confinement model has been the focus of most attention in attempts to understand the photoluminescence of porous Si [1, 57–64]. Changes in the photoluminescence spectra were investigated after altering the size of the Si nanocrystals in porous Si and its surface composition. Size reduction was carried out by various methods such as changing the porosity of the porous Si by immersing it in an aqueous HF solution after anodization [1] or changing the anodization conditions by using different types of Si wafers [62]. The size and the surface composition of the porous Si were changed at the atomic layer level by alternately dipping it in an aqueous HF solution and an aqueous H_2O_2 solution (or aqueous HNO_3 solution) [63, 64]. Photoluminescence was measured before and after each step. These measurements confirmed the appropriateness of the quantum confinement model. However, other explanations have also been advanced, suggesting that the optical center may be chemical compounds on the surface of the porous Si [65] or amorphous-like structures in the porous Si [66].

More detailed information was obtained using selective excitation spectroscopy [67–71] and observing the lifetime of the photoluminescence [69, 72]. Two groups independently found a steplike vibronic structure using selective excitation spectroscopy, which indicates that in the photoluminescence of porous Si, phonon-assisted indirect transitions occur in

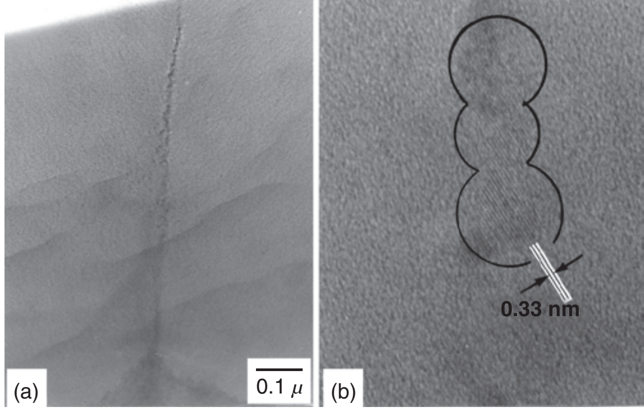


Figure 4. (a) TEM micrograph of threadlike structures consisting of Si nanocrystals with a width of a few nanometers in porous Si; (b) shows the micrograph at a higher resolution. Reprinted with permission from [61], A. Nakajima et al., *Appl. Phys. Lett.* 62, 2631 (1993). © 1993, American Institute of Physics.

Figure 5. TEM micrograph (dark-field) of spherical Si nanocrystals in porous Si. Reprinted with permission from [61], A. Nakajima et al., *Appl. Phys. Lett.* 62, 2631 (1993). © 1993, American Institute of Physics.

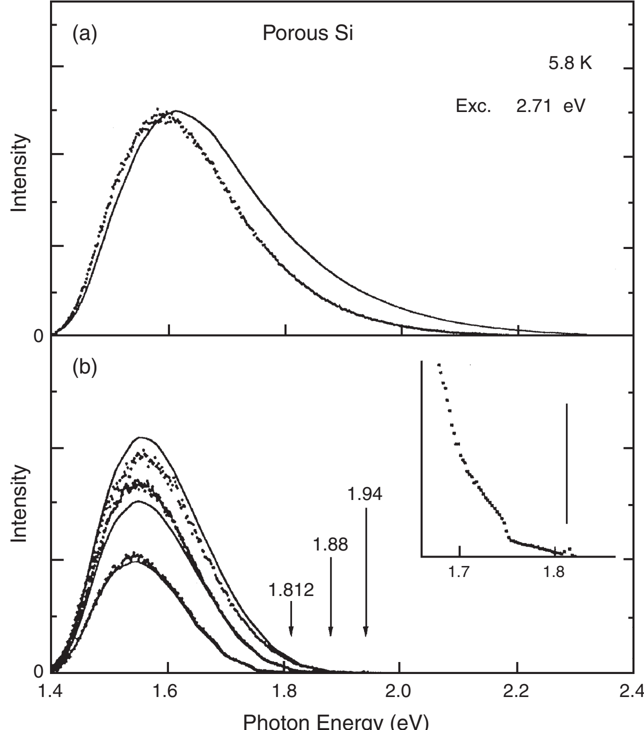


Figure 6. Photoluminescence spectra of porous Si taken at 5.8 K. The ordinate scale is proportional to the photon number per unit energy interval. (a) The dotted curves are the spectra excited by 2.71-eV line of an Ar laser. The solid curve is the distribution function $P(E)$ of the Si nanocrystal in the porous Si which has a bandgap of E deduced from this spectrum. (b) The dotted curves are the spectra excited at 1.812, 1.88, and 1.94 eV, and the solid curves are the best calculated fits to these spectra taking into account the multi-phonon processes. The inset shows the details of the spectrum with phonon structures near the laser line for 1.812-eV excitation. Reprinted with permission from [67] and [69], T. Suemoto et al., *Phys. Rev. Lett.* 70, 3659 (1993); *J. Phys. Soc. Jpn.* 63, suppl. B, 190 (1994). © 1993, 1994, American Physical Society.

the Si nanocrystal, with the bandgap increasing due to quantum confinement. This suggests that the optical centers in the porous Si retain the indirect bandgap characteristics of bulk crystalline-Si.

Figure 6 shows the results of Suemoto et al. [67]. Figure 6a shows the photoluminescence spectra of porous Si excited by an Ar laser at 2.71 eV; this corresponds to "nonselective" excitation and shows a broad peak at around 1.6 eV. Figure 6b shows the spectra under a selective excitation condition, where the excitation energy is within the luminescence band. Several notable features can be seen in Figure 6b: (1) with decreasing excitation photon energy, the intensity decreases substantially and the peak shifts to a lower energy, indicating the selection of luminescent species that emit light at energies lower than the excitation energy; (2) the line-narrowed resonant component is absent, suggesting that fast relaxation of the photoexcited state to a lower energy state occurs before light is emitted; (3) a steplike structure is visible, most clearly at 1.81 eV excitation (the spectrum near the laser line is shown in detail in the inset). Similar steplike structures were observed by Calcott et al. [71]. In regard to (2), they found that the fast relaxation is due to exchange splitting of excitons in porous Si based on the quantum confinement model. They observed that the onset of the luminescence does not occur exactly on the laser line in site-selective measurements, but shifts to a lower energy (the value varies from 23 meV under 2.410-eV excitation to 5 meV around 1.78 eV). They ascribed this energy shift to exchange splitting. In light absorption, the transition occurs mainly in the higher spin-allowed singlet state, but the fast exciton relaxation occurs from the singlet state to the lower triplet state at low temperatures. The emission occurred from the lower triplet state.

However, the steplike structure of the luminescence spectra in selective excitation spectroscopy can be basically interpreted by suggesting that multi-phonon-assisted indirect transitions in the Si nanocrystals occur with an increase in the bandgap of Si due to the quantum confinement of the electron and hole. The positions of the steplike features correspond quite closely to the energies of the momentum-conserving phonons in bulk crystalline Si [73]. The solid curves in Figure 6b are the best calculated fits assuming that the multi-phonon process involves up to two phonons. The calculated curves reproduce the experimental photoluminescence spectra well for all excitation energies (1.81, 1.88, and 1.94 eV). The calculated results reproduce not only the excitation energy dependence but also the temperature dependence of the experimental photoluminescence spectra [67, 69].

There have been a number of reports on the decay dynamics of luminescence at various time scales ranging from picoseconds to milliseconds [72, 74, 75]. Also, the

decay behavior is complicated by sample dependence. Therefore, there was no phenomenological model to provide consistent understanding of all the aspects of the reported decay dynamic until Suemoto et al. proposed a model based on the recombination of electron-hole pairs confined in Si nanocrystals in porous Si with radiative and nonradiative processes [69, 72]. In regard to the nonradiative process, the model takes into account both the tunneling and thermally activated escape processes of an electron (and/or hole) through a barrier. The model also explains the observed temperature dependence well along with the lifetime and nonexponential decay profiles of the photoluminescence [69, 72].

For a single Si nanocrystal, the luminescence intensity at time t after a pulsed excitation is given by

$$I(t) = C_o R_r \exp(-(R_r + R_n)t)$$

Here, R_r and R_n are radiative and nonradiative decay rates, respectively, and C_o is a constant. From the equation, $I(0) = C_o R_r$ is the initial amplitude of the luminescence and it is proportional to R_r, whether or not the nonradiative process exists. Suemoto et al. extracted the form of variations in the radiative decay rate with temperature from the measurement of this initial luminescence amplitude. They found that the decay rate obtained from the near-exponential gradient of the luminescence decay at long times agrees well with that obtained from the initial amplitude above. They concluded that the decay rate at long times represents radiatively dominated decay and that the decrease in the decay rate below the temperature where maximum luminescence intensity occurs is due to variation in the radiative decay rate. The activation energy obtained from the temperature dependence of R_r can be easily understood by considering the exchange splitting of excitons in porous Si as described previously.

In regard to the nonradiative process of electrons, the tunneling and thermally activated processes are shown in Figure 7. Here, E is the kinetic energy of an electron, and E_a is the barrier height from the lowest energy level of the confined electron. The total nonradiative recombination rate R_n becomes

$$R_n = f_0 p + f \exp(-E_a/k_B T)$$

where p is the tunneling probability of an electron through the barrier, and f_0 and f are, respectively, the attempt frequencies for the tunneling and thermally activated processes. Assuming a Gaussian distribution of the barrier, the calculated decay curves using this equation with appropriate fitting parameters show good agreement with the experimental results for every temperature, as shown in Figure 8. From the above analysis, it seems that thermal activation is dominant in carrier escape as the temperature increases above the maximum luminescence intensity, and tunneling becomes significant at very low temperatures with a low rate of radiative decay. The mean barrier height, E_a, obtained is 0.259 eV at a luminescence energy of 1.459 eV, which is one order of magnitude lower than the band offset between Si and SiO_2. They considered two possible reasons for this: (1) the existence of some material that differed from the simple oxide (a compound of Si, O, and H) on the surface of the Si nanocrystal; or (2) undulation in the width of the quantum wire. As the explanation mentioned above, the temperature dependence of the luminescence, and the lifetime and nonexponential decay profiles, can be successfully interpreted in terms of the quantum confinement model for porous Si that is not heavily passivated by oxygen. For heavily oxygen-passivated porous Si, coupling of excitons and local vibration at the surface have been suggested by several groups [76, 77].

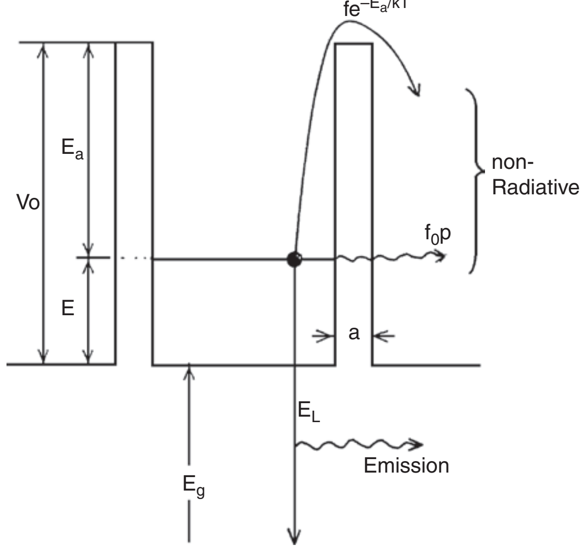


Figure 7. Decay dynamics for Si nanocrystals in porous Si. Reprinted with permission from [69], T. Suemoto et al., *J. Phys. Soc. Jpn.* 63, suppl. B, 190 (1994). © 1994, Physical Society of Japan.

3.3. Applications

Optoelectronic integrated circuits (OEICs) could have a significant impact on display and optical transmission in electronic data processing systems. OEICs with optical interconnections have already been realized in computer

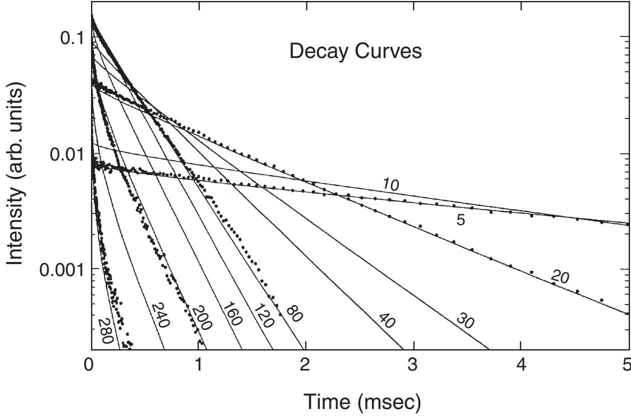


Figure 8. Typical observed decay curves of porous Si (dotted curves) taken at 5, 19, 80, 208, and 271 K. The solid curves are the calculated decay profiles. Reprinted with permission from [72], T. Suemoto et al., *Phys. Rev. B* 49, 11005 (1994). © 1994, American Physical Society.

networks and between computer boards. Many studies are now underway on optical signal transmission both between LSI chips on a board and within individual LSI chips.

An important factor in producing OEICs will be the ability to easily and inexpensively integrate optical and electronic devices on Si wafers. In this respect, the most satisfactory solution would be to create OEICs entirely from Si-based materials. Porous Si and Si nanocrystals have obvious potential in these light-emitting devices since they are forms of Si and emit intense visible photoluminescence. Preliminary versions of OEICs (including optical detectors and waveguides) have been already realized using Si-based materials. However, the major deficiency in current Si-based optoelectronic devices is the lack of suitable light-emitting devices such as light-emitting diodes (LEDs) and semiconductor lasers on Si wafers. At present, these light-emitting semiconductor devices are fabricated from direct-bandgap compound semiconductors. However, direct integration of compound semiconductor devices on a Si wafer has proved to be both complicated and expensive.

In recent years, the optical properties of Er-doped Si have generated intense interest because of the possibility of integrating electronic and optical devices using mature Si technology. Er emission at 1.54 μm, which arises from an internal 4f transition ($^4I_{13/2} \rightarrow {}^4I_{15/2}$) of Er^{3+} ions, is of particular interest because it overlaps the minimum loss region of the optical fibers employed in optical communication.

Recently, to improve the luminescence efficiency of Er-doped bulk Si, Er has been introduced into porous Si through ion implantation [78–81], electrolysis [82], or spin-on techniques [83]. The major advantages of porous Si as a host for Er ions are (1) the strong Er luminescence at 1.54 μm, (2) the weak temperature dependence of the Er luminescence, and (3) the long lifetime of the upper level $^4I_{13/2}$, which makes population inversion possible. The strong Er luminescence at 1.54 μm of Er-doped porous Si arises from the fact that the Er^{3+} ions are excited due to the energy transfer that results from the recombination of excess carriers confined in Si nanoscale grains in the porous Si near the incorporated Er. The quantum confinement of the carriers in Si nanoscale grains appreciably widens the Si bandgap and reduces the thermal quenching of Er^{3+} luminescence at 1.54 μm. The long lifetime of the upper level $^4I_{13/2}$ makes Er-doped porous Si very promising for a wide range of devices such as light-emitting diodes, optical amplifiers, and lasers.

Er-doped silica glass is already used in optical telecommunication technology as an amplification medium. Optical amplification at the 1.54-μm wavelength can be achieved if sufficient Er can be brought into the first excited state of $^4I_{13/2}$. Recently, Fujii et al. reported that the presence of Si nanocrystals in Er-doped SiO_2 considerably enhances the 1.54-μm luminescence of Er^{3+} [84]. Following this observation, a number of reports showed similar results for samples having Si nanocrystals and Er ions prepared using various means of Er doping including co-sputtering [85] and ion implantation [86–88]. All of these observations strongly suggest energy transfer from the Si nanocrystals to the Er. The currently accepted excitation model for Er^{3+} is as follows. First, a photon is absorbed by the Si nanocrystal, which results in the generation of an exciton inside the nanocrystal. This exciton can recombine radiatively, emitting a photon with an energy that depends on the size of the nanocrystal. If an Er ion is present close to the nanocrystal, the exciton can recombine nonradiatively by bringing Er into one of its excited states.

As described previously, the high luminescence efficiency of porous Si under optical excitation is quite remarkable. However, for its practical application as a light emitter, it will be necessary to obtain similarly high efficiency under electrical injection of carriers. Not surprisingly, attempts to observe electroluminescence (EL) began as soon as the high photoluminescence efficiency of porous Si had been reported [89, 90]. Initial results for devices with solid-state contacts were not very encouraging, while high EL efficiency was reported for liquid [91–93] or polymer contacts [94, 95]. Often, the reported quantum efficiency lay in the 10^{-6}–10^{-8} range for solid-state contacts in the initial stage [90]. A further complication hampering the use of EL devices is their poor stability. EL devices of porous Si with solid-state contacts typically fall into two categories: (1) Schottky type structures in which a metal contact deposited on the top provides a junction to the porous Si/bulk-Si structure [89, 90, 96, 97]. These devices often show limited rectification [89, 97]. (2) Vertical PN and PIN diode structures, on the other hand, yield EL only under forward bias [98–101]. This suggests the injection of both types of carriers. Several groups have recently reported the high quantum efficiency of such structures. Steiner et al. demonstrated a quantum efficiency of $\sim 10^{-4}$ in 1993 [100] and Linnros and Lalic demonstrated an even higher quantum efficiency of 2×10^{-3} in 1995 [101]. To improve the stability of EL devices, Tsybeskov et al. carried out light oxidation of porous Si and fabricated a stable PIN EL diode in 1996 [102]. They demonstrated a high quantum efficiency over 10^{-3}. EL was observed only in forward bias and at a voltage as low as 2 V. In addition, there was no degradation of EL during one month of pulsed operation. These results demonstrate fulfilment of many of the conditions required to make EL devices compatible with Si electronics. In 1996, Hirschman et al. demonstrated test structures that integrated Tsybeskov's EL devices (described above) with electronic devices [103]. There was also an attempt to fabricate EL devices consisting of Si nanocrystals by Toyama et al., who fabricated EL diodes showing visible EL emission with a peak energy of 1.8 eV in 1996 [104].

As mentioned above, porous Si, Si nanocrystals embedded in SiO_2, and Si nanocrystals in Er-doped SiO_2 are expected to play a key role in Si-based microphotonic application. However, there are still some obstacles to realizing this goal: (1) the need to further enhance the intensity of the luminescence; (2) the absence of directionality in the emission; and (3) the broad spectral shape of the luminescence (in the case of porous Si and Si nanocrystals). However, all of these problems can be overcome by microcavities. When a photon is confined in a microcavity in resonance with the emission of the active medium, the light emission becomes spectrally sharp and is strongly enhanced in the direction of confinement. Optical microcavities have been demonstrated for porous Si [105], for Si nanocrystals embedded in SiO_2, and for Si nanocrystals in Er-doped SiO_2 [106]. Pavesi et al. demonstrated porous Si microcavities based on distributed Bragg reflectors fabricated by alternating porous Si layers of

different porosities [105]. Iacona et al. fabricated Si nanocrystals embedded within a Si/SiO_2 microcavity [106]. They also realized Si nanocrystals in Er-doped SiO_2 sandwiched in Si/SiO_2 microcavities [106].

To realize laser operation, optical gain has to be achieved. Pavesi et al. demonstrated light amplification using Si nanocrystals dispersed in a SiO_2 matrix [107]. They realized a planar waveguide structure by Si nanocrystals formed by ion implantation into a SiO_2 matrix. A net optical gain was seen in both the waveguide and transmission configuration, with the material gain being of the same order as that of direct-bandgap quantum dots. They explained their observations using a model based on the population inversion of radiative states associated with the Si/SiO_2 interface.

4. ELECTRICAL PROPERTIES AND APPLICATIONS

4.1. Quantum Size Effects

From electrical measurement of the small Si dots sandwiched between tunneling barriers, it is difficult to assess the energy level due purely to the quantum size effect. This is because the contribution of the charging effect cannot be neglected in these small Si dots as quantum confinement energy is inversely proportional to the square of the dot size, while single-electron charging energy is inversely proportional only to dot size. Ye et al. investigated the discrete quantum levels of Si nanocrystals with SiO_2 barriers by observing their resonant tunneling characteristics at 77 K and at room temperature [108]. They fabricated diode structures with a layer consisting of laterally distributed Si nanocrystals. Compared with the theoretical results, they suggest it is necessary to recognize the charging effect to explain the large voltage separation between the resonant peaks.

Optical measurements are useful for assessing quantum energy levels only. Alsmeier et al. successfully obtained the quantum energy levels of Si dots using far-infrared transmission measurements [42]. They fabricated MOS devices with periodic arrays of electron dots on Si, which were widely tunable in diameter, employing electrostatic confinement with dual gates, as described in Section 2.2.

Several groups [109–111] reported the electrical measurements for diodes consisting of a SiO_2/Si nanocrystals/SiO_2 structure after Alsmeier's report; however, accurate separation of the effects of quantum size and single-electron charging has not yet been achieved.

The quantum size effect in a transistor structure is described in the section on single-electron transistors (Section 4.2.2) as it appears with single-electron charging effects in the transistor characteristics.

4.2. Single-Electron Effect

4.2.1. Single-Electron Charging Effect

Studies on the single-electron charging effect in solids originated in attempts to clarify the electrical conductivity of thin metal nanocrystal films. The importance of single-electron charging effects in the temperature and electrical field dependence of resistance in the plane of thin films composed of very small metal particles was first recognized in 1951 by Gorter [3]. In 1962, Neugebauer and Webb [112] confirmed these results experimentally. They measured electrical conduction in the film plane of ultrathin, evaporated metal films and obtained results that could be interpreted on the basis of the single-electron charging effect, that is, activation-type conductance, constant conductance in low electrical fields, and an increase in conductance in high electrical fields. In 1968, Giaever and Zeller [4] performed an interesting set of experiments designed to more clearly show the effect of single-electron charging on electron transport through extremely small metal particles in solids. They fabricated a diode structure of Sn particles with a diameter of about 20 nm and a double tunnel junction, and measured the electrical conductivity along the vertical direction to the plane at low temperature. They interpreted the high-resistance zero-bias anomalies by reasoning that the Coulomb-charging energies for the individual nanoparticles acted in such a way as to suppress electron tunneling to them. A detailed transport theory was later developed by Kulik and Shekhter in 1975 [113]. Much of our present understanding of the single-electron charging effect was developed in these early studies.

However, one drawback of these studies was inadequate control of the size and position of the nanocrystals. Consequently, the results obtained were the averages for the nanocrystals, leading to indirect proof of the existence of a single-electron charging effect. The lack of uniformity in their size and position also rendered these nanocrystals unsuitable for device application.

4.2.2. Single-Electron Transistor

Inversion-Type Si Single-Electron and Single-Hole Transistor Since the theoretical proposal of single-electron transistors (SETs) [6], there have been extensive experimental and theoretical studies on single-electron devices (SEDs). As previously mentioned (Section 1), SETs were initially fabricated using metal or compound semiconductors and measured at low temperatures to investigate the physics of the devices [7–9]. Reports of the use of Si in SEDs began appearing in the mid-1990s. The major reason for this timing was that the study of SETs was then more or less entering the practical phase of the research. From the point of view of down-scaling, the feature size of CMOS transistors would be comparable to that of SETs. Si-based SEDs have an advantage over those based on metal or compound semiconductors because of their compatibility with Si LSI technology. Another reason for the timing was that researchers were becoming aware of the potential to use the mature technologies used to fabricate Si LSI circuits (such as lithography, etching, and oxidation) in the fabrication of the extremely small structures required for high-temperature operation of SEDs. In addition, the ability to use fabrication technologies that required minimal changes to existing LSI device production facilities was a great advantage.

In the beginning, even for Si SETs, many of the fabricated Coulomb islands were rather large and single-electron effects were observed only at low temperatures. Coulomb oscillations and/or Coulomb gaps were observed for Si metal-oxide-semiconductor field-effect transistors

(MOSFETs) with a dual gate structure at 100 mK in 1994 [114], and at 4.2 K in 1995 [115] and 1996 [116]. In these devices, the lower gate formed a narrow inversion channel and the upper gate, which was made up of crossbars insulated by thick SiO_2, introduced potential barriers at constant intervals in a narrow channel. A Coulomb island was formed between two adjacent barriers in the channel. Another method of forming a Coulomb island is to use thin Si-on-insulator (SOI) materials since the thickness of the top Si layer of the SOI wafer can easily be reduced by oxidation and removal of the oxide. In 1994, SETs operating at up to 4 K were fabricated in thin SOI wafers using EB lithography and an isotropic dry-etching process [117]. Two EB lithography steps were used to form the Coulomb island. First, EB lithography and dry etching formed a narrow channel wire, and in the second stage, dry etching was applied for a short time to produce a narrow region in the channel for tunneling barriers. Also, using a thin poly-Si film on SiO_2 makes it easy to realize a size reduction in the vertical direction [118]. Using EB lithography and dry etching for patterning in the lateral direction, a Coulomb island was formed and operation of the SET was observed at 4.2 K (in 1995) [118].

From the point of view of its practical application, an important report was published by Takahashi et al. in 1994 [11]. They realized room-temperature operation of Si SETs using a pattern-dependent oxidation [PADOX] technique to achieve a very thin Si layer on SiO_2 [11, 119]. Conductance oscillations were observed at room temperature. Self-aligned processes usually have the drawback of insufficient controllability of the size and position of the structure formed. However, since the PADOX technique is a mixture of artificial and self-aligned fabrication methods, the size of the Si Coulomb island and its arrangement can be controlled to some extent. Using this PADOX technique or a modified PADOX technique (V-PADOX) [120], their group fabricated a Si SED that showed current switching at 30 K [121], two-input exclusive-OR gate function [122], and a CMOS-type inverter [123], for application such as in single-electron pumps and single-electron binary-decision-diagram (BDD) circuits. As well, a universal literal gate and a quantizer [124], which are basic components of multiple-value logic (MVL), were realized using the SETs fabricated by the above technique and MOSFETs fabricated on the same wafer. Since these MVL components are extremely compact, the number of elements can be greatly reduced by using merged SET-MOSFET devices.

After the report of room-temperature operation of a SET in 1994, various SETs operating at high temperatures were reported. By optimizing EB lithography and dry-etching techniques, Leobandung et al. [125] fabricated lithographically defined Si quantum dots with a diameter of about 20 nm. SETs and single-hole transistors (SHTs) that used the small island as a Coulomb island demonstrated quantum effects as well as Coulomb blockade effects at temperatures over 80 K. The SHTs [126] were fabricated using the same fabrication process except that the source, drain, and poly-Si gate were doped with a different type of species from those of the SETs.

Coulomb islands in the above SETs were fabricated (at least partly) with the help of dry etching. Thus, damage caused by dry etching had to be reduced by subsequent oxidation. In addition, the tunnel barriers between the Coulomb island and the source/drain regions were formed in the narrow Si wire where the quantum size effect raises the potential compared with that in the island. Oxidation after dry etching enabled the sizes of the Coulomb island and tunnel barriers to be further reduced. It should be noted that after dry etching, damage can be amended and the size of the Coulomb island and tunnel barriers reduced by using a wet-etching process [43, 127]. Nakajima et al. fabricated SETs showing clear Coulomb oscillation at 4 K using the wet-etching process [127]. In contrast, when Coulomb islands are formed using a deposition process, simple, damage-free fabrication is possible. In this case, the tunnel barrier usually consists of SiO_2 or air. Tunnel barriers made of SiO_2 have the advantage of greater barrier height, which is one of the conditions required for high-temperature operation of SETs [128, 129]. Choi et al. observed a Coulomb staircase and Coulomb oscillation at room temperature [34] using a conventional LPCVD method for Si nanocrystal formation, as described in the previous section [31, 32]. A few Si nanocrystals with a diameter of around 10 nm were positioned between ultrasmall metal pads with a gap <30 nm in the SiO_2 layer. However, with deposition methods, selecting only one Coulomb island and controlling its precise location present a difficult problem that has not yet been solved.

An interesting method has been proposed for realizing a high-temperature operating SET. A SET with a Coulomb island of a size beyond the limits of current state-of-the-art lithography was recently fabricated by Kim et al. [130, 131]. An amorphous Si side-wall was used as an etch mask for the formation of a SOI narrow channel wire. This method of side-wall patterning enables nanoscale patterning using a combination of conventional lithography and process technology. The SET has side-wall depletion gates on the narrow SOI channel, forming a small Coulomb island. The advantage of a side-wall structure is that it can implement a feature size below the limit of EB lithography. Periodic Coulomb oscillations have been observed up to 77 K and basic operation of a dynamic multifunctional SET logic circuit has been successfully demonstrated at 10 K by combining the fabricated SETs and a MOSFET.

Quantum Size Effects Quantum effects have more influence on the electrical characteristics of SETs and SHTs at higher operating temperatures since a smaller sized Coulomb island is required for stable operation at higher temperatures [125, 126, 132–136]. Roughly speaking, the quantum confinement energy is inversely proportional to the square of the dot size while the single-electron charging energy is inversely proportional to the dot size only. Therefore, the effect of quantum-level spacing increases and cannot be neglected as the island size decreases. To calculate the precise quantum-level spacing, it is necessary to know the confinement potential. However, precise potential profiles for the confinement have not yet been determined experimentally for SETs. Although experimentally obtained addition energy has been compared with calculated quantum confinement energy [125, 126, 132], the periodic drain current oscillations cannot be explained by assuming a simple potential well model for the confinement. However, more reliable values for quantum-level spacings were

obtained using the positions of the negative differential conductance and fine structures [134]. It has been suggested that in extremely small devices operating at room temperature it is very hard to predict the Coulomb oscillation peak positions due to these quantum confinement effects as well as the background charge since the energy spectrum fluctuates not only according to the size of the Si island but also according to the shape and orientation of the island [135]. Takahashi et al. proposed a method for adjusting the peak position after device fabrication by injecting charges into the Si nanocrystal floating gates [136].

Effect of High Doping Although various contingent phenomena, such as unintentionally introduced impurities or EB irregularities, can result in size reduction of the Coulomb island or tunnel barriers beyond the limits of state-of-the-art lithography, the controllability and reproducibility of the device fabrication method are extremely important in integrating the devices into circuits. Augke et al. suggested that highly doped SETs, nominally consisting of a single island and two Si tunneling barriers, act either as a SET with multiple Coulomb islands or as a SET with a single island corresponding to the geometric shape of the device, depending on the gate voltage applied [137]. The multiple islands are formed by the additional tunneling barriers introduced into the device as a result of fluctuations in the dopant concentration. These barriers can be removed by raising the Fermi level via the application of an appropriate gate voltage since the height of the barriers is usually less than that of a barrier defined by the geometrical shape. In another view, Sakamoto et al. [138] suggested that when the doping level is extremely high, fluctuations in the dopant concentration may be statistically very small, resulting in the electrical characteristics expected for a SET with a single island corresponding to its geometric shape.

System with Multiple Coulomb Islands Conversely, Coulomb islands can be self-assembled utilizing the dopant fluctuations discussed earlier. Frequently, Coulomb blockade phenomena have been observed in the conduction of highly doped narrow Si wires. Smith and Ahmed [139] observed the Coulomb blockade phenomena in highly doped (phosphorus) narrow Si wires (60–70 nm wide, 40 nm high, and 4–6 μm long) with a side-gate in a SOI configuration. They observed Coulomb oscillations with multiple periodicities as a function of the gate voltage and a minimum blockade size greater than zero. These phenomena are characteristic of single-electron tunneling in multiple-tunnel junctions of varying Coulomb island sizes, since the charging energy levels in each island in such a multiple island system can never be aligned with each other and the source-drain voltage at the same time. In this system, the gate coupling to the Coulomb island and the charging energies vary. There is speculation that the Coulomb islands and tunnel barriers are formed due to the dopant distribution. It is noted that in these high-doped narrow Si wires, equidistant peaks in the Coulomb oscillation have been reported. Tilke et al. [140] reported that 50-nm-wide Si wire fabricated on highly doped (As) SOI material shows equidistant peaks in the Coulomb oscillation at temperatures up to 100 K. They speculated that the pattern-dependent two-dimensional oxidation of laterally structured SOI wires is the reason for this. Another possible reason is the segregation effects of As dopants during thermal oxidation at the rough edge of the wire. They emphasize that the advantage of high doping is that it inhibits the separation of discrete energy levels, and that the charging energy for adding one electron to the SET is ideally not perturbed by the single-particle energy states.

Similar Coulomb blockade effects have been observed in the conduction of narrow poly-Si wires. The structural characteristics of the Si wires very much depend on the deposition method and conditions, and directly influence the electrical properties.

Ng et al. [141] examined the effect of wire length on the single-electron charging effect in ultrathin ($\sim$7 nm) wires of partly recrystallized poly-Si fabricated using electron-beam annealing of hydrogenated amorphous Si deposited by PECVD. In their work, long wires of micron-scale length and 50 nm in width were likely to exhibit Coulomb blockade effects at high temperatures of 70 K. Within the wires, Coulomb islands are speculated to be tapping sites in amorphous regions at grain boundaries. Another possibility is that a confinement due to potential fluctuations arising from dopant distribution forms Coulomb islands. In addition, the poly-Si grains may act as tunnel barriers. The larger Coulomb gaps observed with increasing wire length can be interpreted as quasi-one-dimensional multiple Coulomb island systems.

Solid-phase crystallized narrow poly-Si wires have exhibited single-electron charging effects. The wire was formed using thermal recrystallization of amorphous Si deposited by the PECVD method. After heavily doped (phosphorous) wire had been defined, thermal oxidation was performed, resulting in a minimum wire cross-section of 20×30 nm^2 [142]. In these wires, finer peaks are superimposed on the main peak in the Coulomb oscillation, suggesting the presence of a multiple tunnel junction along the poly-Si wires. The Coulomb islands are considered to be the poly-Si grains and the tunneling barriers are SiO_x created at the grain boundaries [143]. In the narrow wires, poly-Si grains form a one-dimensional chain of Coulomb islands, creating a multiple tunnel junction device.

Kawamura et al. [144] investigated narrow ($\geq$95 nm) and extremely thin ($\sim$7 nm) heavily phosphorous-doped poly-Si wires fabricated using LPCVD. The wire was fabricated in a two-step process: amorphous Si was deposited on the SiO_2 film at 520 °C by LPCVD and polycrystallized by annealing at 650 °C. The increase observed in the peak number of Coulomb oscillations with increased temperature (from 5 to 10 K) indicates that multiple Coulomb islands are connected in series in the narrow wire (120 nm wide). This is because, for electrical conduction to occur in a multiple Coulomb island system, the electrochemical potentials of each island must be at the same level within the limits imposed by thermal smearing. That is, kT determines the allowable mismatch between the electrochemical potentials of all of the islands. An increase in the peak number of Coulomb oscillations was also observed with increasing drain voltage in the narrow wire at 5 K, which indicates that the tunneling process occurs in the nonresonant inelastic tunneling mode. In nonresonant inelastic tunneling, an electron can tunnel through the multiple barriers between two adjacent islands nonresonantly by emitting

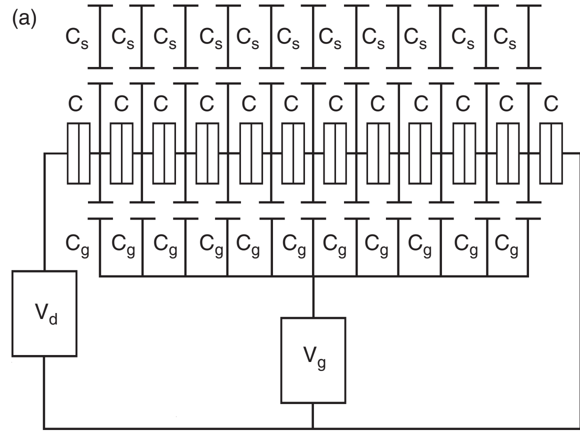
acoustic phonons. In the wires, the Coulomb islands are considered to be the small poly-Si grains and the tunneling barriers are the high potential barrier layer created at the grain boundaries due to the segregation of impurity atoms or the presence of a large number of trapping states in that region. The barrier height was estimated to be 26 meV from the activation energy of the conductance. The temperature dependence of the conductance within and outside of the Coulomb gap, suggested the generation of a charge soliton.

There are several advantages in using multiple tunnel junctions (MTJs). One is the ability to raise the operating temperature of the device in an MTJ system since the effective junction capacitance decreases due to the capacitances connected in series. This leads to a decrease in the total capacitance of the island inside the MTJ system and an increase in the charging energy of the island and a rise in the operating temperature. Another advantage of MTJs is the variety of their potential applications, such as to electron pumps and quantum dot cellular automata (QCA). To use such MTJ systems in SETs for practical purposes requires detailed investigation of the conduction mechanism.

Detailed descriptions of the fabrication procedure and measurements for SETs with one- and two-dimensional regular arrays of MTJs with metallic islands have been previously reported [8, 145, 146]. In such arrays, the presence (or absence) of one electron on an island induces the polarization of neighboring islands, described as a charge soliton (or antisoliton). However, for Si SETs with a MTJ system, most of the investigated arrays were naturally formed ones where junction size and position fluctuated, and they have not been evaluated precisely, as described earlier [34, 140, 144, 147]. Since the number of junctions is also unknown, the possibility of the contribution of solitons to conduction is suggested by Kawamura et al. [144], as mentioned above, only in these naturally grown systems.

To examine the conduction mechanism and the effect of solitons in detail requires the fabrication of a regular array of MTJs with numerous islands, in which the concept of solitons is important. Nakajima et al. [148] fabricated a Si SET with a one-dimensional regular array of MTJs with 11 islands for high-temperature operation and investigated the conduction mechanism in relation to solitons. A one-dimensional regular array of MTJs consisting of 11 islands was fabricated using EB direct writing in a SOI layer (Fig. 9). The tunnel barriers were formed in constrictions of silicon in between the islands due to the quantum size effect. The top layer of silicon and the buried oxide on the SOI wafers were 20 and 400 nm thick, respectively. Doping of the top silicon layer was carried out using $POCl_3$ diffusion (2×10^{20} cm^{-3}). The fabrication process included EB lithography and dry etching using an electron cyclotron resonance (ECR) etcher with the resist pattern as a mask. Subsequent isotropic wet etching in a $NH_4OH/H_2O_2/H_2O$ solution [43] reduced the dimensions of the device and amended damage that occurred during the dry-etching process. The widths of the barrier and island region were about 40 and 60 nm, respectively, as seen in Figure 9b.

For application to a QCA, Single et al. investigated a SET that included double islands with a regular array of junctions



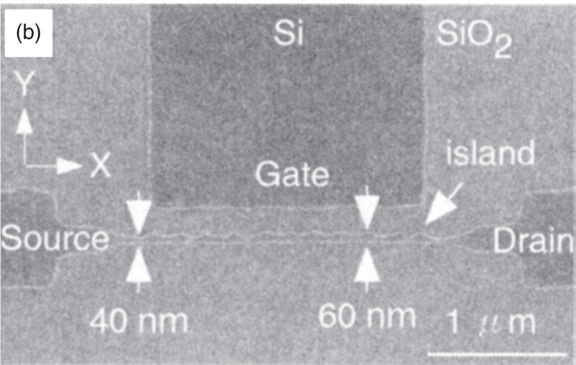


Figure 9. (a) Equivalent circuit of fabricated single-electron transistor (SET) with multiple tunnel junctions (MTJs) consisting of 11 islands. C_s, C_g, and C represent the substrate capacitance, gate capacitance, and tunnel junction capacitance of a Coulomb island, respectively. Parasitic capacitance C_0 of the island consists of C_s and C_g. (b) Plan-view scanning electron microscopy (SEM) micrograph of the SET. Reprinted with permission from [148], A. Nakajima et al., *Appl. Phys. Lett.* 81, 733 (2002). © 2002, American Institute of Physics.

connected in series using EB lithography [149]. They confirmed the charge distribution between the dots on the system could be changed by the amount of one electron by measuring the drain current as a function of the two side-gates for each dot at 4.2 K.

4.2.3. Single-Electron Memory

Yano et al. reported the first room-temperature operation of a Si single-electron memory (SEM) in 1993 (Fig. 10) [10]. The SEM was based on a similar operating principle to that of a flash memory, except that the single-electron charging effect was utilized in the write/erase process [10, 150]. These reports were of great significance because they were the first to point out the potential application of single-electron devices to ultra-LSIs (ULSIs). The SEM uses very thin Si film (only about 3 nm thick), in which nanoscale grains form nanoscale Coulomb islands and a series of somewhat larger nanoscale grains forms the channel. Since there is a high statistical probability that one or more Coulomb islands are in the vicinity of the channel, the trapped charge in the Coulomb islands modulates the threshold voltage of the transistor section of the SEM. This concept overcomes

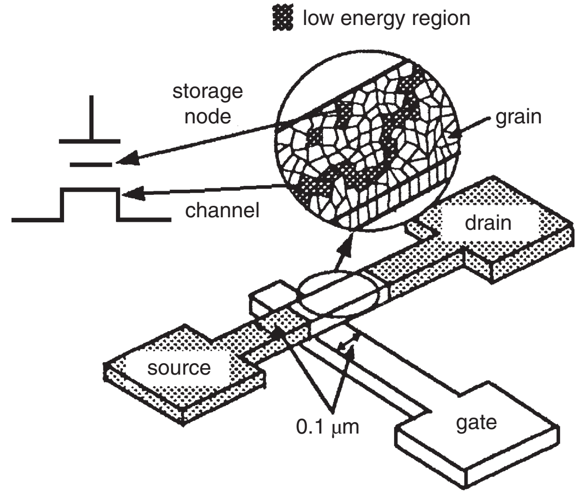


Figure 10. A schematic device structure of Si single-electron memory fabricated by Yano et al. Reprinted with permission from [10], K. Yano et al., "IEEE International Electron Devices Meeting," 1993, p. 541. © 1993, Institute of Electrical and Electronics Engineers, Inc.

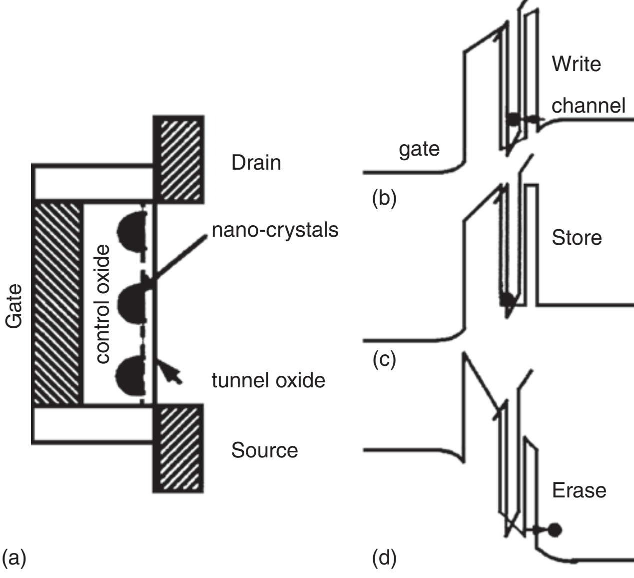


Figure 11. A schematic cross section of Si single-electron memory fabricated by Tiwari et al. (a), and band diagram during injection (b), storage (c), and removal (d) of an electron from a nanocrystal. Reprinted with permission from [153], S. Tiwari et al., *Appl. Phys. Lett.* 68, 1377 (1996). © 1996, American Institute of Physics.

the main difficulty for SEMs in reading out small charges trapped in Coulomb islands at room temperature.

The most difficult problem in SEMs with naturally formed structures is the suppression of cell-to-cell scattering of the threshold voltage, which is essential for an acceptable error-rate memory system. To solve this problem, Ishii et al. introduced the concept of a verify process [151], which is described in detail later. A prototype of an ultralarge integrated SEM circuit (128 Mb) was demonstrated [152], in which the verify process was applied after both the writing and erasing processes, and a double-stacked memory cell structure was used. Utilizing the concept proposed by Yano et al., a Si single-electron memory with a structure more similar to that of flash memory was reported in 1995 by Tiwari et al. [30]. This device has a conventional MOSFET with Si nanocrystals acting as floating gates that are separated from the channel by a thin tunnel SiO_2 (Fig. 11) [30, 153–155]. In this device, single-electron transfer through the oxide was realized between the wide channel region and the Si nanocrystals by direct tunneling (DT) at 77 K. This SEM has several promising features over those of conventional flash memories, although the improvement of the SEM, such as the size of the Si nanocrystals, their uniformity, and density, is necessary for application to ULSIs. The advantages are as follows: (1) simple, low-cost device fabrication due to none of the complications associated with the dual poly-Si process used in a conventional flash memory; (2) better retention owing to the suppression of leakage currents between the nanocrystals and channel due to Coulomb blockade and quantum confinement effects [154]; and (3) again, better retention owing to the suppression of the leakage current between the nanocrystals and channel due to lack of conduction between the nanocrystals [153]. These improvements in retention time enable thinner tunneling oxides and lower operating voltages.

The next key research target for the nanocrystal memory described above is to further improve the threshold voltage characteristics and retention time. Refining the physical properties of the Si nanocrystals, such as crystal size and size distribution, and nanocrystal densities in the tunneling dielectrics, will directly contribute to this improvement. For this purpose, various tunneling dielectrics have been studied because the physical properties of the underlying layer largely influence the properties of the deposited Si nanocrystals. Kim et al. proposed ultrathin LPCVD Si_3N_4/SiO_2 double layers as tunneling dielectrics for use in a DT regime [156]. Because LPCVD Si_3N_4 has a rougher surface than thermal SiO_2, smaller and more uniform Si nanocrystals, at a higher density, can be deposited on LPCVD Si_3N_4 compared with those obtained on gate SiO_2. As a result of these features, clearer single-electron tunneling effects in the drain current versus gate voltage characteristics were observed for SEMs with double-layer tunneling dielectrics than for those with SiO_2 tunneling dielectrics at room temperature. In addition, Fernandes et al. proposed using atomic-layer-deposited (ALD) Al_2O_3/SiO_2 double layers as ultrathin tunneling dielectrics in a DT regime [157]. Better control of Si nanocrystal density against deposition conditions (SiH_4 pressure) was obtained for the deposition of nanocrystals on Al_2O_3 than on Si_3N_4. Better charge retention was also observed for double-layer tunneling dielectrics than for SiO_2 tunneling dielectrics.

To improve the physical properties of Si nanocrystals, a new formation process was introduced to SEM fabrication. De Blauwe et al. adopted a three-step aerosol process for Si nanocrystal fabrication and fabricated a floating-gate FET that acts in a Fowler–Nordheim (FN) tunneling regime [158]. The three-step process separates formation of Si nanocrystals, their surface oxidation, and their deposition onto the tunneling dielectrics.

The above reports of using SEM in a DT regime are all for *n*-channel nanocrystal memory. In contrast, the feasibility of *p*-channel nanocrystal memory was demonstrated

by Han et al. [159]. They observed V_{th} shifts due to hole tunneling in the Si nanocrystals through 4-nm-thick SiO_2 tunneling dielectrics. Since the barrier height of SiO_2 tunneling dielectrics for holes (about 4.7 eV) is larger than that for electrons (3.1 eV), the FN tunneling component due to holes is much smaller than that of electrons, leading to difficulties in using *p*-channel EEPROM devices in a FN regime. However, in a DT regime, the difference in gate tunneling current is less between *n*-channel MOSFETs and *p*-channel MOSFETs than in a FN regime. This suggests the possibility of using flash-type *p*-channel devices in a DT regime. They also observed better retention characteristics for holes than for electrons due to the higher barrier height of SiO_2 for holes than for electrons.

Interestingly, Ishii et al. proposed a new concept for utilizing a SEM with Si nanocrystals to replace a conventional nonvolatile memory with floating gates [160]. The devices they proposed are used in a FN regime with improvements in write/erase and retention characteristics. As charges are stored separately in multiple nanometer-size dots, only a limited quantity of nanocrystals are affected by weak spots in the surrounding insulator. This suppresses the generation of cells with an abnormally short retention time. Average write/erase operation between the nanocrystals sharpens the characteristic distribution of the device.

To attain long charge-retention characteristics without losing the high-speed write/erase characteristics of a DT regime, Ohba et al. proposed and fabricated a floating gate structure consisting of self-aligned double-stacked Si dots (Fig. 12) [161]. They showed that a SEM with double-stacked Si dots exhibited more than 10^2 times longer retention than a conventional SEM with a single-layer Si nanocrystal floating gate. This improvement in retention was attributed to quantum confinement and a Coulomb blockade in the lower Si dot. The trade-off between retention and write/erase speed is resolved by applying a high gate voltage for write/erase processes. Applying a high voltage enables a high write/erase speed. Since charge retention is carried out in the low gate voltage region, it is possible to obtain a long retention time without losing high-speed write/erase capability.

It should be noted that there have been attempts to use nanocrystals other than Si nanocrystals as floating dots in SEMs. Hanafi et al. fabricated a SEM with Ge nanocrystals

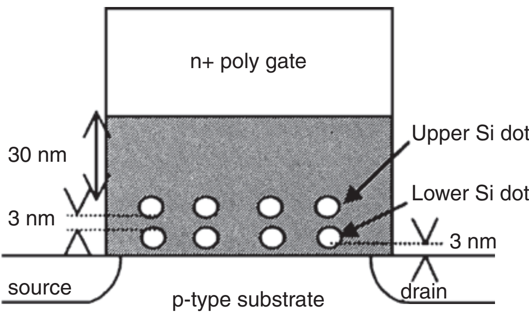


Figure 12. A schematic diagram of Si self-aligned double-dot memory fabricated by Ohba et al. Reprinted with permission from [161], R. Ohba et al., "IEEE International Electron Devices Meeting," 2000, p. 313. © 2000, Institute of Electrical and Electronics Engineers, Inc.

formed using a Ge ion implantation technique [155]. A high dose of Ge was ion implanted into thermal SiO_2 followed by annealing to introduce Ge nanocrystals into the oxide as floating gates. The drawback of this technique is the broader spatial distribution of the nanocrystals in the SiO_2, although it has the advantage that some of the nanocrystals are only a short oxide thickness away from the channel. King et al. developed a novel technique for fabricating Ge nanocrystals as floating gates that permitted more precise position controllability and uniformity of depth direction in SiO_2 [162]. The technique, which is based on thermal oxidation in combination with Ge segregation phenomena, provides excellent control of thickness for the top and bottom oxide layers that sandwich the nanocrystals. Taking another approach, Nakajima et al. developed a simple technique for fabricating metal nanocrystals as floating gates, which enables excellent uniformity of size and position, especially in depth direction [163–168]. The advantage of using metal nanocrystals as floating gates instead of semiconductor ones is the identical addition energy they provide because the contribution of the quantum confinement energy is negligible. Sn nanocrystals have been fabricated in thin thermally grown SiO_2 layers using low-energy ion implantation followed by thermal annealing [163–165]. The average diameter of the resulting Sn nanocrystals is 4.2 nm with a standard deviation of 1.0 nm. Sn nanocrystals preferentially reside in the SiO_2 layer at about 2 nm from the SiO_2/Si interface. This stable depth and excellent size uniformity are probably related to the transition region near the SiO_2/Si interface where compressive strain due to density mismatch occurs.

There may be an optimum nanocrystal depth and size that minimizes strain energy. Sb nanocrystals have also been fabricated in thin thermally grown SiO_2 layers using the same technique of low-energy ion implantation followed by thermal annealing [166–168]. Strictly speaking, the proposed method, which uses Sn in ion implantation, is not completely compatible with LSI fabrication since Sn is not typically used in conventional LSI device fabrication. However, Sb is one of the typical doping species used in conventional LSI fabrication. In contrast to the formation of Sn nanocrystals described above, when Sb nanocrystals are formed in a thin thermally grown SiO_2 layer, not only the effect of the compressive strain near the SiO_2/Si interface, but also the narrow as-implanted profile resulting from low-energy ion implantation contribute to uniform size and depth. When a narrow as-implanted profile is achieved with low-energy ion implantation, it results in an almost constant ion concentration in only a narrow region. If this narrow ion distribution can be adequately maintained during thermal annealing due to the slow diffusion of Sb in the SiO_2 region, the narrow profile will result in very uniform depth and size.

However, these devices used naturally grown Si nanocrystals as floating dot gates [10, 30] or even channels [10], leading to difficulties in controlling size, number, and position. This, in turn, led to poor uniformity and difficulties in analyzing the electrical characteristics. Hence, there was an urgent need for technology that could fabricate nanoscale features with precise control over size, number, and position. Thus, the next step in improving these devices was to artificially fabricate nanoscale floating dots and channels using

EB lithography. Consequently, in 1996 [169], a Si single-electron memory that could operate at room temperature was reported by Nakajima et al. with a nanoscale floating dot gate artificially fabricated using EB lithography and subsequent etching as described in Section 2.2. In this memory device [127, 169–171], a new technique was developed using a self-aligning process whereby a nanoscale floating dot gate was stacked exactly above a channel of nanoscale width (Fig. 13). This device thus allowed precise control over the size and the position of the floating dot gate, with the result that quantized threshold voltage (V_{th}) shifts due to single-electron transfer between the floating dot and channel were observed at room temperature (Fig. 14). Similar devices were reported around the same time [172].

Interestingly, the channel region beneath the floating dot is slightly wider than other channel regions in this device due to the lateral etching property of dry etching [127, 171]. This slightly wider channel region has been found to act as a Coulomb island at low temperatures. Consequently, this device can realize single-electron transfer not only between the floating dot and channel but also along the channel. If the Coulomb oscillation due to the single-electron transfer along the channel is combined with the quantized V_{th} shifts due to single-electron transfer between the floating dot and the channel, the possibility arises of producing a memory with ultralow power consumption.

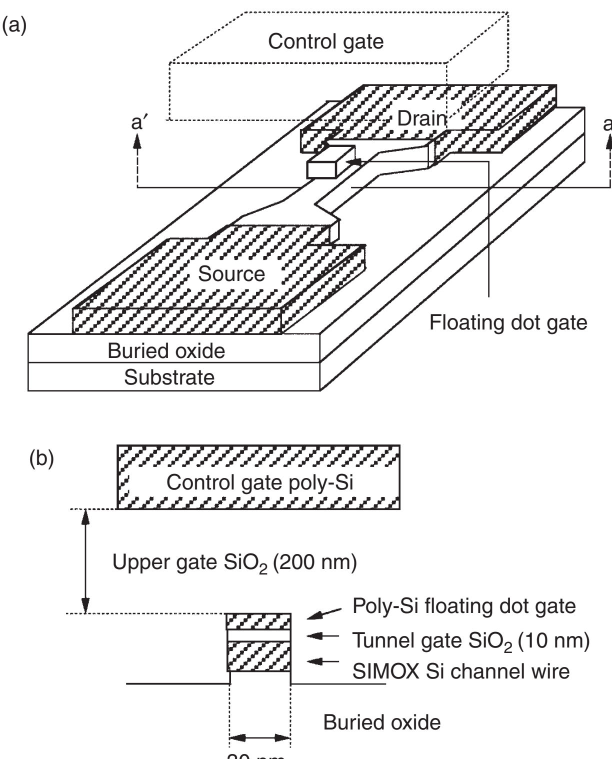
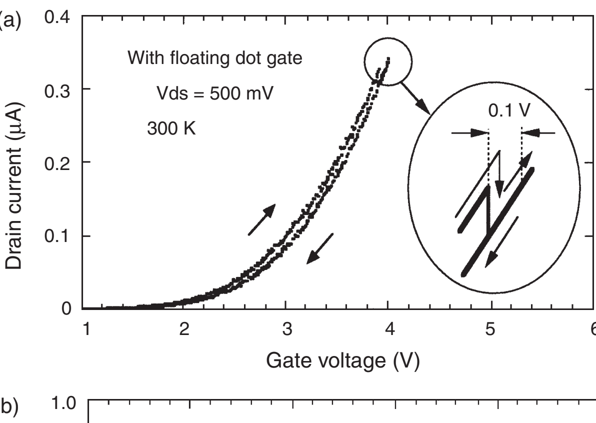
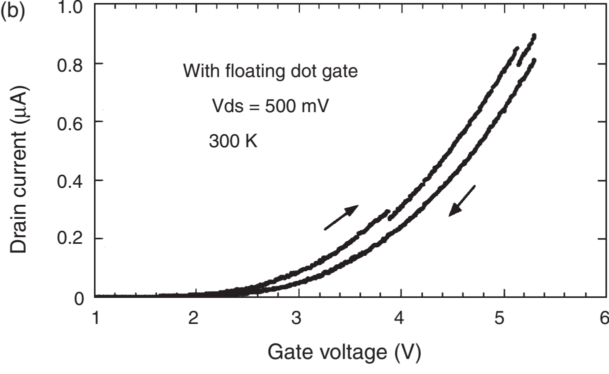


Figure 13. (a) A schematic diagram of Si single-electron memory with a self-aligned nanoscale floating dot gate fabricated by Nakajima et al. The shaded source and drain regions are heavily implanted with As^{+}. (b) Cross-sectional view along the a-a′ line. Reprinted with permission from [170], A. Nakajima et al., *Appl. Phys. Lett.* 70, 1742 (1997). © 1997, American Institute of Physics.





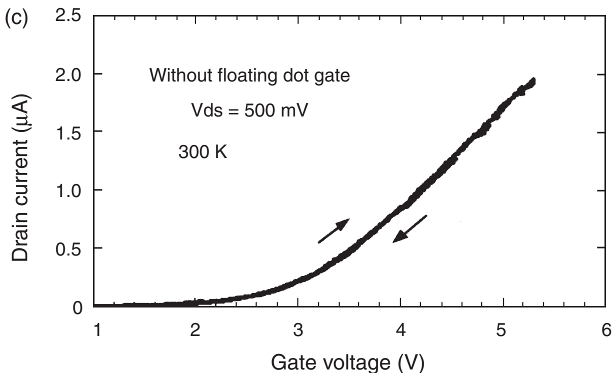


Figure 14. The drain current (I_d) vs. control gate voltage (V_g) characteristics at room temperature for a FET with a floating dot gate (a) and (b) and for a FET with no floating dot gate (c). Reprinted with permission from [170], A. Nakajima et al., *Appl. Phys. Lett.* 70, 1742 (1997). © 1997, American Institute of Physics.

It should be pointed out that the results of simulations have been published for the electrical characteristics of single-electron memories with a nanoscale floating dot gate [173, 174], and that the results obtained by Amakawa et al. provide a good explanation for the experimental results described in this section [174].

5. CONCLUSIONS

When the device size of MOSFETs decreases, the decrease in the mean number of carriers results in an increase in the thermal fluctuations of the carrier number relative to the mean value. This fluctuation in carrier numbers lowers

the reproducibility of the electrical characteristics for individual MOSFETs and between MOSFETs. Single-electron devices, which employ the Coulomb blockade effect, have the advantage of suppressing this fluctuation in carrier numbers [175].

To be considered as candidates for use in future LSI circuits, single-electron devices must be highly integrated and capable of operating at room temperature. It is therefore necessary to examine the reliability of the system at room temperature when single-electron devices are highly integrated. However, only a few reports have been published on this issue.

Two groups have anticipated the reliability of integrated single-electron memories at room temperature [175, 176]. Yano et al. [175] reported experimentally and theoretically on the time dependence of the probability distribution of the numbers of stored electrons for a single-electron memory with a floating dot gate. They showed that the probability distribution was clearly affected by the Coulomb blockade effect: the probability of a second-electron injection into the floating dot is much lower than the value expected for a Poisson distribution, which appears in the electron-tunneling process when the Coulomb blockade effect is absent. Due to the suppression of the second-electron injection into the floating dot gate (which was 10 nm in size), the probability of storing two electrons less than the average number of stored electrons is lower than the probability required by each memory cell for the operation of a 1-gigabit (Gb) memory chip with a 1000-fit error-rate. From this, they suggested that the average number of electrons required to achieve reliable operation is at least five, assuming that a 20% electron-count deviation is tolerable for each memory cell.

Usuki et al. [176] analyzed theoretically the basic trade-off between write errors and the number of stored electrons in a single-electron memory with a nanoscale floating dot gate at room temperature. They calculated the partition function from the charging energy of the floating dot, and then evaluated the mean number of stored electrons in the dot and the write error rate at room temperature. The write error rate was defined as the probability that the number of stored electrons is less than half the mean number of stored electrons. For a floating dot with a diameter of 10 nm, it was found that more than seven electrons are necessary at room temperature to maintain an error rate of 10^{-15}, which is the rate required by each memory cell to ensure reliable writing for a 1-Gb memory. The energy consumption of the memory cell was also estimated and found to remain low, even though each cell uses more than seven stored electrons. These results indicate that a memory with a small number n of stored electrons ($n > 7$) exhibits no disadvantages relative to single-electron memories with one stored electron.

These reports suggest that it is difficult to employ an integrated system when the write operation of the single-electron memory uses only a single-electron injection into a floating dot bigger than 10 nm. A simple calculation of the thermal fluctuation in the number of stored electrons (neglecting the discreteness of the quantum level in the floating dot) [176] shows that a floating dot with a diameter of less than 1 nm is required to obtain reliable write operation in an integrated system with only one stored electron in each memory. Since it is extremely difficult to reproducibly and uniformly fabricate a floating dot with a size of 1 nm with current fabrication technology, the development of single-electron memories that use a small number of stored electrons is probably a suitable objective for the next stage of research.

Another difficult problem encountered in the search for a memory that uses a small number of electrons is threshold voltage scattering due to background charges. To resolve this, Ishii et al., [151] proposed a new concept of applying the verify process, which has already been used in the operation of flash memory, to a single-electron memory. This process reduces not only the threshold voltage scattering due to stochastic variation and thermal fluctuation of the stored charge in the floating dot gate but also the scattering due to background charge. Threshold voltage and write time scattering between devices caused by the size and the positional fluctuation of nanoscale structures will also be suppressed by this verify process.

To produce memories that use few electrons, it is also necessary to investigate read and retention errors, since these also affect actual memories. It is equally necessary to continue efforts to improve the fluctuation in size and the productivity of nanoscale structures, and to control the distribution of background charge.

Unfortunately, the prospects for integration of the logic circuits of single-electron transistors (SETs) are not as bright as those for the memory devices described above. Kuwamura et al. [177] simulated the operation of complementary-SET (CSET) inverter circuits and pointed out that reliable operation at room temperature requires the total capacitance of the Coulomb island for each SET to be less than 0.16 aF (which corresponds to the self-capacitance of a Si dot with a diameter of around 1 nm in SiO_2) when CSET inverter circuits are highly integrated (a system of 10^{10} CSET inverters with an effective operating rate of 1% operates without error during a 10-year period with a clock rate of 100 MHz). When Shimano et al. [178] carried out simulations to test the reliability of CSET inverter circuits, they found that unless a 0.1-nm square tunnel junction of the Coulomb island is fabricated for each SET, at room temperature, it seems to be impossible for the CSET inverter circuit to meet the system specifications for reliability required by conventional Si ULSI systems (the system bit error rate is 10^{-10} during a 10-year operation period with a gate number of 100 kgates and a clock rate of 100 MHz).

Although several other logic circuits with SETs have been proposed [179, 180], no evaluation of their reliability in a highly integrated system have been performed at room temperature. The situation is further complicated by the fact that changes in the peak and valley positions of the Coulomb oscillation due to background charges increase the error rate of logic circuits with SETs [181]. In contrast to the case for memory (the verify process cannot be used for logic circuits), at present there are no reasonable solutions to the problem of background charge. Therefore, it is extremely important to clarify whether reliability can be maintained at room temperature when the logic circuits of SETs are highly integrated.

GLOSSARY

Addition energy Energy required to add or remove an electron to/from a dot. This energy generally includes both the contributions of charging energy and of quantum level spacing.

Charge soliton (antisoliton) In one- or two-dimensional regular array of multiple islands with multiple junctions, the presence (or absence) of one electron on an island induces a polarization of neighboring islands. The transfer of the charge distribution is called charge soliton (antisoliton).

Electron pump One of the single-electron devices categorized as a single-charge transfer device. Electron pump transfers exactly one electron from the source to the drain during one cycle of ac gate biases, even when the drain is reversely biased. Hence the name "pump."

Flash memory Electrically erasable and programmable read-only memory in which all the memory cells are erased in a lump.

Indirect bandgap semiconductor Semiconductor in which the minimum in the conduction band and the maximum in the valence band occur at different wave vectors.

PIN diode A diode in which a thin but lightly doped region is introduced between the *p*- and the *n*-side of a *p*-*n* junction. The lightly doped region can be assumed to be intrinsic.

Porous Si A Si layer consisting of many nanoscale structures and pores, formed by anodization of bulk Si in an aqueous HF solution.

Quantum cellular automata A machine consisting of identical cells, where each cell stores one bit of information and only talks to its nearest neighbors. A line of these cells can be used to transmit binary information and these cells can be arranged to form logic circuits.

Quantum size effect An effect in which the free electrons of holes occupy discrete quantum levels and have a discrete excitation spectrum when they are confined in a small region where their de Broglie wavelength is comparable to the size of the region.

Single-electron device A device that utilizes Coulomb blockade effect for the electron transfer.

Single-electron transistor A transistor having one or more Coulomb island(s) with tunnel junctions at both sides of each island. The electron tunnel through the island(s) is controlled by the gate voltage utilizing Coulomb blockade effect.

REFERENCES

1. L. T. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
2. R. A. Millikan, *Phys. Rev.* 2, 109 (1913).
3. C. J. Gorter, *Physica* 17, 777 (1951).
4. I. Giaever and H. R. Zeller, *Phys. Rev. Lett.* 20, 1504 (1968).
5. J. Lambe and R. C. Jaklevic, *Phys. Rev. Lett.* 22, 1371 (1969).
6. D. V. Averin and K. K. Likharev, *J. Low Temp. Phys.* 62, 345 (1986).
7. T. A. Fulton and G. J. Dolan, *Phys. Rev. Lett.* 59, 109 (1987).
8. L. S. Kuzmin, P. Delsing, T. Claeson, and K. K. Likharev, *Phys. Rev. Lett.* 62, 2539 (1989).
9. L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 64, 2691 (1990).
10. K. Yano, T. Ishii, T. Hashimoto, T. Kobayashi, F. Murai, and K. Seki, "IEEE International Electron Devices Meeting," 1993, p. 541.
11. Y. Takahashi, M. Nagase, H. Namatsu, K. Kurihara, K. Iwadate, Y. Nakajima, S. Horiguchi, K. Murase, and M. Tabe, "IEEE International Electron Devices Meeting," 1994, p. 938.
12. S. Furukawa and T. Miyasato, *Phys. Rev. B* 38, 5726 (1988).
13. H. Takagi, H. Ogawa, Y. Yamazaki, A. Ishizaki, and T. Nakagiri, *Appl. Phys. Lett.* 56, 2379 (1990).
14. Y. Sun, R. Nishitani, and T. Miyasato, *Jpn. J. Appl. Phys.* 33, L1645 (1994).
15. Q. Zhang, S. C. Bayliss, and D. A. Hutt, *Appl. Phys. Lett.* 66, 1977 (1995).
16. V. G. Baru, A. P. Chernushich, V. A. Luzanov, G. V. Stepanov, L. Y. Zakharov, K. P. O'Donnell, I. V. Bradley, and N. N. Melnik, *Appl. Phys. Lett.* 69, 4148 (1996).
17. H. Z. Song, X. M. Bao, N. S. Li, and X. L. Wu, *Appl. Phys. Lett.* 72, 356 (1998).
18. H. Morisaki, F. W. Ping, H. Ono, and K. Yazawa, *J. Appl. Phys.* 70, 1869 (1991).
19. H. Morisaki, H. Hashimoto, F. W. Ping, H. Nozawa, and H. Ono, *J. Appl. Phys.* 74, 2977 (1993).
20. D. Nesheva, C. Raptis, A. Perakis, I. Bineva, Z. Aneva, Z. Levi, S. Alexandrova, and H. Hofmeister, *J. Appl. Phys.* 92, 4678 (2002).
21. Y. He, C. Yin, G. Cheng, L. Wang, X. Liu, and G. Y. Hu, *J. Appl. Phys.* 75, 797 (1994).
22. X. Liu, X. Wu, X. Bao, and Y. He, *Appl. Phys. Lett.* 64, 220 (1994).
23. D. Zhang, R. M. Kolbas, P. D. Milewski, D. J. Lichtenwalner, A. I. Kingon, and J. M. Zavada, *Appl. Phys. Lett.* 65, 2684 (1994).
24. Z. Ma, X. Liao, G. Kong, and J. Chu, *Appl. Phys. Lett.* 75, 1857 (1999).
25. S. P. Withrow, C. W. White, A. Meldrum, J. D. Budai, D. M. Hembree, Jr., and J. C. Barbour, *J. Appl. Phys.* 86, 396 (1999).
26. S. Guha, S. B. Qadri, R. G. Musket, M. A. Wall, and T. Shimizu-Iwayama, *J. Appl. Phys.* 88, 3954 (2000).
27. K. A. Littau, P. J. Szajowski, A. J. Muller, A. R. Kortan, and L. E. Brus, *J. Phys. Chem.* 97, 1224 (1993).
28. F. Huisken, B. Kohn, and V. Paillard, *Appl. Phys. Lett.* 74, 3776 (1999).
29. L. Torrison, J. Tolle, D. J. Smith, C. Poweleit, J. Menendez, M. M. Mitan, T. L. Alford, and J. Kouvetakis, *J. Appl. Phys.* 92, 7475 (2002).
30. S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan, "IEEE International Electron Devices Meeting," 1995, p. 521.
31. A. Nakajima, Y. Sugita, K. Kawamura, H. Tomita, and N. Yokoyama, *Jpn. J. Appl. Phys.* 35, L189 (1996).
32. A. Nakajima, Y. Sugita, K. Kawamura, H. Tomita, and N. Yokoyama, *J. Appl. Phys.* 80, 4006 (1996).
33. A. Kohno, H. Murakami, M. Ikeda, S. Miyazaki, and M. Hirose, *Jpn. J. Appl. Phys.* 40, L721 (2001).
34. B. H. Choi, S. W. Hwang, I. G. Kim, H. C. Shin, Y. Kim, and E. K. Kim, *Appl. Phys. Lett.* 73, 3129 (1998).
35. A. Masson, J. J. Metois, and R. Kern, *Surf. Sci.* 27, 463 (1971).
36. J. J. Metois, M. Gauch, A. Masson, and R. Kern, *Thin Solid Films* 11, 205 (1972).
37. M. Zacharias, J. Blasing, P. Veit, L. Tsybeskov, K. Hirschman, and P. M. Fauchet, *Appl. Phys. Lett.* 74, 2614 (1999).
38. L. Tsybeskov, K. D. Hirschman, S. P. Duttagupta, M. Zacharias, P. M. Fauchet, J. P. McCaffrey, and D. J. Lockwood, *Appl. Phys. Lett.* 72, 43 (1998).
39. M. Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Blasing, *Appl. Phys. Lett.* 80, 661 (2002).
40. F. Gourbilleau, X. Portier, C. Ternon, P. Voivenel, R. Madelon, and R. Rizk, *Appl. Phys. Lett.* 78, 3058 (2001).
41. S. Tong, X. Liu, and X. Bao, *Appl. Phys. Lett.* 66, 469 (1995).
42. J. Alsmeier, E. Batke, and J. P. Kotthaus, *Phys. Rev. B* 41, 1699 (1990).

43. A. Nakajima, H. Aoyama, and K. Kawamura, *Jpn. J. Appl. Phys.* 33, L1796 (1994).
44. H. Fukuda, J. L. Hoyt, M. A. McCord, and R. F. W. Pease, *Appl. Phys. Lett.* 70, 333 (1997).
45. T. Takagahara and K. Takeda, *Phys. Rev. B* 46, 15578 (1992).
46. I. Sagnes, A. Halimaoui, G. Vincent, and P. A. Badoz, *Appl. Phys. Lett.* 62, 1155 (1993).
47. W. L. Wilson, P. F. Szajowski, and L. E. Brus, *Science* 262, 1242 (1993).
48. L. Patrone, D. Nelson, V. I. Safarov, M. Sentis, W. Marine, and S. Giorgio, *J. Appl. Phys.* 87, 3829 (2000).
49. B. G. Fernandez, M. Lopez, C. Garcia, A. Perez-Rodriguez, J. R. Morante, C. Bonafos, M. Carrada, and A. Claverie, *J. Appl. Phys.* 91, 798 (2002).
50. B. Garrido, M. Lopez, O. Gonzalez, A. Perez-Rodriguez, J. R. Morante, and C. Bonafos, *Appl. Phys. Lett.* 77, 3143 (2000).
51. L. N. Dinh, L. L. Chase, M. Balooch, W. J. Siekhaus, and F. Wooten, *Phys. Rev. B* 54, 5029 (1996).
52. F. Iacona, G. Franzo, and C. Spinella, *J. Appl. Phys.* 87, 1295 (2000).
53. Y. Kanemitsu, T. Ogawa, K. Shiraishi, and K. Takeda, *Phys. Rev. B* 48, 4883 (1993).
54. Y. Kanemitsu, *Phys. Rev. B* 53, 13515 (1996).
55. Y. Kanemitsu, S. Okamoto, M. Otobe, and S. Oda, *Phys. Rev. B* 55, R7375 (1997).
56. J. Linnros, N. Lalic, A. Galeckas, and V. Grivickas, *J. Appl. Phys.* 86, 6128 (1999).
57. D. J. Lockwood, *Solid State Commun.* 92, 101 (1994).
58. A. G. Cullis, L. T. Canham, and P. D. J. Calcott, *J. Appl. Phys.* 82, 909 (1997).
59. R. T. Collins, P. M. Fauchet, and M. A. Tischler, *Phys. Today*, Jan. 24 (1997).
60. A. G. Cullis and L. T. Canham, *Nature* 353, 335 (1991).
61. A. Nakajima, Y. Ohshima, T. Itakura, and Y. Goto, *Appl. Phys. Lett.* 62, 2631 (1993).
62. V. Lehmann and U. Gosele, *Appl. Phys. Lett.* 58, 856 (1991).
63. A. Nakajima, T. Itakura, S. Watanabe, and N. Nakayama, *Appl. Phys. Lett.* 61, 46 (1992).
64. A. Nakajima, Y. Nara, Y. Sugita, T. Itakura, and N. Nakayama, *Jpn. J. Appl. Phys.* 32, 415 (1993).
65. M. S. Brandt, H. D. Fuchs, M. Stutzmann, J. Weber, and M. Cardona, *Solid State Commun.* 81, 307 (1992).
66. T. George, M. S. Anderson, W. T. Pike, T. L. Lin, R. W. Fathauer, K. H. Jung, and D. L. Kwong, *Appl. Phys. Lett.* 60, 2359 (1992).
67. T. Suemoto, K. Tanaka, A. Nakajima, and T. Itakura, *Phys. Rev. Lett.* 70, 3659 (1993).
68. T. Suemoto, K. Tanaka, A. Nakajima, and T. Itakura, *J. Lumin.* 60&61, 324 (1994).
69. T. Suemoto, K. Tanaka, and A. Nakajima, *J. Phys. Soc. Jpn.* 63, suppl. B, 190 (1994).
70. P. D. J. Calcott, K. J. Nash, L. T. Canham, M. J. Kane, and D. Brumhead, *J. Phys.: Condens. Matter* 5, L91 (1993).
71. P. D. J. Calcott, K. J. Nash, L. T. Canham, M. J. Kane, and D. Brumhead, *J. Lumin.* 57, 257 (1993).
72. T. Suemoto, K. Tanaka, and A. Nakajima, *Phys. Rev. B* 49, 11005 (1994).
73. K. L. Shaklee and R. E. Nahory, *Phys. Rev. Lett.* 24, 942 (1970).
74. T. Matsumoto, T. Futagi, H. Mimura, and Y. Kanemitsu, *Phys. Rev. B* 47, 13876 (1993).
75. S. Gardelis, J. S. Rimmer, P. Dawson, B. Hamilton, R. A. Kubiak, T. E. Whall, and E. H. C. Parker, *Appl. Phys. Lett.* 59, 2118 (1991).
76. Y. Kanemitsu and S. Okamoto, *Phys. Rev. B* 56, R1696 (1997).
77. M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allen, and C. Delerue, *Phys. Rev. Lett.* 82, 197 (1999).
78. J. H. Shin, G. N. van den Hoven, and A. Polman, *Appl. Phys. Lett.* 66, 2379 (1995).
79. F. Namavar, F. Lu, C. H. Perry, A. Cremins, N. M. Kalkhoran, and R. A. Soref, *J. Appl. Phys.* 77, 4813 (1995).
80. X. Wu, U. Hommerich, F. Namavar, and A. M. Cremins-Costa, *Appl. Phys. Lett.* 69, 1903 (1996).
81. U. Hommerich, F. Namavar, A. Cremins, and K. L. Bray, *Appl. Phys. Lett.* 68, 1951 (1996).
82. T. Kimura, A. Yokoi, H. Horiguchi, R. Saito, T. Ikoma, and A. Sato, *Appl. Phys. Lett.* 65, 983 (1994).
83. A. M. Dorofeev, N. V. Gaponenko, V. P. Bondarenko, E. E. Bachilo, N. M. Kazuchits, A. A. Leshok, G. N. Troyanova, N. N. Vorosov, V. E. Borisenko, H. Gnaser, W. Bock, P. Becker, and H. Oechsner, *J. Appl. Phys.* 77, 2679 (1995).
84. M. Fujii, M. Yoshida, Y. Kanzawa, S. Hayashi, and K. Yamamoto, *Appl. Phys. Lett.* 71, 1198 (1997).
85. K. Watanabe, M. Fujii, and S. Hayashi, *J. Appl. Phys.* 90, 4761 (2001).
86. P. G. Kik, M. L. Brongersma, and A. Polman, *Appl. Phys. Lett.* 76, 2325 (2000).
87. P. G. Kik and A. Polman, *J. Appl. Phys.* 88, 1992 (2000).
88. F. Priolo, G. Franzo, D. Pacifici, and V. Vinciguerra, *J. Appl. Phys.* 89, 264 (2001).
89. A. Richter, P. Steiner, F. Kozlowski, and W. Lang, *IEEE Electron Device Lett.* 12, 691 (1991).
90. N. Koshida and H. Koyama, *Appl. Phys. Lett.* 60, 347 (1992).
91. A. Halimaoui, C. Oules, G. Bomchil, A. Bsiesy, F. Gaspard, R. Herino, M. Ligeon, and F. Muller, *Appl. Phys. Lett.* 59, 304 (1991).
92. P. M. M. C. Bressers, J. W. J. Knapen, E. A. Meulenkamp, and J. J. Kelly, *Appl. Phys. Lett.* 61, 108 (1992).
93. W. H. Green, E. J. Lee, J. M. Lauerhaas, T. W. Bitner, and M. J. Sailor, *Appl. Phys. Lett.* 67, 1468 (1995).
94. N. Koshida, H. Koyama, Y. Yamamoto, and G. J. Collins, *Appl. Phys. Lett.* 63, 2655 (1993).
95. K. Li, D. C. Diaz, Y. He, J. C. Campbell, and C. Tsai, *Appl. Phys. Lett.* 64, 2394 (1994).
96. P. Steiner, F. Kozlowski, and W. Lang, *IEEE Electron Device Lett.* 14, 317 (1993).
97. H. Shi, Y. Zheng, Y. Wang, and R. Yuan, *Appl. Phys. Lett.* 63, 770 (1993).
98. F. Namavar, H. P. Maruska, and N. M. Kalkhoran, *Appl. Phys. Lett.* 60, 2514 (1992).
99. Z. Chen, G. Bosman, and R. Ochoa, *Appl. Phys. Lett.* 62, 708 (1993).
100. P. Steiner, F. Kozlowski, and W. Lang, *Appl. Phys. Lett.* 62, 2700 (1993).
101. J. Linnros and N. Lalic, *Appl. Phys. Lett.* 66, 3048 (1995).
102. L. Tsybeskov, S. P. Duttagupta, K. D. Hirschman, and P. M. Fauchet, *Appl. Phys. Lett.* 68, 2058 (1996).
103. K. D. Hirschman, L. Tsybeskov, S. P. Duttagupta, and P. M. Fauchet, *Nature* 384, 338 (1996).
104. T. Toyama, T. Matsui, Y. Kurokawa, H. Okamoto, and Y. Hamakawa, *Appl. Phys. Lett.* 69, 1261 (1996).
105. L. Pavesi, C. Mazzoleni, A. Tredicucci, and V. Pellegrini, *Appl. Phys. Lett.* 67, 3280 (1995).
106. F. Iacona, G. Franzo, E. C. Moreira, and F. Priolo, *J. Appl. Phys.* 89, 8354 (2001).
107. L. Pavesi, L. D. Negro, C. Mazzoleni, G. Franzo, and F. Priolo, *Nature* 408, 440 (2000).
108. Q. Ye, R. Tsu, and E. H. Nicollian, *Phys. Rev. B* 44, 1806 (1991).
109. S. Y. Chou and A. E. Gordon, *Appl. Phys. Lett.* 60, 1827 (1992).
110. M. Fukuda, K. Nakagawa, S. Miyazaki, and M. Hirose, *Appl. Phys. Lett.* 70, 2291 (1997).
111. M. Otobe, H. Yajima, and S. Oda, *Appl. Phys. Lett.* 72, 1089 (1998).
112. C. A. Neugebauer and M. B. Webb, *J. Appl. Phys.* 33, 74 (1962).
113. I. O. Kulik and R. I. Shekhter, *Sov. Phys. JETP* 41, 308 (1975).
114. H. Matsuoka, T. Ichiguchi, T. Yoshimura, and E. Takeda, *Appl. Phys. Lett.* 64, 586 (1994).

115. H. Matsuoka and S. Kimura, *Appl. Phys. Lett.* 66, 613 (1995).
116. H. Matsuoka and H. Ahmed, *Jpn. J. Appl. Phys.* 35, L418 (1996).
117. D. Ali and H. Ahmed, *Appl. Phys. Lett.* 64, 2119 (1994).
118. A. Ohata, H. Niiyama, T. Shibata, K. Nakajima, and A. Toriumi, *Jpn J. Appl. Phys.* 34, 4485 (1995).
119. Y. Takahashi, H. Namatsu, K. Kurihara, K. Iwadate, M. Nagase, and K. Murase, *IEEE Trans. Electron Devices* 43, 1231 (1996).
120. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, “IEEE International Electron Devices Meeting,” 1998, p. 123.
121. A. Fujiwara, Y. Takahashi, K. Yamazaki, H. Namatsu, M. Nagase, K. Kurihara, and K. Murase, “IEEE International Electron Devices Meeting,” 1997, p. 163.
122. Y. Takahashi, A. Fujiwara, K. Yamazaki, H. Namatsu, K. Kurihara, and K. Murase, “IEEE International Electron Devices Meeting,” 1998, p. 127.
123. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, “IEEE International Electron Devices Meeting,” 1999, p. 367.
124. H. Inokawa, A. Fujiwara, and Y. Takahashi, “IEEE International Electron Devices Meeting,” 2001, p. 147.
125. E. Leobandung, L. Guo, Y. Wang, and S. Y. Chou, *Appl. Phys. Lett.* 67, 938 (1995).
126. E. Leobandung, L. Guo, and S. Y. Chou, *Appl. Phys. Lett.* 67, 2338 (1995).
127. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *Appl. Phys. Lett.* 71, 353 (1997).
128. Y. Ito, T. Hatano, A. Nakajima, and S. Yokoyama, *Appl. Phys. Lett.* 80, 4617 (2002).
129. T. Hatano, Y. Ito, A. Nakajima, and S. Yokoyama, *Jpn. J. Appl. Phys.* 40, 2017 (2001)
130. D. H. Kim, S.-K. Sung, K. R. Kim, B. H. Choi, S. W. Hwang, D. Ahn, J. D. Lee, and B.-G. Park, “IEEE International Electron Devices Meeting,” 2001, p. 151.
131. D. H. Kim, S.-K. Sung, K. R. Kim, J. D. Lee, B.-G. Park, B. H. Choi, S. W. Hwang, and D. Ahn, *IEEE Trans. Electron Devices* 49, 627 (2002).
132. L. Zhuang, L. Guo, and S. Y. Chou, “IEEE International Electron Devices Meeting,” 1997, p. 167.
133. H. Ishikuro, T. Fujii, T. Saraya, G. Hashiguchi, T. Hiramoto, and T. Ikoma, *Appl. Phys. Lett.* 68, 3585 (1996).
134. H. Ishikuro and T. Hiramoto, *Appl. Phys. Lett.* 71, 3691 (1997).
135. H. Ishikuro and T. Hiramoto, “IEEE International Electron Devices Meeting,” 1998, p. 119.
136. N. Takahashi, H. Ishikuro, and T. Hiramoto, “IEEE International Electron Devices Meeting,” 1999, p. 371.
137. R. Augke, W. Eberhardt, C. Single, F. E. Prins, D. A. Wharam, and D. P. Kern, *Appl. Phys. Lett.* 76, 2065 (2000).
138. T. Sakamoto, H. Kawaura, and T. Baba, *Appl. Phys. Lett.* 72, 795 (1998).
139. R. A. Smith and H. Ahmed, *J. Appl. Phys.* 81, 2699 (1997).
140. A. Tilke, R. H. Blick, H. Lorenz, J. P. Kotthaus, and D. A. Wharam, *Appl. Phys. Lett.* 75, 3704 (1999).
141. V. Ng, H. Ahmed, and T. Shimada, *J. Appl. Phys.* 86, 6931 (1999).
142. A. C. Irvine, Z. A. K. Durrani, H. Ahmed, and S. Biesemans, *Appl. Phys. Lett.* 73, 1113 (1998).
143. Y. T. Tan, Z. A. K. Durrani, and H. Ahmed, *J. Appl. Phys.* 89, 1262 (2001).
144. K. Kawamura, T. Kidera, A. Nakajima, and S. Yokoyama, *J. Appl. Phys.* 91, 5213 (2002).
145. P. Delsing, T. Claeson, K. K. Likharev, and L. S. Kuzmin, *Phys. Rev. B* 42, 7439 (1990).
146. T. S. Tighe, M. T. Tuominen, J. M. Hergenrother, and M. Tinkham, *Phys. Rev. B* 47, 1145 (1993).
147. R. A. Smith and H. Ahmed, *Appl. Phys. Lett.* 71, 3838 (1997).
148. A. Nakajima, Y. Ito, and S. Yokoyama, *Appl. Phys. Lett.* 81, 733 (2002).
149. C. Single, F. E. Prins, and D. P Kern, *Appl. Phys. Lett.* 78, 1421 (2001).
150. K. Yano, T. Ishii, T. Hashimoto, T. Kobayashi, F. Murai, and K. Seki, *IEEE Trans. Electron Devices* 41, 1628 (1994).
151. T. Ishii, K. Yano, T. Sano, T. Mine, F. Murai, and K. Seki, “IEEE International Electron Devices Meeting,” 1997, p. 171.
152. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Kure, and K. Seki, “IEEE International Solid-State Circuits Conference,” 1998, p. 344.
153. S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan, *Appl. Phys. Lett.* 68, 1377 (1996).
154. S. Tiwari, F. Rana, K. Chan, L. Shi, and H. Hanafi, *Appl. Phys. Lett.* 69, 1232 (1996).
155. H. Hanafi, S. Tiwari, and K. Chan, *IEEE Trans. Electron Devices* 43, 1553 (1996).
156. I. Kim, S. Han, H. Kim, J. Lee, B. Choi, S. Hwang, D. Ahn, and H. Shin, “IEEE International Electron Devices Meeting,” 1998, p. 111.
157. A. Fernandes, B. DeSalvo, T. Baron, J. F. Damlencourt, A. M. Papon, D. Lafond, D. Mariolle, B. Guillaumot, P. Besson, P. Masson, G. Ghibaudo, G. Pananakakis, F. Martin, and S. Haukka, “IEEE International Electron Devices Meeting,” 2001, p. 155.
158. J. De Blauwe, M. Ostraat, M. L. Green, G. Weber, T. Sorsch, A. Kerber, F. Klemens, R. Cirelli, E. Ferry, J. L. Grazul, F. Baumann, Y. Kim, W. Mansfield, J. Bude, J. T. C. Lee, S. L. Hillenius, R. C. Flagan, and H. A. Atwater, “IEEE International Electron Devices Meeting,” 2000, p. 683.
159. K. Han, I. Kim, and H. Shin, “IEEE International Electron Devices Meeting,” 2000, p. 309.
160. T. Ishii, T. Osabe, T. Mine, F. Murai, and K. Yano, “IEEE International Electron Devices Meeting,” 2000, p. 305.
161. R. Ohba, N. Sugiyama, K. Uchida, J. Koga, and A. Toriumi, “IEEE International Electron Devices Meeting,” 2000, p. 313.
162. Y.-C. King, T.-J. King, and C. Hu, “IEEE International Electron Devices Meeting,” 1998, p. 115.
163. A. Nakajima, T. Futatsugi, N. Horiguchi, and N. Yokoyama, *Appl. Phys. Lett.* 71, 3652 (1997).
164. A. Nakajima, T. Futatsugi, N. Horiguchi, H. Nakao, and N. Yokoyama, “IEEE International Electron Devices Meeting,” 1997, p. 159.
165. A. Nakajima, T. Futatsugi, H. Nakao, T. Usuki, N. Horiguchi, and N. Yokoyama, *J. Appl. Phys.* 84, 1316 (1998).
166. A. Nakajima, T. Futatsugi, N. Horiguchi, and N. Yokoyama, *Jpn. J. Appl. Phys.* 36, L1552 (1997).
167. A. Nakajima, H. Nakao, H. Ueno, T. Futatsugi, and N. Yokoyama, *Appl. Phys. Lett.* 73, 1071 (1998).
168. A. Nakajima, H. Nakao, H. Ueno, T. Futatsugi, and N. Yokoyama, *J. Vac. Sci. Technol. B* 17, 1317 (1999).
169. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, “IEEE International Electron Devices Meeting,” 1996, p. 952.
170. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *Appl. Phys. Lett.* 70, 1742 (1997).
171. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *J. Vac. Sci. Technol. B* 17, 2163 (1999).
172. L. Guo, E. Leobandung, and S. Y. Chou, “IEEE International Electron Devices Meeting,” 1996, p. 955.
173. G. Iannaccone, A. Trellakis, and U. Ravaioli, *J. Appl. Phys.* 84, 5032 (1998).
174. S. Amakawa, K. Kanda, M. Fujishima, and K. Hoh, *Jpn. J. Appl. Phys.* 38, 429 (1999).
175. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, and K. Seki, “IEEE International Electron Devices Meeting,” 1995, p. 525.

176. T. Usuki, T. Futatsugi, and A. Nakajima, *Jpn. J. Appl. Phys.* 37, L709 (1998).
177. N. Kuwamura, K. Taniguchi, and C. Hamaguchi, *Trans. IEICE* J77-C-II, 221 (1994) [in Japanese].
178. S. Shimano, K. Masu, and K. Tsubouchi, *Jpn. J. Appl. Phys.* 38, 403 (1999).
179. K. Uchida, K. Matsuzawa, and A. Toriumi, "Extend. Abst. Int. Conf. on Solid State Devices and Materials," 1998, p. 188.
180. T. Yamanaka, T. Morie, M. Nagata, and A. Iwata, "Extend. Abst. Int. Conf. on Solid State Devices and Materials," 1998, p. 190.
181. A. N. Korotkov, R. H. Chen, and K. K. Likharev, *J. Appl. Phys.* 78, 2520 (1995).

www.aspbs.com/enn

Silicon Surface Nanooxidation

D. Stiévenard, B. Legrand
IEMN, UMR8520, CNRS, Lille, France

CONTENTS

1. Introduction
2. Silicon Nanooxidation
3. Modelization of the Oxidation Process
4. Nanostructures
5. Comparison of Techniques
6. Conclusion
Glossary
References

1. INTRODUCTION

In order to fabricate nanostructures, a technological process is needed, involving mask patterning followed by a surface machining. The final size of a device is mainly related to the mask size. Up to now, this size has regularly decreased from 15 to 0.18 μm over the last three decades. However, the manufacturing cost has followed the same type of variation, but in the way of an increase, leading to econonical limits toward 2010–2015. So, new low-cost approaches are under study: an-organic self-assembled layer for molecular electronics, nanoprinting, or scanning probe microscope (SPM) assisted lithography. SPM includes atomic force microscopy (AFM) and scanning tunneling microscopy (STM), two techniques based on a tip–surface interaction, with a nanometer range resolution. For the STM, the average tip–surface distance is on the order of 1 nm, and for AFM, in contact or noncontact mode (see other chapters of this encyclopedia for details on AFM and STM techniques), the tip is continuously or periodically in contact with the surface. If a few volts of polarization are applied between the tip and the surface, an electric field on the order of a few megavolts per centimeter is generated, leading to local electric-field-assisted oxidation of the surface. The subsequent oxide can be used as a mask in order to pattern the surface, under wet or dry etching. GaAs, SiGe, and Si semiconductors have been studied. Up to now, silicon has been the most popular, and this chapter is devoted to this material.

The aim of this chapter is to: (1) to present AFM and STM results for oxide generation, with the different physical parameters governing the oxidation kinetics; (2) to discuss the proposed models used to explain the oxidation process; (3) to show different nanodevices obtained using SPM-based lithography; and finally, (4) to discuss the throughput of SPM lithography, based on recent parallel approaches.

2. SILICON NANOOXIDATION

The first evidence of local nanooxidation of a silicon surface was presented by Dagata et al. in 1990 [1]. The authors used an STM on an Si(111) surface, preliminary passivated in dilute hydrofluoric acid after RCA cleaning [2]. The minimum definable feature was on the order of 100–200 nm size for an applied tip–surface voltage of 3.5 V and a tip velocity of 1 $\mu m \cdot s^{-1}$. Since STM images contain a convolution of topographic and electronic information, the STM-patterned features appear as 1–20 nm depressions. As silicon oxide is typically 1 nm high with respect to the silicon surface, this observation is in apparent contradiction with the observation of a locally oxidated region. In fact, as the silicon dioxide resistivity is higher than the silicon one, when the STM tip is over the oxide, the STM current tends to decrease, and in order to maintain a constant current, the STM feedback loop decreases the tip–surface distance. As the displacement of the tip perpendicular to the surface is interpreted as the local topography, the oxide thus appears as a depression. This observation shows us that the STM can realize local oxidation, but is not an adequate tool to accurently measure a local topography. However, as claimed by Dagata et al., this pioneer work shows that "STM-based nanotechnology is possible."

Modification of a silicon surface using an AFM with a conductive tip was demonstrated by a number of authors [3–6]. Figure 1 gives a schematic representation of the nanooxidation process. Day and Allee [7] used buffered hydrofluoric acid after the AFM oxidation process in order to reveal trenches in the silicon, consistent with silicon consumption in SiO_2 formation. By measuring the oxide height with the AFM, the authors showed that the oxidation is enhanced by the electric field and limited by the diffusion of oxidizing species. Fay et al. [8] examined a modified area of a silicon surface using Auger, electron spectroscopy, scanning

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.
Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (859–864)

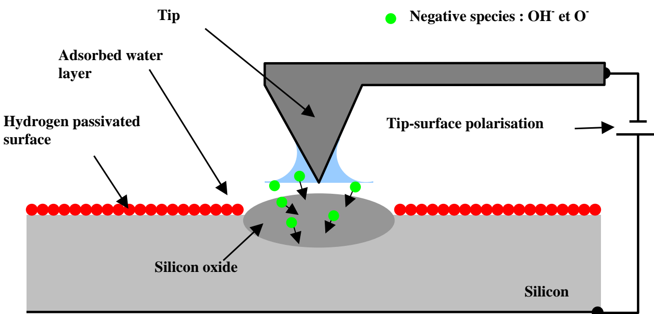


Figure 1. Schematic representation of the nanooxidation process.

electron microscopy, and atomic force microscopy. A comparison of these characterization techniques indicates that the features are both chemical and topographical in nature, and are the result of local oxidation of the silicon substrate.

These works have stimulated the research in the silicon nanooxidation field. A systematic study of the influence of the voltage and time [9–13] for STM and AFM in the contact mode has begun. The main results are that the oxidation is a tip-induced anodization and an electric-field-enhanced reaction.

As the tip–surface interaction is strong under STM or AFM in the contact mode, the tip reliability is poor. In 1995 and 1996, Pérez-Murano et al. [14] and Servat et al. [15] published a new approach using an AFM working in the tapping mode. They showed that a local oxidation is also possible, and realizes dots and lines. As the tip–surface interaction is smoother than in the case of the contact mode, the noncontact mode is more reliable than the contact mode, with high resolution (10 nm) and high speed (0.1 mm/s).

In order to clarify the performances of SPM techniques, Fontaine et al. in 1998 [16] investigated the three different SPM modes. They evaluated the efficiency of each technique in terms of reproducibility, spatial resolution, speed, and masking capability. Their results are summerized in Table 1.

The STM mode, despite reaching an optimum resolution of 10 nm, is not a suitable tool for nanolithography on silicon. In contrast, AFM in the contact mode allows high speed (up to 1 mm/s), and can provide thick oxide (8 nm). However, these performances are achieved at the expense of loss of spatial resolution, mainly associated with the tip shape. AFM in the tapping mode appears to be the best compromise between STM and CM–AFM, offering fine patterns as narrow as 20 nm with relatively high speed (10 μm/s). Similar results have been recently published by Tello and Garcia [17] in 2001.

In 1999, Legrand and Stiévenard [18] introduced an original method for better reliability of the oxidation process. Up to then, a continuous bias voltage was applied between the tip and the surface. Starting from the observation that the time required to obtain 1 nm oxide thickness is on the order of a few microseconds, they concluded that a continuous bias voltage is not necessary, and they proposed a pulsed voltage technique. In such a case, the interaction time of the tip with the surface under the electric field decreases. The frequency oscillation (in the noncontact mode) of the cantilever is taken as a reference, and pulsed voltages with variable phase and duty cycle are used. They show that the variation of the phase allows a 100% modulation of the oxide width. This work was later reinforced by similar results from Pérez-Murano et al. [19] and Calleja and Garcia [20].

The role of oxygen is obviously important for the oxidation process. By choosing suitable and controlled experimental conditions, Marchi et al. [21] proved that the oxide growth process is limited both thermodynamically (as evidenced by the voltage threshold) and kinetically (as shown by a considerable increase of the oxide growth by an ozone supply in the environmental gas). They proposed a qualitative three-step mechanism: (1) surface depassivation by desorption of hydrogen atoms assisted by the electric field above a threshold voltage; (2) the formation of oxyanions in the water meniscus between the tip and the surface, a complex step which may involve both electrolytic and dissolution processes, and (3) diffusion of the oxyanions assisted by the electric field toward the SiO_x–Si interface. More quantitatively refined models are discussed in the Section 3 of this chapter.

With regard to the role of water in the oxidation process, it has been mainly studied for the tapping mode. The formation of a water bridge between the tip and the surface is driven by the electric field [22–24]. Once a liquid bridge is formed, it provides the ionic species and the spatial confinement to pattern the silicon surface. The method was applied to write arrays of several thousand dots with a periodicity of 40 nm and an average width of 10 nm. Finally, Snow et al. [25] reported measurements of the kinetics of scanned probe oxidation under conditions of high humidity and pulsed bias. They concluded that the oxidation reaction is dominated by the production of OH ions for the case of moderate oxide thickness (a few nanometers) and exposure times (typically less than 10 ms). For long exposure times (typically more than 100 ms) and thick oxides, the effects of ion diffusion, space charge, and stress (as pointed out in Section 3) can eventually dominate the kinetics. In fact, for a hydrophobic silicon surface, the oxidation rate for short pulse duration is governed by the density of H_2O molecules

Table 1. Comparison of nanooxidation efficiency of STM, contact mode AFM, and tapping mode AFM.

Factors	STM	Contact mode AFM	Tapping mode AFM
Reproducibility	<50%	>50%	>75%
Maximum speed (μm/s)	8	128[a]	20
Better resolution (nm)	10	30	10
Maximum thickness oxide (nm)[b]	1.5	8	4

[a] In this case, limited by SPM equipment.
[b] Limited by a maximum 12 V polarization associated with the equipment.

in the ambient humidity surrounding the tip–surface interface. For a longer pulse duration, liquid H_2O bridges this interface, and the maximum oxidation rate increases by a factor of $\sim 10^4$ according to the increased density of H_2O molecules.

In the next part of this chapter, we discuss the different models proposed to explain the oxidation process.

3. MODELIZATION OF THE OXIDATION PROCESS

As the modelization of the oxidation process is a key parameter to predict and control the fabrication of nanostructures, the mechanism of tip-induced oxidation has been adressed by several authors since 1995.

The first attempt to understand the physical mechanism and its relevance to anodization layer and thickness was proposed by Gordon et al. [26], who pointed out that the initial density of the surface OH group is rate limiting. Moreover, using degenerated samples, they demonstrated that the presence of holes at the surface to recombine with the negative species (O^- or OH^-) is requested. Indeed, under illumination, the hole concentration increases, and consequently, an increase of the oxidation rate is observed.

Teuschler et al. [5] and Ley et al. [13] proposed an empirical power law ($h \propto t^{1/4}$) which fits experimental data well, but which is incompatible with classical models for oxide growth. However, they showed the existence of a doping-dependent threshold voltage under which no oxidation occurs.

The first quantitative model based on a physical basis was proposed by Stiévenard et al. in 1997 [27]. Their basic proposed idea was to use a Cabrera and Mott model [28] developed for field-induced oxidation, and to extrapolate it for very thin oxide. An analytical formalism fits a linear relation between h and V and $1/h$ versus $\log(t)$ well. Finally, the model allows the estimation of an activation energy associated with the oxidation process. An energy of 0.15 eV is found, far from the 4 eV involved for a high-temperature thermal oxidation process and not far from the one (0.15–0.4 eV) associated with plasma oxidation.

The previous approach is available only for the initial oxide layer, and cannot explain the formation of a thicker oxide layer. The same year, Avouris et al. presented new information [29]: first, he measured an electrochemical current during the oxidation. Its low value (a few 10^{-15} A) is associated with ionic transport, which confirms the role played by OH^- groups in the oxidation mechanism. Second, the rate of oxidation was found to decrease rapidly as the oxide films grow due to the self-limiting influence of the decreasing field strengths and the build up of the stress. A logarithmic time dependence of h versus t is found to fit the data for t ranging from 0.01 to 1000 s and h from a few nanometers to 50 nm well. An apparent contradiction arises from Stiévenard and Avouris's conclusions concerning the variation of h with time, between $1/h \propto -\log(t)$ and $h \propto \log(t)$, respectively. In fact, the two models are not comparable since the time ranges are not the same. In the case of Stiévenard et al., the oxidation time ranges from a few microseconds to a few milliseconds, whereas in the case of Avouris, the oxidation time ranges from a few milliseconds to a fraction of an hour. So, it is difficult to compare the two analyses since the physical events taken into account in the models are different (stress versus cutoff thickness, for example), and the Stiévenard model is only available for the initial thin oxide layer.

Dagata et al. [30, 31] present new data on the growth rate and electrical characteristic of nanostructures produced by scanned probe oxidation by integrating an *in-situ* electrical force characterization technique and a scanning Maxwell-stress microscopy during the fabrication process. Simultaneous topographical, capacitance, and surface potential data were obtained for oxide features patterned on *n*- and *p*-type silicon and titanium thin-film substrates. They proposed a model for the SPM oxidation reaction which complements existing kinetics descriptions derived from the Cabrera–Mott theory (Stiévenard et al.'s model). They show that space-charge effects are consistent with a rapid decline of high initial growth rates, and that oxide thickness and stress considerations alone cannot account for the observed doping and voltage–pulse dependencies, density variations, and SPM oxide growth rate enhancement by voltage modulation and other dynamic pulse-shaping techniques. The space-charge form leads to a 0.1 eV activation energy process, in agreement with the one (0.15 eV) deduced previously with the Cabrera–Mott model. However, their conclusion is that stress and space charge simply describe the same final outcome—an amorphous oxide—from two different points of view.

The more refined model was proposed by Dagata et al. [32] in 2000, followed by a similar approach of Dubois and Bubendorff [33] one month later. Dagata proposed a model that includes a temporal crossover of the system from transient to steady-state growth, and a spatial crossover from predominantly vertical to coupled lateral growth. Starting from a direct-log form for the SPM oxidation, they examined the assumptions used to derive it by Uhlig in 1956. He assumed that the rate-controlling step involves interaction of oxyanions with electronic species—holes or electrons—and defects at a metal–oxide interface. In particular, using the Uhlig model, it is possible to unify key concepts introduced by Stiévenard (a cutoff field associated with the potential at which space charge in the oxide nulls the externally applied field) and Avouris (by replacing the stress by space charge for a thickness-related change in activation energy) with Dagata's conjectures about the influence of space charge on SPM oxide growth. According to this approach, the production of charged defects leads to a build up of space charge within the oxide, inhibiting further growth over long pulse times. The power law of time for the variation of the oxide height is then derived. Dubois and Bubendorff [33] proposed a similar model based on the presence of a space charge due to nonstoichiometric states located close to the bulk/oxide interface. The main assumption in the model derivation is the occurrence of mutual Coulomb repulsions between traps and the control of the ionic transport by the region of lowest electric field. In fact, these assumptions are commonly used to treat transport phenomena in dielectrics. The presence of large trap density during the formation of the very thin oxide layer is questionable, and depends on the presence of a native oxide, which in principle does not exist on

a hydrogenated silicon surface. However, the observed fast oxidation rates are associated with the presence of a large amount of defects. This observation leads to the conclusion of a poor-quality oxide (plasma-like), in agreement with the low oxidation activation energy.

To return to the proposed models, the main parameters to take into account in order to describe the oxidation process are: (1) a field effect with a cutoff field, (2) a diffusion-limiting process, and (3) a space-charge effect associated with defects located at the Si/SiO_2 interface.

In the next section, we give some illustrations of silicon nanostructures obtained after a dry or wet chemical etching of a silicon surface. The nanostructure is obtained thanks to the SPM oxide, which is robust enough to act as a mask during the etching process.

4. NANOSTRUCTURES

One interesting aspect of the SPM-induced nanooxidation is the possibility to obtain low-cost low-dimensional structures. One of the first silicon nanostructures was obtained by Snow et al. [34] using an STM operating in air. The process involves the direct chemical modification of an H-passivated Si(100) surface and a subsequent wet etching. The chemically modified zones of the surface can withstand a deep (greater than 100 nm) wet etching of the unmodified regions with no degradation. Patterns with lateral feature sizes on the order of 25 nm were reliably fabricated. Typical results are given in Figures 2, 3 and 4. Figure 2 shows AFM images of oxide lines and silicon wires after KOH etching, with the associated cross section. Figure 3 shows a 3-D view of silicon nanowires and Figure 4 is a 3-D view of a 5.5 nm high wire on SOI substrate. The true full width at half maximum (FWHM) is estimated to be about 8 nm. The same results were obtained by Snow et al. [36] using an electron cyclotron resonance source for etching. Later, in 1994, the same authors presented a method for fabricating silicon nanostructures [37] with an air-operating AFM (also published by Morimoto et al. [38] and Sugimura and Nakagiri [39]). An electrically conducting AFM tip is used to oxidize regions of size 10–30 nm of an H-passivated Si(100) surface. This oxide acts as a robust mask for pattern transfer into the substrate by selective wet etching (KOH solution). In 1995, a step toward the fabrication of a nanodevice was achieved by Minne et al. [40]. Using an AFM, they fabricated a metal oxide semiconductor field-effect transistor on silicon with an effective 100 nm channel length. The lithography at the gate level was performed with the AFM tip: the gate was defined by an electric-field-enhanced selective oxidation of the amorphous silicon gate electrode. The device has reasonable characteristics (a transconductance of 279 mS/mm with a threshold voltage of 0.55 V). The same kind of result was obtained by Fayfield and Higman [41] and Ishi and Matsumoto [42], and a silicon field-effect transistor was fabricated by Campbell et al. [43] with critical features as small as 30 nm. Snow and Campbell [44] presented an improvement in the fabrication process: the width of the wires and the resistance of the junctions were controlled by real-time, *in-situ* measurement of the device resistance during fabrication. Because the properties of the nanometer-scale devices are very sensitive to size variations, such measurements bring a more accurate method of controlling device properties than by controlling geometry alone *a posteriori*. Structures with critical dimensions of less than 10 nm were fabricated.

In the last section, a comparison of the performances of SPM, optical, and *e*-beam lithographies is presented.

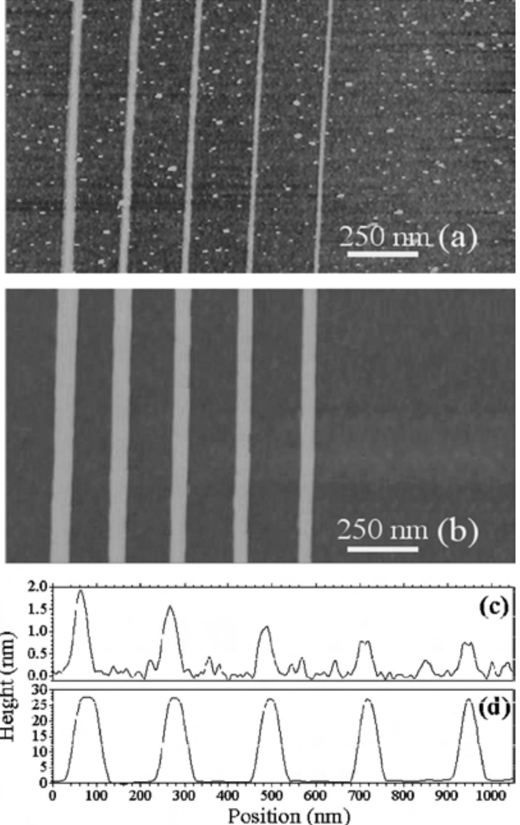


Figure 2. AFM images. (a) Top view of a set of five oxide lines of various widths and heights. (b) Resulting silicon wires after KOH etching. (c) Section across oxide lines of (a). (d) Section across silicon wires of (b) [35].

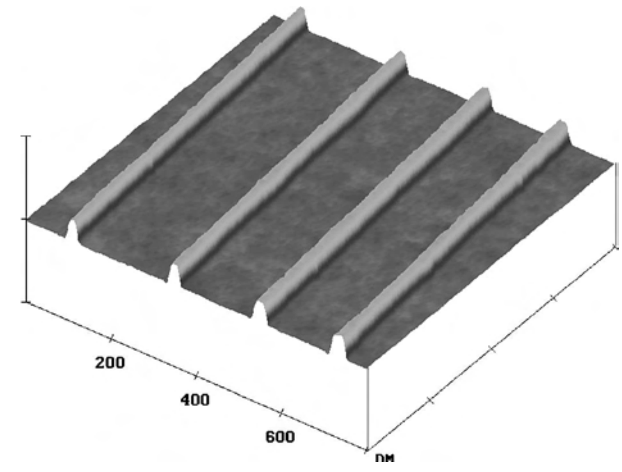


Figure 3. 3-D AFM view of silicon wires (height: 6 nm, width: 25 nm) [35].

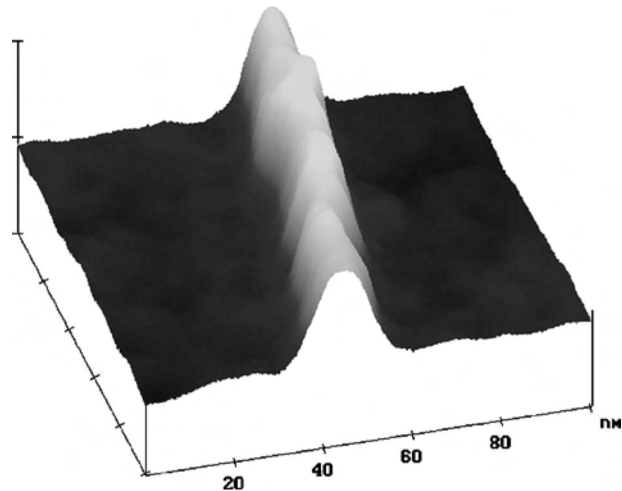


Figure 4. 3-D view of an AFM image (100 nm × 100 nm) showing a 5.5 nm high silicon wire with an FWHM estimated to be about 8 nm.

5. COMPARISON OF TECHNIQUES

Optical lithography has been used ubiquitously at the industrial level to define the patterns of integrated circuits for decades, and allows 0.18 μm resolution today. Projection optical lithography fulfills the five requirements for mass production: fine pattern formation capability, alignment accuracy, depth of focus, high throughput, and low cost. It will be used as long as possible, but theoretical limitations are on the verge of being resolved in the next decade. Indeed, it would be extremely difficult to delineate sub-100 nm fine patterns even by using short-wavelength light.

e-beam lithography could be an alternative since it is already employed for optical mask generation, and is at present the primary lithographic technique at the research level for nanometric device fabrication with 20 nm resolution. However, *e*-beam lithography suffers from a too slow wafer throughput. The exposure rate of an optical stepper is typically 10 cm^2/s, whereas it is less than 0.01 cm^2/s in the case of *e*-beam lithography, which is below the practical rate of 1 cm^2/s [45].

SPM lithography allows delineating down to 10 nm fine patterns, as shown before. The major drawback for practical applications is the very low exposure rate, on the order of 10^{-4} cm^2/s, for a single tip. Therefore, the only solution to fulfill the high-throughput requirement for mass production is the parallel lithography system. Using micromachining and microfabrication techniques, SPM tip arrays were built in the late 1990s. Following the work of Manalis et al. [46], who presented, in 1996, AFM cantilevers with an integrated actuator and sensor, Minne et al. [47] demonstrated lithography over a 100 mm^2 area using an array of 50 cantilevers. Up to a 1000 AFM tip array, called the "millipede," was built by IBM nearly at the same time [48]. Expected applications mainly concern ultrahigh-density data storage: using a thermomechanical process, each tip can write 30–40 nm sized bit indentations in a polymer film, and read them afterwards. On the other hand, Wada proposed in 1998 a lithography system including 10^4–10^6 micromachine SPMs in parallel [45]. Such a number of tips would allow reaching the practical exposure rate and the throughput required for mass production with a pattern minimum dimension of 10 nm. In order to ascertain the feasibility of the proposed system, these authors fabricated micro-STMs using technologies based on standard integrated-circuit fabrication technology. The total micro-STM size is about 100 μm, which is small enough to be integrated on a large scale.

6. CONCLUSION

As shown in this chapter, SPM-induced silicon surface oxidation is of prime interest for nanometer-scaled lithography. Many devices have already been demonstrated using SPM-based lithography, and several theoretical works have been devoted to the understanding and modelization of the SPM-induced oxidation mechanism(s). To realize prototypes (dots, wires, single electron transistors, etc.) with a nanometric size, such a technique is very low cost and powerful. It offers a unique way to study the physical properties of nanostructures. In the case of mass production, the low throughput is a major drawback. However, it can be partly overcome using a parallel system with up to a few thousand tips on the same support. Nevertheless, the poor tip reliability, even using pulsed voltage techniques, cannot be sufficiently increased for industrial applications. Only thermomechanical writing on polymer films is a possible route to low-cost and high-density data storage.

GLOSSARY

Atomic force microscope (AFM) Allows measurement of the topography of a surface with a nanometer resolution.

Etching During a technology process, the etching, wet or dry, allows the removal of materials in order to obtain 3-D structures.

Nanotechnology Technology using sub-100 nm resolution.

Oxydation Chemical reaction involving a material and oxygen.

Scanning probe microscope (SPM) Generic name for AFM or STM.

Scanning tunneling microscope (STM) Allows measurement of the topography and the local electronic structure of a surface.

Silicon on insulator (SOI) From bottom to top, materials made of a silicon substrate, an insulator, and a thin silicon top layer.

REFERENCES

1. J. A. Dagata, J. Schneir, H. H. Harary, C. J. Evans, M. T. Postek, and J. Bennet, *Appl. Phys. Lett.* 56, 2001 (1990).
2. J. A. Dagata, J. Schneir, H. H. Harary, J. Bennett, and W. Tseng, *J. Vac. Sci. Technol. B* 9, 1384 (1990).
3. S.-T. Yau, X. Zheng, and M. H. Nayfeh, *Appl. Phys. Lett.* 59, 2457 (1991).
4. N. Barniol, F. Pérez-Murano, and X. Aymerich, *Appl. Phys. Lett.* 61, 462 (1992).
5. T. Teuschler, K. Mahr, S. Miyazaki, M. Hundhausen, and L. Ley, *Appl. Phys. Lett.* 67, 3144 (1995).
6. H. Sugimura, T. Yamamoto, N. Nakagiri, M. Miyashita, and T. Onuki, *Appl. Phys. Lett.* 65, 1569 (1994).

7. H. C. Day and D. R. Allee, *Appl. Phys. Lett.* 62, 2691 (1993).
8. P. Fay, R. T. Brokenbrough, G. Abeln, P. Scott, S. Agarwala, and I. Adesida, *J. Appl. Phys.* 75, 7545 (1994).
9. M. Yasutake, Y. Ejiri, and T. Hattori, *Jpn. J. Appl. Phys.* 32, L1021 (1993).
10. H. Sugimura, N. Kitamura, and H. Masuhara, *Jpn. J. Appl. Phys.* 33, L143 (1994).
11. T. Hattori, Y. Ejiri, K. Saito, and M. Yasutake, *J. Vac. Sci. Technol. A* 12, 2586 (1994).
12. T. Teuschler, K. Mahr, S. Miyazaki, M. Hundhausen, and L. Ley, *Appl. Phys. Lett.* 66, 2499 (1995).
13. L. Ley, T. Teuschler, K. Mahr, S. Miyazaki, and M. Hundhausen, *J. Vac. Sci. Technol. B* 14, 2845 (1996).
14. F. Pérez-Murano, G. Abadal, N. Barniol, and X. Aymerich, *J. Appl. Phys.* 78, 6797 (1995).
15. J. Servat, P. Gorostiza, F. Sanz, F. Pérez-Murano, N. Barniol, G. Abadal, and X. Aymerich, *J. Vac. Sci. Technol. A* 14, 1 (1996).
16. P. A. Fontaine, E. Dubois, and D. Stiévenard, *J. Appl. Phys.* 84, 1776 (1998).
17. M. Tello and R. Garcia, *Appl. Phys. Lett.* 79, 424 (2001).
18. B. Legrand and D. Stiévenard, *Appl. Phys. Lett.* 74, 4049 (1999).
19. F. Pérez-Murano, K. Birkelund, K. Morimoto, and J. A. Dagata, *Appl. Phys. Lett.* 75, 199 (1999).
20. M. Calleja and R. Garcia, *Appl. Phys. Lett.* 76, 3427 (2000).
21. F. Marchi, V. Bouchiat, H. Dallaporta, V. Safarov, and D. Tonneau, *J. Vac. Sci. Technol. B* 16, 2952 (1998).
22. R. Garcia, M. Calleja, and F. Pérez-Murano, *Appl. Phys. Lett.* 72, 2295 (1998).
23. M. Luna, J. Colchero, and A. M. Baro, *Appl. Phys. Lett.* 72, 3461 (1998).
24. R. Garcia, M. Calleja, and H. Rohrer, *J. Appl. Phys.* 86, 1898 (1999).
25. E. S. Snow, G. G. Jernigan, and P. M. Campbell, *Appl. Phys. Lett.* 76, 1782 (2000).
26. A. E. Gordon, R. T. Fayfield, D. D. Litfin, and T. K. Higman, *J. Vac. Sci. Technol. B* 13, 2805 (1995).
27. D. Stiévenard, P. A. Fontaine, and E. Dubois, *Appl. Phys. Lett.* 70, 3272 (1997).
28. N. Cabrera and N. F. Mott, *Rep. Prog. Phys.* 12, 163 (1948).
29. P. Avouris, T. Hertel, and R. Martel, *Appl. Phys. Lett.* 71, 285 (1997).
30. J. A. Dagata, T. Inoue, J. Itoh, and H. Yokoyama, *Appl. Phys. Lett.* 73, 271 (1998).
31. J. A. Dagata, T. Inoue, J. Itoh, K. Matsumoto, and H. Yokoyama, *J. Appl. Phys.* 84, 6891 (1998).
32. J. A. Dagata, F. Pérez-Murano, G. Abadal, K. Morimoto, T. Inoue, J. Itoh, and H. Yokoyama, *Appl. Phys. Lett.* 76, 2710 (2000).
33. E. Dubois and J. L. Bubendorff, *J. Appl. Phys.* 87, 8148 (2000).
34. E. S. Snow, P. M. Campbell, and P. J. McMarr, *Appl. Phys. Lett.* 63, 749 (1993).
35. B. Legrand, D. Deresmes, and D. Stievenard, *J. Vac. Sci. Technol. B* 20, 1 (2002).
36. E. S. Snow, W. H. Juan, S. W. Pang, and P. M. Campbell, *Appl. Phys. Lett.* 66, 1729 (1995).
37. E. S. Snow and P. M. Campbell, *Appl. Phys. Lett.* 64, 1932 (1994).
38. K. Morimoto, K. Araki, K. Yamashita, K. Morita, and M. Niwa, *Appl. Surf. Sci.* 117/118, 652 (1997).
39. H. Sugimura and N. Nakagiri, *Nanotechnol.* 6, 29 (1995).
40. S. C. Minne, H. T. Soh, P. Flueckiger, and C. F. Quate, *Appl. Phys. Lett.* 66, 703 (1995).
41. T. Fayfield and T. K. Higman, *J. Vac. Sci. Technol. B* 13, 1285 (1995).
42. M. Ishii and K. Matsumoto, *Jpn. J. Appl. Phys.* 34, 1329 (1995).
43. P. M. Campbell, E. S. Snow, and P. J. McMarr, *Appl. Phys. Lett.* 66, 1388 (1995).
44. E. S. Snow and P. M. Campbell, *Science* 270, 1639 (1995).
45. Y. Wada, *Microelectron. J.* 29, 601 (1998).
46. S. R. Manalis, S. C. Minne, and C. F. Quate, *Appl. Phys. Lett.* 68, 871 (1996).
47. S. C. Minne, J. D. Adams, G. Yaralioglu, S. R. Manalis, A. Atalar, and C. F. Quate, *Appl. Phys. Lett.* 73, 1742 (1998).
48. P. Vettiger, M. Despont, U. Dreschler, U. Dürig, W. Häberle, M. I. Lutwyche, H. E. Rothuizen, R. Stutz, R. Widmer, and G. K. Binning, IBM *J. Res. Develop.* 44, 323 (2000).

www.aspbs.com/enn

Single-Electron Devices

Konstantin Likharev

Stony Brook University, Stony Brook, New York, USA

CONTENTS

1. Introduction
2. Basic Single-Electron Devices
3. Scaling and Implementation
4. Possible Applications
5. Conclusion
Glossary
References

1. INTRODUCTION

The phenomenal success of semiconductor electronics during the past three decades was based on the scaling down of silicon field effect transistors (MOSFET) and the resulting increase of density of logic and memory chips. The most authoritative industrial forecast, the *International Technology Roadmap for Semiconductors* [1] predicts that this exponential ("Moore-Law") progress of silicon MOSFETs and integrated circuits will continue at least for the next 15 years. By the end of this period, devices with 10-nm minimum features (gate length L) should become commercially available. Recent experimental and results theoretical studies give grounds to believe that (given a proper economic climate) this frontier may really be reached. (For recent reviews, see [2, 3].)

However, prospects to continue the Moore law beyond the 10-nm frontier are much more uncertain. On one hand, recent theoretical modeling [4] of the most promising MOSFET species (double-gate transistors with an ultrathin undoped channel) indicates that the devices with a gate as short as ~3 nm still may provide performance high enough for use in most logic and memory circuits. On the other hand, the same studies show that such nanoscale MOSFETs would be extremely sensitive to very small variations of their physical dimensions. Figure 1 shows a typical result of such modeling; it implies that to keep variations of the transistor threshold voltage V_t below an acceptable limit (~50 mV), the gate length of a 5-nm MOSFET should be controlled better than ~0.25 nm, much tighter than the farthest ITRS projection of 0.7 nm for the critical dimension control accuracy at the 3σ level. Such high sensitivity to parameter fluctuations, and, hence, to the fabrication process as a whole, may lead to skyrocketing of chip fabrication facilities costs, very high even now. As a result, the Si-MOSFET-based Moore law may stop at $L \sim 10$ nm, i.e., long before fundamental physical limits have been reached.

These prospects give a strong motivation for the search for alternative devices that could replace MOSFETs beyond the 10-nm frontier, providing comparable performance but allowing less expensive fabrication. This search should start with a careful examination of the basic physics of transistor-like electron devices. Such an examination has to address, first of all, the famous particle-wave duality. Quantum mechanics show that electrons may behave either as discrete particles or continuous de Broglie waves, depending on experimental conditions. (Surprisingly enough, a sufficiently clear understanding of these conditions for conduction electrons in solids was achieved not so long ago, in the 1980s; for reviews see, e.g., [5, 6].)

Consider, for example, a generic situation where two parts of a conductor are separated by some interface, and ask whether the electric charge of each conductor is a multiple of the fundamental charge e at any instant (this is natural for the particle picture) or may be continuous (the wave picture allows this, because the wavefunction of each electron may be split between the two parts). The answer to this question turns out to depend on whether the effective tunnel resistance R of the interface is larger or smaller than the natural quantum unit of resistance

$$R_Q \equiv \hbar/e^2 \approx 4.1 \text{ k}\Omega \tag{1}$$

If the resistance is low, $R \ll R_Q$, the charge of each conductor may be continuous, but, in the opposite limit, it may be only a multiple of e.

This relation may be derived and explained in numerous ways [4–6]; perhaps the simplest interpretation is as follows. In a closed ("Hamiltonian") quantum system, the characteristic energy of quantum fluctuations per degree of freedom is $E_Q \sim \hbar\omega/2$, where ω is a characteristic frequency. In contrast, each part of the conductor we are discussing, concerning its electric charge degree of freedom, is an "open" system, strongly interacting with its environment (in classics,

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (865–884)

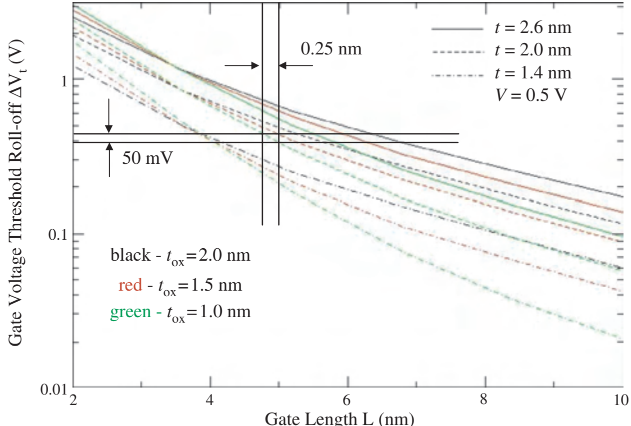


Figure 1. Threshold voltage roll-off (relative to that for $L \to \infty$) as a function of gate length L, for various combinations of oxide thickness t_{ox} and channel thickness t, calculated for a simple model of an "ultimate silicon MOSFET." The source-to-drain voltage V is fixed at a level typical for ON current saturation. Reprinted with permission from [4], V. A. Sverdlov et al., *IEEE Trans. Electron Devices* (2003). © 2003, IEEE.

corresponding to an RC relaxator, rather than an LC oscillator). For such a system, $\hbar\omega$ should be replaced by $\hbar/\tau$, with $\tau = RC$, where C is the capacitance between the two conducting parts. Transfer of a single electron between the parts causes an electrostatic energy change of the order of $E_C \sim e^2/2C$. Comparing E_C and E_Q, we notice that C cancels, and see that if $R \ll R_Q$, quantum fluctuations smear out the electrostatic energy difference that tries to keep the electric charge of each part constant.

To comprehend the importance of this result, let us combine it with the so-called Landauer formula for the interface conductance $G \equiv 1/R$ [5, 6]:

$$G = (e^2/\pi\hbar)\Sigma_i D_i \tag{2}$$

where D_i is the interface transparency (i.e., the probability of electron transmission) for a particular transversal mode of electron propagation; the sum is over all the modes. Comparison of Eqs. (1) and (2) shows that if the conductor cross section is so narrow that quantum confinement makes only one propagating mode possible, the condition of electric charge discreteness takes a simple and natural form: $D \ll 1$. However, for devices with a larger cross-section A, the restriction on the average transparency is much more severe: $D \ll 1/N$, where N is the transversal mode number. For a degenerate conductor, it is of the order of λ_F^2/A, where λ_F is the Fermi wavelength, typically of the order of 1 nm.

A MOSFET transistor is a good example of a device where the number n of electrons in the channel is never quantized, because the boundaries between the electrodes and channel are typically highly transparent ($D \sim 1$), and, in addition, the number of transversal modes is high ($N \gg 1$). As a result, these transistors do not exhibit single-electron charging effects (e.g., the Coulomb blockade) even if the average number of electrons in the channel is small. An adequate understanding of such devices may be achieved by using the wave language.

Theoretically, for nanoscale "wave" devices with $R \ll R_Q$, the FET-type control of transport is not the only possible mode of operation: the effects of quantum interference of electron de Broglie waves can be, as a matter of principle, used for this purpose as well. In the 1980s and early 1990s, numerous publications were devoted to the analysis and sometimes enthusiastic advertising of such "quantum electronic devices." However, later, the prospects for their practical applications (perhaps, with very few exceptions) have been recognized as rather dim, mostly for the following reason. In contrast to optical phonons (which obey Bose statistics), charge carriers in solids are fermions and, in particular, obey the Pauli principle: each of them must have a different energy, and, hence, a different de Broglie wavelength. Therefore, high-contrast interference patterns require operation with either a single transversal mode or a small number of modes.

The accuracy δL of size definition of the nanostructures that single out such a mode from a continuum, and then handle its interference, should be much better than the de Broglie wavelength $\lambda = h/(2mE)^{1/2}$ of the used electrons. Energy E should be well above the thermal fluctuation scale (typically, $\sim k_B T$) to avoid interference pattern smearing by thermal fluctuations. Combining these two requirements and plugging in fundamental constants, we may see that for room temperature, δL should be, as in nanoscale MOSFETs, well below 1 nm. This estimate shows that quantum interference devices do not have any substantial advantage over FETs for sub-10-nm scaling.

To summarize, besides some interesting memory cell ideas (which will be briefly discussed in Section 4.), in the category of "wave" electron devices ($R \ll R_Q$), we are still left essentially with field-effect transistors. In nanoscale structures with high impedance ($R \gg R_Q$), single-electron charging effects dominate and form the basis for single-electron devices. Their brief review will be the main subject of this article. (For earlier reviews see, e.g., [7–10].) Due to the space and time restrictions, I will not be able to cover the closely related superconductor, single-Cooper-pair devices [8–10] that may have some interesting applications in metrology and quantum computing, but can hardly make a large impact on broader electronics, due to the necessity of deep refrigeration (below the transition temperature of the superconductor material used).

2. BASIC SINGLE-ELECTRON DEVICES

2.1. Single-Electron Box

Physics of single-electron tunneling may be well explained on the example of the generic single-electron device, frequently called the "single-electron box" (Fig. 2a). The basic physics of this device was understood by Lambe and Jaklevic [11] as early as in 1969, on the basis of their experiments with disordered granular structures, while the first quantitative theory of the box was developed by Kulik and Shekhter [12].

The device consists of one small conductor ("island") separated from an external electrode by a tunnel barrier with high resistance,

$$R \gg R_Q. \tag{3}$$

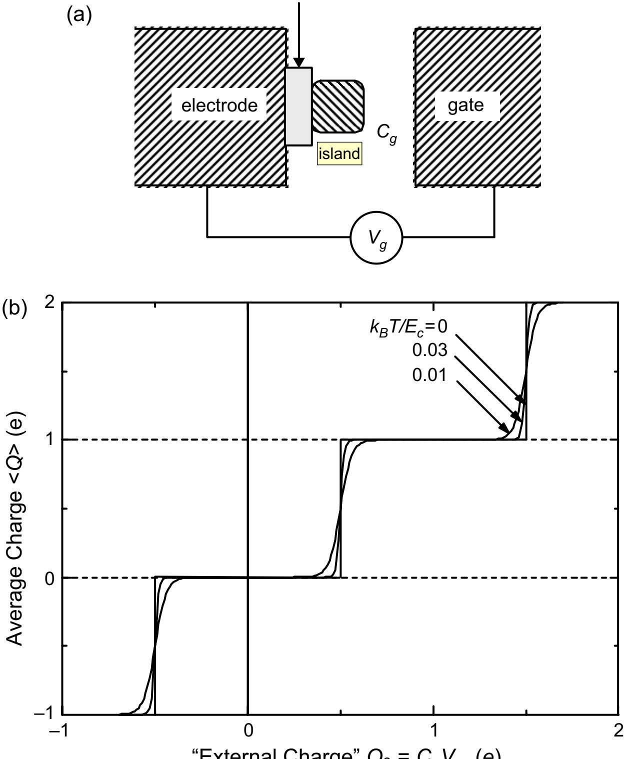


Figure 2. (a) Single-electron box and (b) its "Coulomb staircase," i.e., the steplike dependence of the average charge Q of the island on the gate voltage V_g, for several values of temperature.

An external electric field may be applied to the island by using a capacitively coupled gate electrode. The field changes the local Fermi level of the island and thus determines the conditions of electron tunneling. Elementary electrostatics shows that the energy of the system may be presented as

$$W = Q^2/2C_\Sigma + (C_0/C_\Sigma)QV_g + \text{constant} \quad (4)$$

where $Q = -ne$ is the island charge (n is the number of uncompensated electrons), C_g is the island-gate capacitance, while C_Σ is the total capacitance of the island (including C_g). It is convenient to rewrite Eq. (4) in another form,

$$W = (Q_0 - ne)^2/2C_\Sigma + \text{constant}, \quad Q_0 \equiv C_gV_g \quad (5)$$

where parameter Q_0 usually is called the "external charge." From its definition, it is evident that in contrast with the discrete total charge Q of the island, the variable Q_0 is continuous and may be a fraction of the elementary charge e.

At sufficiently low temperatures,

$$k_BT \ll E_C, \quad E_C \equiv e^2/C_\Sigma \quad (6)$$

the stationary number n of electrons in the island corresponds to the minimum of W; an elementary calculation using Eq. (5) shows that Q is a steplike function of Q_0, i.e., of the gate voltage (Fig. 2b), jumping by e when

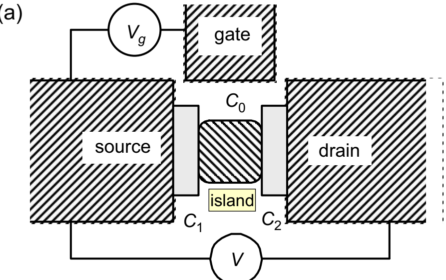
$$Q_0 = e\left(n + \frac{1}{2}\right), \quad n = 0, \pm 1, \pm 2, \ldots \quad (7)$$

If the temperature is increased to $k_BT \sim E_c$, the system has nonvanishing probability p_n will be in other states as well, with $p_n = \exp\{-W(n)/k_BT\}/\Sigma_n \exp\{-W(n)/k_BT\}$. A straightforward calculation of the average charge $\langle Q\rangle = -\Sigma_n enp_n$ yields the pattern shown in Fig. 2b: the steplike dependence of charge on gate voltage is gradually smeared out by thermal fluctuations. This is typical for all single-electron devices, so that the operation temperature of most of them should satisfy Eq. (6). (Notable exception are single–electron temperature standards—see Section 4.1.)

The physics of this "Coulomb staircase" is very simple: increasing gate voltage V_g tries to attract more and more electrons to the island. The discreteness of electron charge tunneling through low-transparency barriers (with $R \gg R_Q$, see the Introduction) necessarily makes this increase steplike.

2.2. Single-Electron Transistor

The single-electron box per se is still not a valuable device, but its simple modification, splitting its electrode into two parts, source and drain (Fig. 3a), turns it into a very important device, the single-electron transistor. This device, that was first suggested in 1985 [13, 14] and first implemented 2 years later [15], is clearly reminiscent of a MOSFET but with a small conducting island limited by two tunnel barriers, instead of the usual channel, connecting source and drain.



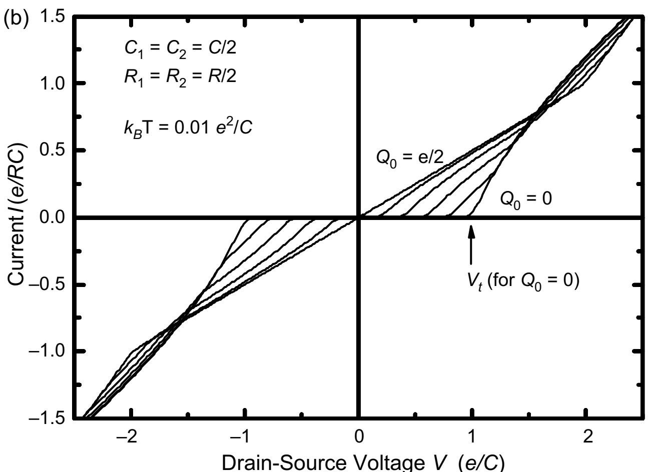


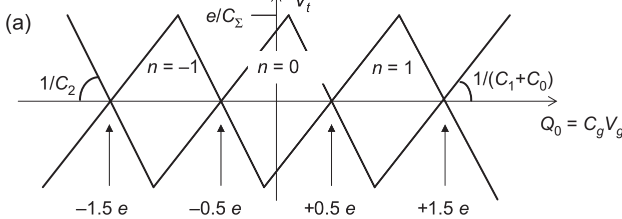
Figure 3. (a) Single-electron transistor and (b) a typical set of source–drain I–V curves of a symmetric transistor for several values of Q_0, i.e., of the gate voltage V_g, calculated by using the "orthodox" theory of single-electron tunneling.

Figure 3b shows a typical set of direct current (dc) I–V curves of such a transistor, for several values of the "external charge" Q_0, defined in the same way as in the single-electron box [see Eq. (5)]. One can see that at small drain-to-source voltage V, there is virtually no current, besides the special values of Q_0 defined by Eq. (7). The physics of this phenomenon (the "Coulomb blockade") is easy to understand: even if $V > 0$, and, thus, it is energy-advantageous for a electron to go from source to drain, on its way the electron has to tunnel into the island first and change its charge Q it by $\Delta Q = -e$. Such charging would increase the electrostatic energy W of the system

$$W = (Q_0 - ne)^2/2C_\Sigma - eV(n_1C_2 + n_2C_1)/C_\Sigma + \text{constant},$$
$$C_\Sigma \equiv C_0 + C_1 + C_2 \qquad (8)$$

(where n_1 and n_2 are the number of electrons passed through the tunnel barriers 1 and 2, respectively, so that $n = n_1 - n_2$), and, hence, at low enough temperatures ($k_BT \ll E_C$), the tunneling rate is exponentially low.

At a certain threshold voltage V_t, the Coulomb blockade is overcome, and current starts to grow with V. The most important property of the single-electron transistor is that V_t is a periodic function of V_g, vanishing at special values of gate voltage, given by Eq. (27) (see Fig. 4). The reason for these so-called "Coulomb blockade oscillations" is evident from the previous discussion of the single-electron box: in the special points (7), one electron may be transferred to the island from either drain or source without changing the electrostatic energy of the system, even at $V = 0$. Hence, an electron can tunnel from the source to the island and then to the drain even at negligible V, so that $V_t = 0$. As can be readily shown from Eq. (8), at low temperatures the dependence of V_t on V_g is piecewise-linear, with its lower and upper branches forming the so-called "diamond diagram" (Fig. 4).



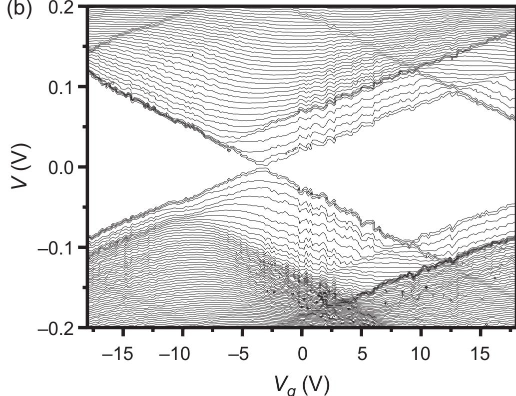


Figure 4. Threshold characteristics of single-electron transistors: (a) as given by the "orthodox" theory for $T \to 0$, and (b) experimental contour plots of constant current for an Al/AlO$_x$ transistor with $e/C_\Sigma \approx 0.1$ V, measured at $T = 4.2$ K. Reprinted with permission from [16], Y. A. Pashkin et al., *Appl. Phys. Lett.* 76, 2256 (2000). © 2000, American Institute of Physics.

Since the I–V curves of the transistor are continuous (Fig. 3b), if a small current is fixed by an external circuit (say, by a large resistor in series with the SET), V is close to V_t and also follows the diamond diagram (e.g., Fig. 4b). Thus, the voltage gain and transconductance of single-electron transistors may change sign depending on the gate voltage—an important difference in comparison with the usual field-effect transistors. The last mentioned property is very convenient for the design of SET analogs of (CMOS) circuits (see Section 5.3).

On the other hand, the same diamond diagram shows that the voltage gain is limited by a capacitance ratio: $(G_V)_{\max} = C_0/C_2$ [14]. It may be larger than unity (see, e.g., experiments [17–19]), but hardly much larger than that, especially in room-temperature transistors.

2.3. Single-Electron Trap

Another key device, the "single-electron trap," also may be understood as a generalization of the single-electron box. Let us replace the single tunnel junction in Fig. 2a with a one-dimensional (1D) array of $N > 1$ islands separated by tunnel barriers (Fig. 5a). (This device was first discussed explicitly in 1991 [20, 21], but, in fact, it may be considered just a particular operation mode of a more complex device, the single-electron turnstile, invented earlier [22].) The main

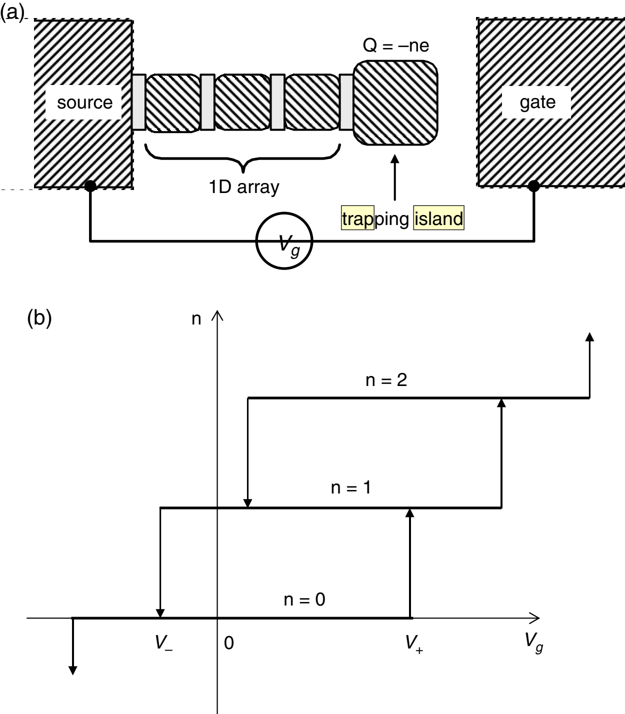


Figure 5. (a) Single-electron trap and (b) its static characteristic at $T \to 0$ (schematically).

new feature of this system is its internal memory (bi- or multistability): within certain ranges of applied gate voltage V_g, the system may be in one of two (or more) charged states of its edge island (Fig. 5b).

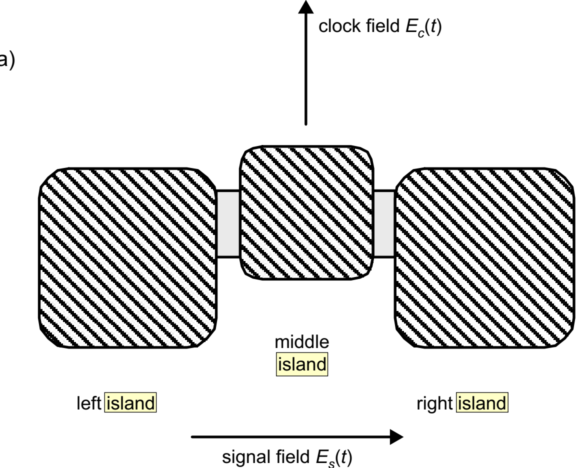
The reason for this multistability stems from the peculiar properties of an electron inside a 1D array; an electron located in one of the islands of the array extends its field to a certain distance [23, 24] and, hence, may interact with the array edges (is attracted to them). As a result, the electrostatic self-energy of the electron has a maximum in the middle of the array. By applying sufficiently high gate voltage $V_g = V_+$, the energy profile may be tilted enough to drive an electron into the edge island; if the array is not too long, other electrons feel its repulsion and do not follow. If the gate voltage is subsequently decreased to the initial level (0), the electron is still trapped in the edge island, behind the energy barrier. To remove the electron from the trap, the voltage has to be reduced further, to $V_g < V_- < 0$. As a result, the $n(V_g)$ dependence exhibits regions of bi- or multistability, in which the charge state of the trap depends on its prehistory (Fig. 5b).

The retention time of a certain charge state within the multistability region is fundamentally limited by the thermal activation over the energy barrier and a higher-order quantum process, cotunneling [25]. The first effect is exponentially low in E_c/k_BT, while the second effect falls exponentially with the array length N. As a result, electron retention time may be very long; experimentally, single electron trapping for more than 12 hr (at temperatures below 1 K) has been demonstrated [26–28].

2.4. Single-Electron Parametron

Single-electron parametron [29] is essentially a short segment of a 1D array of islands, galvanically detached from electron sources. The simplest version of the device uses three small islands separated by two tunnel barriers (Fig. 6a). For simplicity, I will describe its operation for the case when the system is charged by one additional electron, though such precharging is, in fact, unnecessary [30].

Let the parametron be biased by a periodic "clock" electric field $\boldsymbol{E}_c(t)$, oriented vertically, for example, by a slight vertical shift of the central island. (The same effect may be achieved in a strictly linear array, by using a special gate located closer to the central island [30].) This field keeps an extra electron in the central island during a part of the clock period. At some instant, the field E_c reaches a certain value E_t (Fig. 6b) at which electron transfer to one of the edge islands becomes energy advantageous. If the system were completely symmetric, the choice between the two edge islands would be random, i.e., the system would undergo what is called "spontaneous symmetry breaking." However, even a small additional field $\boldsymbol{E}_s$ applied by a similar neighboring device(s) may determine the direction of electron tunneling at the decision-making moment. Once the energy barrier created by the further change of the clock field has become large enough, the electron essentially is trapped in one of the edge islands, and the field $\boldsymbol{E}_s$ may be turned off. Now the device itself may serve as a source of the dipole signal field $\boldsymbol{E}_s$ for the neighboring cells. The sign of this field



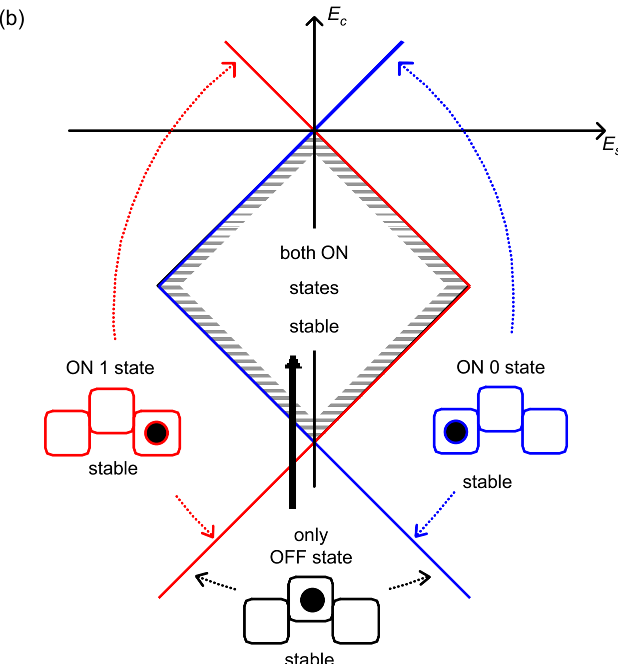


Figure 6. (a) Single-electron parametron (adapted from [29]) and (b) its phase diagram (adapted from [30]).

(i.e., of the electric dipole moment of the device) presents one bit of information.

Figure 6b shows the phase diagram of the parametron [30], which gives a quantitative description of the properties described previously. Each of "ON" states (with the extra electron in one of the edge islands) is stable within an angle-limited region. These regions overlap, providing the (hatched) diamond-shaped region of cell bistability. The route from "OFF" state (with the electron in the central island) to one of the ON states, as described, in shown by the bold arrow.

Recently, the parametron was experimentally demonstrated by two groups [31, 53] (albeit in one case under a different name).

3. SCALING AND IMPLEMENTATION

The most important conditions for operation of single-electron devices are given by Eqs. (3) and (6). The former condition, $R \gg \hbar/e^2 \approx 10$ kΩ, is relatively easy to satisfy experimentally, since resistance R of a tunnel barrier grows exponentially with its thickness. This relation shows that the output resistance of single-electron devices is unavoidably high, practically above ~100 kΩ. This feature frequently is cited as a drawback. However, what really is important for applications (e.g., interconnect recharging speed) is the maximum current density per unit width. For a room-temperature SET with $V \sim V_t \sim E_C/e \sim 1$ V, the total current would be below 10 μA. However, since the width of such transistor has to be very small, ~1 nm (see following), the available current density may be well above 1000 μA/μm, i.e., even higher than that of standard silicon MOSFETs [1].

The latter condition (6) makes the practical implementation of single-electron transistors operating at room temperature rather problematic. Figure 7 shows the energy E_a necessary to put an additional electron on a transistor island of a certain radius, calculated within the framework of a simple model. This energy is crudely a sum of the electrostatic contribution $E_C = e^2/C$, where C is the island capacitance, and quantum confinement energy E_k. While E_C dominates for relatively large islands, for 1-nm-scale islands with a size comparable with the electron de Broglie wavelength, E_k becomes substantial. Notice that at $E_k \gtrsim E_C$, the basic ("orthodox") theory of single-electron tunneling [8] (used in particular to calculate plots in Fig. 3b) should be modified [32, 33] (see also review [34]), although the device implications do not change significantly.

Both theory and experiment show that single-electron tunneling effects (i.e., some current modulation by gate voltage in single-electron transistors) become visible at $E_a \sim 3k_BT$.

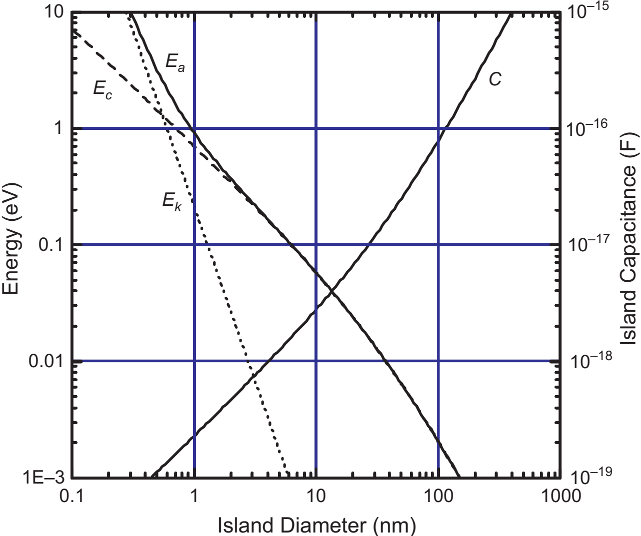


Figure 7. Single-electron addition energy E_a (solid line), and its components: charging energy E_C (dashed line) and electron kinetic energy E_k (dotted line), for a simple model of a conducting island. Reprinted with permission from [10], K. K. Likharev, *Proc. IEEE* 87, 606 (1999). © 1999, IEEE.

It means that to notice these effects at $T \sim 100$ mK (the temperature routinely reached in standard helium-dilution refrigerators), E_a should be above ~25 μeV, corresponding to the island capacitance $C \sim 5 \times 10^{-15}$ F and an island size of the order of 1 micron, with a tunnel junction area ~0.1 × 0.1 μm^2. Such dimensions can be reach routinely by several methods, including, notably, the now-classical technique of metal evaporation from two angles through a hanging resist mask (most typically formed by direct e-beam writing on a double-layer resist). This method was used, in particular, for the first experimental demonstration of a single-electron transistor [15]; later this technique, complemented with subsequent island oxidation, has been advanced to increase E_a to ~0.1 eV in certain samples [16]. This is sufficient to see a slight current modulation at room temperature.

However, for reliable operation of most digital single-electron devices (for exceptions, see Section 4.3.), the single-electron addition energy should be at least 100 times larger than k_BT. This means that for room temperature operation, E_a should be as large as ~3 eV. According to Fig. 7, this value corresponds to island size about 1 nm. Reproducible fabrication of integrated circuits with features so small presents quite a challenge. Most claims of success in this direction are based on studies of evidently irreproducible structures, for example, arrays of nanoparticles deposited between rather distant source and drain. (By chance, one of these particles may be tunnel-coupled to its neighbors much more weakly than are others, forming a single-electron transistor island, while strongly coupled particles effectively merge, forming its source and drain.) Of course, the parameters of such transistors are unpredictable, and there is no hope of using them in integrated circuits.

A more interesting option is fabrication of discrete transistors with scanning probes, for example, by nano-oxidation of metallic films [35, 36] or manipulation with carbon nanotubes [37, 38]. For the former devices, single-electron addition energies as high as 1 V have been reached [36], but unfortunately, the current in these transistors is very low (below 10^{-11} A). Moreover, these techniques are so slow that prospects for their use in the fabrication of circuits of any noticeable integration scale are very slim, though for discrete devices (e.g., Section 4.4.), this approach may be promising.

Several methods that are closer to standard CMOS technology also have been used to fabricate single-electron transistors, mostly by the oxidation of a thin silicon channel until it breaks into one or several tunnel-coupled islands (see the pioneering work [39] and recent results [40–43]). The highest values of E_a reached in those efforts (~250 mV [43]) are still not high enough for digital applications, but the necessary values $E_a \sim 3$ eV will probably be reached soon. A much larger problem with this approach is that the parameters of the resulting transistors are highly irreproducible, and there is little chance of changing this situation without the development of patterning technologies with a sub-nm resolution—a very distant goal indeed. (The use of carbon nanotubes [44] can hardly solve this problem.)

I believe that virtually the only promising way toward integrated circuits when using single-electron tunneling is the chemically directed self-assembly [45, 46] of molecular single-electron devices on prefabricated nanowires.

First, single-molecule single-electron transistors operating at room temperature already have been demonstrated experimentally [47–49]. Figure 8a shows the molecules used in experiments [49]. The obtained I–V curves (Fig. 8c) closely resemble the theoretical curves shown in Fig. 3b (though the adequate qualitative theory of these structures should depart considerably from the "orthodox" theory, by taking into account the electron motion quantization effects [32–34]). The main problem now is to synthesize molecules that would combine suitable device characteristics with the ability to self-assemble, with high yield, on a few nm gaps between prefabricated nanowires.

This approach can provide very high areal density of devices only if the wire pitch (wire width plus distance to the next wire) is of the order of the molecular device size. This would be a very tough task for usual lithography; however, if we put severe restrictions on the wire geometry, e.g., limit it by several layers of parallel wires, then special lithographic methods may be us, e.g., nanoimprinting [50] or an extension of the laser-assisted deposition techniques [51] to shorter wavelength (either EUV [52] or soft X-rays).

The nanowires may allow molecular single-electron devices to connect to larger and, hence, more sparse silicon nano-MOSFET devices (Fig. 9 [3]). The general idea of such "CMOL" (CMOS/nanowire/molecular) hybrid circuits is to combine the advantages of single-electron devices (very high density) with those of advanced field-effect transistors (high voltage gain and high reproducibility). A detailed discussion of this opportunity may be found in [3].

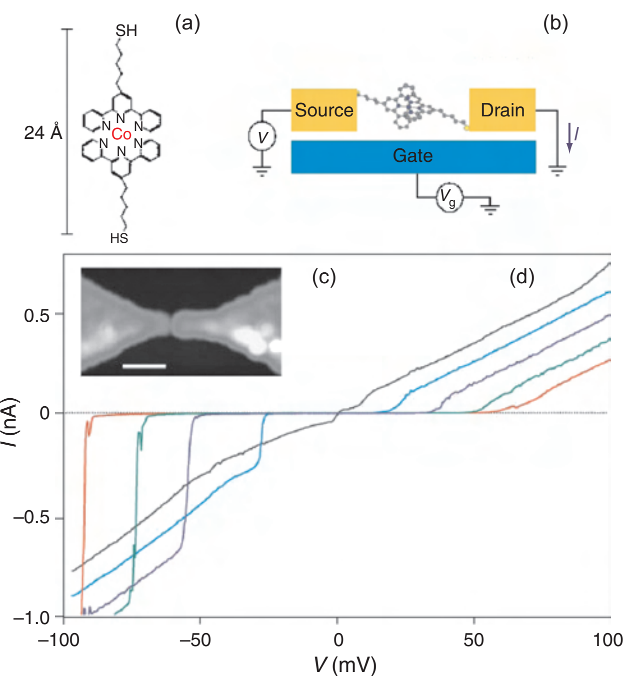


Figure 8. A single-molecular implementation of the single-electron transistor: (a) the used molecules; (b) general scheme of the experimental structure, (c) SEM picture of the gold electrodes, and (d) family of dc I–V curves recorded for several values of the gate voltage. Reprinted with permission from [49], J. Park et al., *Nature* 417, 722 (2002). © 2002, Macmillan Magazines Ltd.

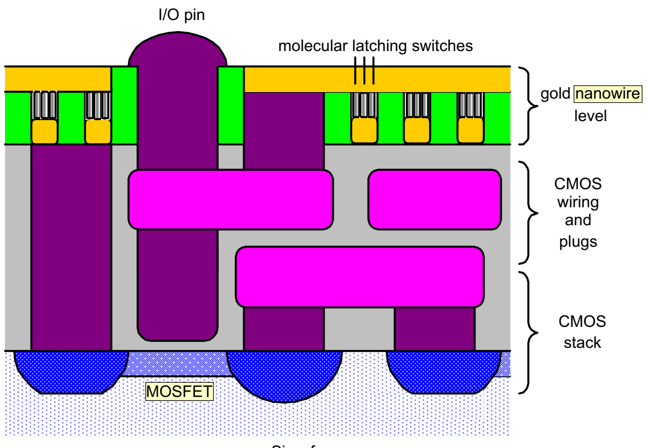


Figure 9. The CMOL circuit concept [3].

4. POSSIBLE APPLICATIONS

4.1. Analog and Metrology Applications

4.1.1. Supersensitive Electrometry

If the source-drain voltage V applied to a single-electron transistor is slightly above its Coulomb blockade threshold V_t, source-drain current I of the device is extremely sensitive to the gate voltage V_g. In fact, Figures 3 and 4 show that even the changes δV_g, corresponding to sub-single-electron variations δQ_e of the external charge, lead to measurable variations of I. This extremely high sensitivity has been suggested as a basis for supersensitive electrometry [14]. Calculations based on the orthodox theory have shown that the sensitivity of such an electrometer is not very impressive if the internal capacitance C_i of the signal source is large, on the scale of e^2/k_BT (at sub-1-K temperatures, above $\sim$0.1 pF) [66, 67]. On the other hand, if the source is so small and so close to the single-electron transistor that C_i, of the order of the capacitances of the transistor $C_{1,2}$, white noise limits the charge sensitivity at an extremely low level, $(\delta Q_e)_{\min}/\Delta f^{1/2} \sim 10^{-6} e/\sqrt{\text{Hz}}$, for the devices with routinely achieved 100-nm-scale islands (here Δf is the measurement bandwidth). This is some 7 orders of magnitude better than sensitivity of the best commercially available instruments, and about 4 orders of magnitude more sensitive than specially designed low-temperature MOSFETs (e.g., [83]).

Tunnel barriers and the electrostatic environment of single-electron devices always contain electron trapping centers and other two-level systems, each capable of producing "telegraph noise"—discrete, random low-frequency variations of the barrier conductance. An ensemble of these variations, with exponentially broad distribution of parameters, produces excess $1/f$-type noise; in single-electron transistors, such noise may be very high, typically limiting the charge resolution at the level of the order of $10^{-4} e/\sqrt{\text{Hz}}$ for a few Hertz signal frequencies. The sensitivity may be reduced radically (to a level below $10^{-5} e/\sqrt{\text{Hz}}$ [68]) by using special stacked geometry in which the single-electron island is lifted over the substrate that apparently hosts most of the $1/f$ noise sources. Another way to reach a similar sensitivity is to sense the transistor parameter change at a GHz-range

frequency [69, 70]. Such modulation cannot beat the white-noise imposed limitation of sensitivity [71] but helps to avoid most of the $1/f$ noise.

The high sensitivity of single-electron transistors has already enabled several groups to use them as electrometers in unique physical experiments. For example, they have made possible unambiguous observations [72–75] of parity effects in superconductors. The transistors also have been used in the first measurements of single-electron effects in single boxes and traps (e.g., [20, 26–28]). A modified version of the transistor has been used for the first proof of the existence of fractional-charge excitations in the fractional quantum Hall effect [76]. Another example is the sensitive measurement of local variations in the chemical potential of two-dimensional (2D) electron gas in GaAs/AlGaAs heterostructures [77, 78].

Further applications of these unique electrometers in scientific experimentation are imminent. For example, it has been suggested to use them, together with superconductor sensors, for submillimeter-wave photon counting [79]. A prototype detector of this kind for far-infrared photons (wavelength around 200 μm) has already been demonstrated [80].

Another exciting field is the fabrication of a single-electron transistor on the tip of a scanning probe and, with its help, implementation of a new type of scanning microscopy, combining submicron spatial resolution with sub-single-electron sensitivity [81]. This novel technique has already been used to observe single-charged impurities in GaAs/AlGaAs heterostructures.

4.1.2. Single-Electron Spectroscopy

Another useful spin-off of single-electron electrometry is the possibility of measuring the electron addition energies (and, hence, the energy-level distribution) in quantum dots and other nanoscale objects, pioneered by McEuen et al. [82].

There are two natural ways to carry out such measurements. The first is to use the quantum dot as the island of the single-electron box, capacitively coupled to the single-electron transistor or other sensitive electrometer [83]. The second is to use the quantum dot directly as the island of a weakly biased single-electron transistor ($V \to 0$) and measure the specific values V_n of the gate voltages V_g, which provide peaks of the source-drain conductance G (see Eq. (7) and its discussion). The latter method is technically simpler (only one small island is needed). Moreover, the height of the conduction peak gives additional information on the corresponding electron wavefunction distribution in the dot (e.g., [84]).

Single-electron spectroscopy has been applied to numerous objects, including laterally confined "puddles" of 2D electron gas in semiconductor heterostructures with lateral or vertical tunneling to source and drain, nanometer metallic clusters, naturally formed nanocrystals, C_{60} buckyballs, carbon nanotubes, and organic macromolecules and clusters. Presently single-electron spectroscopy is developing virtually independently of other applications of single-electron devices (for a detailed review, see [34]).

4.1.3. dc Current Standards

Another possible application of single-electron tunneling is fundamental standards of dc. The first suggestion for such a standard [13] followed the prediction that simple single-electron systems (e.g., a single small tunnel junction biased by a fixed dc I from a high-resistive source) should exhibit the so-called (SET) oscillations with frequency $f_{SET} = I/e$. Phase locking of these relaxation oscillations with an external radio frequency (rf) source of a well characterized frequency f would provide the transfer of a certain integer number m of electrons per period of external rf signal and thus generate dc, which is fundamentally related to the frequency as $I = mef$.

Despite the experimental confirmation of the existence of SET oscillations and the possibility of their phase locking [85], it turned out that better dc accuracy may be provided by other devices, the so-called single-electron turnstiles [22] and pumps [86], which do not exhibit oscillations in the autonomous mode. The following theoretical and experimental work has shown that better dc accuracy may be provided by the pumps, since these devices allow a more "gentle" (quasi-reversible) single-electron transfer. The best experimental effort so far gave a relative current accuracy of 1.5×10^{-8} [87], presumably limited by the photo-excitation of extra electrons by some electromagnetic radiation creeping into the system.

Unfortunately, the output current of single-electron standards is very limited (presently in the pA range), preventing a much broader application of these devices in metrology. A direct increase of the drive frequency f in pump-type devices runs into the problem of rapidly growing dynamic error rate. One way [10] to solve this problem is to use nanometer-scaled versions of charge-coupled devices [88] that would carry just a small, fixed number of electrons in each potential well, due to their Coulomb repulsion [89]. The first experimental step toward implementation of such devices has already been made [90].

4.1.4. Temperature Standards

An absolute temperature standard based on properties of 1D single-electron arrays has been proposed by J. Pekola and his collaborators [91–94]. At low temperatures, arrays with $N \gg 1$ islands exhibit dc I–V curves generally similar to those of single-electron transistors (see, e.g., Fig. 3b), with a clear Coulomb blockade of tunneling at low voltages ($|V| < V_t$) and approaching the linear asymptote $V = NRI+$ constant at $|V| \gg V_t$. If the temperature is raised above E_c/k_B, thermal fluctuations smear out the Coulomb blockade, and the I–V curve is almost linear at all voltages: $G \equiv dI/dV \approx G_n \equiv 1/NR$. The only remaining artifact of the Coulomb blockade is a small dip in the differential conductance around $V = 0$, with FWHM width $\Delta V = 5.44\ Nk_BT/e$.

Theoretical analyses have shown [92, 93] that this relation is surprisingly stable with respect to variations of almost any of the array parameters, providing a remarkable opportunity to use the arrays for absolute thermometry, since the fundamental constants are known with high accuracy. Each particular array may give a high ($\sim$1%) accuracy of T within less than 1 decade of temperature variations, but for arrays

with different island size (and, hence, different E_C), these ranges may be shifted and overlap. Thus it is possible to have an absolute standard of temperature with a very broad (say, 2 decade) total range from several circuits fabricated on a single chip.

This development is very encouraging, but further engineering work is needed to see whether these new devices will be able to compete with (or even replace) the more established temperature standards.

4.2. Random Background Charge Problem

The analog applications of single-electron tunneling previously discussed are very interesting, but, in terms of contemporary (predominantly digital) electronics, these are minor-niche applications at most. From the very formulation of single-electronics as an applied discipline [7, 95], the most intriguing question was whether single-electron devices will be able to compete with semiconductor transistors for use in digital integrated circuits. Any realistic discussion of this opportunity cannot ignore one more serious problem of single-electron devices.

Let a single-charged impurity be trapped in the insulating environment, for example on the substrate surface, at a distance r from a single-electron island, comparable to its size a. The impurity will polarize the island, creating on its surface a polarization ("image") charge of the order of $e(a/r)$. This charge effectively is subtracted from the external charge Q_0 [e.g., Eq. (5)]. This charge affects all characteristics of single-electron devices, for example, in the single-electron transistor, it determines the Coulomb blockade threshold V_t (see Fig. 4). For $r \sim a$, this shift may be large, in the order of $(V_t)_{\max}$, even from a simple impurity. By using even the most optimistic estimate compatible with experimental data, 10^9 cm^{-2} [54], for the minimum concentration of charged impurities, and assuming 1-nm island size, we can estimate that at least a 10^{-3} part of the chip area would be "poisoned," so that the same fraction of single-electron devices will have an unacceptably large background charge fluctuation, $\delta Q_0 \gtrsim 0.1\ e$.

A possible way to circumvent this problem is the use of single-electron transistors with resistive (rather than capacitive) coupling [14, 55, 56], which are insensitive to background charge. These devices require "ohmic" resistors with very high resistance (above $\sim$1 MΩ) and a quasi-continuous ("subelectron") transfer of charge that would provide the compensation for the fractional part of the random background charge. Such resistors have been demonstrated (e.g., [57]), however, their implementation for room temperature operation presents a problem. In fact, theoretical analyses [58–60] show that to provide the continuous transfer of charge, a diffusive conductor has to be much longer than the electron–phonon interaction length. For most materials at room temperature, this length is well above 10 nm (e.g., [61]), i.e., much larger than the desirable size of the whole device. Moreover, the stray capacitance of such a resistor would be much larger than that of the island itself, reducing its single-electron charging energy and making room-temperature operation impossible. As a matter of principle, electron hopping in quasi-insulators may ensure higher resistance at smaller resistor length; however, results of a recent analysis of quasi-continuous transport in this regime [62] also leads to pessimistic conclusions concerning the resistor size. For these reasons, prospects for room-temperature operation of resistively coupled single-electron devices do not look optimistic.

4.3. Memory Cells

The effect of random background charge randomness is easier to avoid in regular circuits, such as matrices of semiconductor memories [63, 64], that allow an access to each cell by selecting the corresponding "horizontal" (word) and "vertical" (bit) lines.

The published suggestions for single-electron memory cells are based mainly on two approaches:

- By using various modifications of the single-electron trap (Fig. 5) with either MOSFET or single-electron transistor readout,
- Direct scaling down of the cells of usual nonvolatile (e.g., flash) memories [64].

The latter approach includes an interesting proposal [96, 97] (see also [98–104]) for relatively large (multi-electron) memory cells, by using silicon nanocrystals rather than a single floating gate, in usual nonvolatile memories. The main advantage of this idea is that a single leaking defect in the tunneling barrier would not ground the whole stored charge but only its minor ($\sim 1/N$) part. This may allow very thin tunnel barriers (which would be unreliable in the ordinary case) to be used and thus decrease the characteristic time of Fowler-Nordheim tunneling, which presents the lower bound for the write–erase cycle. A potential drawback of nanocrystalline floating gates is that the electric field of the surface of each crystal, which determines the Fowler-Nordheim tunneling rate, depends on the size and exact shape of the crystal and is basically unpredictable. This may provide an undesirable broadening of the statistics of the electric field at the surface of nanocrystal surfaces and, hence, of write–erase thresholds, especially at any attempt to scale the cells to 10-nm scale size.

Returning to the two basic options outlined above, the physics of these two approaches is not much different: in both of them, insertion and extraction of a single electron to the trapping island ("floating gate"), i.e. write and erase operations, are achieved by the field suppression of the potential barrier separating the island from the electron source (word line). The only difference is that in the former case, the barrier is created by the Coulomb repulsion of electrons in a short 1D tunnel junction array, while, in the latter case, the barrier physics is the usual conduction band offset a single thick tunnel barrier (e.g., $\sim$8 nm of SiO_2).

Numerous experiments with single-electron cells of both types were useful for the development of the field; in particular, room-temperature operation of single cells has been demonstrated by several groups [65, 105–107]. However, because of the background charge randomness (see Section 4.2.), these cells can hardly be sufficiently reproducible. In fact, a single-charge impurity near the floating gate has an effect equivalent to an addition of (positive or negative) external charge $\Delta Q_0 \sim e$, and thus shifts the threshold for both write–erase, and readout operations from their

nominal values rather considerably. In the trap-type cells, the same effect, in addition, may change the array switching thresholds randomly. (Memory cells based on single- [or few-] electron trapping in grains of nanocrystalline MOSFET channels [108, 109] have even larger threshold spreads due to random locations of the grains and random transparency of tunneling barriers.)

However, due to the regular structure of memory arrays, several ways of avoiding the random background charge effects are available. The idea suggested first was to use the periodic character of the threshold characteristic of single-electron transistors for the cell contents readout [110]. In this approach, the memory structure may be very simple (Fig. 10a). Binary 1 is stored in a relatively large floating gate in the form of a positive charge $Q = Ne$, with $N \sim 10$, while binary 0 is presented by a similar negative charge. (Since $N \gg 1$, the effect of a random background charge on the floating gate is negligible.) The write–erase process is achieved by Fowler-Nordheim tunneling through a barrier separating the floating gate from the word line. Readout is destructive and combined with WRITE 1 operation—if the cell contents was 0, during the WRITE 1 process the injected electrons ramp up the electric potential U of the floating gate, so that the external charge Q_0 of the readout single-electron transistor is ramped up by $N'e$, with $1 < N' < N$. Due to the fundamental periodicity of the transconductance (Fig. 4), this ramp-up causes N' oscillations of the transistor current. (If the initial charge of the floating gate corresponded to binary 1, the transistor output is virtually constant.) The current creates oscillations of voltage between two bit lines connected to SET source and drain. These oscillations are picked up, amplified, and rectified by an FET sense amplifier; the resulting signal serves as the output. The main idea behind this device is that the random background charge will cause only an unpredictable shift of the initial phase of the current oscillations, which does not affect the rectified signal.

This concept has been verified experimentally [111] by using a low-temperature prototype of the memory cell. A very attractive feature of such SET–FET hybrid approach is a relatively mild minimum feature requirement: room temperature operation is possible with an electron addition energy of about 250 meV. Figure 7 shows that this level requires a minimum feature (SET island) size of about 3 nm, i.e., much larger than that required for purely single-electron digital circuits. The reason for this considerable relief is that in this hybrid memory, the single-electron transistor works essentially in the analog mode, as a sense preamplifier–modulator, and can tolerate a substantial rate of thermally activated tunneling events. One drawback of this memory is the need for an FET sense amplifier–rectifier. However, estimates show that since its input signals have already been pre-amplified with the SET, one FET amplifier may serve several hundred memory cells within the ordinary NOR architecture [63, 64], and, hence, the associated chip real estate loss per bit is minor. The next drawback is more essential: the signal charge swing $\Delta Q_0 = N'e$ at the single-electron island should exceed e. Since the gate oxide should not be too leaky to sustain acceptable retention time of the floating gate, its thickness should be at least a few nanometers. For silicon dioxide, this gives a specific capacitance

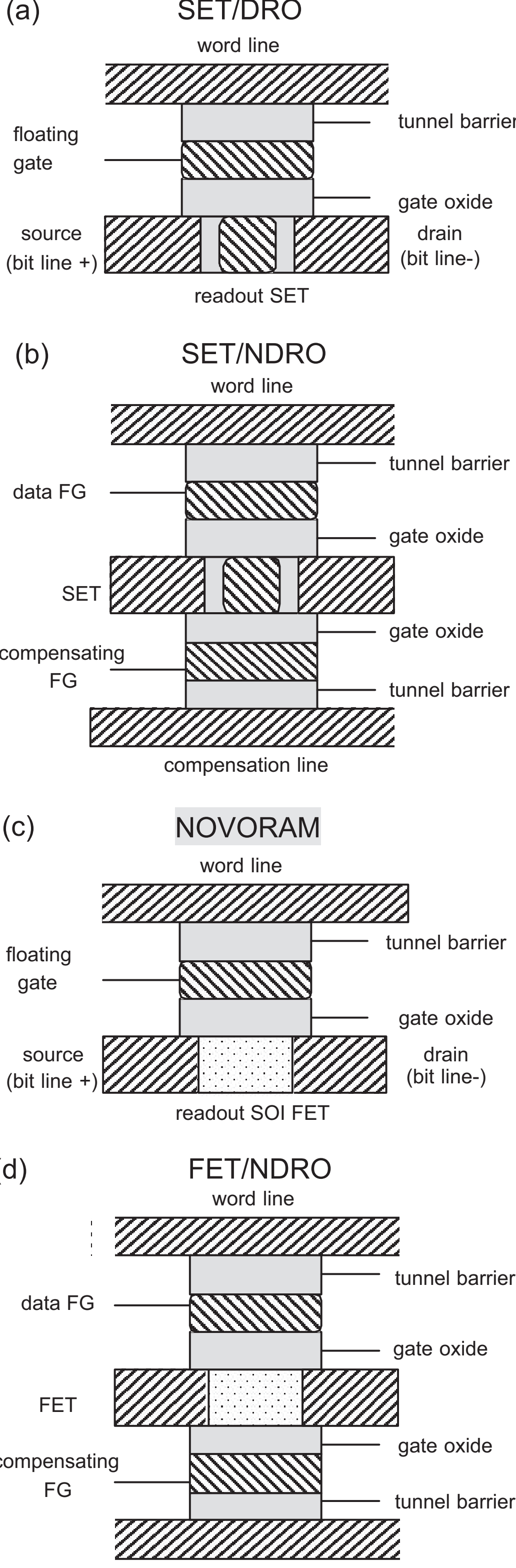


Figure 10. Suggested few-electron memory cells: (a) cell with destructive SET readout (SET/DRO, [110]); (b) SET-readout cell with background charge compensation, making non-destructive readout possible (SET/NDRO, [112]), (c) cell with nondestructive FET readout which may also be used in NOVORAM [113, 114], and (d) a similar cell with background charge compensation [3].

of approximately 1 μF/cm^2, corresponding to $\sim$10 nm^2 area per one electron (at 1 V). This prevents scaling of the island well below its maximum size ($\sim$3 nm).

Both these drawbacks may be avoided by using a more complex memory cell (Fig. 10b, [112]) in which the random background charge Q_0 of the island is compensated by weak capacitive coupling with the additional ("compensating") floating gate. The necessary few-electron charge of this gate may be inserted from an additional ("compensation") word line before the beginning of the memory operation, when special peripheral CMOS circuit measures Q_0 of each single-electron transistor (this may be done in parallel for all bits on a given word line) and develops an adequate combination of bit line signals. During the actual memory operation, the compensating charge is constant but may be adjusted periodically if necessary. Due to the compensation, the SET may be biased reliably at the steep part of its control characteristic (Fig. 14) and, hence, used for nondestructive readout of the cell state. Moreover, the signal charge swing may be relatively small, $\Delta Q_0 \sim (k_B T/E_a)e$, conveniently decreasing with the island size. This may allow the single-electron island to be scaled down to $\sim$1 nm, with the $\sim$3 nm floating gate storing just 2 to 3 electrons.

An alternative approach to few-electron memories is to scale down the generic structure of nonvolatile memory cells [64] by using a nanoscale MOSFET for readout (Fig. 10c). As was already mentioned in the Introduction, advanced MOSFETs with an ultrathin silicon-on-insulator channel may be scaled down to $\sim$10 nm, sustaining high performance and reproducibility. Further scaling down would lead to large random fluctuations of the threshold voltage, which again may be corrected by using a back floating gate (Fig. 10d).

Estimates of the maximum density of the SET–FET hybrid memories may be carried out under the assumption that the SET islands are fabricated by directed self-assembly, so their size is independent of the wiring line half-pitch F. In this case, the area of the cells shown in Fig. 10 is essentially independent of the island size and ranges between $6F^2$ and $8F^2$, where F is the minimum feature size of a given technology. The cell type determines the possible limits of this scaling. For the SET–DRO (Fig. 10a), it is confined in a narrow range around 3 nm, while for the SET–NDRO (Fig. 10b), F may be scaled down to $\sim$1.5 nm (line patterning permitting). For the FET–NDRO cells shown in Fig. 10c, *d*, the range of F does not have an upper bound; for the simple cell shown in Fig. 10c, the lower bound for F is about 5 nm, while the more complex cell with background charge compensation can be scaled down to $F \sim 2$ nm, i.e., to the cell area of about 30 nm^2. This should allow memories with a density well beyond 10^{12} bits/cm^2, enabling chips with a multiterabit integration scale. These exciting prospects are, however, contingent on the development of nm-scale fabrication technologies.

A possible problem with all the memory cells described is the slowness of the write–erase operations. In fact, they rely on the process of Fowler-Nordheim tunneling, similar to that used in the ordinary nonvolatile, floating-gate memories [64]. The standard 8-to-10-nm SiO_2 barriers used in such memories do not change transparency fast enough to allow floating gate recharging than in $\sim$1 μs even if the applied electric field is as high as $\sim$12 MV/cm, close to a breakdown threshold. (A reduction of the barrier thickness or height makes retention time too short.) Such a long write–erase cycle is acceptable for typical applications of flash memories [64] but is too long to replace (DRAM) in bit-addressable memories.

This problem can be solved by using special layered ("crested") barriers [113]. Calculations show [113–115] that trilayer crested barriers may combine a 1-ns-scale write time with a $\sim$10-year retention time, at apparently acceptable electric fields (about 10 MV/cm). This gives hope that such barriers may have the extremely high endurance necessary for RAM applications. Moreover, the ratio of the necessary write voltage (V_W) to the highest retention voltage (V_R) for such barriers may be well below 3. This condition ensures that the disturb effects on self-selected cells are small enough even in the simplest memory architectures [115].

Thus, if CMOS-compatible crested barriers are implemented, the few-electron cells shown in Fig. 10 may be used as random-access memories, replacing the (generically unscalable) DRAM. Moreover, crested barriers may enable the so-called nonvolatile random access memory (NOVORAM [115]) with a very simple cell structure (Fig. 10c) and architecture [105], which may be able to compete with DRAM and possibly SRAM even at the current technology level (minimum feature size F of the order of 100 nm), and be scaleable down to $F \sim 3$ nm. Some further decrease of F is possible by dropping the nonvolatility requirement and organizing cell refresh with period $\sim$1 s, because, in this case, the crested barrier and gate oxide may be thinned by $\sim$50%.

An alternative way to speed up the write–erase operation in floating gate memories is to replace the tunnel barrier with a transistor. Such two-transistor (2-T) memory cells were suggested long ago [116]; for a while they could not compete with DRAM because of the much larger cell size. The apparent impossibility to scale DRAM much below 100 nm has led to a recent revival of research and development work in this direction, and several new versions of the 2-T memories have been suggested (and some explored experimentally):

1) Single-electron trap with an additional gate controlling the potential profile high of the array and readout by using either a single-electron transistor [117] or nanoscale MOSFET [118–121]. Unfortunately, neither approach addresses the problem of random background charge. Notice that in experiments with semiconductor single-electron arrays, the Coulomb blockade mechanism of electron transport control coexists (and was repeatedly confused) with the usual field-effect mechanism. In this case, we are speaking about the effective replacement of the single-electron array by an MOSFET with very thin (and, hence, not very uniform) channel.
2) The authors of a recent work [122] have made this replacement conscientiously, arguing that their 2-nm-thick silicon channel should have lower parasitic source-to-drain leakage, because quantum confinement increased the effective bandgap. Indeed, their

MOSFETs exhibited leakage current as low as 10^{-19} A, sufficient to keep the cell retention time above 100 ms, i.e., the typical DRAM refresh time. A similar memory, but with a special "stacked" vertical FET [123] for fast write–erase (phase-state low electron drive memory [124]) (see also [125]). The channel of such a transistor incorporates three horizontal silicon nitride tunnel barriers partitioning the channel into several parts connected in series. Such separation improves transport control of the channel potential (and, hence, of the electron transport) by the surrounding gate; the considerable loss of ON current caused by these barriers is tolerable because the MOSFET should only recharge a very small capacitance of the charging node. Unfortunately, in such vertical MOSFETs, with their relatively large channel cross section, reaching acceptably small leakage current (and, hence, retention time) may be a difficult problem.

Though the last two approaches look interesting, I believe that the successful implementation of silicon-compatible crested barriers will make NOVORAM, with its simple structure, and small cell area, a more promising option.

The single- and few-electron memories will have to compete not only with each other, but with several other prospective memory concepts.

In ferroelectric memories, information is stored as a sign of electric polarization of a layer of a ferroelectric material. This polarization may be read out either destructively (as in DRAM) or nondestructively (e.g., if it controls a readout FET) [126]. Strong features of these memories include simple cell structure (and, as a result, small cell area), potential nonvolatility, and fast write–erase time (some materials have internal repolarization time well below 1 ns). In terms of an immediate practical introduction, the complexity of ferroelectric materials and their compatibility with the generic CMOS process is the main challenge. However, in the long term, scalability of these memories may be a larger problem. With the decrease in area, the height of the energy barrier $\Delta U \approx P_s E_c V$ separating two polarization states (where P_s is the saturation polarization; E_c, the coercive electric field; and V, the ferroelectric layer volume) decreases and should finally become comparable with the thermal fluctuation scale $k_B T$, resulting in random cell switching. For typical ferroelectric film parameters $P_s \sim 50\ \mu\text{C/cm}^2$ and $E_c \sim 5\ \text{V}/\mu\text{m}$ [126], such spontaneous switching should become rather noticeable ($\Delta U \lesssim 300\ k_B T$) at $V \sim 20 \times 20 \times 2\ \text{nm}^3$, even if P_s and E_c do not degrade with size. Thus, if no new breakthroughs are made, ferroelectric memory cells can hardly compete for terabit applications, which are the main promise of single- and few-electron memories.

In magnetic memories [127, 128], bits are stored in the form of thin-film magnetization. These memories share almost all the advantages and drawbacks of ferroelectric cells listed above (low 0/1 output signal value is an additional issue). However, because of the essentially similar physics, scaling of such cells to nanoscale may again be an issue. In this case, $\Delta U \sim B_s H_c V$ (where $B_s \sim 2$ T is the saturation magnetization and $H_c \sim 10$ Oe, the coercive magnetic field) drops below 300 $k_B T$ at even larger volume ($\sim 30 \times 30 \times 10\ \text{nm}^3$).

Memory cells based on structural phase transition include notably the Ovonics Unified Memory (OUM) [129, 130]. In an OUM cell, a chalcogenide alloy (GeSbTe) is switched from a conductive crystalline phase to a highly resistive amorphous phase under the effect of heating by current passed through a special heater. Though the chalcogenide materials are relatively complex, considerable progress in their deposition has been made in the course of development of CD-RW and DVD-RW technologies. As a result, OUM cells with surprisingly high endurance (up to 10^{13} cycles) have been demonstrated. Some OUM problems include relatively long write–erase time (reportedly, close to 100 ns, i.e., considerably longer than that of DRAM). Unfortunately, I am not aware of any published experimental data or reliable theoretical results sufficient to evaluate the dependence of the retention time on the storage region volume and, thus, to evaluate prospects of OUM scaling into the terabit range.

Finally, single-molecular memories may be based on various background physics. For example, they may be just the molecular implementation of single-electron memories as discussed previously. However, there is an alternative possibility: to use molecular conformation changes of the molecule for information storage. The studies of such single-molecular devices [131–135] are still in their infancy, and it is difficult to compare their prospects with the more mature approaches previously reviewed. I believe the most important issue to be addressed is the rate of thermally induced, spontaneous switching of a single molecule between its configurations. To keep the rate (and, hence, the soft error rate) sufficiently low, the potential barrier ΔU separating two possible states must be kept rather high, at least 100 $k_B T$. For room-temperature operation, this estimate yields ~3 eV, a pretty high energy scale. This requirement (together with those of fast electrical readout and reproducible directed self-assembly) will probably limit the list of possible candidate molecules rather severely.

4.4. Electrostatic Data Storage

A combination of single-electron transistors with crested barriers may be used not only in terabit-scale memories but also in ultradense electrostatic data storage systems [113] (Fig. 11). In this design, a single-electron transistor, followed by a MOSFET amplifier, a few microns apart, is fabricated on a tip-shaped chip playing the role of a READ–WRITE head. The data bits are stored as a few-electron charge trapped in a group of nanoscale conducting grains deposited on top of a crested tunnel barrier. It is important that since each bit is stored in a few (~10) grains, their exact shape and location are not important, so the storage medium production does not require any nanofabrication.

WRITE is performed by the application of the same voltage V_W to both input terminals, relative to the conducting ground layer of the moving substrate. The resulting electric field of the tip induces rapid tunneling of electrons from the ground through the crested barrier into a ~30-nm-wide group of grains. For READ, the single-electron transistor is activated by source–drain voltage $2V_R \geq V_t$. In this state, it is very sensitive to the electric field created by the group of charged grains it is being flown above.

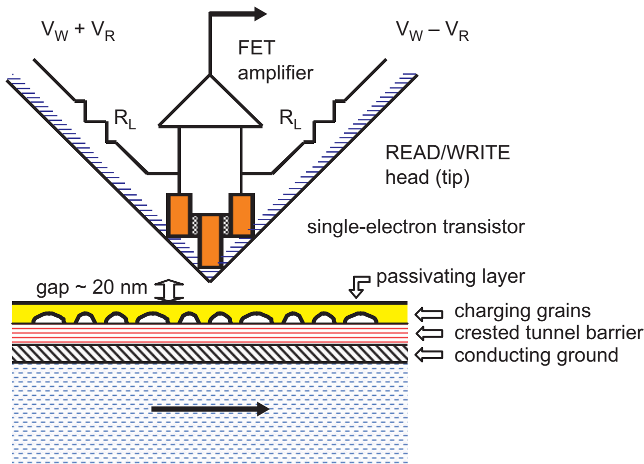


Figure 11. Proposed electrostatic data storage system with hybrid SET/FET readout [113], K. K. Likharev, *Appl. Phys. Lett.* 73, 2137 (1998). © 1998, American Institute of Physics.

Simple estimates show that with a 20-nm tip-to-substrate distance (close to those already implemented in the best present-day magnetic storage systems), the electrostatic system is capable of a density ~10 terabits per square inch, i.e., at least an order of magnitude more than the best prospects for the magnetic competition of which I am aware. The use of crested barriers may provide a very broad bandwidth of both WRITE and READ operations, up to 1 Gbps per channel, possibly quite adequate even for this enormous bit density.

Notice that for this particular application, the difficulty of fabrication of room-temperature single-electron transistors, outlined in Section 3, is not a major concern, because one would need just one (or a few) transistors per system, and slow fabrication techniques (like the scanning probe oxidation) may be acceptable. Moreover, as in single-electron memories, the transistor would work as an analog amplifier, so that the single-electron island size of ~3 nm would be sufficient.

The experiments [81], mentioned in Section 4.1., may be considered as the first step toward the implementation of this idea.

4.5. Logic Circuits

4.5.1. Voltage State Logics

Most suggestions of logic circuits based on single-electron devices may be referred to one of two groups. In circuits of the first, "voltage state", group (first considered in [14]) single-electron transistors (Fig. 3) are used in CMOS-like circuits. This means that the single-electron charging effects are confined to the interior of the transistor, while externally it looks like the usual electronic device switching persisting currents, with binary unity–zero presented with high–low dc voltage levels (physically not quantized). This concept simplifies the circuit design that may ignore all the single-electron physics particulars, except the specific dependence of the drain current I on the drain-to-source voltage V and gate-to-source voltage V_g (Figs. 3 and 4).

Analyses of this opportunity have shown that due to the specific shape of this dependence (oscillating transconductance), both resistively coupled [14] and capacitively coupled [136] single-electron transistors allow a very simple implementation of CMOS-type inverters, without a need for two types of transistors (like n-MOSFETs and p-MOSFETs in the standard CMOS technology). On the other hand, peculiarities of functions $I(V, V_g)$ make the exact copying of CMOS circuits impossible, and, to get substantial parameter margins, even simple logic gates have to be redesigned. The redesigned and optimized circuits (Fig. 12) may operate well within a relatively wide parameter window [137, 138]. Even after such optimization, the range of their operation with an acceptable bit error rate shrinks under the effect of thermal fluctuations as soon as their scale $k_B T$ reaches approximately $0.01E_a$. (For other suggested versions of the voltage state logic [136, 139], the temperature range apparently is even narrower). The maximum temperature may be somewhat increased by replacing the usual single-island single-electron transistors with more complex devices with 1D-array "channels" and distributed gate capacitances [140]. However, this would increase the total transistor area at the given minimum feature size.

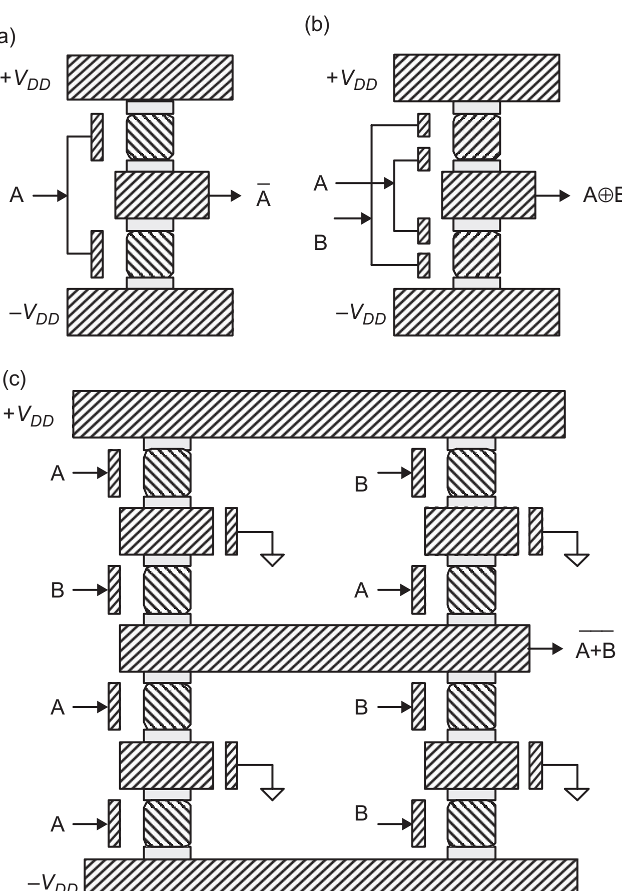


Figure 12. Basic gates of the complementary SET logic family using capacitively-coupled single-electron transistors: (a) inverter; (b) XOR, and (c) NOR/NAND [138], R. H. Chen et al., *Appl. Phys. Lett.* 68, 1954 (1996). © 1996, American Institute of Physics.

A disadvantage of voltage state circuits is that neither of the transistors in each complementary pair is closed too well, so that the static leakage current in these circuits is fairly substantial, of the order of 10^{-4} e/RC [137]. The corresponding static power consumption is negligible for relatively large devices operating at helium temperatures. However, at the prospective room-temperature operation, this power becomes on the order of 10^{-7} watt per transistor. Though apparently low, for the hypothetical circuits that would be dense enough ($>10^{11}$ transistors per cm^2) to challenge the prospective CMOS technology, this number gives an unacceptable static power dissipation density (>10 kW/cm^2).

4.5.2. Charge State Logics

The power dissipation problem may be avoided, to a large extent, by using "charge state" logic circuits in which single bits of information are presented by the presence or absence of single electrons at certain conducting islands throughout the whole circuit. In this case, the static currents and power virtually vanish. This approach has been explored theoretically since 1987 [141], and quite a few families of charge state logic circuits have been suggested [29, 30, 141–167]. In most suggestions, an electron is confined in a cell consisting of one or a few islands, while the logic switching is achieved via electrostatic (or spin [148, 154, 160]) coupling of the cells. Another classification of single electron logics may be based on where they take the energy necessary for logic operations: from dc power supply [141, 150, 158], ac power supply (also playing the role of a global clock) [29–31, 140, 145, 162, 166, 167], or just from the energy of an external signal [143–148, 154, 156, 160, 161, 163]. (The last category includes the so-called quantum (or quantum-dot) cellular automata based on ground-state computing. This concept has been controversial from the very beginning [149, 159–161], and, under the growing criticism [176, 177], it was eventually repudiated by its authors [31].)

Only a few of these concepts have been analyzed in detail, especially at finite temperatures. To my knowledge, the most robust charge-state logic circuits suggested till now are those based on the single-electron parametrons (see Section 2.4.). Figure 13 shows a possible shift register based on the parametrons. The direction of the shift of the central island of each next device is shifted by $\pi/3$ within the yz plane. The circuit is driven by electric field $\boldsymbol{E}_c(t)$ rotating in the same plane and providing the periodic switching on the SET parametrons, with an appropriate phase shift. As a result, each digital bit (one per three cells) is being shifted by 3 cells along the structure each clock period. Majority logic gates, sufficient for arbitrary logic circuits, may be implemented in the same way [30]. Geometric modeling and numerical simulation of these circuits within the framework of the orthodox theory have shown that they may operate correctly within approximately $\pm 20\%$ deviations from the optimal clock amplitude. Estimates show that the maximum operation temperature of these logic circuits is of the order of that of voltage mode circuits, i.e., of the order of $0.01E_c/k_B$, if the bit error rate is in the practically acceptable range (below $\sim 10^{-20}$).

A new, potentially useful feature of the charge state logics is the natural internal memory of their "logic gates" (more proper terms are "finite-state cells" or "timed gates"), thus combining the functions of combinational logic and latches. This feature makes natural the implementation of deeply pipelined ("systolic") and cellular automata architectures. The back side of this advantage is the lack of an effective means of transferring a signal over large distances; crudely speaking, this technology does not allow simple wires, just shift registers.

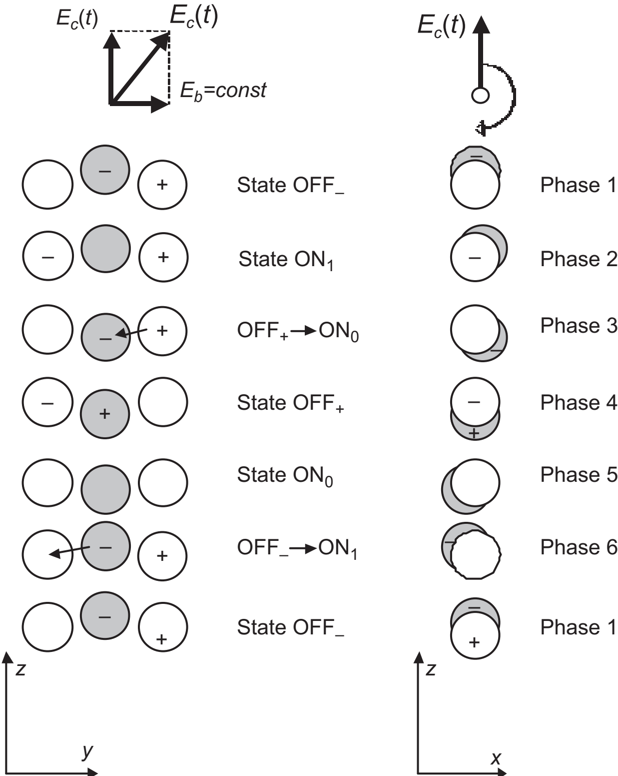


Figure 13. A shift register based on single-electron parametrons.

4.5.3. Phase-Mode Logic

Within the framework of our classification, the so-called phase-mode logic [168–171] should be placed in a separate category, because in these circuits the information-keeping cells are coupled by rf signals, carrying the binary information coded by either of two possible values of the rf signal phase. The elementary cell of this logic essentially is a relaxation oscillator generating SET oscillations with frequency $f_{SET} = I/e$ (see the discussion in Section 4.1.) phase locked by an external reference with frequency $2f_{SET}$. Such "subharmonic" generator has two possible phases of oscillations, which differ by π; the oscillation state with each phase has exactly the same amplitude and is locally stable, i.e., can be used to code a single bit of information. Moreover, the oscillator can impose its phase upon a similar, adjacent oscillator that had been turned off temporarily (e.g., by turning off its power supply current I) and is now being turned on. Thus information may be transferred along a row of oscillators, working as a shift register, very similar to that consisting of single-electron parametrons (see Fig. 13 and its discussion). Majority gates can be organized similarly, enabling the implementation of arbitrary logic functions.

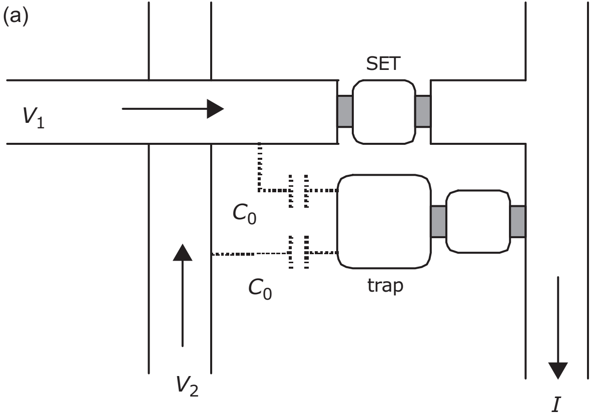
A theoretical advantage of this logic family is the potential insensitivity of its cells to the random background charges, if a resistor with quasi-continuous charge transfer is used for the insertion of dc current I. Unfortunately, as was discussed in Section 2, the known implementations of such resistors lead to very large size, making the implementation of practical (room-temperature) phase-mode logic circuits hardly feasible.

4.5.4. Devices for Defect-Tolerant Architectures

The main problem of all single-electron logic circuits is the effect of random background charge, described in Section 2. Until this problem has been solved, it is difficult to discuss seriously any of numerous suggestions of single-electron implementation of traditional logic circuit architectures (e.g., usual microprocessors, digital signal processors, etc.) based on number crunching.

I believe that the only way to make single-electron devices useful for large-scale electronic circuits is to introduce innovative architectures with high defect tolerance. The design of such architectures is especially natural for hybrid CMOL circuits combining CMOS components and single-molecule single-electron devices connected by nanowires (Fig. 9 and its discussion). In this case, single-electron devices may be used to perform only the functions that are natural for them (i.e., compact memory cells or externally controlled switches), leaving the more difficult functions (e.g., signal amplification with high voltage gain) for more sparse CMOS circuits.

Figure 14 shows a three-terminal version [173] of the earlier suggested [172] device that is a simple combination of the single-electron transistor and the trap, working together as a "latching switch." The device consists of three small islands connected by four tunnel junctions. Island 1, together with input and output wires serving as source and drain, forms a single-electron transistor. Islands 2 and 3 form a single-electron trap (cf., Fig. 5), with trapping island 3 capacitively coupled to the SET island 1, thus playing the role of a single-electron floating gate. If the effective source-to-drain voltage $V = (V_1 + V_2)/2$ applied from two inputs of the device is low, the trap in equilibrium has no extra electrons and its total electric charge is zero. As a result, the transistor remains in the Coulomb blockade state, and input and output wires are essentially disconnected. If V is increased beyond a certain threshold V_{inj} (which should be lower than the Coulomb blockade threshold voltage V_t of the transistor), one electron is injected into the trap. In this charge state, the Coulomb blockade in the transistor is lifted, keeping the wires connected at any V. However, if the node activity (voltage V) is low for a long time, either thermal fluctuations or cotunneling eventually kick the trapped electron out of the trap, and the transistor closes, disconnecting the wires. Figure 14b shows typical results of numerical simulation of the latching switch dynamics with a perfect background charge of the SET island (it may be adjusted by voltage applied to the SET gate). The trap injection threshold may be adjusted by voltage applied to an additional gate. Figure 15 shows a possible single-molecule implementation of such a device [174].



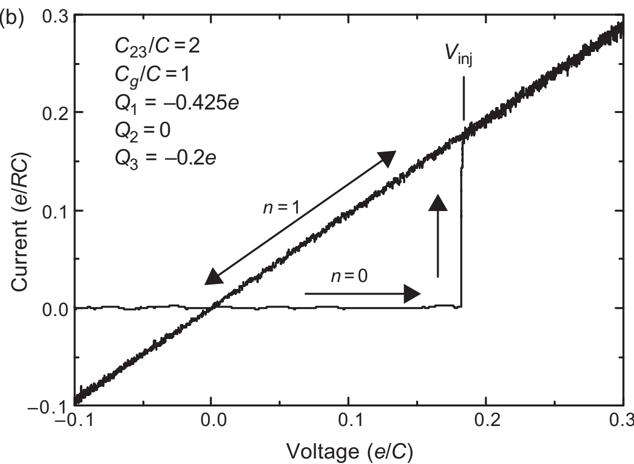


Figure 14. (a) Three-terminal single-electron latching switch and (b) its dc I–V curve calculated using the "orthodox" theory [172, 173].

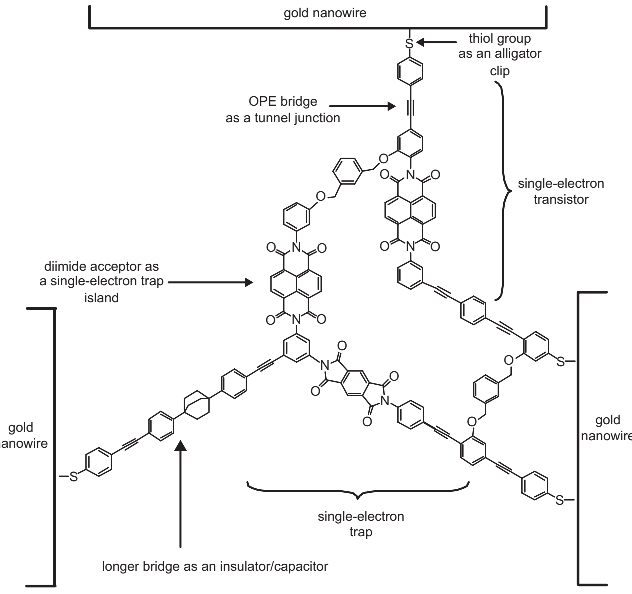


Figure 15. Possible molecular implementation of the three-terminal latching switch [174].

The latching switch is still sensitive to random charge impurities and, thus, should be used in redundant circuits that may tolerate at least a small percent of faulty devices. I believe that the architecture design space for such defect-tolerant circuits is very broad, and I will give just one example based on our recent Stony Brook work [172–175]. Figure 16 shows a possible structure of so-called distributed crossbar arrays for neuromorphic networks ("CrossNets") based on such switches. Cell bodies (which are supposed to be implemented in nanoscale CMOS technology) are embedded sparsely into a 2D array of single-electron latching switches working as single-bit-weight synapses. (Depending on the particular insertion rule, we distinguish several CrossNet species, e.g., FlossBar, InBar, and RandBar [173].) In each CrossNet, each cell body is hard-wired to a limited number $4M$ of other cells, with the discrete synaptic weights controlling which of these connections are currently active. Vice versa, signal activity of the network determines whether the synapses are open or closed, though the state of any particular synapse also is affected by the underlying randomness of single-electron tunneling.

Recent results of numerical simulation of limited fragments of the CrossNets on usual supercomputers have shown [173–175] that they may be taught to perform at least fast image recognition in the so-called Hopfield mode (Fig. 17). These results give hope that such self-adaptive

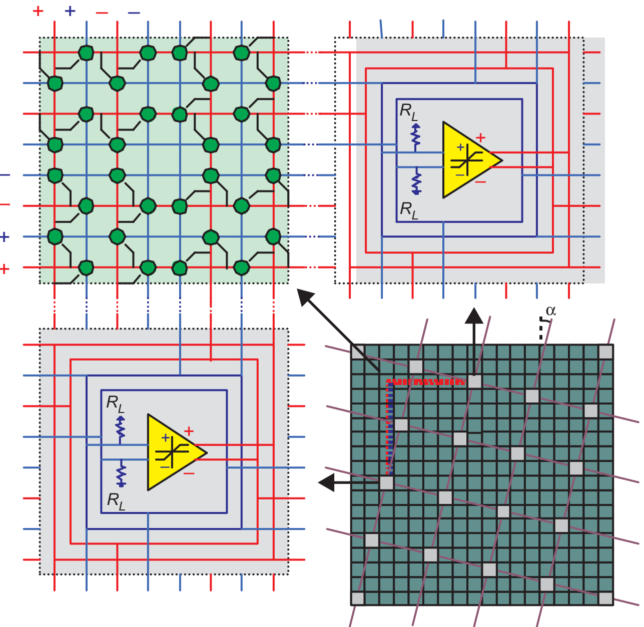


Figure 16. A Distributed Crosspoint Network ("CrossNet") and its basic components: synaptic plaquettes (green cells) and somas (gray cells) connected with axonic (red) and dendritic (blue) wires. On the synaptic plaquette (left upper panel), circles with arrows denote three-terminal latching switches (Fig. 14, 15). Solid points on the somatic cell borders show open-circuit terminations of the dendritic and axonic wires; due to these terminations the neural cells in the same row or column of the InBar do not interact. The bottom right panel shows the most explored CrossNet architecture, the Inclined Crossbar ("InBar"). In that panel, dashed red and blue lines show the particular connections show in the rest of the figure, while this violet lines indicate the square lattice formed by the somatic cells in InBar. These lines may be used for input/output of signals from individual cell.

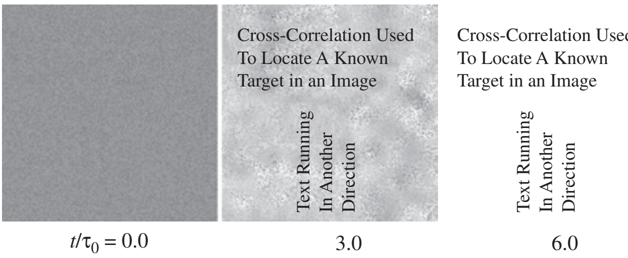


Figure 17. The process of recall (recognition) of one of three trained black-and-white images by an InBar-type CrossNet with 256 × 256 neural cells and connectivity parameter $M = 64$. The initial image (left panel) was obtained from the trained image (identical to the one shown in the right panel) by flipping 40% of pixels. $\tau_0 = MR_L C_0$ is the effective time constant of intercell interaction, where C_0 is the dendrite wire capacitance per one plaquette.

("plastic") networks also can be taught, by using global reinforcement training techniques, to perform more complex information processing. Moreover, it is expected that if a sufficiently large hierarchical, CrossNet-block-based system [174] will be able, after a period of initial training, to self-evolve and to self-improve on the basis of its interaction with the informational environment.

If these expectations are confirmed, CrossNets may have very impressive characteristics. Estimates show [172–175] that at the neural cell connectivity $M = 10^4$ (comparable with that of the human cerebral cortex) and a nanowiring pitch of 4 nm, neural cell density may be as high as 10^7 cm^{-2}, implying that brain-scale systems could be implemented on a silicon area of the order of 30 × 30 cm^2. The estimated speed scaling of this network is even more impressive, with time of signal propagation between the neural cells of the order of 100 ns, at acceptable power dissipation. This speed is nearly 4 magnitudes higher than that in biological neural networks. The systems capable of self-evolving at such a high speed would have a revolutionary impact on all information technology and possibly on the society as a whole.

5. CONCLUSION

Single-electron devices frequently are considered as the most probable candidates for the replacement of silicon MOSFETs as soon as the process of exponential improvement of the latter devices runs out of steam, apparently in about a decade. The most reasonable basis for these expectations is that the physics of single-electron device operation is basically material-insensitive, and, hence, these devices may be fabricated from a broad variety of materials. The most important corollary of this insensitivity is the possibility of self-assembly of chemically synthesized molecular devices on prefabricated metallic nanowire structures. Due to its potentially low cost, this technology may eventually replace lithography as the main road to future nanoelectronics.

However, working of these prospects, one has to take into account substantial problems of single-electron devices; first of all, their low voltage gain and high sensitivity to single charge impurities in the dielectric environment. Both these problems may be most readily circumvented in hybrid,

few-electron memory cells (see Section 4.3.). Such memory cells may be scaled down to a $\sim$30 nm^2 cell area, the range apparently out of reach of DRAM, ferroelectric, and magnetic memories. Complemented with crested tunnel barriers, the few-electron memories may also feature 1-ns-scale write–erase time. Moreover, the feature size requirements for single-electron devices in memory cells are more relaxed (for room-temperature operation, they should not be larger than $\sim$3 nm, instead of $\sim$1 nm for logic circuits). As a result, one may expect practical implementation of room-temperature few-electron memories, even by using the usual lithographic patterning methods or their straightforward future extensions.

For logic circuits, the random background charge effects are much harder to overcome. Nevertheless, the single-electron devices, used in CMOL hybrid circuits with nanoscale MOSFETs, may become the basis for implementation of novel, massively parallel architectures for advanced information processing, e.g., self-evolving neuromorphic networks (see Section 4.5.). In the most optimistic scenarios, such systems may eventually become more important for information processing than the traditional digital processors. Presently, the technologic and social consequences of such a breakthrough cannot be predicted in detail but would certainly be extremely significant.

GLOSSARY

CMOL circuit A hybrid VLSI circuit combining a CMOS stack and several levels of parallel nanowires connected by self-assembled molecular devices.

Coulomb blockade Suppression of electron transport through a system with a single-electron island(s) due to single-electron charging.

CrossNet A distributed crossbar circuit structure compatible with (i) neuromorphic network architectures and (ii) CMOL implementation.

Single-electron addition energy The minimal energy necessary to bring an additional electron from infinity into a single-electron island.

Single-electron box A system of two electrodes (source and gate) and one single-electron island connected to the source via a tunnel junction. Voltage applied to the gate changes electrostatic potential of the island and determines the equilibrium number of electrons in it.

Single-electron island A piece of conductor so small that injection/ejection of a single electron (e.g., through a tunnel junction) causes a substantial of island's electrostatic pontential that may affect transport of the following electrons.

Single-electron parametron A system of three (or more) single-electron islands connected by tunnel junctions but isolated galvanically from environment. External electric fields applied to the parametron may cause either a "soft" (reversible) switching of the electron configuration of the system and, hence, of its electric dipole moment or a "freeze" of the moment in one of two possible metastable states.

Single-electron transistor A system of three electrodes (drain, source, and gate) in which the source and drain are connected via two tunnel junctions and a single-electron island between them. Voltage applied to gate changes electrostatic potential of the island and, hence, the rate of electron transport between source and drain.

ACKNOWLEDGMENTS

The author is grateful to many colleagues for numerous fruitful discussions of the issues mentioned in this article. Research at Stony Brook on topics discussed in this paper is supported in part by AFOSR, DOE, NSF, and SRC.

REFERENCES

1. "International Technology Roadmap for Semiconductors: 2001 Edition, 2002 Update," available on the Web at public.itrs.net.
2. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, *Proc. IEEE* 89, 259 (2001).
3. K. K. Likharev, in "Nano and Giga Challenges in Microelectronics" (J. Greer, Eds.), p. 27. Elsevier, Amsterdam, 2003.
4. V. S. Sverdlov, T. J. Walls, and K. K. Likharev, *IEEE Trans. Electron Devices* 50, 1926 (2003).
5. Y. Imry, "Introduction to Mesoscopic Physics." Oxford Univeristy Press, Oxford, 1997.
6. S. Datta, "Electronic Transport in Mesoscopic Systems." Cambridge Univeristy Press, Cambridge, (MA) 1997.
7. K. K. Likharev, *IBM J. Res. Dev.* 32, 144 (1988).
8. D. V. Averin and K. K. Likharev, in "Mesoscopic Phenomena in Solids" (B. Altshuler, P. Lee, and R. Webb, Eds.), p. 173. Amsterdam, Elsevier, 1991.
9. M. H. Devoret and H. Grabert (Eds.), "Single Charge Tunneling." Plenum, New York, 1992.
10. K. K. Likharev, *Proc. IEEE* 87, 606 (1999).
11. J. Lambe and R. C. Jaklevic, *Phys. Rev. Lett.* 22, 1371 (1969).
12. I. O. Kulik and R. I. Shekhter, *Sov. Phys. JETP* 41, 308 (1975).
13. D. V. Averin and K. K. Likharev, *J. Low Temp. Phys.* 62, 345 (1986).
14. K. K. Likharev, *IEEE Trans. Magn.* 23, 1142 (1987).
15. T. A. Fulton and G. D. Dolan, *Phys. Rev. Lett.* 59, 109 (1987).
16. Y. A. Pashkin, Y. Nakamura, and J. S. Tsai, *Appl. Phys. Lett.* 76, 2256 (2000).
17. G. Zimmerli, R. L. Kautz, and J. M. Martinis, *Appl. Phys. Lett.* 61, 2616 (1992).
18. R. A. Smith and H. Ahmed, *Appl. Phys. Lett.* 71, 3838 (1997).
19. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, *Appl. Phys. Lett.* 76, 3121 (2000).
20. T. A. Fulton, P. L. Gammel, and L. N. Dunkleberger, *Phys. Rev. Lett.* 67, 3148 (1991).
21. D. V. Averin and K. K. Likharev, in [9], p. 311.
22. L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 64, 2691 (1990).
23. K. K. Likharev, N. S. Bakhvalov, G. S. Kazacha, and S. I. Serdyukova, *IEEE Trans. Magn.* 25, 1436 (1989).
24. K. K. Likharev and K. A. Matsuoka, *Appl. Phys. Lett.* 67, 3037 (1995).
25. D. V. Averin and A. A. Odintsov, *Phys. Lett. A* 140, 251 (1989).
26. P. D. Dresselhaus, L. Ji, S. Y. Han, J. E. Lukens, and K. K. Likharev, *Phys. Rev. Lett.* 72, 3226 (1994).
27. L. Ji, P. D. Dresselhaus, S. Y. Han, K. Lin, W. Zheng, and J. E. Lukens, *J. Vac. Sci. Technol., B* 12, 3619 (1994).
28. K. A. Matsuoka, K. K. Likharev, P. D. Dresselhaus, L. Ji, S. Y. Han, and J. E. Lukens, *J. Appl. Phys.* 81, 2269 (1997).
29. K. K. Likharev and A. N. Korotkov, *Science* 273, 763 (1996).
30. A. N. Korotkov and K. K. Likharev, *J. Appl. Phys.* 84, 6114 (1998).

31. A. O. Orlov, R. K. Kummamuru, R. Ramasubramaniam, T. Toth, C. S. Lent, G. H. Bernstein, and G. L. Snider, *Appl. Phys. Lett.* 78, 1625 (2001).
32. D. V. Averin and A. N. Korotkov, *J. Low Temp. Phys.* 80, 173 (1991).
33. D. V. Averin, A. N. Korotkov, and K. K. Likharev, *Phys. Rev. B* 44, 6199 (1991).
34. L. P. Kouwenhoven, C. M. Markus, P. L. McEuen, S. Tarucha, R. M. Westervelt, and. N. S. Wingreen, in "Mesoscopic Electron Transfer" (L. Sohn Eds.), p. 105. Kluwer, Dordrecht, 1997.
35. K. Matsumoto, M. Ishii, K. Segawa, Y. Oka, B. J. Vartanian, and J. S. Harris, *Appl. Phys. Lett.* 68, 34 (1996).
36. K. Matsumoto, Y. Gotoh, T. Maeda, J. A. Dagata, and J. S. Harris, *Appl. Phys. Lett.* 76, 239 (2000).
37. H. W. C. Postma, T. Teepen, Z. Yao, M. Grifoni, and C. Dekker, *Science* 293, 76 (2001).
38. C. Thelander, M. H. Magnusson, K. Deppert, L. Samuelson, P. R. Poulsen, J. Nygard, and J. Borggreen, *Appl. Phys. Lett.* 79, 2106 (2001).
39. Y. Takahashi, M. Nagase, H. Namatsu, K. Kurihara, K. Iwdate, Y. Nakajima, S. Horiguchi, K. Murase, and M. Tabe, *Electron. Lett.* 31, 136 (1995).
40. Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, *IEEE Trans. Electron. Devices* 47, 147 (2000).
41. D. H. Kim, J. D. Lee, and B.-G. Park, *Jpn. J. Appl. Phys., Part 1* 39, 2329 (2000).
42. K. Uchida, J. Koga, R. Ohba, and A. Torumi, in "Techn. Dig. IEDM," p. 863. IEEE Press, Piscataway (NJ), 2000.
43. M. Saitoh, N. Takahashi, H. Ishikuro, and T. Hiramoto, *Jpn. J. Appl. Phys., Part 1* 40, 2010 (2001).
44. J. B. Cui, M. Burghard, and M. Kern, *Nano Letters* 2, 117 (2002).
45. G. M. Whitesides, J. Mathias, and C. T. Seto, *Science* 254, 1312 (1991).
46. J. H. Fendler, Chem. Mater. 13, 3196 (2001).
47. V. V. Shorokhov, P. Johansson, and E. S. Soldatov, *J. Appl. Phys.* 91, 3049 (2002).
48. N. B. Zhitenev, H. Meng, and Z. Bao, *Phys. Rev. Lett.* 88, 226801 (2002).
49. J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruña, P. L. McEuen, and D. C. Ralph, *Nature* 417, 722 (2002).
50. S. Zankovych, T. Hoffmann, J. Seekamp, J. U. Bruch, and C. M. S. Torres, *Nanotechnology* 12, 91 (2001).
51. D. Meschede and H. Metcalf, *J. Appl. Phys.* (2002).
52. R. F. Service, *Science* 293, 785 (2001).
53. E. G. Emiroglu, Z. A. K. Duranni, D. C. Hasko, D. A. Williams, *J. Vac. Sci. Technol., B* 20, 2806 (2002).
54. E. H. Nicollian and J. R. Brews, "MOS Physics and Technology." Wiley, New York, 1982.
55. A. N. Korotkov, *Appl. Phys. Lett.* 72, 3226 (1998).
56. Y. A. Pashkin, Y. Nakamura, and J. S. Tsai, *Appl. Phys. Lett.* 74, 132 (1999).
57. W. Zheng, J. R. Friedman, D. V. Averin, S. Han, and J. E. Lukens, *Solid State Commun.* 108, 839 (1998).
58. K. E. Nagaev, *Phys. Lett. A* 169, 103 (1992).
59. K. E. Nagaev, *Phys. Rev. B* 52, 4740 (1995).
60. Y. Naveh, D. V. Averin, and K. K. Likharev, *Phys. Rev. B* 58, 15371 (1998).
61. M. V. Fischetti, S. E. Laux, and E. Crabbe, *J. Appl. Phys.* 78, 1058 (1995).
62. V. K. Sverdlov, A. N. Korotkov, and K. K. Likharev, *Phys. Rev. B* 63, R081302 (2000).
63. B. Prince, "Semiconductor Memories," 2nd ed. Wiley, Chichester, 1991.
64. W. D. Brown and J. E. Brewer, "Nonvolatile Semiconductor Memory Technology." IEEE Press, Piscataway (NJ), 1998.
65. L. Guo, E. Leobandung, and S. Y. Chou, *Science* 275, 649 (1997).
66. A. N. Korotkov, D. V. Averin, K. K. Likharev, and S. A. Vasenko, in "Single Electron Tunneling and Mesoscopic Devices" (H. Koch and H. Lübbig Eds.), p. 45. Springer, Berlin, 1992.
67. A. N. Korotkov, *Phys. Rev. B* 49, 10381 (1994).
68. V. A. Krupenin, D. E. Presnov, A. B. Zorin, and J. Niemeyer, *Physica B* 284, 1800 (2000).
69. R. J. Schoelkopf, P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, *Science* 280, 1238 (1998).
70. A. Aassime, G. Johansson, G. Wendin, R. J. Schoelkopf, and P. Delsing, *Phys. Rev. Lett.* 86, 3376 (2001).
71. A. N. Korotkov and M. A. Paalanen, *Appl. Phys. Lett.* 74, 4052 (1999).
72. M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, *Phys. Rev. Lett.* 69, 1997 (1992).
73. J. G. Lu, J. M. Hergenrother, and M. Tinkham, *Phys. Rev. B* 53, 3543 (1996).
74. T. M. Eiles, J. M. Martinis, and M. H. Devoret, *Phys. Rev. Lett.* 72, 1862 (1993).
75. P. Lafarge, P. Joyes, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 70, 994 (1993).
76. V. J. Goldman and B. Su, *Science* 267, 1010 (1995).
77. M. Field, C. G. Smith, M. Pepper, K. M. Brown, E. H. Linfield, M. P. Grimshaw, D. A. Ritchie, and G. A. C. Jones, *Phys. Rev. Lett.* 77, 350 (1996).
78. Y. Y. Wei, J. Weis, K. v. Klitzing, and K. Eberl, *Appl. Phys. Lett.* 71, 2514 (1997).
79. R. J. Schoelkopf, S. H. Moseley, C. M. Stahle, P. Wahlgren, and P. Delsing, *IEEE T. Appl. Supercon.* 9, 2935 (1999).
80. S. Komiyama, O. Astafiev, V. Antonov, T. Kutsuva, and H. Hirai, *Nature* 403, 405 (2000).
81. M. J. Yoo, T. A. Fulton, H. F. Hess, R. L. Willett, L. N. Dunkleberger, R. J. Chichester, L. N. Pfeiffer, and K. W. West, *Science* 276, 579 (1997).
82. P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, *Phys. Rev. Lett.* 66, 1926 (1991).
83. R. C. Ashoori, H. L. Stormer, J. C. Weiner, L. N. Pfeiffer, K. W. Baldwin, S. J. Pearton, and K. W. West, *Phys. Rev. Lett.* 68, 3088 (1992).
84. J. A. Folk, S. R. Patel, S. F. Godijn, A. G. Huibers, S. M. Cronenwett, C. M. Markus, K. Campman, and A. C. Gossard, *Phys. Rev. Lett.* 76, 1699 (1996).
85. P. Delsing, L. S. Kuzmin, K. K. Likharev, and T. Claeson, *Phys. Rev. Lett.* 63, 1861 (1989).
86. H. Pothier, P. Lafarge, P. F. Orfila, C. Urbina, D. Esteve, and M. H. Devoret, *Physica B* 169, 1598 (1991).
87. M. W. Keller, J. M. Martinis, N. M. Zimmerman, and A. H. Steinbach, *Appl. Phys. Lett.* 69, 1804 (1996).
88. D. F. Barbe (Ed.), "Charge-Coupled Devices." Springer, Berlin, 1980.
89. D. V. Averin and K. K. Likharev, in "Nanostructures and Mesoscopic Systems" (W. Kirk and M. Reed, Eds.) p. 283. Academic Press, Boston, 1992.
90. A. Fujiwara and Y. Takahashi, *Nature* 410, 560 (2001).
91. J. P. Pekola, K. P. Hirvi, J. P. Kauppinen, and M. A. Paalanen, *Phys. Rev. Lett.* 73, 2903 (1994).
92. K. P. Hirvi, M. A. Paalanen, and J. P. Pekola, *J. Appl. Phys.* 80, 256 (1996).
93. S. Farhangfar, K. P. Hirvi, J. P. Kauppinen, J. P. Pekola, J. J. Toppari, D. V. Averin, and A. N. Korotkov, *J. Low Temp. Phys.* 108, 191 (1997).
94. J. P. Pekola, J. K. Suoknuuti, J. P. Kauppinen, M. Weiss, P. van der Linden, and A. G. M. J. Jansen, *J. Low Temp. Phys.* 128, 263 (2002).
95. K. K. Likharev, "Dynamics of Josephson Junctions and Circuits," Chap. 16. Gordon and Breach, New York, 1986.

96. S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan, *Appl. Phys. Lett.* 68, 1377 (1996).
97. H. I. Hanafi, S. Tiwari, and I. Khan, *IEEE Trans. Electron. Devices* 43, 1553 (1996).
98. Y. Shi, K. Saito, H. Ishikuro, and T. Hiramoto, *J. Appl. Phys.* 84, 2358 (1998).
99. Y. C. King, T. J. King, and C. M. Hu, *IEEE Electron Device Lett.* 20, 409 (1999).
100. I. Kim, S. Han, K. Han, J. Lee, and H. Shin, *IEEE Electron Device Lett.* 20, 630 (1999).
101. K. Han, I. Kim, and H. Shin, *IEEE Trans. Electron Devices* 48, 874 (2001).
102. R. Ohba, N. Sugiyama, J. Koga, K. Uchida, and A. Torumi, *Jpn. J. Appl. Phys., Part 1* 39, 989 (2000).
103. A. Fernandes, B. DeSalvo, T. Baron, J. F. Damlencourt, A. M. Papon, D. Lafond, D. Mariolle, B. Guillaumot, P. Besson, P. Masson, G. Ghibaudo, G. Pananakakis, F. Martin, and S. Haukka, in "Techn. Dig. IEDM," p. 155. IEEE Press, Piscataway (NJ), 2001.
104. T. Ishii, T. Osabe, T. Mine, F. Murai, and K. Yano, in "Techn. Dig. IEDM," p. 305. IEEE Press, Piscataway (NJ), 2001.
105. A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, *Appl. Phys. Lett.* 70, 1742 (1997).
106. J. J. Welser, S. Tiwari, S. Rishton, K. Y. Lee, and Y. Lee, *IEEE Electron Device Lett.* 18, 278 (1997).
107. T. Tsutsumi, K. Ishii, E. Suzuki, H. Hiroshima, M. Yamanaka, I. Sakata, S. Kanemaru, S. Hazra, T. Maeda, and K. Tomizawa, *Electron. Lett.* 36, 1322 (2000).
108. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Hashimoto, T. Kobayashi, T. Kure, and K. Seki, *Proc. IEEE* 87, 633 (1999).
109. K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Kure, and K. Seki, in "ISSCC'98 Digest of Technical Papers," p. 344. IEEE Press, New York, 1998.
110. K. K. Likharev and A. N. Korotkov, in Proc. of the 1995 Int. Semicond. Device Res. Symp. (ISDRS)," p. 355. University of Virginia, Charlottesville, 1995.
111. C. D. Chen, Y. Nakamura, and J. S. Tsai, *Appl. Phys. Lett.* 71, 2038 (1997).
112. A. N. Korotkov, *J. Appl. Phys.* 92, 7291 (2002).
113. K. K. Likharev, *Appl. Phys. Lett.* 73, 2137 (1998).
114. A. Korotkov and K. Likharev, in "Techn. Dig. IEDM," p. 223. IEEE Press, Piscataway (NJ), 1999.
115. K. K. Likharev, *IEEE Circuits and Devices.* 16, 16 (2000).
116. H. Shichijyo, S. D. S. Malhi, A. H. Shah, G. P. Pollack, W. F. Richardson, M. Elahy, S. Banerjee, R. Womack, and P. K. Chaterjee, in "Ext. Abstr. of Int. Conf. on Solid State Devices and Materials," p. 265. 1984.
117. N. J. Stone and H. Ahmed, *Appl. Phys. Lett.* 73, 2134 (1998).
118. K. Nakazato and H. Ahmed, *Appl. Phys. Lett.* 66, 3170 (1995).
119. Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, *Appl. Phys. Lett.* 74, 1293 (1999).
120. Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, *IEEE Trans. Electron Devices* 47, 2334 (2000).
121. A. C. Irvine, Z. A. K. Durrani, and H. Ahmed, *J. Appl. Phys.* 87, 8594 (2000).
122. T. Osabe, T. Ishii, T. Mine, F. Murai, and K. Yano, in "Techn. Digest IEDM," p. 301. IEEE Press, Piscataway (NJ), 2000.
123. K. Nakazato, K. Itoh, H. Mizuta, and H. Ahmed, *Electron. Lett.* 35, 848 (1999).
124. K. Nakazato, K. Itoh, H. Ahmed, H. Mizuta, T. Kisu, M. Kato, and T. Sakata, in "Dig. of Techn. Papers of ISSCC," p. 132. IEEE Press, Piscataway (NJ), 2000.
125. J. H. Yi, W. S. Kim, S. Song, Y. Khang, H.-I. Kim, J. H. Choi, H. H. Lim, N. I. Lee, K. Fujihara, H.-K. Kang, J. T. Moon, and M. Y. Lee, in "Techn. Dig. IEDM," p. 787. IEEE Press, Piscataway (NJ), 2001.
126. J. F. Scott, "Ferroelectric Memories." Springer, New York, 2000.
127. S. Tehrani, B. Engel, J. M. Slaughter, E. Chen, M. DeHerrera, M. Durlam, P. Naji, P. Whig, J. Janesky, and J. Calder, *IEEE Trans. Magn.* 36, 2752 (2000).
128. K. Inomata, *IEICE T. Electron.* E84C, 740 (2001).
129. S. Lai and T. Lowrey, in "Techn. Dig. IEDM," p. 803. IEEE Press, Piscataway (NJ), 2001.
130. Web site of Ovonyx, (http://www.ovonyx.com/ovonyxtech.html).
131. M. Reed, *Proc. IEEE* 87, 652 (1999).
132. C. P. Collier, E. W. Wong, M. Belogradský, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, and J. R. Heath, *Science* 285, 391 (1999).
133. J. Chen, M. Reed, A. M. Rawlett, and J. M. Taur, *Science* 286, 1550 (1999).
134. C. P. Collier, G. Mattersteig, E. W. Wong, Y. Lyo, K. Beverly, J. Sampaio, F. M. Raymo, J. F. Stoddart, and J. R. Heath, *Science* 289, 1172 (2000).
135. Z. J. Donhauser, B. A. Mantooth, K. F. Kelly, L. A. Bumm, J. D. Monnell, J. J. Stapleton, D. W. Price Jr., A. M., Rawlett, D. L. Allara, J. M. Tour, and P. S. Weiss, *Science* 292, 2303 (2001).
136. J. R. Tucker, *J. Appl. Phys.* 72, 4399 (1992).
137. A. N. Korotkov, R. H. Chen, and K. K. Likharev, *J. Appl. Phys.* 78, 2520 (1995).
138. R. H. Chen, A. N. Korotkov, and K. K. Likharev, *Appl. Phys. Lett.* 68, 1954 (1996).
139. H. Iwamura, M. Akazawa, and Y. Amemiya, *IEICE T. Electron.* E81–C, 42 (1998).
140. R. H. Chen and K. K. Likharev, *Appl. Phys. Lett.* 72, 61 (1998).
141. K. K. Likharev, and V. K. Semenov, in Ext. Abstr. of Int. Supercond. Electronics Conf., Tokyo (1987), p. 128.
142. S. V. Vyshenskii, S. P. Polonsky, and K. K. Likharev, "Single-electron logic circuits," 1990, unpublished (described in [21]).
143. P. Bakshi, D. A. Broido, and K. Kempa, *J. Appl. Phys.* 70, 5150 (1991).
144. C. S. Lent, P. D. Tougaw, W. Porod, and G. Bernstein, *Nanotechnology* 4, 49 (1993).
145. K. Nomoto, R. Ugajin, T. Suzuki, and I. Hase, *Electron. Lett.* 29, 1380 (1993).
146. D. V. Averin, L. F. Register, K. K. Likharev, and K. Hess, *Appl. Phys. Lett.* 64, 126 (1994).
147. P. D. Tougaw and C. S. Lent, *J. Appl. Phys.* 75, 1818 (1994).
148. S. Bandyopadhyay, B. Das, and A. E. Miller, *Nanotechnology* 4, 113 (1994).
149. A. N. Korotkov, *Appl. Phys. Lett.* 67, 2412 (1995).
150. M. G. Ancona, *J. Appl. Phys.* 78, 3311 (1995).
151. K. Nomoto, R. Ugajin, T. Suzuki, and I. Hase, *J. Appl. Phys.* 79, 291 (1996).
152. C. S. Lent and P. D. Tougaw, *J. Appl. Phys.* 80, 4722 (1996).
153. S. Bandyopadhyay and V. Roychowdhury, *Jpn. J. Appl. Phys., Part 1* 35, 3350 (1996).
154. A. V. Krasheninnikov and L. A. Openov, *JETP Lett.* 64, 231 (1996).
155. S. N. Molotkov and S. S. Nazin, *JETP* 83, 794 (1996).
156. C. S. Lent and P. D. Tougaw, *Proc. IEEE* 85, 541 (1997).
157. N.-J. Wu, N. Asahi, and Y. Amemiya, *Jpn. J. Appl. Phys.* 36, 2621 (1997).
158. N. Asahi, M. Akazawa, and Y. Amemiya, *IEEE Trans. Electron Devices* 44, 1109 (1997).
159. V. P. Roychowdhury and M. P. Anantram, in "Proceedings IEEE Conference on Application-Specific Systems, Architectures, and Processors" (L. Thiele, Eds.), p. 14. IEEE Comp. Soc., Los Alamitos (CA), 1997.
160. A. M. Bychkov, L. A. Openov, and I. A. Semenihin, *JETP Lett.* 66, 298 (1997).
161. Y. Fu and M. Willander, *J. Appl. Phys.* 83, 3186 (1998).
162. A. Gin, P. D. Tougaw, S. Willams, *J. Appl. Phys.* 85, 8281 (1999).

163. C. Ungarelli, S. Francaviglia, and M. Macucci, and G. Iannaccone, *J. Appl. Phys.* 87, 7320 (2000).
164. J. R. Pasky, L. Henry, and P. D. Tougaw, *J. Appl. Phys.* 87, 8604 (2000).
165. M. T. Niemier and P. M. Kogge, *Int. J. Circ. Theor. App.* 29, 49 (2001).
166. L. Bonci, G. Iannaccone, and M. Macucci, *J. Appl. Phys.* 89, 6435 (2001).
167. J. Timler and C. S. Lent, *J. Appl. Phys.* 91, 823 (2002).
168. R. A. Kiehl and T. Ohshima, *Appl. Phys. Lett.* 67, 2494 (1995).
169. T. Ohshima, and R. A. Kiehl, *J. Appl. Phys.* 80, 912 (1996).
170. T. Ohshima, *Appl. Phys. Lett.* 69, 4059 (1996).
171. F. Y. Liu, F. T. An, and R. A. Kiehl, *Appl. Phys. Lett.* 74, 4040 (1999).
172. S. Fölling, Ö. Türel, and K. K. Likharev, in "Proc. of the 2001 Int. Joint Conf. on Neural Networks," p. 216. International Neural Network Society, Mount Royal (NJ), 2001.
173. Ö. Türel, and K. K. Likharev, *Int. J. Circ. Theor. App.* 31, 37 (2003).
174. K. Likharev, A. Mayr, I. Muckra, and Ö. Türel, in Molecular Devices III (Annals of New York Acad. Sci., Vol. 1000), p. 146 (2003).
175. Ö. Türel, I. Muckra, and K. Likharev, in 'Proc. of the Int. Joint Conf. on Neural Networks", p. 365 (2003).
176. L. Bonci, M. Gattobigio, G. Iannaccone, and M. Macucci, *J. Appl. Phys.* 92, 3169 (2002).
177. M. Macucci "Critical Assessment of the QCA Architecture as a Viable Alternative to Large Scale Integration", preprint (2003).

www.aspbs.com/enn

Single-Electron Dynamics

Toshimasa Fujisawa

NTT Corporation, Atsugi, Japan, and Tokyo Institute of Technology, Tokyo, Japan

CONTENTS

1. Introduction: Single-Electron Dynamics
2. Static Characteristics of a Single Electron
3. Single-Electron Dynamics in Nanostructures
4. Toward Quantum Information Processing
Glossary
References

1. INTRODUCTION: SINGLE-ELECTRON DYNAMICS

Electrical current is microscopically a flow of large numbers of electrons, each of which carries an elementary charge, *e*. For instance, 1 A corresponds to a flow of approximately 10^{19} electrons per second. We generally do not care much about the transport of each electron for such a large current. However, as the electrical current becomes smaller and smaller, the flow of individual electrons becomes important. A commercially available current meter can measure down to a few femtoamperes (noise floor of a few fA/$\sqrt{Hz}$), which corresponds to about 10^4 electrons per second. Usually, we cannot observe the flow of individual electrons in an electrical conductor. However, recent developments in nanotechnology allow us to control and measure single-electron transport very accurately. An electron pump device, which carries exactly one electron during one cycle of voltage modulation, is considered a possible current standard with extremely high accuracy. A single-electron transistor has an extremely high charge sensitivity (about 10^{-5} electrons/$\sqrt{Hz}$) on a small conductive island. This device would detect individual electron transport of a current on the order of a few nanoamperes. With these techniques, various dynamical behaviors of single electrons can also be studied. Usually each electron moves randomly in a conductor. However, electron transport through a small conductive island, known as single-electron tunneling, can be somewhat correlated. An electron that has entered the small island leaves it before another electron is allowed to enter. Moreover, in a one-dimensional array of small islands, each electron is expected to transport more regularly. This research on single-electron transport has been started by understanding the dc and low-frequency transport characteristics, but now they are moving to more high-frequency dynamical single-electron transport characteristics, which we call *single-electron dynamics*.

Another important aspect of a single electron is that its behavior is governed by quantum mechanics. One of the nonclassical characteristics is the particle–wave duality. Although a single electron is a particle located somewhere when it is measured, it can behave as if it was located in two (or more) different places at one time. For instance, when an electron (the same is true for a photon or other single particles) is transmitted through and diffracted from a pair of small holes (a double slit), it appears randomly at a position on the screen behind the double slit. The probability of the appearance on the screen shows an interference pattern that is described by the wave characteristics of an electron (the quantum mechanical picture of an electron). One cannot determine the path (slit) an electron has taken without causing the interference pattern to disappear. This is an example of spatial interference, but similar interference can also appear in the time domain. Consider two conductive islands (a double quantum dot) located close to each other. When an electron is added to this double dot system, it occupies one of the two dots in the classical picture. However, quantum mechanics allows it to occupy both dots simultaneously. One cannot determine which dot the electron occupies at a given time, but one can determine the probability of finding the electron in each dot. When an appropriate electromagnetic field is applied to the system, the probability starts to oscillate. This *coherent dynamics of single electrons* is another interesting topic covered in this chapter.

A single electron rotates by itself (electron spin). Electron spin usually does not play an important role in nonmagnetic materials. However, it can determine the transport characteristics of a small island (a quantum dot) that contains a small number of electrons. Coulomb interactions of electrons confined in a small region induce different spin states, which give rise to spin-dependent transport. Single-electron spin in a quantum dot shows a long energy relaxation time and is expected to have a long decoherence time, which promises to provide the *coherent dynamics of single-electron spin*.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

One challenging area of research in coherent single-electron dynamics is to realize quantum computing hardware, in which information processing is performed in the time evolution of quantum states. Quantum computing is expected to perform some sorts of calculations efficiently at an amazing speed that is unavailable in a classical computing scheme. The single-electron charge state and the single-electron spin state in quantum dots are good candidates for basic building blocks of a quantum computer.

2. STATIC CHARACTERISTICS OF A SINGLE ELECTRON

A *small conductive island* that accommodates a tunable number of electrons is often used to manipulate a single electron [1]. The electrons in the island can be controlled one by one. When the island is so small that the energy quantization is significant, the island is often called a *quantum dot* [2]. Electrons occupy discrete levels, each of which has twofold spin degeneracy. When a few electrons are confined in a well-defined potential, the quantum dot is often referred to as an *artificial atom*. Electrons occupy well-defined orbitals, as in a normal atom.

These islands, quantum dots, and artificial atoms are usually connected to the source and drain electrodes through tunneling barriers to allow transport measurements. A couple of gate electrodes are also used to control the electrostatic potential of the island and the tunneling barriers. In this section, we summarize the fabrication techniques of the devices and their dc transport characteristics.

2.1. Fabrication Techniques

The fabrication techniques for single-electron devices have been developed together with large-scale integration technology for solid-state devices. Electron beam lithography, which patterns arbitrary fine structures with a few nanometers resolution, is often used to fabricate small structures. Here, some typical fabrication processes are summarized.

Well controllable quantum dots are often fabricated from two-dimensional electron gas (2DEG) formed in a modulation-doped heterostructure [3, 4]. The heterostructure consists of a thick high-quality undoped GaAs layer, an undoped $Al_xGa_{1-x}As$ layer (an Al content of $x \sim 0.3$), and a Si-doped $Al_xGa_{1-x}As$ layer. The 2DEG accumulates at the interface of the GaAs and AlGaAs layers. Since the 2DEG is spatially separated from the doping layer, it shows a high electron mobility of 10^5–10^7 cm^2/Vs, with a typical sheet carrier concentration of $n = 2$–5×10^{11} cm^{-2} at low temperature (<4 K). The quality of the 2DEG has been improved for its application to high-electron mobility transistors as well as for the physical interest in integer and fractional quantum Hall effects. Small structures, like quantum dots, can be fabricated by processing the heterostructure. For a typical fabrication procedure, the first step is a wet etching process for device isolation and defining a conductive channel. Next, some Ohmic contacts (an AuGeNi alloy) are fabricated by metal evaporation, which is followed by a thermal annealing process. Then, fine Schottky gates (Au/Ti) are evaporated on the surface with the aid of electron beam lithography. Additional etching processes using a fine electron beam lithographic pattern can define the conductive channel more precisely. Considering that the 2DEG is depleted approximately 50–200 nm away from the edge of the etched pattern and about 10–20 nm away from the moderately biased gate electrode, a small conductive island (quantum dot) of 100–1000 nm can be routinely fabricated. This technique provides high-quality nanostructures with multiple gates.

Extremely clean quantum dots containing a very few electrons have been fabricated in a pillar structure of a resonant tunneling heterostructure [5]. An undoped InGaAs potential well separated from n-type GaAs electrodes by AlGaAs tunneling barriers can be designed to accommodate 2DEG even at zero voltage. A pillar structure with a diameter of $\sim$0.5 μm and a height of $\sim$0.5 μm is patterned and fabricated by using electron beam lithography and dry and wet etching. Then a gate electrode is deposited around the pillar structure. Electrons flow from the top of the pillar to the bottom of the structure. The electrons are confined in a square well potential in the vertical direction and in a two-dimensional harmonic potential in the lateral direction. The advantage of this structure is that the electron number can be reduced to zero. Since the confinement potential is well defined, the electronic state in the quantum dot can be predicted from theoretical calculations.

Another fabrication process often used for superconductor and normal metal islands is shadow evaporation [6, 7]. By using a double (or triple) layer of resists having different sensitivities for electron beam lithography, suspended or overhanging resist layers can be fabricated. Then metals are evaporated from different angles to the sample to form multiple metal layers that overlap in some regions. A thin insulating layer can be inserted between metal layers by using an oxidation process. For example, two aluminum evaporations with oxidation between provide $Al/AlO_2/Al$ tunneling junctions. Small metal islands, tunneling barriers, gate electrodes, and source and drain leads can be fabricated on a substrate by properly designing the lithographic pattern. High-quality Al islands show superconducting characteristics at low temperature and in a low magnetic field.

Much smaller islands, or quantum dots, can be fabricated by using various nanotechnologies. For instance, crystal growth of thin InAs on a GaAs substrate results in the formation of nanometer-scale islands (10–50 nm). This is known as the Stranski–Krastanov growth mode, which occurs when there is a large difference in the crystal lattice constants. Source and drain electrodes, as well as gate electrodes, can also be fabricated for transport measurements [8]. Moreover, a variety of materials, such as nanoparticles of magnetic, superconductor, and semiconductor materials as well as molecules of carbon nanotubes and fullerenes, can also be used as small quantum dots. However, as the islands becomes smaller, fabrication processes for transport measurement generally become difficult.

Generally, larger quantum dots have smaller characteristic energies and work only at lower temperatures. However, they can be fabricated together with multiple gate electrodes to control the characteristics of the dot independently. Therefore, the fundamental characteristics of quantum dots are investigated in relatively large quantum dots at

very low temperature (<0.1 K), while better characteristics are obtained in smaller quantum dots.

2.2. Coulomb Blockade and Single-Electron Tunneling

Here we discuss how an electron can be manipulated in a small island. Consider a conductive island connected to the source and drain electrodes via tunneling barriers and connected to a gate electrode with a small capacitor, as schematically shown in Figure 1a and in the equivalent circuit diagram of Figure 1b. We introduce the orthodox Coulomb blockade theory, which describes the Coulomb blockade (CB) effect and single-electron tunneling (SET) behavior [1]. For simplicity we assume zero bias voltage is applied between the source and drain electrodes, $V_{SD} = 0$. The total energy, $U(N)$, of the system, in which an island containing N electrons is affected by a gate voltage, V_g, via a capacitance, C_g, is given by [9]

$$U(N) = \frac{(-Ne + C_g V_g + q_0)^2}{2C_\Sigma} + E_{int}(N). \qquad (1)$$

The first term is the electrostatic energy approximated by the constant capacitance model, that is, constant Coulomb interaction in the island. In parentheses is the sum of the electron charge on the dot, the induced charge by the gate, and an offset charge, q_0. C_Σ is the total capacitance of the dot. The second term, $E_{int}(N)$, is the sum of the energies of the N occupied electron levels, measured relative to the Fermi energy of the leads, accounting for the internal degrees of freedom of the QD. Other corrections to many-body interactions can be included in E_{int}. We neglect the second term in this section, but discuss it in detail in Section 2.3. Ignoring this term corresponds to considering a relatively large island containing many electrons that occupy the continuum density of states. Even if an electron is excited to a higher energy state, the electron can relax quickly to the minimum energy, $E_{int}^{(\min)}$, which is almost independent of N.

Therefore, by neglecting E_{int}, the total energy of the system changes with V_g as shown in Figure 1c. The number of electrons, N, is determined to minimize the total energy and therefore becomes a well-defined integer ($N = N_0$, $N_0 + 1$, $N_0 + 2, \ldots$). The energy gap to neighbor charge states can be maximized to $e^2/2C_\Sigma$, for instance, at $V_g = V_{g0}$. When excitation energies, such as the thermal energy, are much smaller than this energy, an electron can neither enter nor leave the island. Therefore, transport through the island is blocked (Coulomb blockade). The energy gap can be made 0 by adjusting the gate voltage at $V_g = V_{g1}$, where N can fluctuate between N_0 and $N_0 + 1$ only by 1, but not more than 1. This means that electrons tunnel through the island one by one. This single-electron tunneling scheme is maintained unless the excitation energy exceeds the charging energy, $E_c \equiv e^2/C_\Sigma$. The Coulomb blockade and single-electron tunneling appear alternately by sweeping the gate voltage with a period given by e/C_g (CB oscillation).

When the source–drain voltage, V_{SD}, is applied, the CB region shrinks and the SET region expands, as shown in the charging diagram of Figure 1d. The maximum width of a CB diamond region is given by E_c/e in the V_{SD} direction and e/C_g in the V_g direction. The CB and SET appear in diamond (parallelogram) regions, as shown in Figure 1d. The number of electrons in the dot, N, can be controlled by changing V_{SD} or V_g.

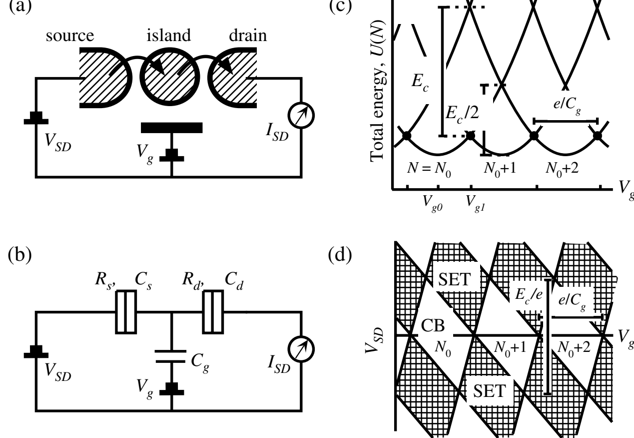


Figure 1. Coulomb blockade (CB) and single-electron tunneling (SET) in a single island. (a) Schematic diagram of the conductive region of a typical SET device (hatched regions). (b) Corresponding circuit diagram. The island is connected to the source and the drain via tunneling barriers and is connected to the gate electrode by a capacitor. (c) Total energy, $U(N)$, of the system as a function of the gate voltage, V_g. (d) The diagram of the CB and SET regions in the V_g–V_{SD} plane. Slightly asymmetric tunneling capacitances are assumed.

2.3. Energy Quantization and Many-Body Effects

When the size of an island (quantum dot) is so small that an electron wave can interfere, only the constructively interfering wave (electron orbital) can exist in the quantum dot, forming quantized energy levels [2, 10]. Each level can accommodate two electrons with spin-up and spin-down. In a very simple picture, electrons occupy levels starting from the lowest and making spin pairs. Therefore, the total spin, S, of the system is 0 for even N and $\frac{1}{2}$ for odd N. $E_{int}(N)$ in Eq. (1) is the sum of the occupied energy levels for all N electrons. For this quantum dot regime, the characteristic energy for CB and SET is the addition energy, E_{add}, which is the energy required to add one electron to the quantum dot. When an electron is added to an odd-N quantum dot, the addition energy is identical to the charging energy, $E_{add}(\text{odd-}N) = E_c$. However, when an electron is added to an even-N quantum dot, excess energy equivalent to the energy spacing, Δ, between quantized levels is required, $E_{add}(\text{even-}N) = E_c + \Delta$. The difference in the addition energy is reflected in the width of CB regions, which now show even–odd asymmetry (wider width for even N).

However, when N is relatively small, electrons in a quantum dot interact, and the width of the CB region exhibits more complicated variations with N. This comes from the Coulomb interactions between the electrons in a quantum dot. Electrons occupy different orbitals that have different spatial distributions in the dot. Electrons prefer to occupy

different orbitals in order to reduce Coulomb repulsion (the direct Coulomb interaction). Two electrons occupy different orbitals, rather than make a spin pair, if the energy gain is larger than the energy cost for taking a higher energy orbital. If electrons occupy different orbitals, their spins prefer to maximize the total spin in order to gain the exchange Coulomb energy. These Coulomb interactions determine the electron filling in a quantum dot (many-body effects). The width of CB regions changes with N in a complicated manner. In other words, one can study the many-body effects from the N-dependent addition energy.

The many-body effects can be clearly studied in a quantum-dot disk, in which electrons are confined in a two-dimensional harmonic potential [10–12]. In this case, the orbitals can be expressed analytically and can be identified as 1s, 2p, 3s, and 3d,... from the lowest energy to higher energy. Such a quantum dot is referred to as an artificial atom. The ground state of the one-electron quantum dot (artificial hydrogen atom) is one electron in the 1s orbital. This configuration does not change with the magnetic field. However, the ground state of the two-electron quantum dot (artificial helium atom) is the spin-singlet state having two anti-parallel-spin electrons in the 1s orbital, or the spin-triplet state having parallel-spin electrons in the 1s and 2p orbitals, depending on the strength of the Coulomb interactions and the level spacing. The transition from the spin-singlet state to the spin-triplet state occurs by applying a magnetic field ($B = 2$–6 T), which increases the Coulomb interactions and decreases the level spacing. Note that similar singlet–triplet transition in a normal helium atom is expected to occur at an extremely high magnetic field of 4×10^5 T. One can study simple atomic physics in moderate parameter spaces using an artificial atom (see Section 3.10).

2.4. Double Quantum Dot

A double quantum dot comprises two quantum dots separated by a tunneling barrier [13–16]. If a quantum dot is an artificial atom, a double quantum dot can be regarded as an artificial diatomic molecule. The two quantum dots can be coupled electrostatically (an ionic bond in the language of chemistry) as well as quantum mechanically (a covalent bond). Consider two quantum dots connected through a tunneling barrier with a capacitor, C_c, and tunneling coupling, T_c. Figure 2a is the schematic diagram of the double dot connected between the source and drain in series. For a moment, we neglect the effect of the source and drain contacts for simplicity. Figures 2b–2d show the charging diagram of the double quantum dot in different regimes [14]. The gate voltages, V_{gl} and V_{gr}, are swept to shift the potential of the respective dots; (n, m) represents the stable charge state in which n and m electrons occupy the left and the right dots, respectively. When there are no interactions between the two dots ($C_c \sim 0$ and $T_c \sim 0$), the charging diagram is just Coulomb blockade oscillations of the two dots [see Fig. 2b]. If only the electrostatic coupling is turned on, ($C_c > 0$, $T_c \sim 0$), charging of an electron on one quantum dot lifts up the potential of the other dot. And the gate voltage for one dot also affects the potential of the other dot

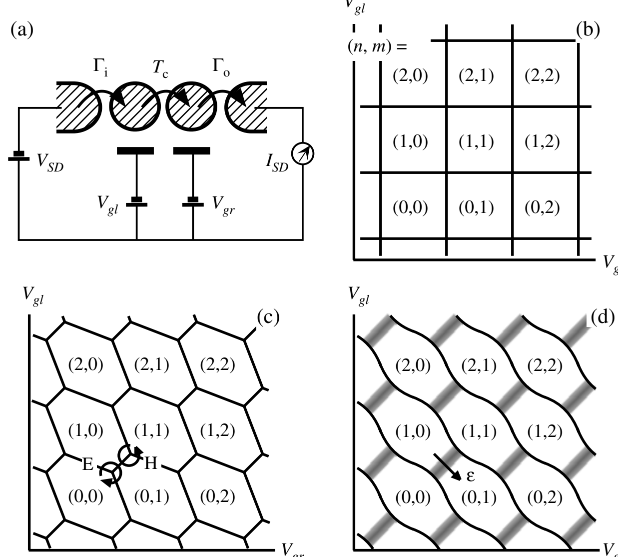


Figure 2. Charging diagram of a double quantum dot. (a) Schematic circuit diagram of a double dot. (b–d) Charging diagrams in the different regimes. (a) Both electrostatic and tunneling couplings are neglected. (c) Only electrostatic coupling is considered. Single-electron transport through the double dot with an electron-like process at the triple point, E, and with a hole-like process at H. Circled arrows are typical gate–voltage trajectories in the single-electron pumping operation. (d) Both electrostatic and tunneling couplings are considered. The charge state is no longer a good quantum state in the hatched regions.

via the coupling capacitor. In this case, the charging diagram becomes a hexagonal honeycomb structure, as shown in Figure 2c. Each crossing point in Figure 2b splits into two triple points [see E and H in Fig. 2c], whose spacing corresponds to the interdot coupling energy defined by

$$E_{id} = e^2 C_c / C_{\Sigma l} C_{\Sigma r} \tag{2}$$

Here $C_{\Sigma l}$ and $C_{\Sigma r}$ are the total capacitance of the left and the right dots, respectively. The transport through the double dot from the source to the drain is allowed only at these triple points, where three charge states are energetically degenerated. At the triple point indicated by E, an electron tunnels from the left to the right (the clockwise direction in the charging diagram) or the right to the left (the counterclockwise direction) sequentially. However, the tunneling sequence at the triple point H is somewhat peculiar. The electron tunneling process propagates in the reverse direction of the electron flow. It is easy to conceive that a hole, rather than an electron, tunnels sequentially.

Now we discuss the amplitude of current through a double quantum dot. We assume that electrons can tunnel between the two dots elastically, that is, only when the discrete energy states of the two quantum dots are aligned approximately within the lifetime broadening of the states. Then we expect a very sharp peak at the resonance, whose current profile is basically independent of the thermal distributions in the leads. In the vicinity of a triple point, only one energy state in each dot contributes to the transport [17]. At a large source–drain voltage, which saturates the current, the current profile is determined by the energy difference between

the two charge states, ε, and it is given by a Lorentzian function as

$$I(\varepsilon) = \frac{e}{\hbar} \frac{\Gamma_o T_c^2}{\varepsilon^2 + \Gamma_o^2/4 + T_c^2(1 + \Gamma_o/\Gamma_i)} \quad (3)$$

where Γ_i and Γ_o are the tunneling rate of the incoming and outgoing barriers, respectively [15]. Note that Γ_i and Γ_o appear asymmetrically, so the three parameters, T_c, Γ_i, and Γ_o, can be deduced from the current profiles obtained at different current directions. This is a convenient way to determine the parameters experimentally, if the inelastic current (see Section 3.11) and other currents from excited states can be neglected.

If the tunneling coupling, T_c, is made larger than Γ_i and Γ_o, coherent tunneling coupling (covalent bond) between the two dots is expected [14, 18, 19]. The charging diagram deforms as shown in Figure 2d. In the gray regions, one cannot distinguish whether a single electron is in the left or right dot. The localized states in each dot are mixed into bonding and antibonding states, which are energetically separated by $2\hbar T_c$ at the original triple points. The current peak is broadened and elongated into a crescent shape in the vicinity of the original triple point. Finally, if the tunneling coupling becomes too strong, the double dot becomes effectively a large single dot, in which just single CB oscillations are expected. In the intermediate and intriguing coupling regime, two dots are coupled strongly and coherently, but are weakly coupled to the source and drain electrodes just to measure the current. This double dot can be considered a tunable two-level system, which is discussed in Section 2.5.

2.5. Two-Level Systems—Electron Spin and Pseudo-Spin

The two-level system is the simplest system in which quantum mechanical characteristics can appear and provides an important model for the study of the quantum dynamics. After describing a single electron spin as a typical two-level system, we define a tunable two-level system in a double quantum dot.

An electron spin is fully described by quantum mechanics. The direction of an electron spin can be arbitrary in principle. However, electron spin takes spin-up, denoted here by $|0\rangle$, or spin-down, $|1\rangle$, when the z component of the spin is measured. An arbitrary spin state before the measurement is described by a linear superposition of these two states (bases) as

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \quad (4)$$

where θ and φ are mixing angles that determine the probability of measurement results. This representation is chosen to point out the spin orientation in spherical coordinate, and it is convenient for describing the state in the Bloch sphere of Figure 3.

Now, consider an electron spin in the presence of magnetic field $B = (B_x, B_y, B_z)$. The Hamiltonian of the system can be described as

$$H = -\frac{1}{2}\mu_s(\sigma_x B_x + \sigma_y B_y + \sigma_z B_z) \quad (5)$$

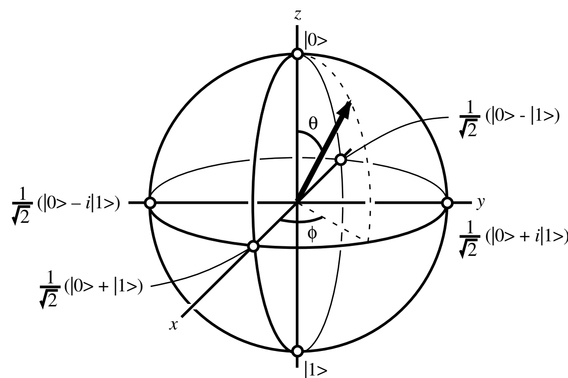


Figure 3. Bloch sphere of a spin 1/2 system. Any two-level system can be mapped onto the Bloch sphere. The arrow, θ and ϕ in the polar coordinate, indicates the state, $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$.

where σ_x, σ_y, and σ_z are the Pauli matrices, and μ_s is the magnetic moment of a spin. The stationary eigenstates of the system can be obtained by solving the time-independent Schrödinger equation, $H|\psi\rangle = E|\psi\rangle$. This gives the eigen energies

$$E_\pm = \pm\frac{1}{2}\mu_s\sqrt{B_x^2 + B_y^2 + B_z^2} \quad (6)$$

and the corresponding eigenstates of

$$|\psi_+\rangle = \sin\frac{\theta}{2}|0\rangle - e^{i\varphi}\cos\frac{\theta}{2}|1\rangle \quad (7)$$

$$|\psi_-\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \quad (8)$$

where θ and φ are the directions of the magnetic field. This argument looks roundabout, since the z direction may be chosen as the direction of the magnetic field for simplicity. However, Eqs. (6) and (7) give general eigenstates for any two-level system under a Hamiltonian, which can be expressed by fictitious magnetic fields (Eq. 5).

Now, we define a two-level system in a double quantum dot [16, 18]. We choose the ground state of the charge state (1, 0) as one state $|0\rangle$ and the ground state of the charge state (0, 1) as the other state $|1\rangle$. Other excited states and other charge states can be ignored, if they are energetically much higher than the lowest two energy states, $|0\rangle$ and $|1\rangle$, that is, if the double quantum dot is properly adjusted somewhere in between a pair of triple points (gray regions in Fig. 2d). In addition, the double dot can be effectively isolated from the source and drain electrodes because the interdot charging energy can block the transport (see Section 2.4). The energy difference, ε, between the two states can be changed by the gate voltages, as shown by the arrow in Figure 2d. The coherent tunneling coupling, T_c, allows the system to change between $|0\rangle$ and $|1\rangle$. In this case, the Hamiltonian of the system is written as

$$H = \frac{1}{2}\varepsilon\sigma_z + \hbar T_c\sigma_x \quad (9)$$

where $\frac{1}{2}\varepsilon$ and $\hbar T_c$ can be regarded as fictitious magnetic fields in the z and x directions, respectively (compare with Eq. 5). Therefore, we can map the two-level

system of the double quantum dot onto the spin system (pseudo-spin model). Similarly any two-level system can be mapped onto the spin system, and its superposition can be represented in the Bloch sphere. The eigen energies of the double quantum dot are given by

$$E_{\pm} = \pm\frac{1}{2}\sqrt{\varepsilon^2 + 4(\hbar T_c)^2} \tag{10}$$

which is schematically shown in Figure 4a [18]. The corresponding wavefunctions, which are schematically shown in Figure 4b–d, are symmetric for energy E_- (often called the symmetric state or bonding state) and anti-symmetric for energy E_+ (antisymmetric state or antibonding state). And the eigenstates are always in the x–z plane ($y = 0$) of the Bloch sphere.

3. SINGLE-ELECTRON DYNAMICS IN NANOSTRUCTURES

In this section, we discuss the dynamical behavior of a single electron. In most of the phenomena, various types of potential modulations are included in the discussion of time-integrated tunneling current. When the modulation is relatively slow compared to the tunneling rate, the system changes adiabatically. We can still use the stationary states of quantum dots in the absence of the modulation. However, if the modulation frequency becomes higher than the tunneling rate, the system changes nonadiabatically, and the state is no longer the original stationary state. We have to use time-dependent Schrödinger equations, $i\hbar\, d\psi/dt = H\psi$,

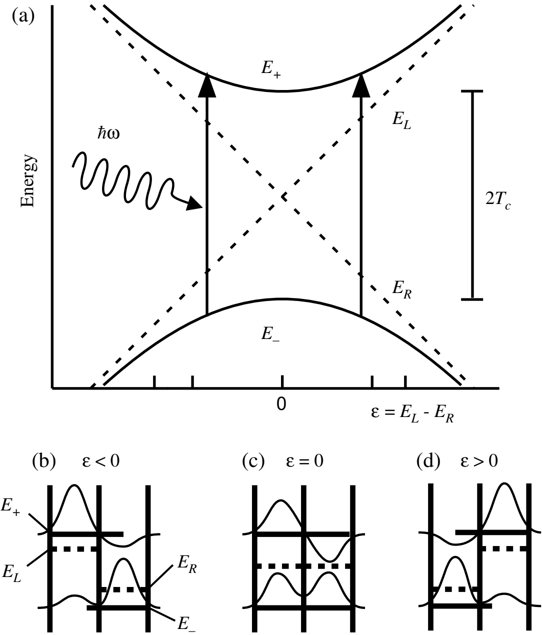


Figure 4. (a) Schematic energy diagram of a double quantum dot. E_L and E_R are the energies of the localized charge states (1, 0) and (0, 1), respectively. When the tunneling coupling is turned on, the bonding state with the energy E_- and the antibonding state with the energy E_+ are formed. (b–d) Schematic energy diagram and wavefunctions of the bonding and antibonding states for (b) $\varepsilon < 0$, (c) $\varepsilon = 0$, and (d) $\varepsilon > 0$.

to describe the dynamics. This section also describes the decoherence mechanisms, which are very important characteristics in the dynamics. Finally, measurement techniques for single electron states in a relatively short time domain are discussed.

3.1. Single-Electron Turnstile and Pump

Single-electron turnstiles and pumps are single-electron devices that carry exactly one electron from one electrode (the source) to the other electrode (the drain) during one cycle of the periodic potential modulation [20]. During the first half of the modulation period one electron tunnels from the source to the island, and the electron tunnels out to the drain during the second half of the cycle. By repeating the modulation at a sufficiently high frequency, typically $f \sim 10$ MHz, a reasonable electrical current, $I = ef \sim 1.6$ pA, is obtained. These devices have been discussed as a possible current standard for metrology [21]. The frequency-locked current in a single electron pump, the frequency-locked voltage generated by the Josephson effect, and the quantum resistance realized by the quantum Hall effect are expected to close the quantum metrology triangle formed by voltage, current, and frequency.

Current in a turnstile device is driven by an external dc bias voltage applied between the source and drain, and the current is regulated so that only one electron is allowed to pass through the device during one cycle [22]. The current direction is given by the polarity of the bias voltage, and the current vanishes at zero bias voltage. On the other hand, the current in single-electron pumps is generated by a pumping process from the source to the drain. The direction of the pumping current is independent of the applied bias voltage, and it works even at zero bias voltage [23]. Usually, the turnstile devices are operated with a single modulation voltage, while the pump devices are operated with more than two gate voltages modulated with different phases.

Consider pumping processes in a double quantum dot. The charge state of a double quantum dot is represented by (n, m), where n and m are the number of excess electrons in the double dot [see Fig. 2c]. If the system is initially in the (0, 0) state, modulation voltages are designed to move the charge state to the (1, 0) state, followed by the (0, 1) state, and return it to the initial state (0, 0). During this cycle, an electron moves from the left to the right sequentially. This trajectory is schematically shown by an arrow enclosing the triple point E in Figure 2c. The other trajectory in the reversed direction enclosing the other triple point H carries a hole (0 state) from the right to the left, which gives the same current direction. In this way, exactly one electron, or hole, is carried from one lead to the other in one cycle of the potential modulation.

For metrology, an extremely high accuracy of less than 10^{-8} errors per cycle is required. The deviation from the frequency-locked current arises from thermal activation, Poisson tunneling statistics, the cotunneling process, and so on. The probability of thermal excitation to other charge states can be suppressed below 10^{-8} at a dilution-refrigerator temperature of less than 50 mK for a typical double dot with interdot charging energy of about 100 μeV

[20]. The current error also arises from the Poisson statistics of tunneling events, whose interval is widely distributed. Even when an electron is allowed to tunnel, the tunneling may not happen during the finite period of the cycle. This error can be reduced by making the modulation frequency much smaller than the tunneling rate of the barriers.

The cotunneling process, which is the simultaneous tunneling over two (or more) tunneling junctions via an intermediate virtual state [24], is a serious problem. The cotunneling current survives even at zero temperature. Although the cotunneling current is much smaller than the frequency-locked current, cotunneling can easily degrade the accuracy of the current standard, especially when higher current is required. The cotunneling current can be reduced either by driving the pumping device with a high resistor located very close to the device or by using multiple tunneling junctions connected in series.

Other practical problems arise from low-frequency electrical noise and background charge fluctuations, which shift the effective operating gate voltage randomly, and high-frequency (microwave to far-infrared region) noise, which gives rise to photon-assisted tunneling processes [25]. By adjusting the operating parameters properly, a single-electron pump with seven tunneling junctions works very nicely with an error per pumped electron of 1.5×10^{-8} [26]. Single-electron pumps provide extremely accurate current of the order of $\sim$1 pA. Another way to achieve a possible current standard at a relatively large current is the moving quantum dot, which is discussed in Section 3.3.

The single-electron pump is a fundamental tool that allows us to manipulate electrons one by one. In principle, one can design the voltage waveform so that a single electron is injected into or extracted from an island *on demand* within the accuracy of the inverse of the tunneling rate. One can add any operation or manipulation during the interval.

3.2. Single-Electron Tunneling Oscillation

Single electron tunneling, in which electrons travel through an island one by one, is a sort of correlated tunneling process over the two junctions. In normal SET devices, tunneling events are uncorrelated *in time*, and the interval between them is distributed according to Poisson statistics. However, the distribution of the tunneling interval can be made narrower than Poisson statistics (sub-Poisson statistics), and each electron tunneling can take place more regularly (SET oscillation).

A very simple scheme for SET oscillation is a small tunneling junction driven by a constant current. When the tunnel junction with a small tunneling capacitance is driven by a small constant current, the voltage across the junction increases linearly in time due to the charging. Single-electron tunneling is allowed only when the voltage exceeds the charging energy of the junction and the voltage drops suddenly by discharging (tunneling of) an electron. Therefore, the voltage across the junction shows a sawtooth pattern in time. The charging and discharging should take place almost regularly as long as the junction is driven by a constant current. The constant current source can in principle be generated by a high-resistance resistor connected to the junction [27, 28]. However, it is not easy to fabricate such a resistor. A small parasitic capacitor in the vicinity of the junction smears the sawtooth pattern of the voltage difference, and the junction is then effectively driven by a constant voltage.

Another way to achieve SET oscillation is with a one-dimensional (1D) array of small islands connected by tunneling barriers. When an excess electron is injected into the array, potential distribution inside the array shows soliton characteristics (charge soliton). The soliton has a characteristic size, $M \sim \sqrt{C_T/C_0}$, in the unit of an island, where C_T and C_0 are the tunneling capacitance and the self-capacitance of an island, respectively. If the length of the array is longer than $2M$, transport through it is influenced by Coulomb repulsion between the solitons. The solitons line up in a Wigner lattice inside the array and propagate in a regular way (SET oscillation). This type of SET oscillation has been indirectly observed under microwave irradiation, where the SET oscillation is locked at the applied frequency [29]. The appearance of nonlinear characteristics under microwave irradiation indicates correlated transport in the one-dimensional array of islands.

The correlated tunneling process can also be studied in the shot noise of the single-electron tunneling current. When the current is carried by a Poisson process, the shot noise is given by $2e|I|$ (full shot noise), where I is the current. The suppression from the full shot noise indicates time-correlated tunneling and sub-Poissonian statistics. Even in a conventional single quantum dot device, the shot noise can be suppressed by half if the incoming and outgoing tunneling rates are almost the same [30, 31]. It is theoretically expected that shot noise in resonant Cooper pair tunneling can be suppressed significantly [32].

3.3. Moving Quantum Dots

In this section, we discuss the unique characteristics of moving quantum dots, in which an array of quantum dots each containing exactly one electron propagates from the source to the drain. When a high-frequency voltage (typically a few giga Hertz) is applied to a metal electrode on a piezoelectric material (such as GaAs), a surface acoustic wave (SAW) is generated and propagates on the surface. The propagating SAW deforms the crystal lattice, thus modulating the conduction band in the vicinity of the surface through the piezoelectric interaction between the lattice deformation and the electrons. Therefore, a sinusoidal potential for electrons propagates with the SAW velocity, $v = 2980$ m/s for GaAs. When a one-dimensional channel is prefabricated along the SAW propagation direction by conventional lithographic techniques, a moving 1D array of potential puddles (quantum dots) is formed. Each quantum dot carries a well-defined integer number of electrons, n ($= 1, 2, 3, \ldots$), which can be tuned by external voltages or the intensity of the SAW. The acoustoelectric current is quantized to $I = nef$ with the SAW frequency, f [33–37]. In this case, electrons are transferred with the propagating potential, and no tunneling processes are involved. The current amplitude can be on the order of nanoamperes (I $\sim$ 0.48 nA for $n = 1$ and $f = 3$ GHz), which is about 3 orders of magnitude larger than that in electron pump devices. Therefore, more practical current-standard devices are expected.

In order to obtain a sufficient potential modulation, the SAW is usually generated with a narrow-band interdigital transducer situated a few millimeters from the electrical channel [34]. The typical wavelength of the SAW is about 1 μm for $f_{SAW} = 3$ GHz and should be comparable to the length of the one-dimensional channel. Application of a few milliwatts of microwave power to the transducer gives a typical potential modulation of about 1 meV in the active region. Note that the modulation amplitude is significantly reduced by the screening effect from the existing electrons.

The accuracy of the current is expected to be excellent (less than 10^{-10} errors for a modulation amplitude of about 20 meV), when the electron transport is driven adiabatically [33, 38]. Practical error may come from insufficient potential modulation, background charge fluctuations, the existence of standing wave modes, thermal effects, and so on. Some devices show current accuracy of about 5×10^{-5} [35, 36], but this accuracy is not yet enough for metrology purposes. Further improvements are desired for this device.

3.4. Photon-Assisted Tunneling

In the previous sections, the energy states in quantum dots are assumed to be shifted adiabatically in proportion to the applied modulation voltages. This is a good approximation, provided the modulation frequency is much lower than the tunneling rate. However, when the modulation frequency exceeds the tunneling rate, a quantum mechanical description of the system is required in order to understand the nonadiabatic transport characteristics. In this and the next few subsections, we concentrate on sinusoidal potential modulation, in which the modulation can be regarded as a coherent photon field. Various kinds of electron–photon interactions analogous to those in quantum optics are expected. For instance, electrons can tunnel through a barrier by absorbing or emitting a photon. This photon-assisted tunneling (PAT) has been theoretically developed in the Tien–Gordon theory [39, 40] and experimentally studied in the transport through Josephson junctions [41] and resonant tunneling structures [42, 43], as well as quantum dots [44, 45]. Coherent electron–photon coupling is expected and has actually been demonstrated in double quantum dots and superconductor islands.

When a sinusoidal voltage, $V(t) = V_0 \sin \omega t$, is applied across a tunneling barrier, the wavefunction of an electron on one electrode acquires a phase shift, $\int eV(t)\, dt/\hbar$, relative to the other electrode. The wavefunction can be written as a superposition of a series of photon sidebands as

$$\psi(x, t) = \psi_0(x, t) \sum_{n=-\infty}^{\infty} J_n(\alpha) \exp(-in\omega t) \qquad (11)$$

where $J_n(\alpha)$ is the nth order Bessel function of the first kind, $\alpha \equiv eV_0/\hbar\omega$ is the normalized modulation amplitude, and $\psi_0(x, t)$ is the original stationary wavefunction in the absence of sinusoidal voltage. The energies of the photon sidebands are separated by the photon energy $\hbar\omega$, and the amplitude of each wavefunction is proportional to $J_n(\alpha)$ (as shown schematically in the inset of Fig. 3a). Therefore, the photon-assisted tunneling can be considered a normal tunneling process to one of the photon sidebands. The effective tunneling rate to the nth sideband state is given by

$$\Gamma_n = J_n^2(\alpha)\Gamma \qquad (12)$$

where Γ is the original tunneling rate in the absence of modulation. Here the energy dependence of Γ is neglected for simplicity; $n = \pm 1, \pm 2, \ldots$ corresponds to the tunneling with $|n|$ photon absorption (for positive n) and emission (for negative n). Equation (12) can also be applied in the case of intense microwaves, where nonlinear optical effects can appear. Figure 5a shows the function $J_n^2(\alpha)$, which indicates how the effective tunneling rate (or current) changes with the normalized modulation amplitude α. For a weak amplitude of $\alpha \ll 1$, Γ_n is proportional to the nth power of the photon power, that is $J_n^2(\alpha) \sim (\alpha^2)^n$. For a strong amplitude of $\alpha \gtrsim 1$, however, $J_n^2(\alpha)$ is an oscillating function of α. At a specific amplitude where $J_1(\alpha) = 0$, for instance at $\alpha \sim 2.8$, tunneling through a one-photon sideband should vanish even in the presence of an intense oscillating potential. It should be noted that the zero-photon tunneling rate Γ_0 for $n = 0$ is also affected by the photon field, and the total sum is always conserved, $\sum_{n=-\infty}^{\infty} J_n^2(\alpha) \equiv 1$.

The Tien–Gordon theory explains very well the I–V characteristics of Josephson junctions and other tunneling devices, such as resonant tunneling structures and superlattices, in the presence of microwaves, millimeter waves, THz waves, and far-infrared light [40–43]. Here we focus on the photon-assisted tunneling in single-electron systems

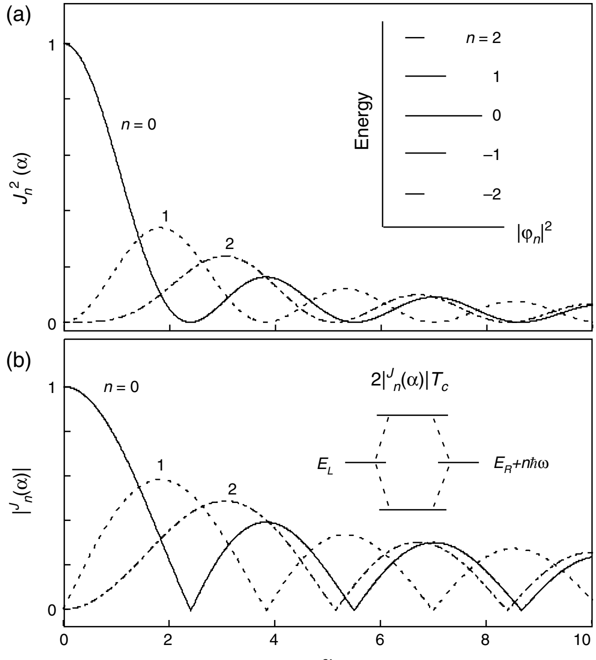


Figure 5. (a) The square of the Bessel function. The photon-assisted tunneling current is proportional to the square of the Bessel function. The inset shows the sideband structure in the presence of oscillating potential. The length of the bars schematically indicates the amplitude of the wavefunctions. (b) The absolute value of the Bessel function. The energy splitting of the two states separated by n times the photon energy is proportional to the absolute value of the Bessel function.

[44, 45]. A microwave (photon energy of 4–200 μeV for microwave frequency of 1–50 GHz) is often used as a coherent photon source. Application of a microwave on a gate or source–drain electrodes induces photon-assisted tunneling on the two tunneling barriers. Since the two tunneling barriers are close to each other, it is technically difficult to control the modulation amplitude across each barrier independently, and photon-assisted tunneling can occur at either barrier. Nevertheless, the flow of current through the device can be quantitatively explained by the Tien–Gordon theory. The photon-assisted tunneling from the continuum density of states in the electrode to a discrete energy state in the quantum dot produces a pumping current. A broad pumping current with a width corresponding to the photon energy appears in the vicinity of the SET current peak.

When a microwave is properly applied to modify the potential of the quantum dot, coherent tunneling through one of the photon sidebands is expected. An electron can tunnel into the dot by absorbing one photon through one barrier and simultaneously tunnel out from the dot by emitting one photon through the other barrier. Distinct resonant PAT current peaks have been observed [46], indicating a coherent PAT process. In this case, the PAT current peaks are separated from the SET current peak by the photon energy. Furthermore, when the photon energy is made higher than the energy spacing in the quantum dot, one can excite an inner electron in the quantum dot to the reservoir. This process is analogous to the photoionization process of atoms.

3.5. PAT in a Double Quantum Dot

A more interesting case is the electron–photon interaction in a double quantum dot, where an oscillating potential is applied between two discrete energy states [47, 48]. Note that a similar situation can be realized in a superconducting island, where photon-assisted tunneling of a Cooper pair is observed between superconducting charge states in the island [49]. In this subsection, we restrict ourselves to a weakly coupled double dot, where an electron tunnels sequentially through three barriers (see Section 3.6 for the strongly coupled double dot). In this case, we can still use the Tien–Gordon theory to describe the pumping current. The photon-assisted tunneling current appears as a very sharp peak because of the discrete energy states. The peak appears when the two energy states are exactly separated by the photon energy. The current profile is given by Eq. (3) by substituting $J_n(\alpha)T_c$ for T_c [47]. In the weak coupling limit, the width of the peak is determined by the tunneling rate of outer barriers, in principle, or by the effective electron temperature in some cases. In a low-temperature cryostat with a low-noise measurement system, the width can be reduced to a few microelectron volts, corresponding to a frequency of about 1 GHz. The double dot is expected to act as a good microwave and millimeter wave spectrometer [50]. The current increases linearly with microwave power in the low-power limit, as expected from Eq. (12) [18, 51].

Furthermore, the double dot is a unique system in which population of each quantum dot can be controlled by external voltages [48]. In the normal population, the occupied state is energetically lower than the empty state, and excitation from the lower state to the higher state takes place by a photon absorption. However, in the *inverted population*, the occupied state has a higher energy than the empty state, and the microwave stimulates the transition from the higher state to the lower state by emitting a photon (*stimulated emission*). One can study microwave emission and absorption spectra just by changing the voltages.

3.6. Coherent Electron–Photon Interaction

Here we discuss the coherent electron–photon interaction in a strongly coupled double quantum dot, where the simple Tien–Gordon theory cannot be applied [52, 53]. In this case, the double dot can be considered a two-level system. In order to generalize the problem, we use the pseudospin model introduced in Section 2.5. We consider that a microwave is applied to the two-level system to modify the energy difference, ε, between the two charge states, (1, 0), denoted by $|0\rangle$, and (0, 1), denoted by $|1\rangle$; $\varepsilon(t) = \varepsilon_0 + \varepsilon_1 \cos \omega t$. Then the Hamiltonian of the system is time dependent and given by

$$H(t) = \frac{1}{2}(\varepsilon_0 + \varepsilon_1 \cos \omega t)\sigma_z + \hbar T_c \sigma_x \tag{13}$$

where T_c is the tunneling coupling. When the Hamiltonian is periodic, $H(t+T) = H(t)$, with the period $T = 2\pi/\omega$, the eigenstates have the form $\psi(t) = e^{-iEt/\hbar}u(t)$, where E is the quasi-energy and $u(t)$ is a periodic function (Floquet function) with the period T. E and $u(t)$ are the eigenenergy and eigenstate of the one-period evolution operator $U(T)$, which can be obtained by integrating the equation

$$i\hbar \frac{\partial}{\partial t} U(t) = H(t)U(t) \tag{14}$$

numerically [54].

Figure 6 shows the numerical calculation of the eigenenergy, E, for different normalized modulation amplitudes,

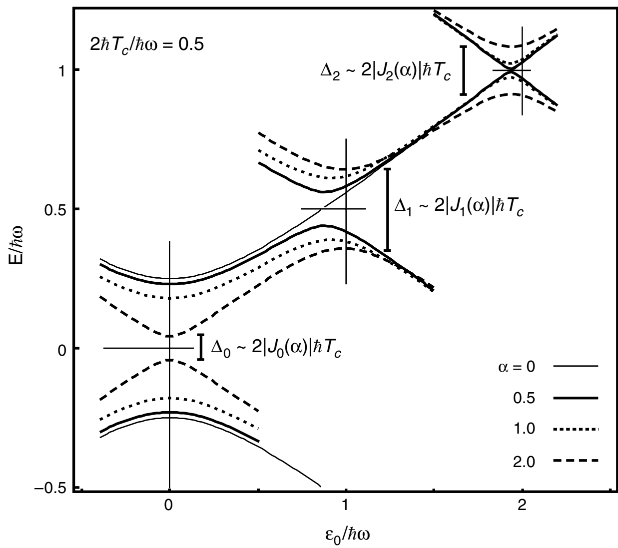


Figure 6. Quasi-energy, E, of the two-level system in a double quantum dot in the presence of microwave modulation. The numerical calculation is done for $2\hbar T_c/\hbar\omega = 0.5$ and $\alpha = \varepsilon_1/\hbar\omega = 0, 0.5, 1$, and 2.

$\alpha = \varepsilon_1/\hbar\omega$. The trace for $\alpha = 0$ is identical to that in Figure 4a, which schematically illustrates photon excitation from the bonding state (lower energy) to the antibonding state (higher energy). When the microwave amplitude is very weak, $\alpha \ll 1$, the application of the microwave does not perturb the energy states of the double dot. A current peak appears when the bonding and antibonding states are exactly separated by the photon energy, $\varepsilon = \sqrt{(\hbar\omega)^2 - 4(\hbar T_c)^2}$ see arrows in Figure 4a. So there is a minimum frequency, T_c/π, to observe the microwave excitation current. This is an accurate measurement to experimentally determine ε and T_c, since it can be deduced from the frequency. Actually, the anticrossing behavior has been demonstrated in a strongly coupled double dot and in a superconducting island [18, 49].

However, eigenstates of the system are influenced by microwave modulation, as shown in Figure 6. As the amplitude increases, the original bonding and antibonding states, $|\psi_+\rangle$ and $|\psi_-\rangle$, are coherently coupled with the aid of n photons, and a new energy gap, Δ_n, appears at the energy spacing close to the integer multiple of the photon energy, $\varepsilon_0 \sim n\hbar\omega$ [53]. The appearance of the gap indicates the coherency of the system, and the absorption and emission can take place coherently and oscillatory (Rabi oscillations, see next subsection). In the limit $T_c \ll \omega$, where analytical solutions can be obtained, the nth energy gap, Δ_n, is approximately given by

$$\Delta_n = 2|J_n(\alpha)|\hbar T_c \tag{15}$$

at $\varepsilon_0 = n\hbar\omega$ [53]. Figure 4b shows the Bessel function, $|J_n(\alpha)|$, to illustrate how the energy gap changes with α (compare with the square of the Bessel function for Tien–Gordon theory in Fig. 4a). As is also seen in Figure 6, the energy gaps change with α. It should be noted that the position of the gap also changes. The energy gap with one photon appears at $\varepsilon = \pm\sqrt{(\hbar\omega)^2 - 4(\hbar T_c)^2}$, in the case of weak microwave amplitude, $\alpha \ll 1$, while it shifts toward $\varepsilon = \hbar\omega$ for strong amplitude. This shift is also observed as a shift of the current peak in a double quantum dot device [55]. Strong nonlinear electron–photon coupling can be studied in quantum dot systems.

3.7. Coherent Oscillations in the Time Domain

In this subsection, we discuss some coherent oscillations in the time domain. First, we describe Rabi oscillations in a double quantum dot under microwave irradiation. Then, a slightly different type of charge oscillation induced by abruptly changing a gate voltage and spin rotation in the microwave magnetic field are discussed.

In order to simplify the problem, we consider that the two levels in the double quantum dot are separated by ε in the limit of $\hbar T_c \ll \varepsilon$ and the microwave is irradiated at the resonant condition, $\varepsilon = \hbar\omega$, with the same notations as in the previous subsection. Therefore, the localized state, in which an electron occupies the left or the right dot, gives a good picture of the system. Suppose an electron occupies the left dot before microwave irradiation at $t < 0$ and the microwave is turned on for the period $0 < t < t_p$. If microwave absorption and emission take place coherently during the microwave irradiation, the probability of finding the electron in the dot should oscillate in time,

$$P_{l/r}(t) = \frac{1}{2}[1 \pm \cos(\Delta_1 t/\hbar)] \tag{16}$$

where the frequency, $\Delta_1/\hbar$, is given by Eq. (15). This is analogous to Rabi oscillations in atomic systems [56]. One cannot directly observe the oscillation by detecting the location of the electron, since the measurement itself causes the decoherence of the system and makes the oscillation cease. Instead, one can measure the location of the electron after turning off the microwave at $t > t_p$. A simple measurement scheme is a current measurement. The gate voltages can be adjusted so that an electron in the right dot, for example, can escape to the lead to measure as a current. This tunneling rate must be much smaller than the oscillation frequency to reduce the decoherence. However, this measurement yields only one electron at most per pulse. Since a typical current meter requires about 10^5 electrons to detect a signal, a reasonable current can be obtained by repeating identical measurements (microwave pulses) many times. Then the current is proportional to the probability, $P_r(t_p)$. The oscillation pattern appears if the current is measured at various pulse lengths, t_p, or strengths of electron–photon coupling, $\Delta_1 = 2|J_1(\alpha)|\hbar T_c$.

Practically, the coherent oscillation does not continue forever due to dissipation and decoherence (Section 3.8) and ends up with a stationary ground state. In order to obtain the oscillations, higher oscillation frequency and less decoherence are required. However, the observation of Rabi oscillations in a single-electron system is difficult, because of technical difficulties in obtaining a short microwave pulse compared to the relatively short decoherence time of single charge states. Successful oscillations have been observed in an exciton (electron–hole pair) system, in which ultrafast optical excitation is used [57] .

A slightly modified technique is coherent charge oscillation induced by a high-speed voltage pulse, instead of a microwave pulse [58, 59]. A localized state is prepared before the pulse, and the two levels are aligned during the pulse ($\varepsilon = 0$) so that the electron can go back and forth between the two dots coherently.

Neglecting the source and drain electrodes, the Hamiltonian of the system, by using the same notation as in Section 2.5, changes from

$$H(t) = \frac{1}{2}\varepsilon_0\sigma_z + \hbar T_c\sigma_x \quad (t < 0) \tag{17}$$

before the pulse, where $|\varepsilon_0| \gg T_c$ is chosen to give a localized state in the steady state, to

$$H(t) = \frac{1}{2}\varepsilon_1\sigma_z + \hbar T_c\sigma_x \quad (0 < t < t_p) \tag{18}$$

during the pulse, where $\varepsilon_1 \sim 0$ is the situation we are interested in. The dynamics of the pseudo-spin can be understood well using the Bloch sphere introduced in Section 2.5. The system is initialized in the ground state of Eq. (17) (at the north pole in the Bloch sphere of Fig. 7e) at $t = 0$ and then precesses about the x axis during the pulse.

In a way similar to the microwave-induced Rabi oscillations, the electron in the right dot can be detected as a current, which is proportional to the probability of finding an electron in the right dot, $P_r(t_p)$. This is the projection of the quantum state in the Bloch sphere onto the z axis and is given by

$$P_{l/r}(t_p) = \frac{1}{2}[1 \pm \cos(2T_c t_p)] \tag{19}$$

at $\varepsilon_1 = 0$.

Figure 7 shows the coherent oscillation pattern in the $t_p - \varepsilon_1$ plane. As $|\varepsilon_1|$ grows larger, the oscillation amplitude and the oscillation period decrease. For $\varepsilon_1 \neq 0$, the Hamiltonian of Eq. (18) contains an additional fictitious magnetic field in z directions. The pseudo-spin rotates about the direction of the total fictitious magnetic field, which is illustrated in the trajectories in Figure 7d–f. It is important that any position (state) in the Bloch sphere can be prepared by adjusting ε_1 and t_p in a single pulse. By using a fast pulse generator with a rise time of 50–60 ps, successful charge oscillations have been observed in superconducting charge states in a single Cooper-pair box [58, 59] and more recently in single electron states in a double quantum dot [60].

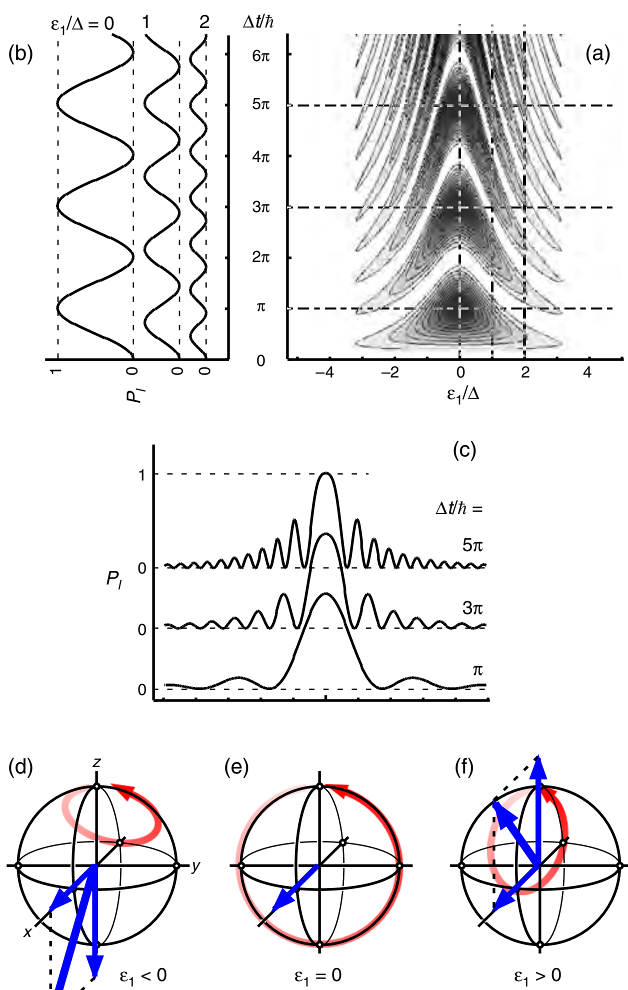


Figure 7. Time evolution of the two-level system in a double quantum dot. (a) Probability of finding the system in the excited state, P_e (white for 0 and black for 1) as a function of the normalized time, $\Delta t/\hbar$, and the normalized offset energy, ε_1/Δ. (b) Oscillations in the time domain appear at $\varepsilon_1/\Delta = 0$, 1, and 2 (they are offset for clarity). (c) The interference fringe in the offset energy for $\Delta t/\hbar = \pi$, 3π, and 5π. (d–f) Schematic trajectories of the oscillation in the Bloch sphere for (d) $\varepsilon_1 < 0$, (e) $\varepsilon_1 = 0$, and (f) $\varepsilon_1 > 0$.

Another important coherent oscillation is the electron spin resonance, in which the real electron spin precesses in the presence of a static magnetic field, B_0, in the z direction and an oscillating magnetic field, B_1, in the x direction. The Hamiltonian of the system is given by

$$H(t) = -\frac{1}{2}\mu_s(B_0\sigma_z + B_1 \cos \omega t \sigma_x) \tag{20}$$

which is similar to Eq. (13) by exchanging the x and z directions. When the frequency of the oscillating magnetic field is adjusted at the Zeeman splitting energy, the spin direction goes up and down with a precession about z direction. This is known as the electron spin resonance, and many kinds of coherent effects have been investigated in many systems. However, because of insufficient measurement sensitivity for electron spin, the electron spin resonance has to be obtained from an ensemble of many spins. Dynamical measurements on a single electron spin are desired (see Section 4.3).

3.8. Decoherence in Quantum Dots

Loss of coherence, which arises from the interaction between a quantum system and the environment, kills the coherent oscillations of the quantum system. There are two or three characteristic times to describe the coherency. Longitudinal relaxation time, which is often called T_1, is the characteristic time by which energy of the system is dissipated into the environment (dissipation). The transverse relaxation time, T_2, is the time by which the phase of the system becomes uncertain (decoherence). If the transverse relaxation time originates from the inhomogenity of an ensemble of the quantum system, it is often referred to as T_2^* (dephasing). Since dissipation also kills the phase information, generally $T_2^* \leq T_2 \leq 2T_1$.

Electrons in a quantum dot have orbital and spin degrees of freedom. These degrees of freedom are actually influenced by the environment surrounding the quantum dot, such as photons in the electrical leads and vacuum, phonons of the crystal lattice, other electrons in the electrodes and impurities, nuclear spin of the crystal, and noise and other fluctuations in the control signal. Moreover, spin and orbital degrees of freedom are coupled with relativistic correction (spin–orbit interactions). The quantum system has to be well isolated from the environment to maintain the coherency of the system.

The characteristic times (T_2^*, T_2, and T_1) of a system depend on the material, devices, and measurements and their parameters. In the next few subsections, we describe some dissipation and dephasing mechanisms that appear in a single-electron system.

3.9. Momentum Relaxation in a Single Quantum Dot

We start from the energy relaxation process from an excited state to the ground state of a single quantum dot without changing spin. The dominant interaction in this case is the coupling with acoustic phonons for a typical quantum dot whose energy spacing is less than ~10 meV. Because of the discrete energy states in a quantum dot, the quantum dot only couples with a phonon whose energy is identical to the energy difference of the ground and excited states. At very low temperature, where no phonons are excited, only spontaneous emission of a phonon is important. It should be noted that the wavelength of the phonon can be close to the size of the QDs. This is a notable difference when compared to electron–photon coupling, in which the wavelength of the photon is usually much longer than the size of the system. The strength of electron–phonon interaction becomes maximum when the half wavelength is about the size of the quantum dot [61, 62].

We have to consider different types electron–phonon couplings to fully understand the mechanisms. In addition to the normal bulk phonon modes, surface acoustic waves have to be taken into account, if the quantum dot is close to the surface within the phonon wavelength, and interface phonon modes if it is close to the interfaces. Moreover, for polar semiconductors, for example, GaAs, there are two types of couplings. One is the deformation type, in which the deformation of the lattice shifts the potential for electrons, and the other is the piezoelectric type, in which the deformation gives an electric field. Generally, the piezoelectric type coupling is more efficient for low energy phonons (less than ~1 meV for GaAs).

The electron–phonon interactions for optical characteristics are intensively studied because inefficient optical properties (the phonon bottleneck effect) are expected for a QD whose energy spacing is relatively large [63]. The phonon bottleneck effect is undesirable with respect to optical characteristics, but it is desired in order to reduce the decoherence of the quantum system. The energy relaxation time is expected to become significantly long, if the energy spacing is not right at the optical phonon energy and if the corresponding phonon wavelength is longer than the size of the QDs. There are still many questions about its influence on the efficiency of luminescence, but recent studies indicate that the phonon bottleneck effect does exist when other relaxation mechanisms, which may be related to holes or other electrons, are well suppressed [64–66]. Optical techniques are often restricted by their generated electron–hole pairs, which open other relaxation channels.

As for the transport measurement, in which no holes are generated, the suppression of phonon emission appears more clearly. Excitation spectra in single-electron tunneling characteristics indicate relatively long relaxation time; however, conventional transport characteristics do not give quantitative information about the relaxation time [67]. The recent discovery of time-dependent transport through a QD in the Coulomb blockade regime allows us to measure the energy relaxation time in a QD that contains just one electron (an artificial hydrogen atom). The energy relaxation process from the 2p orbital (the first excited state) to the 1s orbital (the ground state) is schematically shown in Figure 8a. The relaxation time, $T_1 = 3$–10 ns, slightly increases with increasing energy spacing (1.5–2.5 meV) by changing the magnetic field ($B = 0$–5 T) [68]. This behavior is understood as the spontaneous emission of an acoustic phonon and indicates the phonon bottleneck effect. It should be noted that the observed T_1 time is close to the minimum condition, where the half wavelength is about the size of the quantum dot (maximized electron–phonon interaction). The strength of electron–phonon coupling could be reduced by tailoring the structure of the quantum dot.

3.10. Spin Relaxation in a Quantum Dot

When the energy relaxation involves a spin–flip, simple photon or phonon emission cannot contribute to the relaxation. Spin–flip mechanisms, such as spin–orbit or hyperfine interactions, have to be considered together with the phonon emission that is required for energy conservation. Spin–orbit interaction, which mixes the spin and orbital degrees of freedom, is known to be a dominant spin–flip mechanism in 2D electron systems in semiconductors. Causes of spin–orbit interaction include the lack of crystal inversion symmetry, the electric field of the confinement potential, and impurities and interfaces of the structures. However, the spin–orbit interaction does not contribute to the spin relaxation in a QD to the first order if the energy spacing is much larger than the spin–orbit-induced spin splitting

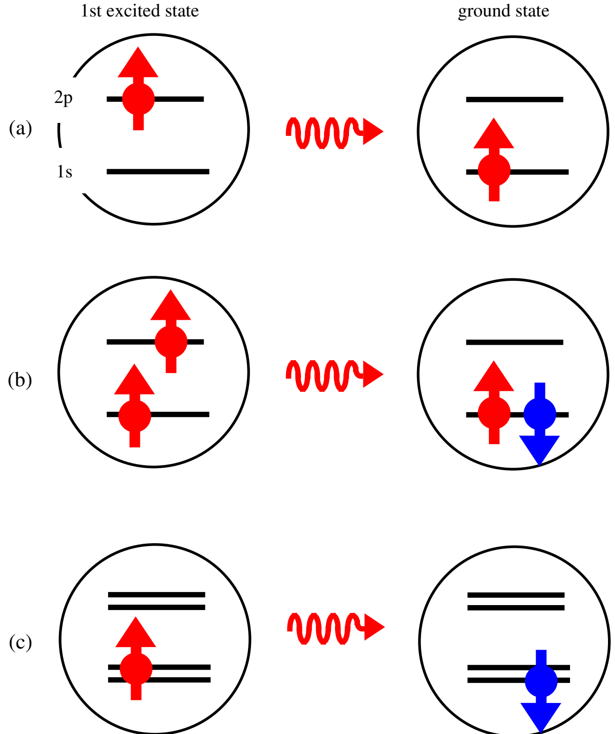


Figure 8. Energy relaxation processes in a quantum dot. The upper and lower horizontal lines in each circle represent the 2p and 1s orbital states, respectively, in a quantum dot. (a) Momentum relaxation from the 2p state to the 1s state. (b) Spin and momentum relaxation in a two-electron quantum dot. (c) Spin relaxation in a one-electron quantum dot. The double line indicates Zeeman splitting in a magnetic field.

energy (a few microelectron volts for GaAs). Higher order spin–orbit interactions give rise to a small effect on the spin–flip energy relaxation, whose relaxation time is expected to be longer than 1 ms for a typical GaAs QD [69–72]. Nevertheless, spin–orbit coupling is theoretically predicted to be a major source of the spin–flip energy relaxation process.

The spin–flip energy relaxation time has also been intensively studied in optical measurements. However, the observed relaxation time strongly depends on the excitation conditions. When the QD is excited resonantly, a very long relaxation time ($\gg$20 ns, which is even longer than the experimental limit of the optical technique) has been reported at low temperatures ($<$20 K) [73, 74]. However, it is not as long as the theoretical predictions. The optical measurement is restricted by its electron–hole recombination lifetime.

Single-electron transport measurements with time-dependent gate voltage allow us to measure extremely long relaxation times (the electrical pump and probe technique) [68, 75, 76]. The energy relaxation from the spin-triplet state to the spin-singlet state in a two-electron quantum dot (artificial helium atom) involves a spin–flip, as schematically shown in Figure 8b. This inelastic spin relaxation time is found to be about 200 μs for a specific device. From a detailed analysis, this relaxation was found to be dominated by the cotunneling process, which exchanges the spin and the energy with the electrode [68]. For instance, a spin-up electron leaves the dot, and simultaneously a spin-down electron enters it. During this process, the QD loses energy and the electrode gains the same energy. Since the strength of the cotunneling process is determined by the tunneling rate, the cotunneling process can be easily reduced by increasing the tunneling barrier. Then, the resulting relaxation time should become longer than 200 μs and may be dominated by spin–orbit interactions.

The inelastic spin relaxation time ($>$200 μs) is more than 5 orders of magnitude longer than when no spin–flip is involved ($\sim$10 ns) [68]. The ratio of the relaxation time can be larger than 3×10^4, which is close to the theoretical expectation of 5×10^5 for this structure [70]. It is interesting to compare this ratio to that of real atoms. Quantum dots, or artificial atoms, couple to lattice vibration modes (phonons), while real atoms couple to electromagnetic fields (photons). The strength of the optical transition in real atoms strongly depends on the characteristics of the state, known as selection rules [77]. The typical transition lifetime, or relaxation time, of allowed transitions for electric dipole coupling is on the order of nanoseconds, e.g., 1.6 ns for the Lyman α transition line from the 2p state to the 1s state in a hydrogen atom. However, some transitions have an extremely long relaxation time, for example, 7860 s for the relaxation from the two-electron spin-triplet excited state to the spin-singlet ground state in a helium atom. In this case, the transition is forbidden by spin conservation and parity [77]. The huge ratio of the relaxation times, more than 10^{12} for hydrogen and helium atoms, indicates the high quality of the quantum numbers, spin, and angular momentum. The ratio decreases with increasing atomic number, because spin–orbit interaction increases and the Russel–Saunders approximation becomes worse. The ratio in a GaAs artificial atom is $>3 \times 10^4$, which is comparable to the ratio of 10^5 for potassium and calcium atoms, which are located in the same row as gallium and arsenide in the periodic table. This crude comparison implies that the spin–orbit interaction in the nanostructure is almost comparable to an atomic property.

Another important spin relaxation process is the transition between Zeeman sublevels in a magnetic field (see Fig. 8c). This relaxation time has been measured with the electrical method, and is also longer than the experimental limit of about 50 μs [78]. However, this spin relaxation time is expected to be longer than 1 ms, based on consideration of only the spin–orbit interaction [68].

The long spin relaxation time in quantum dots gives rise to spin-dependent tunneling. Since an electron has spin 1/2, a single-electron tunneling changes the total spin of the dot by 1/2. Other tunneling transitions that change the total spin by more than 1/2 should be blocked (spin blockade) [79, 80]. Moreover, the long spin relaxation time induces nonequilibrium transport that cannot be explained by the orthodox CB theory [9, 81].

3.11. Energy Relaxation in a Double Quantum Dot

In contrast to single dots, the energy states in a double quantum dot are spatially separated. Here we discuss the energy relaxation between the two charge states in the double dot defined in Section 2.5. Suppose two quantum dots are weakly coupled by a tunneling barrier, and consider an energy relaxation between two localized states, that is, $\varepsilon \gg T_c$. By changing ε, the overlap of the wavefunctions can be modified, as schematically shown in Figure 4(b–d). The transition rate, W, is given by $W = (T_c/\varepsilon)^2 J(\varepsilon)$, where $J(\varepsilon)$ is a spectral function that describes the interaction with its environment. As is the case for momentum relaxation in a single quantum dot, the energy relaxation without spin–flip is dominated by electron–phonon interactions. In the double quantum dot case, the spectral function can be studied by changing ε.

In single-electron transport measurements through a weakly coupled double quantum dot in a series configuration, the current through the device, I, can be directly related to the transition rate, that is, $I = eW$ for $\varepsilon > T_c$, if the outer tunneling barriers are made more transparent than the inelastic transition of interest [82, 83]. Measurements on GaAs double quantum dots indicate a spectral function close to the ohmic interaction in this system. Because of the relatively small energy ($\varepsilon = 4$–100 μeV), piezoelectric interactions with acoustic phonons are the dominant mechanisms.

In some cases, $J(\varepsilon)$ in a double quantum dot shows some structures, which come from the relation between the phonon wavelength and the spacing between the dots [82, 83]. The phonon emission between the two states is enhanced when the spacing is half of the phonon wavelength, but suppressed when the spacing equals the wavelength. The electron–phonon interaction might be strong to renormalize the single-electron tunneling [84]. Even though there is no confinement of phonons, such oscillatory behavior is observable in the current spectrum of a double quantum dot.

3.12. Decoherence of a Double Quantum Dot

In contrast to the energy relaxation processes, decoherence is a loss of phase information over the ensemble of measurements. In the discussions of coherent oscillations in the time domain (see Section 3.7), we completely neglect the decoherence and dissipation. However, in reality, oscillation does not continue forever, but is dampened over a duration characterized by the decoherence time, T_2. It is convenient to assume an exponential decay of the oscillation amplitude, but the amplitude can decay in a Gaussian shape in some cases [85].

The decoherence of the two-level system in a double quantum dot may be dominated by fluctuations of ε and T_c. The fluctuation changes the frequency of the coherent oscillations and smears out the oscillation in the ensemble measurement. According to a detailed study of coherent oscillations in superconducting charge states, the $1/f$ noise in the background charge fluctuations is the decoherence mechanism [85]. Although the power spectrum decreases with a $1/f$ dependence at higher frequency, the integrated contributions come from a wide frequency range. A single-electron state has the advantage of high controllability, but this means that it can easily couple to the environment. The mechanism of the charge fluctuations is not well understood, and the reduction of the noise may not be easy.

One way to avoid the influence of fluctuations is to design the quantum state to be less sensitive. As seen in Figure 4a, the energy spacing between two eigenstates is insensitive to the fluctuation of ε at $\varepsilon = 0$ to the first order, while it is sensitive at $\varepsilon > T_c$. Therefore, longer decoherence time is expected at $\varepsilon = 0$. In a superconducting island with a SQUID interferometer to control the Josephson coupling (equivalent to T_c), the quantum system can be set at the saddle point where the energy spacing is insensitive to the fluctuation of ε and T_c. A very long decoherence time of a few microseconds compared to the precession period of $\sim$50 ps has been realized [86].

3.13. Radiofrequency Single-Electron Transistor

The single-electron transistor is known to work as a highly sensitive electrometer. In principle, the operating frequency can go beyond 1 GHz if it is determined by the intrinsic RC time constant of the tunneling resistance and capacitance. However, the practical operating frequency is limited to a low-frequency range (a few kiloHertz at most) in conventional dc current measurements. The capacitance of the electrodes, including the measurement instruments, is so large that the frequency range is restricted by the RC time constant of the electrode capacitance and the tunneling resistance.

The radiofrequency (rf) single-electron transistor (RFSET), which works as a wide-band and highly sensitive electrometer, is a SET combined with an impedance transformer (LC resonator) [87]. The capacitance of the problem can be canceled by an external inductor located close to the SET device if it is operated at the resonant frequency, f_{res}. The maximum frequency of the RFSET is approximately given by f_{res}/Q_{LC}, where Q_{LC} is the quality factor of the resonator. The external LC resonator placed close to an SET device has a typical bandwidth of about 100 MHz, with $f_{res} \sim 1$ GHz and $Q_{LC} \sim 10$. Better performance (higher f_{res} and larger Q_{LC}) may be obtained using an on-chip resonator. The conductance of the SET is measured by the reflection or the transmission of the rf carrier signal at f_{res} [88, 90]. The transmission amplitude, or the small change in the reflected signal, is, in principle, proportional to the admittance (the inverse of the impedance) of the SET [90].

The sensitivity of a charge respective to the island can be about $10^{-5}\text{e}/\sqrt{Hz}$, which is usually restricted by the noise of the high-frequency amplifier. This means that the intrinsic noise of the RFSET (shot noise) is very low at high frequency (>10 kHz), while it suffers from $1/f$ noise at low frequency. Theoretically, the noise of the RFSET is expected to be very close to that of the conventional SET [89]. The charge sensitivity can be as low as $2 \times 10^{-6}\text{e}/\sqrt{Hz}$ for an optimized device using a practical superconducting island, and better sensitivity is expected by using a smaller island [87].

The RFSET technique is very attractive and can be used as a highly sensitive fast-response electrometer for many applications [91]. If the RFSET is attached to another quantum dot located in another conductive channel, each single-electron tunneling process would be detected with a high sensitivity [92]. This would be an extremely sensitive current meter, in which current flow could be detected by counting tunneling electrons. If a RFSET is attached to a two-level system in a double quantum dot (see Section 2.5 and 3.7), the charge state can be detected in a short time [93]. This is desirable for further investigation of quantum dynamics and correlations in single-electron systems.

4. TOWARD QUANTUM INFORMATION PROCESSING

Quantum information processing is digital data processing with the aid of coherent time evolution of quantum states. Analogous to a bit that is the unit of digital information, the unit of quantum information is the quantum bit (qubit), which is basically realized in any single two-level system. Quantum information processing would provide various advantages that have never been obtained in the conventional classical approach [94, 95]. For instance, quantum cryptography would provide secure telecommunications because any unknown single quantum state cannot be duplicated and because any unknown single quantum state cannot be determined completely. A quantum nondemolition measurement scheme improves the measurement accuracy by avoiding back action. Quantum computation is programmable interferometry in which only one or a few desired answers can be efficiently obtained from an extremely large number of candidates. Recently, quantum information processing has become attractive for realizing very specific tasks.

4.1. Concept of Quantum Computation

First, we stress that quantum information processing is completely different from so-called "quantum devices." Quantum devices are designed to transform a classical input, like a voltage, into another classical output, like a current, with excellent transfer characteristics with the aid of quantum mechanics. The resonant tunneling diode, which is a typical quantum device that works even at room temperature, shows negative differential resistance in the current (output)–voltage (input) characteristics. Although quantum mechanics is required to design the diode, one can use the diode without considering the quantum mechanics, if the characteristics are known. In quantum information processing, however, the quantum state is the carrier of information. The quantum state is transformed from an input state to an output state by applying an external field, for example, an electromagnetic field, in a certain period. The transformation of the quantum state is called a quantum logic gate and should be a unitary transformation to keep coherency. The quantum state changes by applying a series of quantum logic gates necessary for quantum computation from the beginning, at which some classical values are input to the quantum state, until the end of the computation, at which the output states are finally measured as classical values. One may not measure the intermediate state, which would result in the collapse of the quantum state. Quantum computing requires a high degree of quantum coherence for a long enough time to complete the computation.

In principle, any calculation that can be performed in a conventional classical computer can also be performed in a quantum computer. But this is not a good idea because quantum computation requires extremely high accuracy and coherency in the quantum logic gates and because computation errors are unavoidable. However, one can design the algorithm of quantum computation in such a way that a series of data processings are performed at once in a parallel fashion (quantum parallelism). Then, quantum computation is expected to provide extremely efficient calculations for specific problems that cannot be solved efficiently with conventional classical computers. For instance, factorization of a large number is a formidable task for conventional computers. There are no efficient algorithms for this in classical computers, and one has to check sequentially whether the number to factorize is divisible by a number from 2 to the square root of the number to factorize (actually, there is a better way using a probabilistic algorithm). Nevertheless, factorization of a 1000-digit number would require 10^{25} years (longer than the history of the universe) by using 1000 workstations in parallel [96]. In Shor's factoring algorithm for quantum computing, the factorization problem is attributed to how efficiently a Fourier transformation is calculated. The quantum Fourier transform can be constructed from a few kinds of quantum logic gates, and can be calculated very efficiently. Factoring a 1000-digit number would require only a few 10^6 steps in the quantum computation [96]. The factorization (quantum Fourier transformation) is just an example of quantum computation. There are many algorithms, such as database search and quantum simulations, that would solve some problems very efficiently.

Although the potential of quantum computing is fascinating, only a few small-scale quantum computers have been realized. The solution nuclear magnetic resonance (NMR) quantum computer is the most advanced quantum computer. The nuclear spin of an atom is used as a qubit, and a single molecule that contains distinguishable atoms (nuclear spins) works as a quantum computer. The NMR quantum computer works by applying a sequence of radiofrequencies (quantum logic gates) to many molecules (quantum computers) in a solution. Recently, factorization of 15, whose prime numbers are 3 and 5, was demonstrated using a solution NMR QC with 7 qubits [97]. However, the solution NMR QC has problems integrating large numbers of qubits and thus may not be a practical quantum computer. It is generally accepted that solid-state quantum computers are good candidates for scalability. There are many proposals based on the charge state or flux state in a superconducting island, the charge state, electron spin state, or exciton state (electron–hole pair) in a semiconductor quantum dot, and the nuclear spin state of impurities or crystals. One-qubit operations have been demonstrated in some solid-state systems, and two-qubit operations has recently been studied in superconducting system [98].

To construct a quantum computer, single quantum states would have to be prepared physically, manipulated coherently, preserved for a long enough time, measured individually, and integrated in large quantity [99]. Each of these points requires further development. Decoherence is one of the major problems in solid-state systems. Fundamental research has been conducted to reduce the decoherence problems and to devise the best quantum logic gates. In the following two subsections, we briefly summarize the strategies for realizing quantum computers using single-electron dynamics.

4.2. Single-Electron Charge Qubit

As we discussed in Section 2.5, a single electron in a double quantum dot can be used as a two-level system, serving as a qubit (charge qubit) [18, 100]. A similar charge qubit has been realized in a superconducting island, in which the two-level system is represented by an extra Cooper pair occupying or not occupying the island (a superconducting charge qubit). The coherent oscillations induced by microwave irradiation or by a high-speed voltage pulse can be used as a rotation gate for one-qubit operation (Section 3.7). The qubit state can be controlled to any state [θ and ϕ in the Bloch sphere, see Figure (7d–f)] by tailoring the pulse shape. The NOT gate, which reverses the classical information, can be obtained by applying a π pulse (a half cycle of the oscillation at $\varepsilon = 0$). The Hadamard gate, which creates a superposition state from an eigenstate, is achieved by a $\pi/2$ pulse (a quarter cycle of the oscillation).

When two sets of double quantum dots (two qubits) are fabricated to couple electrostatically, any superposition of four bases $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, where the first and second numbers indicate the location of electron in respective double dot, can be prepared. The dipole coupling between the two qubits affects the total energy of the states. The controlled-NOT gate, which is a typical two-qubit operation, can be performed by applying a voltage pulse to degenerate two states, say $|00\rangle$ and $|01\rangle$, for a certain period that exchanges the two states. This means that the state of the

second qubit (target qubit) is reversed (NOT operation) only when the first qubit (control qubit) is 0. If this is performed coherently, the controlled-NOT gate should work for any superposition state as well. For instance, starting from the initial state $|00\rangle$, the Hadamard gate on the fist qubit followed by the controlled-NOT gate brings an entangled state

$$|00\rangle \rightarrow_H |00\rangle + |10\rangle \rightarrow_{CNOT} |01\rangle + |10\rangle \tag{21}$$

in which the first and second qubits are correlated [95].

The qubit state can be measured by an electrical current, as explained in Section 3.7. In this case, ensemble averaging over many measurements is required due to small current sensitivity. In some cases, for example, when the correlation between the two qubits is essential, it is desirable to measure a single qubit state without any ensemble averaging. The RFSET technique discussed in Section 3.13 would provide a single shot measurement without averaging for the charge qubit. If the electrostatic coupling between the qubit and the RFSET can be made sufficiently large, the qubit state can be measured in a relatively short time (hopefully ~10 ns). Of course, the RFSET can be turned off by switching off the carrier rf signal during the quantum computation to minimize the decoherence from the measurement.

4.3. Single-Electron Spin Qubit

The spin degree of freedom is an alternative way to construct a qubit [101]. If the charge qubit is an artificial qubit, electron spin is a natural qubit. The coherency and manipulation of electron spins have been studied in many systems. The spin coherence time of conductive electrons in bulk GaAs crystal can be longer than 100 ns [102], and electron spin bound to a donor in silicon shows $T_2 \sim 300$ μs [103]. Electron spin based quantum computation is motivated by the long decoherence time. However, in contrast to the countless studies of the ensemble of spins, little work has been done on the manipulation of single-electron spin. In order to address each electron spin (qubit) in a quantum computer, single-spin manipulation and measurement techniques are essential.

A simple scheme for one-qubit operation is the electron spin resonance discussed in Section 3.7. The effective g factor of each electron spin can be made different for different quantum dots by using g-factor engineering, so that each qubit is addressed by a corresponding microwave frequency [101]. Or a moderate magnetic field gradient in the device may be useful in changing the Zeeman splitting energy. However, a typical one-qubit operation using an electron spin resonance will require a relatively long time, ~100 ns, because of the weak magnetic dipole transition. Alternative ways using the optical Stark effect in a specific band structure [104] or exchange coupling among three electron spins constituting one qubit [105] are promising for much faster operations Two-qubit operation can also be performed by the exchange coupling between two quantum dots [106].

Single shot spin measurement is a challenging technique for quantum information technology. One proposal is based on the spin-dependent tunneling between two quantum dots combined with an RFSET [107]. When each of the two quantum dots possesses one electron spin before the measurement, tunneling from one dot to the other is allowed if the two electron spins can make a spin pair (spin-singlet state) [108]. This spin-dependent tunneling could be measured with an RFSET in a short time.

GLOSSARY

Adiabatic approximation When the Hamiltonian of a quantum system changes slowly, the wave-function can be approximated by an eigenstate of the instantaneous Hamiltonian (adiabatic approximation). However, when the Hamiltonian changes very fast (nonadiabatically), the wave-function becomes a nonstationary superposition state.

Bloch sphere Any linear superposition of two orthonormal bases can be expressed in a form, Eq. (4), which indicates a point on the unit sphere (Bloch sphere). This representation is very useful for visualizing a quantum state.

Cotunneling Cotunneling is two or more tunneling processes that occur successively in a short time. The intermediate state may have a high energy if the energy cost is within the energy uncertainty given by the interval of the corresponding tunneling processes.

Coulomb interaction Coulomb interaction between two or more electrons can be described by a direct integral, which comes from the direct overlap of the two wavefunctions, and an exchange integral, which depends on the spin state.

Poisson statistics Poisson statistics describe frequency distribution when the probability of an event is very small. The probability for n events happening is given by $p(n) = a^{-n}e^{-a}/n!$, where a is the average number of events. When the distributions are narrower than Poisson statistics, they are described by sub-Poisson statistics.

Rabi oscillation When a coherent electromagnetic field is resonantly applied to a two-level system, emission and absorption take place coherently and with oscillation (Rabi oscillation).

Spontaneous/stimulated emission When a transition from a higher energy state to a lower energy state occurs by the emission of a boson (photon or phonon), it consists of spontaneous emission, which always occurs due to vacuum fluctuation of the bosonic system, and stimulated emission, which is proportional to the number of existing bonons.

1/f noise $1/f$ noise has a power spectrum close to $1/f$ frequency dependence. In electrical noise, $1/f$ noise comes from an ensemble of many electron traps, each of which emits or captures an electron independently (Poisson statistics).

ACKNOWLEDGMENTS

Discussions with R. Aguado, D. G. Austing, H. D. Cheong, T. Hayashi, Y. Hirayama, T. Honda, K. Ishibashi, T. Itakura, A. V. Khaetskii, L. P. Kouwenhoven, G. Lang, Y. Nakamura, T. H. Oosterkamp, S. Tarucha, and Y. Tokura, and W. G. van der Wiel are gratefully acknowledged.

REFERENCES

1. H. Grabert and M. H. Devoret (Eds.), "Single Charge Tunneling, Coulomb Blockade Phenomena in Nanostuctures," NATO ASI series B 294. Plenum Press, New York, 1991.
2. L. P. Kouwenhoven, C. M. Marcus, P. L. McEuen, S. Tarucha, R. M. Westervelt, and N. S. Wingreen, in "Mesoscopic Electron Transport" (L. L. Sohn, L. P. Kouwenhoven, and G. Schön, Eds.), NATO ASI series E 345, pp. 105–214. Kluwer Academic, Dordrecht, 1997.
3. U. Meirav, M. A. Kastner, and S. J. Wind, *Phys. Rev. Lett.* 65, 771 (1990).
4. T. Fujisawa and S. Tarucha, *Appl. Phys. Lett.* 68, 526 (1996).
5. D. G. Austing, T. Honda, and S. Tarucha, *Semicond. Sci. Tecnol.* 11, 212 (1996).
6. T. A. Fulton and G. J. Dolan, *Phys. Rev. Lett.* 59, 109 (1987).
7. Y. Nakamura, C. D. Chen, and J. S. Tsai, *Jpn. J. Appl. Phys.* 35, L1465 (1996).
8. D. G. Austing, S. Tarucha, P. C. Main, M. Henini, S. T. Stoddart, and L. Eaves, *Appl. Phys. Lett.* 75, 671 (1999).
9. T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, *Phys. Rev. Lett.* 88, 236802 (2002).
10. L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, *Rep. Prog. Phys.* 64, 701 (2001).
11. S. Tarucha, D. G. Austing, T. Honda, R. J. van der Hage, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 77, 3613 (1996).
12. L. P. Kouwenhoven, T. H. Oosterkamp, M. W. S. Danoesastro, M. Eto, D. G. Austing, T. Honda, and S. Tarucha, *Science* 278, 1788 (1997).
13. F. R. Waugh, M. J. Berry, D. J. Mar, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Phys. Rev. Lett.* 75, 705 (1995).
14. C. Livermore, C. H. Crouch, R. M. Westervelt, K. L. Campman, and A. C. Gossard, *Science* 274, 1332 (1996).
15. N. C. van der Vaart, S. F. Godijn, Yu. V. Nazarov, C. J. P. M. Harmans, and J. E. Mooij, *Phys. Rev. Lett.* 74, 4702 (1995).
16. W. G. van der Wiel, S. De Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven, *Rev. Mod. Phys* (in press).
17. L. P. Kouwenhoven, *Science* 268, 1440 (1995).
18. T. H. Oosterkamp, T. Fujisawa, W. G. van der Wiel, K. Ishibashi, R. V. Hijman, S. Tarucha, and L. P. Kouwenhoven, *Nature* 395, 873 (1998).
19. R. H. Blick, D. Pfannkuche, R. J. Haug, K. v. Klitzing, and K. Ebel, *Phys. Rev. Lett.* 80, 4032 (1998).
20. D. Esteve, in "Single Charge Tunneling, Coulomb Blockade Phenomena in Nanostuctures" (H. Grabert and M. H. Devoret, Eds.), NATO ASI series B 294, pp. 109–137. Plenum Press, New York, 1991.
21. K. K. Likharev and A. B. Zorin, *J. Low Temp. Phys.* 59, 347 (1985).
22. L. J. Geerling, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 64, 2691 (1990).
23. H. Pothier, P. Lafarge, C. Urbina, D. Esteve, and M. H. Devoret, *Europhys. Lett.* 17, 249 (1992).
24. D. V. Averin and Yu. V. Nazarov, in "Single Charge Tunneling, Coulomb Blockade Phenomena in Nanostuctures" (H. Grabert and M. H. Devoret, Eds.), NATO ASI series B 294, pp. 217–248. Plenum Press, New York, 1991.
25. M. W. Keller, J. M. Martinis, and R. L. Kautz, *Phys. Rev. Lett.* 80, 4530 (1998).
26. M. W. Keller, J. M. Martinis, A. H. Steinbach, and N. M. Zimmerman, *IEEE Trans. Instrum. Measurement*, 46, 307 (1997).
27. D. V. Averin, and K. K. Likharev, *J. Low. Temp. Phys.* 62, 345 (1986).
28. A. B. Zorin, S. V. Lotkhov, H. Zangerle, and J. Niemeyer, *J. Appl. Phys.* 88, 2665 (2000).
29. P. Delsing, K. K. Likharev, L. S. Kuzmin, and T. Claeson, *Phys. Rev. Lett.* 63, 1861 (1989).
30. S. Hershfield, J. H. Davis, P. Hyldgaard, C. J. Stanton, and J. W. Wilkins, *Phys. Rev. B* 47, 1967 (1993).
31. H. Birk, M. J. M. de Jong, and C. Schönenberger, *Phys. Rev. Lett.* 75, 1610 (1995).
32. M. S. Choi, F. Plastina, and R. Fazio, *Phys. Rev. Lett.* 87, 116601 (2001).
33. Q. Niu, *Phys. Rev. Lett.* 64, 1812 (1990).
34. V. I. Talyanskii, J. M. Shilton, M. Pepper, C. G. Smith, C. J. B. Ford, E. H. Linfield, D. A. Ritchie, and G. A. C. Jones, *Phys. Rev. B* 56, 15180 (1997).
35. J. Cunningham, V. I. Talyanskii, J. M. Shilton, M. Pepper, M. Y. Simmons, and D. A. Ritchie, *Phys. Rev. B* 60, 4850 (2000).
36. J. Cunningham, V. I. Talyanskii, J. M. Shilton, and M. Pepper, *Phys. Rev. B* 62, 1564 (2000).
37. T. J. M. M. Janssen and A. Hartland, *Physica B* 284–288, 1790 (2000).
38. P. A. Maksym, *Phys. Rev. B* 61, 4727 (2000).
39. P. K. Tien and J. P. Gordon, *Phys. Rev.* 129, 647 (1963).
40. J. R. Tucker and M. J. Feldman, *Rev. Mod. Phys.* 57, 1055 (1985).
41. J. M. Hergenrother, M. T. Tuominen, J. G. Lu, D. C. Ralph, and M. Tinkham, *Physica B* 203, 327 (1994).
42. B. J. Keay, S. Zeuner, S. J. Allen Jr., K. D. Maranowski, A. C. Gossard, U. Bhattacharya, and M. J. W. Rodwell, *Phys. Rev. Lett.* 75, 4102 (1995).
43. S. Verghese, R. A. Wyss, Th. Schäpers, Q. Hu, A. Förster, and M. J. Rooks, *Phys. Rev. B* 52, 14834 (1995).
44. L. P. Kouwehhoven, S. Jauhar, J. Orenstein, P. L. McEuen, Y. Nagamune, J. Motohisa, and H. Sakaki, *Phys. Rev. Lett.* 73, 3443 (1994).
45. R. H. Blick, R. J. Haug, D. W. van der Weide, K. von Klitzing, and K. Eberl, *Appl. Phys. Lett.* 67, 3924 (1995).
46. T. H. Oosterkamp, L. P. Kouwehhoven, A. E. A. Koolen, N. C. van der Vaart, and C. J. P. M. Harmans, *Phys. Rev. Lett.* 78, 1536 (1997).
47. T. H. Stoof and Yu. V. Nazarov, *Phys. Rev. B* 53, 1050 (1996).
48. T. Fujisawa and S. Tarucha, *Superlattices Microstruct.* 21, 247 (1997).
49. Y. Nakamura, C. D. Chen, and J. S. Tsai, *Phys. Rev. Lett.* 79, 2328 (1997).
50. R. Aguado and L. P. Kouwenhoven, *Phys. Rev. Lett.* 84, 1986 (2000).
51. T. Fujisawa and S. Tarucha, *Japan. J. Appl. Phys.* 36, 4000 (1997).
52. A. N. Korotkov, D. V. Averin, and K. K. Likharev, *Phys. Rev. B* 49, 7548 (1994).
53. C. A. Stafford and N. Wingreen, *Phys. Rev. Lett.* 76, 1916 (1996).
54. M. Holthaus and D. Hone, *Phys. Rev. B* 47, 6499 (1993).
55. W. G. van der Wiel, T. Fujisawa, T. H. Oosterkamp, and L. P. Kouwenhoven, *Physica B* 272, 31 (1999).
56. M. Brune, F. Schmidt-Kaler, A. Maali, J. Dreyer, E. Hagley, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 76, 1800 (1996).
57. H. Kamada, H. Gotoh, J. Temmyo, T. Takagahara, and H. Ando, *Phys. Rev. Lett.* 87, 246401 (2001).
58. Y. Nakamura, Yu. A. Pashkin, and J. S. Tsai, *Nature* 398, 786 (1999).
59. Y. Nakamura, and J. S. Tsai, *J. Low Temp. Phys.* 118, 765 (2000).
60. T. Hayashi, T. Fujisawa, H. D. Cheong, Y. H. Jeong, and Y. Hirayama, *Phys. Rev. Lett.* 91, 226804 (2003).
61. U. Bockelmann and G. Bastard, *Phys. Rev. B* 42, 8947 (1990).
62. U. Bockelmann, *Phys. Rev. B* 50, 17271 (1994).
63. H. Benisty, *Phys. Rev. B* 51, 13281 (1995).
64. R. Heitz, H. Born, F. Guffarth, O. Stier, A. Schliwa, A. Hoffmann, and D. Bimberg, *Phys. Rev. B* 64, 241305 (2001).
65. J. Urayama, T. B. Norris, J. Singh, and P. Bhattacharya, *Phys. Rev. Lett.* 86, 4930 (2001).

66. S. Malik, E. V. Le Ru, D. Childs, and R. Murray, *Phys. Rev. B.* 63, 155313 (2001).
67. J. Weis, R. J. Haug, K. v. Klitzing, and K. Ploog, *Phys. Rev. Lett.* 71, 4019 (1993).
68. T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, *Nature* 419, 278 (2002).
69. D. M. Frenkel, *Phys. Rev. B* 43, 14228 (1991).
70. A. V. Kaetskii and Yu. V. Nazarov, *Phys. Rev. B* 61, 12639 (2000).
71. A. V. Kaetskii and Yu. V. Nazarov, *Physica E* 6, 470 (2000).
72. S. I. Erlingsson, Yu. V. Nazarov, and V. I. Fal'ko, cond-mat/0104148 (2001).
73. Y. Toda, S. Shinomori, K. Suzuki, and Y. Arakawa, *Phys. Rev. B* 58, R10147 (1998).
74. M. Paillard, X. Marie, P. Renucci, T. Amad, A. Jbeli, and J. M. Gerard, *Phys. Rev. Lett.* 86, 1634 (2001).
75. T. Fujisawa, Y. Tokura, and Y. Hirayama, *Phys. Rev. B* 63, 081304(R) (2001); *Physica B* 298, 573 (2001).
76. T. Fujisawa, Y. Tokura, D. G. Austing, Y. Hirayama, and S. Tarucha, *Physica B* 314, 224 (2002).
77. H. A. Bethe, and E. E. Salpeter, "Quantum Mechanics of One- and Two-Electron Atoms." Springer-Verlag, Berlin, 1957.
78. R. Hanson, B. Witkamp, L. M. K. Vandersypen, L. H. Willems van Beveren, J. M. Elzerman, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 91, 196802 (2003).
79. D. Weinmann, W. Häusler, and B. Kramer, *Phys. Rev. Lett* 74, 984 (1995).
80. M. Ciorga, A. S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, S. Jullian, Y. Feng, and Z. Wasilewski, *Phys. Rev. B* 61, R16315 (2000).
81. O. Agam, N. S. Wingreen, B. L. Altshuler, D. C. Ralph, and M. Tinkham, *Phys. Rev. Lett.* 78, 1956 (1997).
82. T. Fujisawa, T. H. Oosterkamp, W. G. van der Wiel, B. W. Broer, R. Aguado, S. Tarucha, and L. P. Kouwenhoven, *Science* 282, 932 (1998).
83. T. Fujisawa, W. G. van der Wiel, and L. P. Kouwenhoven, *Physica E* 7, 413 (2000).
84. T. Brandes and B. Kramer, *Phys. Rev. Lett.* 83, 3021 (1999).
85. Y. Nakamura, Yu. A. Pashkin, T. Yamamoto, and J. S. Tsai, *Phys. Rev. Lett.* 88, 047901 (2002).
86. D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Esteve, and M. H. Devoret, *Science*, 296, 886 (2002).
87. R. J. Schoelkopf, P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, *Science*, 280, 1238 (1998).
88. T. Fujisawa and Y. Hirayama, *Appl. Phys. Lett.* 77, 543 (2000).
89. A. N. Korotkov and M. A. Paalanen, *Appl. Phys. Lett.* 74, 4052 (1999).
90. H. D. Cheong, T. Fujisawa, T. Hayashi, Y. Hirayama, and Y. H. Jeong, *Appl. Phys. Lett.* 81, 3257 (2002).
91. M. H. Devoret and R. J. Schoelkopf, *Nature* 406, 1039 (2000).
92. T. Fujisawa, T. Hayashi, Y. Hirayama, H. D. Cheong, and Y. H. Jeong, *Appl. Phys. Lett.*, to appear.
93. A. Aassime, G. Johansson, G. Wendin, R. J. Schoelkopf, and P. Delsing, *Phys. Rev. Lett.* 86, 3376 (2001).
94. D. Bouwmeester, A. Ekert, and A. Zeilinger, Eds., "The Physics of Quantum Information." Springer-Verlag, Berlin, 2000.
95. M. A. Nielsen and I. L. Chuang, "Quantum Computation and Quantum Information." Cambridge, 2000.
96. S. L. Braunstein, "Quantum Computation: A Tutorial," http://chemphys.weizmann.ac.il/schmuel/comp/comp.html.
97. L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, *Nature* 414, 883 (2001).
98. T. Yamamoto, Yu. A. Pashkin, O. Astafiev, Y. Nakamura, and J. S. Tsai, *Nature* 425, 941 (2003).
99. D. P. DiVincenzo, in "Mesocopic Electron Transport" (L. L. Sohn, L. P. Kouwenhoven, and G. Schön, Eds.), NATO ASI series E 345, pp. 657–677. Kluwer Academic, Dordrecht, 1997.
100. A. Barenco, D. Deutsch, A. Ekert, and R. Jozsa, *Phys. Rev. Lett.* 74, 4083 (1995).
101. D. Loss and D. P. DiVincenzo, *Phys. Rev. A* 57, 120 (1998).
102. J. M. Kikkawa and D. D. Awschalom, *Phys. Rev. Lett.* 80, 4313 (1998).
103. M. Chiba and A. Hirai, *J. Phys. Soc. Jpn.* 33, 730 (1972).
104. J. A. Gupta, R. Knoble, N. Samarth, and D. D. Awschalom, *Science* 292, 2458 (2001).
105. D. P. DiVincenzo, D. Bacon, J. Kempe, G. Burkard, and K. B. Whaley, *Nature* 408, 339 (2000).
106. G. Burkard, D. Loss, and D. P. DiVincenzo, *Phys. Rev. B* 59, 2070 (1999).
107. B. E. Kane, N. S. McAlpine, A. S. Dzurak, R. G. Clark, G. J. Milburn, He Bi Sun, and H. Wiseman, *Phys. Rev. B* 61, 2961 (2000).
108. K. Ono, D. G. Austing, Y. Tokura, and S. Tarucha, *Science* 297, 1313 (2002).

www.aspbs.com/enn

Single-Electron Transistors

Jia Grace Lu

University of California at Irvine, Irvine, California, USA

CONTENTS

1. Single-Charge-Tunneling Mechanisms
2. Superconducting Single-Electron Transistor
3. Implementation Approaches of Nanoscale Oxide Junctions
4. Application of Single-Electron Transistors
5. Summary
Glossary
References

1. SINGLE-CHARGE-TUNNELING MECHANISMS

1.1. Introduction

It is well known that classical mechanics provides an accurate description of the behavior of macroscopic objects. On the other hand, the physics of microscopic systems involving electrons, atoms, and molecules is exclusively quantum mechanical. With the advances of fabrication techniques, the crossover region between these two fundamental regimes has been the subject of intense experimental and theoretical investigations. Systems that fall into this crossover region are called mesoscopic—although they contain a macroscopic number of particles, these systems are small enough in physical size that their behavior reveals quantum-mechanical effects. The characteristic length scale of mesoscopic samples varies from nanometers to tens of micrometers.

Several new effects have been observed in mesoscopic devices. The wave nature of electrons is apparent in the interference effects which produce quantum conductance fluctuations and weak localization effects [1–3]. In addition, it has also given rise to the Aharonov–Bohm effect in ring geometries [4]. In contrast, the ballistic transport of electrons through short and narrow channels results in quantized conductances [5]. The discreteness of the electron energy spectrum has been probed in semiconductor heterostructures [6] and small metal particles [7]. It is in this last category of charging effects that this chapter belongs.

The discreteness of the electronic charge is usually not evident in conventional electron devices, in which the current is regarded as a continuous charge flow. However, if electrons are confined to small isolated regions (called islands) that are weakly coupled to its surroundings, the discreteness of the electronic charge can significantly influence the electrical properties of the system. The capacitance of the island to the external circuit can be so small that the charging energy required to add a single electron becomes the dominant energy. This phenomenon is called the "single-electron-charging effect."

Single-electron-charging effects have been widely studied on double oxide-tunnel-junction systems, also called single-electron transistors (SETs). Such a sample is conventionally fabricated by electron-beam lithography and shadow evaporation techniques. It consists of a metallic island which is weakly coupled to two bias leads through small-capacitance and high-resistance tunnel junctions, and capacitively coupled to a gate electrode. The islands and the leads are separated by thin oxide-tunnel barriers, through which electrons can only be transported by quantum-mechanical tunneling. The gate is used to control the average number of electrons on the island. This chapter will emphasize the current transport mechanisms in SETs with a normal metal island or a superconductor island, and describe the fabrication and application of SETs.

1.2. Theoretical Framework

1.2.1. Electron Tunneling

The field of single-charge tunneling was formulated in the middle of the 1980s. The Orthodox theory [8–12] of the correlated single-charge tunneling was shown to be extremely successful in describing most experimental results. The general operating principle of SET is to control the tunneling of a single charge. Such a system must have small islands that connect to other metallic regions via tunnel barriers, with a total tunneling resistance R_Σ exceeding the quantum resistance $R_Q = h/e^2 \approx 25.8$ kΩ. The criterion for this requirement can be derived by noting that, to observe a Coulomb blockade, the charging energy $E_c = e^2/2C_\Sigma$ must exceed the quantum energy uncertainty $\hbar/R_\Sigma C_\Sigma$ associated with

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

the lifetime due to tunneling. Essentially, it means that the wavefunction of an excess electron on the island is localized so that quantum fluctuations of the electric charge are negligible. Another criterion to observe single-charge-tunneling effects is $E_c \gg k_B T$, so that the energy E_c required to add a charge onto the island far exceeds the available energy of thermal fluctuations. These two conditions ensure that the transport of charges through the island is governed by the Coulomb charging energy.

To understand the electron transport through the SET, let us start by considering the tunneling between two normal metal electrodes separated by a barrier (NIN junction), as illustrated in Figure 1. The Fermi energies of the two electrodes are offset from each other by an amount eV. (The symbol e represents the magnitude of the electron charge, i.e., $e = |e|$.) The basic idea is that there is a nonzero probability of charge transfer by quantum-mechanical tunneling of electrons between two metals separated by a thin insulating barrier, which classically forbids tunneling. This probability falls exponentially with the distance of separation, and it depends on the properties of the insulating material. Electrons can tunnel across the barrier via an energy-conserving horizontal transition, from the filled states in electrode 1 to the empty states in electrode 2. The Hamiltonian describing this system is written as [13]

$$H = H_0 + H_T \tag{1}$$

where H_0 is the unperturbed Hamiltonian of the metals when they are completely isolated from each other, which can be expressed as

$$H_0 = \sum_{k,\sigma} \varepsilon_k c^+_{k\sigma} c_{k\sigma} + \sum_{q,\sigma} \varepsilon_q c^+_{q\sigma} c_{q\sigma} \tag{2}$$

where the first term corresponds to electrode 1 and the second corresponds to electrode 2. ε_k and ε_q are the electron energies measured with respect to the Fermi energies of electrodes 1 and 2, respectively. c^+ and c are the creation and annihilation operators. σ is a spin index, and k and q are the electron wavevectors.

H_T is the perturbing Hamiltonian which is a result of the electron wavefunctions "leaking" from one electrode into the other. It is referred to as the transfer or tunneling Hamiltonian, and is given by [13]

$$H_T = \sum_{k,q} T_{k,q} c_{1k} c^+_{2q} + \text{h.c.} \tag{3}$$

where $T_{k,q}$ is the tunneling matrix element which accounts for the wavefunction overlap between the two electrodes. The first term denotes taking an electron from electrode 1 to electrode 2; the Hermitian conjugate (h.c.) term transfers an electron in the opposite direction.

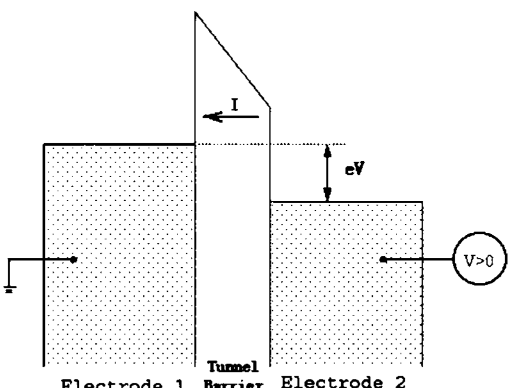


Figure 1. Tunneling diagram for an NIN junction at $T = 0$. An applied voltage produces a difference in the Fermi energies of the electrode. Electrons can tunnel horizontally through the barrier from filled states (shaded) in one electrode to empty states in the other.

Let us assume that the coupling due to tunneling between the two electrodes is sufficiently weak, so that it is reasonable to consider H_T as a first-order perturbation. According to Fermi's golden rule, the transition rate from a single initial state k to a set of final states q is given by [14]

$$\Gamma_k = \frac{2\pi}{\hbar} \sum_q |T_{k,q}|^2 \delta(\varepsilon_k + eV - \varepsilon_q) \tag{4}$$

The Dirac delta function ensures that energy is conserved in the tunneling process, that is, $\varepsilon_q = \varepsilon_k + eV$. One can transform the sum over q states to an integral over the energy ε_q by substituting the density of states (DOS) of electrode 2, $D_2(\varepsilon_k + eV)/2$ at $\varepsilon_k + eV$, and obtain

$$\Gamma_k = \frac{2\pi}{\hbar} |T|^2 \frac{D_2(\varepsilon_k + eV)}{2} \tag{5}$$

Note that the DOS of electrode 2, $D_2(\varepsilon_k + eV)$, has already included the spin degeneracy factor of 2. The division by 2 reflects that only states with the same spin as the original can be tunneled into, that is, no spin flips are allowed. Let us also assume that the average tunneling matrix amplitude $|T|$ is independent of the wavevectors k and q, and therefore, it is independent of the energies ε_k and ε_q.

At thermal equilibrium, the occupation probability of the electronic state is given by the Fermi distribution function

$$f(\varepsilon) = [1 + e^{\beta\varepsilon}]^{-1} \tag{6}$$

where $\beta = 1/k_B T$, and the probability of the unoccupied state is $1 - f(\varepsilon)$. Tunneling can only occur if the initial state is occupied and the final state is unoccupied. Thus, summing over all initial occupied states ε_k, including the factor $D_1(\varepsilon_k)$ for the DOS of electrode 1, one obtains the rate for tunneling from electrode 1 to electrode 2:

$$\Gamma_{1\to 2} = \sum_k \frac{2\pi}{\hbar} |T|^2 \frac{D_2(\varepsilon_k + eV)}{2} [1 - f(\varepsilon_k + eV)] \tag{7}$$

Replacing the sum over initial states k by an integral over ε_k, one gets

$$\Gamma_{1\to 2} = \frac{\pi}{\hbar} |T|^2 \int_{-\infty}^{\infty} D_1(\varepsilon) f(\varepsilon) D_2(\varepsilon_k + eV) \times [1 - f(\varepsilon + eV)]\, d\varepsilon \tag{8}$$

Since the bias energies are much less than the Fermi energies in the normal electrodes, one can assume that the

DOS is independent of energy, that is, $D_1(\varepsilon) = D_1$ and $D_2(\varepsilon) = D_2$. In this way, the forward tunneling rate becomes

$$\Gamma_{1\to 2} = \frac{\pi}{\hbar}|T|^2 D_1 D_2 \int_{-\infty}^{\infty} f(\varepsilon)[1 - f(\varepsilon + eV)]\, d\varepsilon \quad (9)$$

Using mathematical identities for the Fermi functions,

$$f(\varepsilon)[1 - f(\varepsilon + eV)] = \frac{f(\varepsilon) - f(\varepsilon + eV)}{1 - e^{-\beta eV}} \quad (10)$$

one can express Eq. (9) as

$$\Gamma_{1\to 2} = \frac{eV}{e^2 R_T (1 - e^{-\beta eV})} \quad (11)$$

with

$$R_T = \frac{\hbar}{\pi |T|^2 e^2 D_1 D_2} \quad (12)$$

The tunneling rate in the reverse direction $\Gamma_{2\to 1}$ is simply obtained by reversing the sign of the bias voltage, that is,

$$\Gamma_{2\to 1} = \frac{eV}{e^2 R_T (e^{\beta eV} - 1)} \quad (13)$$

The net current through the junction is then obtained by subtracting the reverse tunnel current from the forward tunnel current:

$$I(V) = e(\Gamma_{1\to 2} - \Gamma_{2\to 1}) \quad (14)$$

After inserting the tunneling rates, one gets

$$I(V) = V/R_T \quad (15)$$

recovering the ohmic relationship, with R_T being the normal state resistance of the tunnel junction.

One can also write the tunnel rates in terms of a more general energy change ΔF. Let us define ΔF as the change in the system free energy in going from the initial state to the final state. The electron travels from electrode 1 to electrode 2 in the direction favored by the applied bias, the bias source does an amount of work eV. Assume that the tunneling electron rapidly relaxes to the Fermi level, since the tunneling process is essentially irreversible, it leads to a decrease in the system free energy by this amount. Thus, $\Delta F = -eV$. Then one can express the tunneling rate in terms of ΔF:

$$\Gamma(\Delta F) = \frac{1}{e^2 R_T} \frac{\Delta F}{e^{\beta \Delta F} - 1} \quad (16)$$

Figure 2 plots the tunneling rate as a function of ΔF for various temperatures. This rate equation forms the basis of the Orthodox theory of single-electron tunneling [8–12].

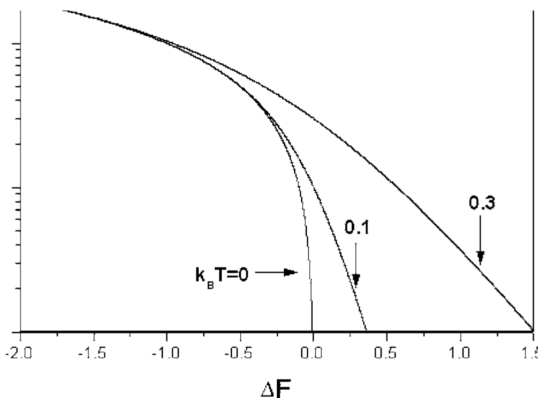


Figure 2. Forward tunneling rate of an NIN junction as a function of ΔF at different values of $k_B T$.

1.2.2. Coulomb Blockade of Tunneling

As discussed before, when the capacitances of the junctions become sufficiently small, and the Coulomb charging energy E_c of a single electron is large enough, then the discreteness of the electronic charge becomes very important. Let us first look at a single tunnel junction with small capacitance. The electrostatic energy of an isolated capacitor C with charges $Q(>0)$ and $-Q$ on the two electrodes is $Q^2/2C$, or $CV^2/2$, where $V = Q/C$. If an electron tunnels from the negative electrode to the positive one, the charge on the capacitor becomes $\pm(Q - e)$, so that the energy of the capacitor becomes $(Q - e)^2/2C$. (Here it is assumed that the single junction does not form a closed circuit, so that the capacitor charges do not come to equilibrium with the rest of the circuit.) This indicates an increase in the system energy unless the initial charge $Q \geq e/2$, that is, $V \geq e/2C$. This implies that electron transfer is energetically forbidden for voltages $V < e/2C$. This regime of zero tunnel current, despite a finite voltage across the junction, is called the Coulomb blockade [1].

Since the relevant time scale for tunneling is set by the tunneling resistance and the capacitance $\tau = RC$, which is on the order of 10^{-10} s, the effective shunting resistance is determined by the high-frequency properties of the leads near the junction, not the dc bias resistance. Hence, unless ultracompact resistors are inserted right at the junction [1], the effective impedance seen will be the high-frequency impedance of the leads, $Z \approx 100\ \Omega$, which is much less than R_Q. Thus, it is difficult to observe the Coulomb blockade effect in a single tunnel junction, no matter how low the temperature gets. One way to circumvent this problem is to make a double-junction system in which each junction is effectively isolated from the low-impedance environment by high tunnel resistance and low capacitance of the other junction (see Fig. 3).

1.2.3. Energy Considerations in the Single-Electron Transistor

In this section, the energetics of the double-junction system made of only normal metals will be considered. The energetic considerations are important because, if one knows how to calculate the change in the system free energy ΔF for a tunneling event, then one can calculate the rate at which

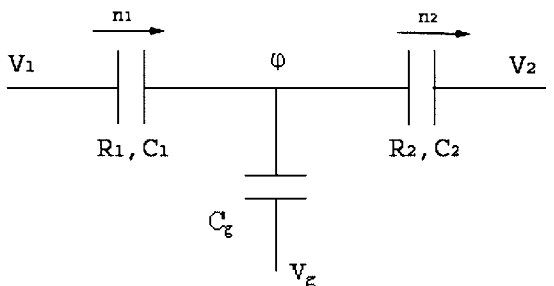


Figure 3. Schematic of a single-electron transistor showing two small capacitance tunnel junctions and a capacitively coupled gate. The region between the two tunnel junctions and the gate capacitor forms the island. By varying the gate voltage, the number of electrons on the island can be controlled.

this particular process occurs. When the leads and the island are normal metals, this rate is determined by Eq. (16). Once all of the tunneling rates are known, the tunneling current through the device can be determined.

The total capacitance of the island to its surroundings is given by the sum of the capacitances to the three electrodes:

$$C_\Sigma = C_1 + C_2 + C_g \tag{17}$$

The self-capacitance of this island is negligibe, and its effects could be included by merely renormalizing the other capacitances. Since charge is quantized in units of e, the total charge on the island is always discrete. However, continuously increasing the gate voltage is mathematically equivalent to continuously adding charge to the island. Thus, the effective charge of the island is continuous. This shows how the gate can tune the strength of the charging energy barrier.

To calculate the system free energy for the single-electron transistor, let us define n_1 and n_2 to be the number of electrons that have tunneled forward across junctions 1 and 2, respectively, and the integer $n = n_1 - n_2$ to be the excess number of electrons on the island. If ϕ is the electrostatic potential of the island, then with the voltages and capacitances defined in the schematic diagram of Figure 3, one has (let $C_3 = C_g$ and $V_3 = V_g$) [15]

$$\sum_{i=1}^{3} C_i(V_i - \phi) = ne \tag{18}$$

From this, one can find the electrostatic potential of the island:

$$\phi = \frac{\sum_{i=1}^{3} C_i V_i - ne}{C_\Sigma} \tag{19}$$

where ϕ is measured with respect to the same reference as V_1, V_2, and V_g. The electrostatic energy U of the system is given by summing the field energies of the three capacitors. Thus,

$$U = \frac{1}{2}\sum_i C_i(V_i - \phi)^2 \tag{20}$$

If one uses Eq. (19) and substitutes for φ in Eq. (20), after some algebra, one obtains

$$\begin{aligned} U &= \frac{(-ne)^2}{2C_\Sigma} + \frac{1}{2C_\Sigma}\sum_i \sum_{j>i} C_i C_j (V_i - V_j)^2 \\ &= \frac{(-ne)^2}{2C_\Sigma} + \Lambda \end{aligned} \tag{21}$$

where Λ is a constant independent of n. Since only the changes in the system free energy associated with tunneling events need to be considered, the exact value of Λ is not important.

When considering the tunneling of an electron onto or off the island, one must include not only the change in the charging energy U from Eq. (20), but also the work done by the bias voltage sources. The total work done by the voltage sources when an electron tunnels onto the island across junction j is

$$W_j = e\sum_i \frac{C_i}{C_\Sigma}(V_i - V_j) \tag{22}$$

In the case of the symmetric voltage bias, that is, $V_1 = -V/2$ and $V_2 = +V/2$, for an electron tunneling onto the island across junction 1, the voltage source does work:

$$W_1 = \frac{e}{C_\Sigma}[(C_2 + C_g/2)V + C_g V_g] \tag{23}$$

and similarly, for junction 2, the voltage source does work:

$$W_2 = -\frac{e}{C_\Sigma}[(C_1 + C_g/2)V - C_g V_g] \tag{24}$$

The system free energy can then be calculated by subtracting the work done by the voltage sources from the electrostatic energy, that is,

$$F_{\text{sys}}(n_1, n_2) = U - n_1 W_1 + n_2 W_2 \tag{25}$$

Substituting Eqs. (23) and (24) into Eq. (25), one obtains

$$\begin{aligned} F_{\text{sys}}(n_1, n_2) = \Lambda + \left[\frac{(-ne)^2}{2C_\Sigma} - n\frac{eQ_o}{C_\Sigma}\right] - n_1\frac{e}{C_\Sigma}(C_2 + C_g/2)V \\ - n_2\frac{e}{C_\Sigma}(C_1 + C_g/2)V \end{aligned} \tag{26}$$

where $Q_o = C_g V_g$ is the induced gate charge, which is a continuous variable as compared to the discrete change ne from tunneling events. This offset can be suppressed by shifting the zero of the gate voltage. In practice, naturally occurring random charged impurities near the island shift the polarization charge by an amount independent of V_g, and may drift or change discontinuously in time.

If the integer $m = n_1 + n_2$ is defined as the total number of electrons that have tunneled forward through either junction, then after completing the square of the term in brackets in Eq. (26), one finds (up to a constant which is independent of n and m)

$$F_{\text{sys}}(n, m) = \frac{(Q_o - ne)^2}{2C_\Sigma} - \left[m + n\frac{C_2 - C_1}{C_\Sigma}\right]\frac{eV}{2} \tag{27}$$

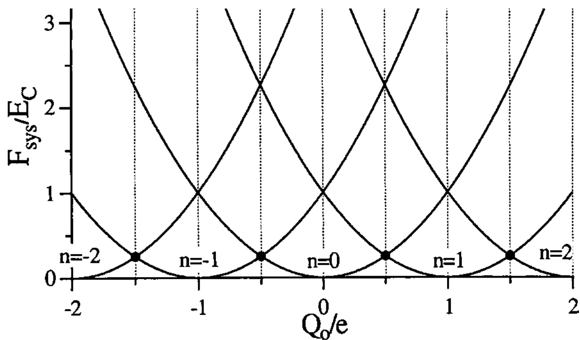


Figure 4. System free energy as a function of Q_o/e for various numbers n of excess electrons at $V = 0$. For fixed n, changing Q_o moves along the appropriate parabola.

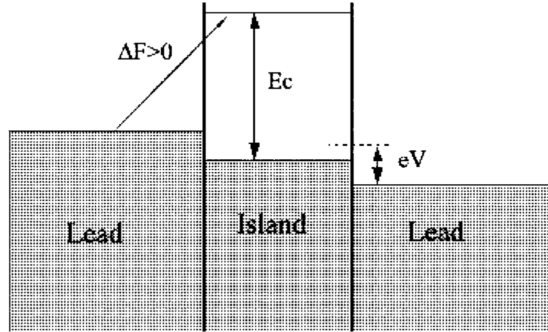


Figure 5. Energy diagram of an NNN SET with symmetric junction capacitances. The Coulomb blockade exists when the tunneling process is energetically unfavorable below the threshold voltage e/C_Σ.

The first term is conventionally called the charging energy, with an effective island charge of $Q_o - ne$. The second term represents the total work done by the voltage sources. Because the effective charge of the island influences the strength of the charging energy barrier, the gate voltage can be used to tune this barrier, and therefore modulate the current through the device. Since n must be an integer, the minimum energy for given Q_o is obtained if n is the integer closest to Q_o/e. That implies that the n giving the lowest charging energy must lie in the range

$$\frac{Q_o}{e} - \frac{1}{2} \leq n \leq \frac{Q_o}{e} + \frac{1}{2} \tag{28}$$

In Figure 4, the system energy is plotted as a series of parabolas for $V = 0$. Each curve corresponds to a different number of n excess electrons on the island. As the gate voltage varies, that is, changing the gate charge Q_o, the system favorable energy level can be changed. The parabolas for n and $n \pm 1$ cross at $Q_o = (n \pm \frac{1}{2})e$ at energy $E_c/4$. At each crossing point, one electron tunnels onto or off the island, changing the number of excess electrons on the island from n to $n \pm 1$. Because the system energy is periodic in Q_o, the current is also periodic in Q_o with period e, and with current peaks occurring at half-integer values of Q_o/e. In other words, in one modulation period of the current, that is, the gate voltage increases by $\pm e/C_g$, a single electron is added or removed from the island. The current through this device is a function of both the bias voltage V and the gate voltage V_g, exhibiting transistor action. Thus, varying the gate voltage can manipulate the tunneling of electrons one by one. Therefore, this system is given the name single-electron transistor.

1.2.4. Energy Diagrams for Single-Electron Tunneling

It is useful to draw simple energy diagrams to illustrate how the Coulomb blockade arises, and how it can be tuned away. Figures 5–7 show the energy diagram for tunneling in a normal metal SET (NNN system) with symmetric junctions, that is, $C_1 = C_2$. The energy that can be supplied by the bias sources is represented by shifting the Fermi level of one of the leads with respect to the other by an amount eV. The change in the charging energy due to tunneling is represented by the distance between the Fermi level in the island and the solid line above it. This solid line represents the position of the Fermi level after the tunneling event.

Consider symmetrically biased junctions at $V_g = 0$; in order to overcome the Coulomb blockade, the bias voltage has to exceed a threshold value at which $\Delta F_{n \to n+1} = 0$. This yields

$$eV \geq (2n+1)\frac{e^2}{C_\Sigma} = 2(2n+1)E_C \tag{29}$$

Figure 5 shows the case for $n = 0$; tunneling from the left lead onto the island increases the charging energy by E_c. For $V < e/C_\Sigma$, this tunneling process is energetically unfavorable ($\Delta F > 0$), and occurs with an exponentially suppressed rate for $T \ll E_c$, so very little current will flow through the system, and thus tunneling is blocked. On the other hand, for $V > e/C_\Sigma$, the bias source provides enough energy to overcome the charging energy barrier, and the Coulomb blockade is overcome and current flows through the system, as shown in Figure 6.

Figure 7 shows that, when $Q_o = e/2$, the Coulomb blockade is completely tuned away, and current will flow through the system for any nonzero bias voltage. The I–V characteristic for $Q_o = 0$ and $Q_o = e/2$ is shown in Figure 8 for two junctions of equal tunnel resistances.

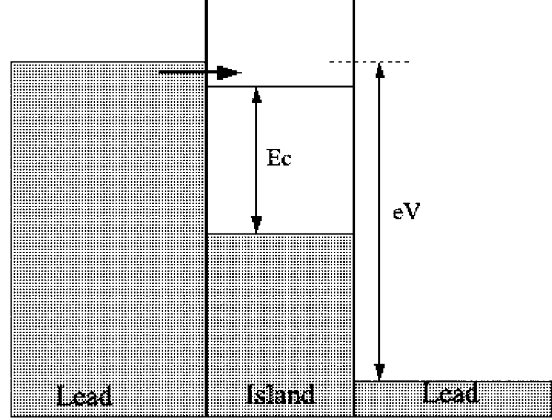


Figure 6. When $V > e/C_\Sigma$, that is, bias voltage sources provide enough energy to overcome the Coulomb barrier, single-electron tunneling occurs.

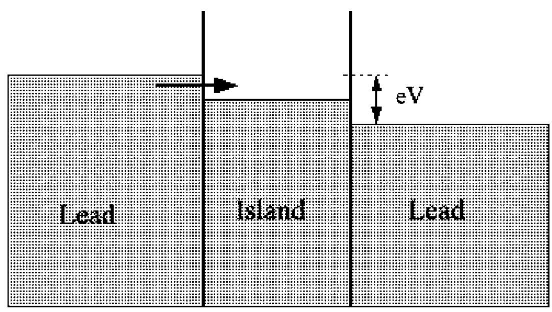


Figure 7. At $Q_o = e/2$, the Coulomb blockade is lifted at infinetesimal voltages as the charging energy barrier diminishes.

Considering electron tunneling across junctions 1 and 2, respectively, one can calculate the energy changes in the transition from the n to $n \pm 1$ state [15]:

$$\Delta E_1^{\pm} = \frac{e^2}{C_\Sigma}\left\{\left[\frac{1}{2} \pm \left(n - \frac{Q_o}{e}\right)\right] \mp \frac{(C_2 + C_g/2)V}{e}\right\} \quad (30)$$

$$\Delta E_2^{\pm} = \frac{e^2}{C_\Sigma}\left\{\left[\frac{1}{2} \pm \left(n - \frac{Q_o}{e}\right)\right] \pm \frac{(C_1 + C_g/2)V}{e}\right\} \quad (31)$$

For a transition to occur at $T = 0$, it is necessary that the relevant energy change ΔE is negative. For net current through the device, both tunneling on across one junction (charging step) and then off across the other (discharging step) must be allowed. The sum of the two ΔE values for a through passage of charge is always exactly eV, so that the energy would always be lowered by $e|V|$. The essence of the Coulomb blockade is that the energy must be lowered on each of the two successive transitions of charging and discharging the island. Actually, as long as the first step becomes energetically favorable, the second step will always become favorable. For the case when an electron tunnels off from the island to the right electrode first, followed by an electron from the left onto the island, it is the same as considering a hole which tunnels from the right electrode onto the island, then off to the left electrode.

From ΔE_1^+ and ΔE_2^-, one can determine the voltage thresholds for the charging step across junction 1 and for the discharging step across junction 2. ΔE_1^+ for the charging step becomes negative for positive bias voltage above the threshold $V_{1,\text{th}}^+$, and ΔE_2^- becomes negative for positive bias voltage above $V_{2,\text{th}}^-$ [15]. The extent of the Coulomb blockade is determined by whichever is the lesser of the two thresholds:

$$V_{1,\text{th}}^+ = \frac{e}{C_2 + C_g/2}\left[\frac{1}{2} + \left(n - \frac{Q_o}{e}\right)\right] \quad (32)$$

$$V_{2,\text{th}}^- = \frac{e}{C_1 + C_g/2}\left[\frac{1}{2} - \left(n - \frac{Q_o}{e}\right)\right] \quad (33)$$

Therefore, the Coulomb blockade can be overcome by a combination of the tunnel voltage V and gate voltage V_g. In the case of symmetric junctions (i.e., $C_1 = C_2$), tunneling is blocked when the following condition is fulfilled:

$$C_\Sigma |V|/e < (2n + 1) - 2C_g|V_g|/e \quad (34)$$

For $n = 0$, one sees that the Coulomb blockade occurs for the area within the diamond of Figure 9. If the junctions are asymmetric, then the diamond shape will be skewed. From the slopes, the capacitance of each junction can be extracted.

1.2.5. Current Through the Single-Electron Transistor

This section will describe how one can calculate the tunneling current through an SET in which the leads and the island are normal metals. Let us first assume that the electrons cannot tunnel across both junctions simultaneously. This process is referred to as cotunneling, which will be discussed in the next section. One can calculate the ΔF associated with transitions from charge state n to charge state n' using Eq. (27). Once ΔF are known, all tunneling rates can be calculated using Eq. (16).

Let us define a steady-state probability $\sigma(n)$ to be the ensemble distribution of the number of electrons on the island. $\sigma(n)$ can be viewed as the classical analog of the diagonal elements of the quantum-mechanical density matrix. Assuming no charge accumulation on the island at steady state, one can determine $\sigma(n)$ by requiring the total probability of tunneling into a state to be equal to the total probability of tunneling out of it.

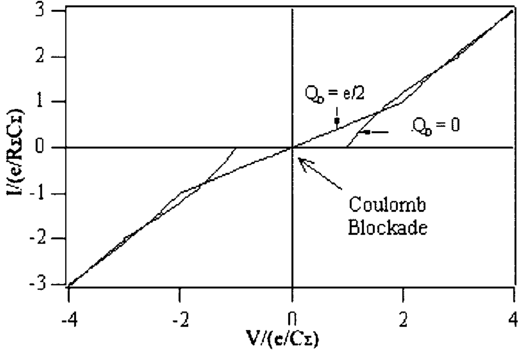


Figure 8. The current through a normal metal SET with identical junctions, calculated as a function of gate-induced charge for the cases of $Q_o = 0$ and $Q_o = e/2$ ($T = 0$).

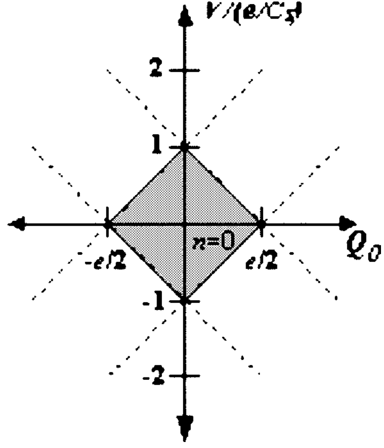


Figure 9. V–Q_o diagram illustrating Coulomb blockade for $n = 0$ state (shaded area).

The master equation to determine $\sigma(n)$ is [10–12]

$$\sum_n \sigma(n)\Gamma(n \rightarrow n') = \sum_{n'} \sigma(n')\Gamma(n' \rightarrow n) \tag{35}$$

Here, $\Gamma(n \rightarrow n')$ includes both the rates $\Gamma_1(n \rightarrow n')$ and $\Gamma_2(n \rightarrow n')$ for going from state n to state n' by tunneling through either junction, and the summation runs over all possible island charges. In the case of normal metals, only adjacent n states are connected via tunneling, and the distribution satisfies detailed balance:

$$\sigma(n)\Gamma(n \rightarrow n+1) = \sigma(n+1)\Gamma(n+1 \rightarrow n) \tag{36}$$

Thus, $\sigma(n)$ can be solved subject to the normalization condition $\sum_n \sigma(n) = 1$. When either the lead or the island is superconducting, the current mechanism changes to two-electron tunneling at low bias voltages, that is, from n to $n \pm 2$. In these cases, the full matrix equation of (35) must be solved.

From the determination of the island charge distribution function $\sigma(n)$, one can calculate the steady-state current through the transistor by considering transitions across a single junction for any possible island charge. For normal metals, taking account of the forward and reverse tunneling across each junction, the net current can be written as

$$I(V) = -e\sum_n \sigma(n)[\Gamma_1(n \rightarrow n+1) - \Gamma_1(n \rightarrow n-1)] \tag{37}$$

$$= -e\sum_n \sigma(n)[\Gamma_2(n \rightarrow n-1) - \Gamma_2(n \rightarrow n+1)] \tag{38}$$

Since there is no charge accumulation on the island in the steady state, the currents in the two junctions are equal. The finite probablity of occupying higher energy n states for voltages above threshold results in a nonlinear I–V curve called the "Coulomb staircase," which is most pronounced in the case of asymmetric junction resistance. This overall approach to calculate the current is referred to as the Orthodox theory of single-electron tunneling [10–12].

1.2.6. Cotunneling in the Single-Electron Transistor

Consider a Coulomb blockade region, where the sequential tunneling of an electron from the left electrode to the island is energetically unfavorable ($\Delta F > 0$), and thus forbidden, even in the presence of a finite bias voltage. Therefore, from what has been previously discussed, it is impossible to transport an electron from the left to the right electrode by sequential single-electron tunneling across the junctions. However, the cotunneling or macroscopic quantum tunneling processes make the transport possible. There are two types of cotunneling processes: [16, 17] (1) an inelastic process, in which two different electron states are involved, and there is an electron–hole excitation left on the island; and (2) an elastic process, in which the same electron state is involved in both tunneling processes, and there is no excitation left over.

Figure 10 illustrates the inelastic cotunneling process. The energy–time uncertainty relation of quantum mechanics results in the quantum fluctuation of electric charge on the island within a short period of time associated with the tunneling event. This implies that, when one electron tunnels onto the island and a second electron tunnels off the island within the time $\hbar/E_c$, the net result is that an electron has tunneled through the whole system. This process involves electrons tunneling simultaneously (here, "simultaneous" is used to distinguish cotunneling from sequential tunneling, in which two tunneling events are separated by a time interval great than $\hbar/E_c$) through the two junctions via a virtual intermediate state with increased electrostatic energy, and leaving an electron–hole excitation on the island [16]. This inelastic cotunneling process transfers the electrostatic energy of the system into the energy of the electron–hole excitations. Since the process is of the second order, the rate of the cotunneling is small compared to the rate of the sequential tunneling, but nevertheless, this process is dominant in the Coulomb blockade regime, where the sequential tunneling is suppressed.

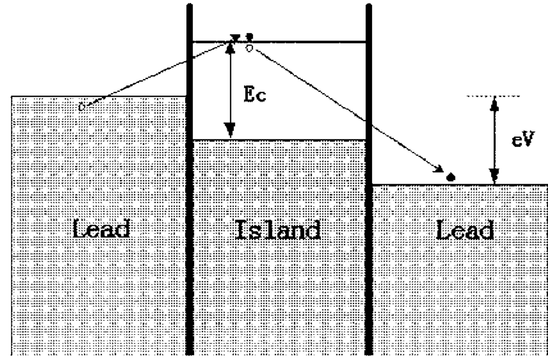


Figure 10. Schematic of the inelastic cotunneling process. Although single-electron tunneling is blocked, charge can be still transported through the system by simultaneous tunneling of electrons across both junctions.

The rate for inelastic cotunneling can be calculated by applying Fermi's golden rule for the higher order transitions. At low temperatures, the current due to this process is [17]

$$I_{\text{elastic}} = \frac{\hbar}{12\pi e^2 R_1 R_2}\left(\frac{1}{\Delta E_1} + \frac{1}{\Delta E_2}\right)^2 \times \left[(eV)^2 + (2\pi k_B T)^2\right]V \tag{39}$$

where ΔE_1 and ΔE_2 are the energy changes associated with the sequential tunneling in the first and second junctions as given by Eqs. (30) and (31). Here, $\Delta E_1, \Delta E_2 \gg eV$. The current varies as the third power of the voltage V, and it causes rounding in the I–V curve right near the voltage onset of the single-electron tunneling.

Cotunneling is elastic if the same electron state is involved in both tunneling processes, and the coherence of the tunneling electron is maintained in the tunneling event. Elastic cotunneling can also be viewed as a process in which an electron tunnels onto the island, virtually diffuses through it, and tunnels off across the opposite junction [18].

Inelastic cotunneling is an important charge-transport mechanism in the SET. It was first observed experimentally by Geerligs et al [19] in a lithographically patterned SET. In contrast, elastic cotunneling has only been observed

in tunneling through extremely small metal particles (~1–10 nm) in a double-junction system using a scanning tunneling microscope. Elastic cotunneling occurs with a negligible rate in lithographically patterned SETs because it takes an electron much longer than $\hbar/E_c$ to virtually diffuse from one junction to the other.

2. SUPERCONDUCTING SINGLE-ELECTRON TRANSISTOR

2.1. Subgap Charge-Transport Mechanism in an NSN System

Let us first consider the charge-transport mechanisms in an SET with normal metal leads and a superconducting (Al) island (NSN system). Charge transport at low temperatures and at low bias voltages in the NSN system occurs by two sequential Andreev steps, which is refered to as the Andreev cycle. (Andreev considered a normal metal and a superconductor in good metallic contact. But his physical picture can be generalized to tunnel junctions.) The first is the charging step, in which one electron incident on junction 1 is Andreev reflected into a hole on the same side of the junction. It can be regarded as an event in which two electrons tunnel onto the island to form a Cooper pair [20, 21]. The second is the discharging step, in which a hole incident on junction 2 is Andreev reflected into an electron. This can be regarded as an event in which a Cooper pair breaks up and forms two electrons tunneling off the island, and the island returns to the original state. The two steps of the Andreev cycle are analogous to the sequential single-electron tunneling, except that two electrons now tunnel in each step. Andreev reflection is subject to the same energetic considerations as is single-electron tunneling in the NNN system. Thus, it also exhibits a Coulomb blockade.

To calculate the Andreev reflection rate, Hekking et al. [21] model each step of the Andreev cycle as a second-order process, and break it down as follows. In the first transition of step 1 (step 2), an electron (hole) tunnels onto the island, and takes the system from the initial state to an intermediate state in which a virtual quasiparticle exists on the island. In the second transition, another electron (hole) tunnels onto the island through the same junction to a state paired with the existing quasiparticles. This leads to immediate recombination, and returns the system to the BCS ground state. By summing over appropriate initital, intermediate, and final states, the Andreev reflection rate across junction i is

$$\Gamma^{i}_{\text{Andreev}}(\Delta E_i) = \frac{G^{i}_{\text{Andreev}}}{e^2} \frac{\Delta E_i}{e^{\beta \Delta E_i} - 1} \tag{40}$$

where ΔE_i corresponds to the energy change in transfering both electrons and G^{i}_{Andreev} is the Andreev conductance of junction i. Note that this rate equation is analogous to that of single-electron tunneling in an NNN SET [refer to Eq. (16)], except that the Andreev conductance of the junction appears in place of the normal tunnel conductance of that junction. Unlike the normal conductance, however, the Andreev conductance depends on the specific geometry and the electronic mean free path in the leads [22–24]. Since Andreev reflection is a second-order process and the junctions are tunnel barriers with very low transparency, the Andreev conductances are expected to be much smaller than the normal tunnel conductance. A first-order estimate, which is based on ballistic electron motion near the junctions and ignores the effects of phase coherence between the tunneling electrons [21], gives Andreev conductances on the order $5 \times 10^{-12}\ \Omega^{-1}$, which are indeed much smaller than $1/R_{\Sigma}$. However, the experimentally determined Andreev conductances are approximately 10^3 times larger than this value. Since the elastic mean-free path in the leads is much smaller than the junction dimension, the ballistic picture assumed is not applicable, and one must consider directly the effects of phase coherence [25]. The magnitude of enhancement of the Andreev conductance by phase coherence is expected to depend on the precise geometry near the tunnel junctions, as well as the location of impurities and other scattering sites.

2.2. Subgap Charge-Transport Mechanisms in an SSS System

Figure 11 shows experimental data at low temperature and zero magnetic field for a sample when both the island and the leads are in the superconducting state (SSS system) [26]. In this figure, the current is measured while slowly sweeping V and quickly sweeping V_g, forming the envelope of all possible $I(V, V_g)$ curves. The upper graph in Figure 11 shows a sharp rise at 960 mV. Since this is an SSS system, this rise should occur at $V = 4\Delta/e$, which yields superconducting gap $\Delta = 240$ meV. The current peaks appearing at approximately 660 mV arise from a "$2e$–e–e" Josephson-quasiparticle (JQP) cycle in which the tunneling of a Cooper

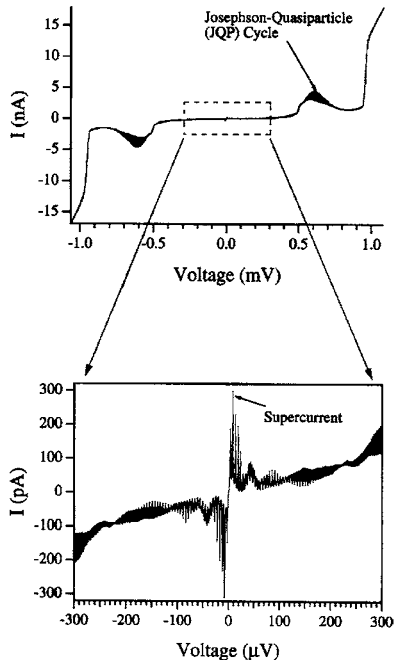


Figure 11. $I(V, V_g)$ data for an all superconducting SET. The upper graph shows the JQP peaks, and the lower graph shows the supercurrent peak which depends strongly on the value of Q_o.

pair in one junction is followed by two sequential quasiparticle tunneling events across the other junction [27–29]. Here, let us focus on the behavior of the system around zero bias voltages, shown enlarged in the lower graph of Figure 11.

In the SSS system, another energy scale, the Josephson coupling energy E_j, becomes important in addition to E_c and Δ. There are three distinguishable regimes [15]. If $E_c \ll E_j$, the charging energy effects are small, and the system has essentially the classical Josephson effect. If $E_c \gg \Delta$, single-electron charging energies dominate, observed phenomena show e periodicity, and Cooper-pair tunneling is unimportant. The most interesting regime is $E_c < E_j < \Delta$, in which there exists a supercurrent that is modulated by the gate charge Q_o with period $2e$.

If one considers junctions in which $R_\Sigma \gg R_Q$, then $E_j \approx (R_Q/R_\Sigma)\Delta \ll \Delta$, so there is a considerable range of E_c for which the two inequalities hold. The fact that Δ is larger than all other energies allows us to restrict our attention (at $T = 0$) to states of the island containing only an even number of electrons, which form Cooper pairs. At zero bias voltage, one can express the total energy of the system by the Hamiltonian [15]

$$H = \frac{(-en + Q_o)^2}{2C_\Sigma} - E_{j1}\cos\varphi_1 - E_{j2}\cos\varphi_2 \tag{41}$$

where n is the number of electrons, E_{j1} and E_{j2} are the Josephson coupling energies of the two junctions which couple the island to the superconducting leads, and φ_1 and φ_2 are the phase differences across each junction. A high conductance of the voltage source makes the phase sum $\vartheta = \varphi_1 + \varphi_2$ behave as a classical variable. On the contrary, the factor $\varphi = (\varphi_1 - \varphi_2)/2$ measuring the departure of the phase on the island from the midpoint between the phases of the two leads can behave as a quantum variable [10].

If one uses ϑ and φ to eliminate φ_1 and φ_2 in Eq. (41), and to simplify with trigonometric identities, one can reduce this equation to [15]

$$H = \frac{(-en + Q_o)^2}{2C_\Sigma} - E_j(\vartheta)\cos(\varphi - \chi) \tag{42}$$

where the parameters

$$E_j(\vartheta) = \sqrt{E_{j1}^2 + E_{j2}^2 + 2E_{j1}E_{j2}\cos\vartheta} \tag{43}$$

and

$$\chi = \arctan\left[\frac{E_{j1} - E_{j2}}{E_{j1} + E_{j2}}\tan\frac{\vartheta}{2}\right] \tag{44}$$

are determined by the fixed values of E_{j1}, E_{j2}, and ϑ. Since the eigenvalues of Eq. (42) do not depend on the reference phase χ, one can suppress it in the following equation, and rewrite it as

$$H = \frac{(-en + Q_o)^2}{2C_\Sigma} - E_j(\vartheta)\,\cos(\varphi) \tag{45}$$

This expression is then the same as the Hamiltonian of a single junction.

In the case when C_Σ is so large that the charging energy term in Eq. (45) can be neglected, the ground state has $\varphi = 0$, so that the supercurrent can be calculated [15]:

$$I_s = -\frac{2e}{\hbar}\frac{\partial E_j}{\partial\vartheta} = \frac{2e}{\hbar}\frac{E_{j1}E_{j2}\sin\vartheta}{E_j} \tag{46}$$

For identical junctions, that is, $E_{j1} = E_{j2} = E_i$, Eq. (46) reduces to

$$I_s = \frac{2e}{\hbar}E_i\sin\frac{\vartheta}{2} \tag{47}$$

which agrees with the classical result for two identical Josephson junctions in series, splitting equally the total phase difference ϑ.

In the other limit, $E_c > E_j$, in which the charging energy term is dominant, it has been shown [8] that the supercurrent at the degeneracy points between even-n states (i.e., at odd-integer values of Q_o/e) is

$$I_s = \frac{2e}{\hbar}\frac{E_{j1}E_{j2}\sin\vartheta}{2E_j} \tag{48}$$

This is exactly half of the value found in Eq. (46) for the classical case in which E_c is negligible [15]. And for midway between the degeneracy points, the supercurrent decreases to a minimum value of

$$I_s = \frac{2e}{\hbar}\frac{E_{j1}E_{j2}\sin\vartheta}{4E_c} \tag{49}$$

In summary, the Josephson current varies periodically with the gate charge. At odd-integer values of Q_o/e, it reaches a maximum value given by Eq. (48), comparable to the classical values for the supercurrent in individual junctions. And in between the degeneracy points, the current is depressed by a factor on the order of $E_j/E_c \ll 1$. It should be noted that thermal and quantum fluctuations can result in phase diffusion, characterized by a finite resistance of the supercurrent. As a result, it leads to a switching current lower than the nominal critical current.

2.3. Parity Effects in a Superconducting Island

One knows that thermodynamic properties of small systems depend on the parity of the number of particles N in the system. The best known example of such a dependence is provided by atomic nuclei, where the binding energy of nuclei with an even number of protons and neutrons is systematically larger than the binding energy of nuclei with an odd number of nucleons [30]. A similar phenomenon is expected to occur in small metallic clusters and tubes. Because of the spin degeneracy of the single particle states, the properties of clusters and tubes with even and odd numbers of electrons are quite different [31–33].

In both of these examples, however, the even–odd asymmetry, specifically the ground state energy difference, decreases with an increasing number of particles and vanishes in the thermodynamic limit. It has been found [34, 35] that superconductivity amplifies the parity effect in such a

way that it takes place even in a superconducting island with a macroscopically large number (10^9) of conduction electrons. Before the experiments described in this section were performed, it was widely believed that the strength of even–odd effects in superconductors would be suppressed as $1/N$. Thus, it was generally agreed that a superconductor with $N \gg 1$ would not exhibit parity effects in its macroscopic properties. However, the strength of the parity effect at very low temperatures is not governed by $1/N$, but rather, is in principle independent of N. Its strength is directly related to the difference between the number of quasiparticle excitations in the island when N is even and when is N odd. Since this difference is in principle equal to 1 at $T = 0$ regardless of the size of the system, at very low temperatures, the parity effect should not weaken as the system is made larger.

2.4. Experimental Observation of the Parity Effect

This section will show the temperature dependence of the parity effect in an NSN SET system. When the bias voltage V is close to zero, the system free energy is simply the electrostatic energy

$$U = \frac{(-en + Q_o)^2}{2C_\Sigma} \tag{50}$$

which depends on n, the number of excess electrons on the island. As V_g is swept, n changes by unity every time Q_o passes through a half-integer value. This leads to a variation of the populations, and hence the current $I(V_g)$ at fixed bias V is e-periodic with peaks at half-integer values of Q_o/e, corresponding to single-electron tunneling.

When the island is superconducting, it was found that the I–Q_o curve taken under similar conditions becomes $2e$-periodic. Figure 12 shows a series of I–Q_o curves for an NSN sample (with $T_c \approx 1.5$ K) at different temperatures and zero magnetic field [36]. At the lowest temperatures, the curves are strongly $2e$-periodic, with two-electron tunneling (Andreev peaks) occurring at odd-integer values of Q_o/e. This $2e$ periodicity indicates that the number of electrons on the island changes by two as Q_o is swept through one current peak. This in turn suggests that the island prefers to have a total number of conduction electrons of a certain parity. As the temperature rises, the Andreev peaks decrease in size, and gradually, each Andreev peak splits apart into a subtle double maximum. The two maxima eventually move farther apart in Q_o, and develop into well-separated peaks, with the current slightly higher at odd-integer values of Q_o/e than at even-integer values. Finally, the current becomes completely e-periodic with equal peaks at all half-integer values of Q_o/e, and the magnitude of the current modulation grows with a further increase in the temperature. From observation, these I–Q_o curves in Figure 12 are $2e$-periodic at temperatures up to and including 275 mK. The dashed line on the I–Q_o curve at 275 mK is drawn to assist viewing the existence of $2e$ periodicity. At 285 mK and above, only e periodicity can be distinguished in these curves. Thus, the crossover temperature from $2e$ to e periodicity can be estimated, $T^* \approx 285$ mK, which is surprisingly only about one fifth of the critical temperature T_c of the sample.

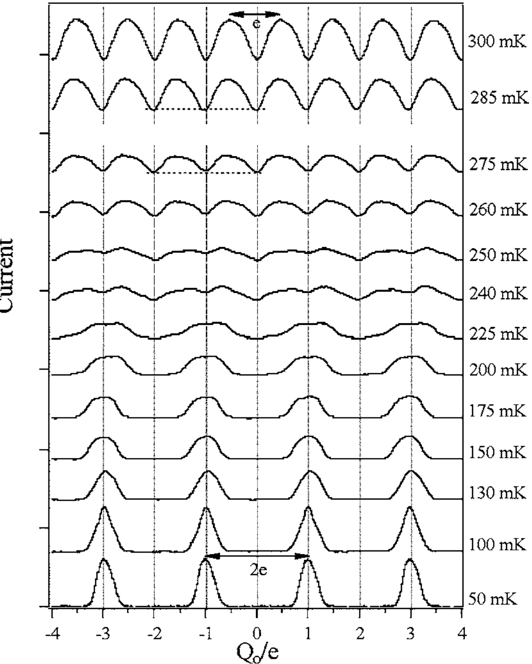


Figure 12. Experimental I–Q_o curves for an NSN sample at a small bias voltage ($V = 125$ μV) with temperatures ranging from 50 to 300 mK. The curves are displaced upward successively for clarity. The current axis is 1 pA per division. At low temperatures, the curves are strongly $2e$-periodic. As the temperature is gradually increased, the curves evolve until they become completely e-periodic above a crossover temperature T^* ($\approx$285 mK for this sample).

2.5. Equilibrium Model

To thoroughly understand the parity effect, it is necessary to make a kinetic calculation [37, 38], solving a master equation to find the self-consistent steady-state nonequilibrium populations of all relevant states, and the resulting current as a function of the bias voltage and the gate voltage. However, in the limit of low bias voltage, state populations will be near the $V = 0$ equilibrium values. At sufficiently low bias voltages, it is expected that the current through the device is proportional to V with a coefficient which is a function of V_g and T, dependent on the equilibrium populations. Thus, the period (e or $2e$) of the current will be determined by the period with which the populations vary with V_g. Therefore, one can simply use the periodicity of the equilibrium populations as a proxy for the periodicity of the current at low bias voltages. The simple approach is very useful, even if it is limited to finding the period of $I(V_g)$, without finding the exact magnitude and waveform.

The equilibrium model explains the origin of the parity effect and the temperature T^* at which it disappears. The basic idea of this model is that, since the BCS ground state of a superconductor consists of a coherent superposition of Cooper pairs, there is necessarily exactly one unpaired electron which appears as a quasiparticle excitation if the number of conduction electrons N is odd. The minimum energy of this excitation at $T = 0$ is given by the superconducting energy gap Δ. Thus, the ground-state energy of the system is higher by an amount Δ when N is odd (one quasiparticle excitation) compared to when N is even (all electrons

are paired). As a result, the states of even N become energetically favorable over those of odd N, giving rise to the even–odd parity effect. Averin and Nazarov [34] introduced to the system free energy an explicitly additive energy term, which has the value Δ in odd-N states and zero in even-N states.

The energy arguments of Averin and Nazarov apply only for the case of $T = 0$ and $H = 0$. At finite temperatures, a number of states become available for quasiparticle excitation. This gives entropy contributions which diminish the free energy difference between the even and odd states. On the other hand, for nonzero magnetic fields, the reduction of the minimum quasiparticle excitation energy also yields a smaller energy difference. Thus, one can generalize the model of Averin and Nazarov by defining a parity (or even–odd)-dependent free energy difference $F_0(T, H)$. The system free energy for $V = 0$ is thus expressed as

$$F_{\text{sys}} = \frac{(-en + Q_o)^2}{2C_\Sigma} + p_n F_0(T, H) \tag{51}$$

with

$$p_n = \begin{cases} 0, & \text{if } n \text{ is even} \\ 1, & \text{if } n \text{ is odd.} \end{cases} \tag{52}$$

2.5.1. Even–Odd Free Energy Difference

The even–odd free energy difference F_0 is calculated using equilibrium statistical mechanics. The grand canonical partition function Z for quasiparticle excitations on the superconducting island is [39]

$$Z = \prod_k (1 + e^{-\beta\xi_k}) \tag{53}$$

where ξ_k is the energy of a quasiparticle excitation in state k relative to the Fermi energy. In zero magnetic field,

$$\xi_k = \sqrt{\varepsilon_k^2 + \Delta^2} \tag{54}$$

where ε_k is the individual electron energy in the normal state relative to the Fermi energy. The grand canonical partition function can be separated algebraically into terms corresponding to an even or an odd number of quasiparticle excitations:

$$\begin{aligned} Z &= \left(1 + \frac{1}{2!}\sum_k \sum_{l\neq k} e^{-\beta\xi_k} e^{-\beta\xi_l} \ldots\right) \\ &\quad + \left(\sum_k e^{-\beta\xi_k} + \frac{1}{3!}\sum_k \sum_{l\neq k} \sum_{m\neq k,l} e^{-\beta\xi_k} e^{-\beta\xi_l} e^{-\beta\xi_m}\right) \\ &= Z_{\text{even}} + Z_{\text{odd}} \end{aligned} \tag{55}$$

If the number of electrons on the island is fixed, only Z_{even} or Z_{odd} is used since quasiparticle excitations can only be created two at a time. Only the tunneling of a quasiparticle from the leads onto the island can change the parity. The even–odd free energy difference is then defined as the difference in free energy calculated for the two parity cases:

$$F_0 = k_B T \ln\left(\frac{Z_{\text{even}}}{Z_{\text{odd}}}\right) \tag{56}$$

2.5.2. Low-Temperature Approximation of $F_0(T)$

At low temperatures, there are rarely any thermally excited quasiparticles. Consequently, the first term of Z_{even} and the first term of Z_{odd} dominate, yielding [39]

$$F_0 \approx -k_B T \ln\left[\sum_k e^{-\beta\xi_k}\right] \tag{57}$$

Making the continuum approximation and using the quasiparticle density of states $\rho_s(\xi)$ which includes the spin degeneracy, one gets

$$F_0 \approx -k_B T \ln\left[2V_I \int_0^\infty \rho_s(\xi) e^{-\beta\xi} d\xi\right] \tag{58}$$

V_I is the volume of the superconducting island, and the factor of 2 on the right-hand side is needed to count quasiparticles with $|k| < k_F$ and $|k| > k_F$. It is expected that $F_0(T = 0, H) = \Omega_G(H)$ since, at zero temperature, the remaining quasiparticle will reside at the minimum excitation energy, that is, the spectroscopic gap $\Omega_G(H)$, which is a decreasing function of the magnetic field. Thus, it is useful to rewrite Eq. (58) as

$$\begin{aligned} F_0 &\approx -k_B T \ln\left[e^{-\beta\Omega_G} 2V_I \int_{\Omega_G}^\infty \rho_s(\xi) e^{-\beta(\xi-\Omega_G)} d\xi\right] \\ &= \Omega_G - k_B T \ln(N_{\text{eff}}) \end{aligned} \tag{59}$$

with

$$N_{\text{eff}} = 2V_I \int_{\Omega_G}^\infty \rho_s(\xi) e^{-\beta(\xi-\Omega_G)}\, d\xi \tag{60}$$

Here, N_{eff} is the number of statistically significant states available for quasiparticle excitation. It roughly equals the number of quasiparticle states within $k_B T$ of the superconducting gap edge. The corresponding entropy is $k_B \ln(N_{\text{eff}})$. Note that the average number of thermally excited quasiparticles in the superconducting island is given by

$$\langle N_{qp}(T)\rangle = 2V_I \int_0^\infty \rho_s(\xi) f(\xi, T)\, d\xi \tag{61}$$

where $f(\xi, T)$ is the Fermi distribution function. For $k_B T \ll \Omega_G$, one can show that

$$\langle N_{qp}(T)\rangle \approx 2V_I \int_{\Omega_G}^\infty \rho_s(\xi) e^{-\beta\xi}\, d\xi = N_{\text{eff}} e^{-\beta\Omega_G} \tag{62}$$

In the low-temperature approximation, Eq. (59) shows that F_0 falls nearly linearly with temperature, becoming zero at a temperature determined by the criterion

$$T^* = \frac{\Omega_G}{k_B \ln(N_{\text{eff}}(T_0))} \tag{63}$$

This can be regarded as a first-order approximation of the $2e$ to e periodicity crossover temperature. The temperature dependence of N_{eff} only contributes weakly because it appears within a logarithm.

2.5.3. Crossover Temperature in Zero Magnetic Field

For zero magnetic field and low tempeartures, one can ignore the temperature dependence in Δ, and set $\Omega_G = \Delta$, so it is appropriate to use the BCS form of the superconducting density of states:

$$\rho_s(\xi) = \begin{cases} 0, & \text{for } \xi < \Delta \\ \rho_n(0)\frac{\xi}{\sqrt{\xi^2-\Delta^2}} & \text{for } \xi > \Delta \end{cases} \tag{64}$$

where $\rho_n(0)$ is the normal density of states in the island per unit energy and per unit volume (including spin). Substituting these zero-field expressions into the low-temperature approximations of Eqs. (59) and (60) gives

$$F_0 \approx \Delta - k_B T \ln(N_{\text{eff}}) \tag{65}$$

$$N_{\text{eff}} \approx V_I \rho_n(0)\sqrt{2\pi\Delta k_B T} \times \left[1 + \frac{3}{8}\frac{k_B T}{\Delta} + O\left(\frac{k_B T}{\Delta}\right)^2\right] \tag{66}$$

As an example, for the NSN sample (whose $I - Q_o$ curves are shown in Fig. 12) with $\Delta \approx 240\ \mu$eV, $E_c \approx 100\ \mu$eV, and island volume $V_I \approx 3.2 \times 10^{-15}$ cm^3, one can estimate N_{eff} to be approximately 10^4. Using the BCS value of $\Delta = 1.76 k_B T_c$, the first-order estimate of the $2e$ to e periodicity crossover temperature is

$$T^*(H = 0) = \frac{\Delta}{k_B T \ln(N_{\text{eff}})} \approx T_c/5 \tag{67}$$

which is in good agreement with the experimental data which gives the crossover temperature to be around 285 mK.

In summary, the nonzero F_0 results in a system free energy which is $2e$-periodic in Q_o. As the temperatures or magnetic fields increase, F_0 gradually decreases, causing a transition from $2e$ to e in the I–Q_o characteristics. To illustrate this point, Figure 13(a) shows the system free energy diagram of the NSN system at $T \approx 0$ and $T \geq T^*$. Physically, two distinct energies are competing with the thermal energy: the energy gap Δ opposes creation of quasiparticles, and the Coulomb energy E_c tries to make the electron number match the gate charge Q_o [36]. When $\Delta > E_c$, the ground state contains the even number of electrons closest to that specified by Q_o. The only accessible degeneracy points between even-n states occur at odd-integer values of Q_o/e. These degeneracy points are associated with the Andreev peaks as in Figure 13(b). As the temperature increases, F_0 decreases according to Eq. (65), and the odd-n states move downward in the system free energy diagram. Because the even–even crossing is the degeneracy point of the lowest energy, the Andreev peak maintains a single maximum as long as $F_o(T) > E_c$. When the temperature becomes T^*, the free energy difference between the even and odd states disappears, that is, $F_o(T^*) = 0$, as illustrated in Figure 13(c). At the degeneracy points where adjacent n states cross at half-integer values of Q_o/e, single-electron tunneling dominates, giving rise to the e-periodic I–Q_o curve as shown in Figure 13(d).

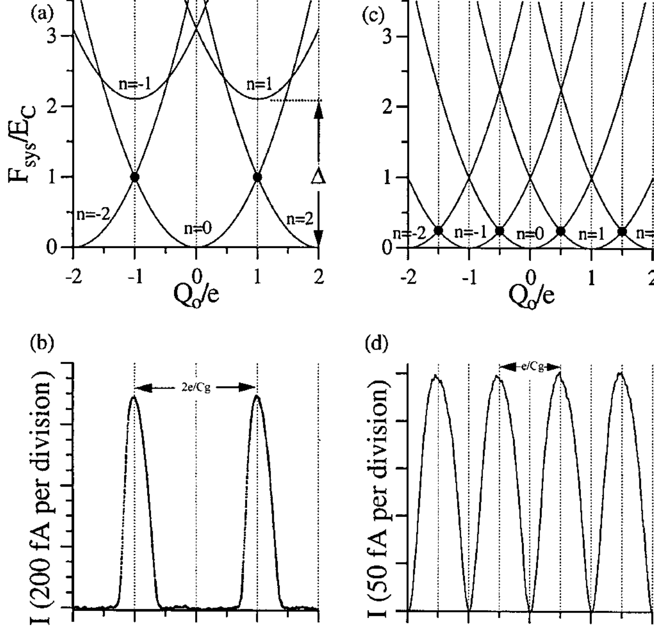


Figure 13. System free energy diagrams and corresponding I–Q_o curves measured at 125 μV for an NSN SET at $T = 50$ and 300 mK, respectively. (a),(b) For $T = 50$ mK, at the lowest lying degeneracy points where adjacent even-n states cross, two-electron tunneling occurs giving rise to the Andreev peaks that are $2e$-periodic in Q_o with peaks at odd-integer values of Q_o/e. (c),(d) For $T = 300$ mK $> T^*$, at the degeneracy points where adjacent n states cross, single-electron tunneling occurs giving rise to the current that is e-periodic with peaks at half-integer values of Q_o/e.

2.5.4. Calculation of a General Form of $F_0(T)$

For T near or above T^*, the presence of a thermally excited quasiparticle in the island becomes important. As a result, in order to accurately calculate the even–odd free energy difference in this temperature regime, one must consider higher order terms in Eq. (55). To simplify the calculation, let us model the quasi-continuum of excited states above the superconducting gap by a single level at energy Ω_G with degeneracy N_{eff}; then the partition sums of Eq. (53) can be written as

$$Z = (1 + e^{-\beta\Omega_G})^{N_{\text{eff}}}. \tag{68}$$

Let us assume that the island is in weak tunneling contact with a particle reservoir at $V = 0$, so that the island can contain either an even or odd number of electrons. By constructing forms in which either the even or odd terms cancel, one can then write partial partition sums as [36]

$$Z_{\text{even}} = \frac{[(1 + e^{-\beta\Omega_G})^{N_{\text{eff}}} + (1 - e^{-\beta\Omega_G})^{N_{\text{eff}}}]}{2} \tag{69}$$

and

$$Z_{\text{odd}} = \frac{[(1 + e^{-\beta\Omega_G})^{N_{\text{eff}}} - (1 - e^{-\beta\Omega_G})^{N_{\text{eff}}}]}{2} \tag{70}$$

At this point, it is helpful to make use of the approximation

$$\left(1+\frac{a}{b}\right)^{b} \approx e^{a} \tag{71}$$

which is valid for $b \gg 1$ and $|a| \ll b$. Applying this relation, for $e^{-\beta\Omega_G} \ll 1$ and $N_{\text{eff}} \gg 1$, one obtains

$$(1+e^{-\beta\Omega_G})^{N_{\text{eff}}} \approx e^{N_{\text{eff}}e^{-\beta\Omega_G}} \tag{72}$$

and

$$(1-e^{-\beta\Omega_G})^{N_{\text{eff}}} \approx e^{-N_{\text{eff}}e^{-\beta\Omega_G}} \tag{73}$$

One can thus write

$$Z_{\text{even}} \approx \cosh\left(N_{\text{eff}}e^{-\beta\Omega_G}\right) = \cosh(\langle N_{qp}\rangle) \tag{74}$$

$$Z_{\text{odd}} \approx \sinh\left(N_{\text{eff}}e^{-\beta\Omega_G}\right) = \sinh(\langle N_{qp}\rangle) \tag{75}$$

Substituting Z_{even} and Z_{odd} into Eq. (56), one finds that

$$\begin{aligned} F_0(T) &\approx k_B T \ln\left[\coth\left(N_{\text{eff}}e^{-\beta\Omega_G}\right)\right] \\ &= k_B T \ln[\coth\langle N_{qp}\rangle] \end{aligned} \tag{76}$$

which asymptotically approaches zero. Figure 14 plots the low-temperature approximation and the high-order form of $F_0(T)$ at zero magnetic field. The temperature dependence of N_{eff} gives the slight downward bending of $F_0(T)$ for temperatures below T^*.

Now, it is appropriate to describe the physical origin of the crossover temperature. According to T^* in Eq. (63), Eq. (62) at $T = T^*$ becomes

$$\langle N_{qp}(T^*)\rangle = N_{\text{eff}}(T^*)e^{-\Omega_G/k_B\left[\frac{\Omega_G}{k_B \ln(N_{\text{eff}}(T^*))}\right]} = 1 \tag{77}$$

Thus, in simple physical terms, the parity effect disappears at the temperature T^* (which is much less than T_c), at which there is on average one thermally excited quasiparticle in the entire island.

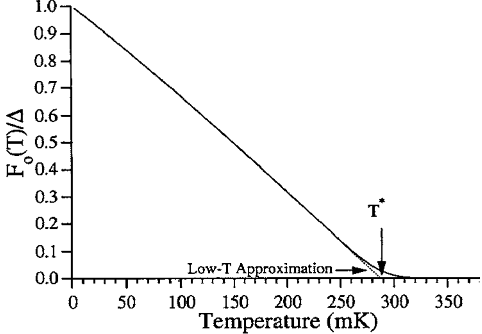


Figure 14. Temperature dependence of the even–odd free energy difference for the NSN sample ($T^* \approx 285$ mK). Dotted line shows the first-order approximation for low temperatures.

3. IMPLEMENTATION APPROACHES OF NANOSCALE OXIDE JUNCTIONS

SETs can be made using a wide variety of metals, semiconductors, or conducting polymers. This section will discuss the implementation of SETs using a lithographically fabricated ultranarrow thin film as an island in an $Al/Al_2O_3/Al$ three-layer process. Since aluminum is a superconductor, this device can be operated either in the superconducting state or in the normal state.

Typically, this kind of SET device is fabricated by electron-beam (e-beam) lithography, followed by a shadow evaporation technique. This fabrication technique was pioneered by Dolan [40]. The fabrication procedure consists of a series of steps. The first is to start with a polished Si wafer, which has a thin layer (~30 Å) of native oxide. The Si wafers are first coated by a bilayer resist, which is an e-beam-sensitive polymer. A scanning electron microscope (SEM) is used to write the device patterns on the wafer chips. Then they are developed so that the electron beam exposed areas are removed. After metal evaporation, the remaining resist gets removed in the liftoff step in acetone. In the following sections, we will take a closer look at the fabrication steps of resist film preparation, electron-beam writing, and metal evaporation.

3.1. Resist Film Preparation

In preparation for resist coating, the Si wafer is ultrasonically cleaned in trichloro-ethylene (TCE), followed by acetone and methanol. The wafer is then blown dry with compressed, dry N_2 gas, ensuring that no solution droplets remain. It is then placed on a spinner for resist coating.

E-beam resist is a polymer consisting of a long linear chain to which a number of side groups are attached. Often positive resist is used, which means that, when writing, the high-energy electrons break chemical bonds and sever the side groups of the resist. These side groups are easily removed, leaving behind a highly porous mass which the developer can infiltrate and dissolve.

The resist has two layers, a sensitive bottom layer and a less sensitive top layer. When coating each resist layer, a few drops of resist solution are applied to the wafer and spun to produce a thin and uniform film. The bottom resist layer consists of a copolymer of methyl methacrylate (MMA) and methacrylic acid (MAA), which has 8.5% by weight MAA. The top resist layer consists of a 1.5% by weight solution of 950K PMMA/MAA copolymer in chlorobenzene. As the electrons travel through the resist and reach the substrate, they backscatter into the resist, interacting with the resist molecules, and broadening the exposed area to form an undercut in the bottom layer. This undercut serves two important purposes. First, a good undercut profile helps to avoid tearing the deposited metal thin films in the liftoff phase of the fabrication. More importantly, sufficient undercut is required in order to create a suspended resist bridge (illustrated as RB in Fig. 17), enabling the formation of tunnel junctions by shadow evaporation. In the small gap between the lithographic lines of the island and the lead, the bottom resist layer is completely developed away, leaving only the top resist layer bridging the gap, forming the resist

bridge. Usually, the bottom layer should be at least three times as thick as the total thickness of the metal deposited. Figure 15 shows a schematic of the resist profile.

3.2. Electron-Beam Lithography—SET Patterning

E-beam lithography is a technique which takes advantage of the extremely small electron probe size in an SEM to create high-resolution patterns. The electron beam is finely focused to produce a probe as small as 50–100 Å in diameter. Such a small probe supplies a current in the range of 1–10 pA, suitable for submicron features. The probe is scanned along the resist to write the device pattern.

To facilitate e-beam lithography, an SEM running the Nanometer Pattern Generation System (NPGS, from Nabity Lithography Systems, Inc.) is often used to control the x and y deflections and the beam blanking. Using digital-to-analog converters connected to the SEM scan coils, the NPGS program controls the electron beam to write each pattern. It takes the designed pattern as input, and the writing parameters include beam current, center-to-center spacing, line spacing, dosage, and magnification. Figure 16 shows a design pattern for e-beam writing.

An accelerating voltage is applied to the SEM filament. Decreasing the accelerating voltage increases the fraction of primary electrons that are backscattered near the surface of the substrate. Since the intensity of the exposure due to backscattered electrons is higher in the bottom resist layer than in the top layer, the backscattered electrons affect the amount of undercut. Thus, a relatively low accelerating voltage is chosen in order to increase the amount of undercut. It is found that 12 kV is sufficiently low for electrons backscattered from the substrate to produce an adequately large undercut.

When e-beam writing is complete, the wafer is developed in an MIBK:IPA solution. MIBK stands for methyl isobutyl ketone, and IPA stands for isopropyl alcohol. The developers usually consist of two ingredients. The first is the solvent, and the faster it can infiltrate the resist mass, the faster it will dissolve it. The second is a nonsolvent which affects the rate at which the resist goes from a gel to a liquid, by adjusting the enthalpy of the solution. After developing, the chip is

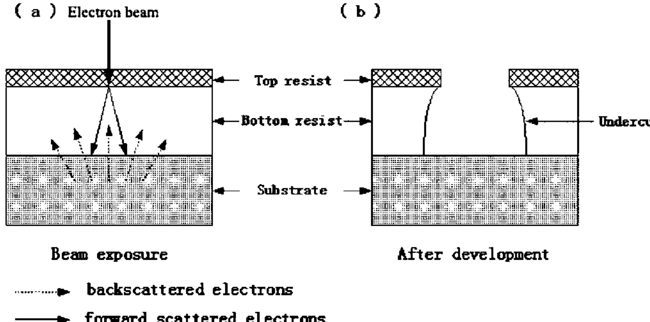


Figure 15. Resist profile. (a) View of the resist being exposed by the electron beam. (b) Because of the backscattered electrons and the higher sensitivity of the thick bottom layer, more of the bottom layer is removed in the developing procedure, thus creating a large undercut.

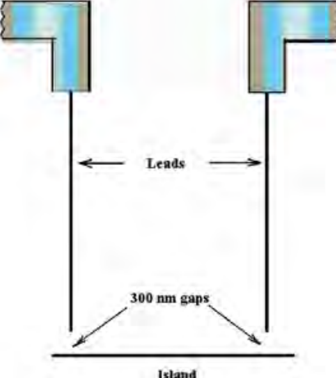


Figure 16. Example of an SET design pattern for e-beam patterning. The island has a length of 2 μm. The 300 nm spacing between the island and the leads produces resist bridges.

gently agitated in an IPA solution stop bath, and blown dry with dry N_2 gas.

3.3. Metal Evaporation and Liftoff

The samples are mounted on a rotatable sample stage in the evaporator. Here, we will show an example of a double-angle evaporation procedure for an SSS SET sample consisting of only aluminum. The first evaporation is straight down, which deposits approximately 250 Å of Al onto the substrate. As shown in Figure 17, this forms the island. This layer of Al is then oxidized in dry O_2 at ~50 mtorr for 3–5 min. The oxide insulating layer is about 1 nm thick. A second evaporation is done at a 45° angle along the two lead channels in the direction of the island. In this evaporation, 550 Å of Al are deposited.

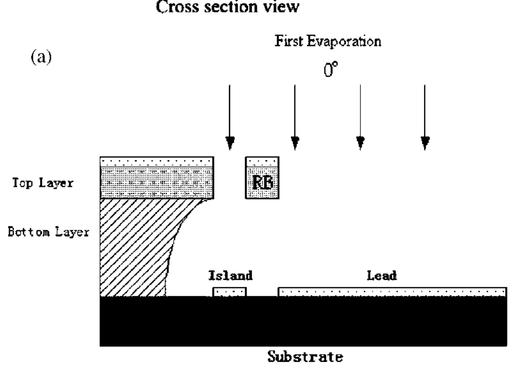
Following the evaporation step, the samples are placed in acetone for liftoff. This dissolves most of the remaining resist. The samples are then rinsed with acetone and methanol to remove the unwanted metal that was originally deposited on top of the resist.

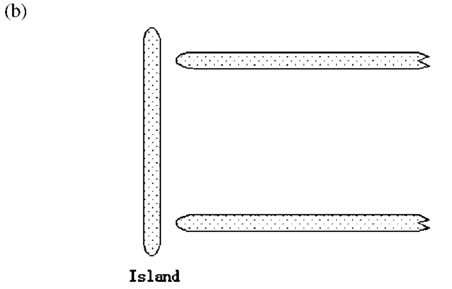
3.4. Sample Measurement

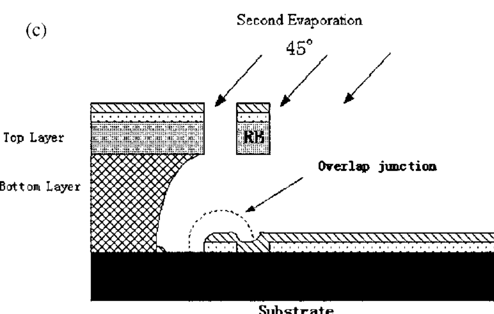
In order to observe the single-electron effects, the SETs must be measured at low temperatures. This is because the energies of these phenomena are lower than 0.1 meV for a system capacitance on the order of 10^{-15} F, which approximately equals the thermal energy at 1 K. Typically, four-probe current transport measurements are performed on samples which are cooled in dilution refrigerators.

SETs are usually very sensitive to noise [41], so a measurement is often taken inside an electromagnetically shielded room, and sufficient lead filtering is needed. High-frequency noise in the leads must be attenuated before it reaches the sample. Otherwise, the noise can lead to an RF heating of the sample, enabling photon-assisted tunneling which can dramatically change the current transport behavior. And for superconducting SETs, it can change the phase dynamics of the Josephson junctions.

Double Angle Evaporation







Top view of metal deposits

(d)

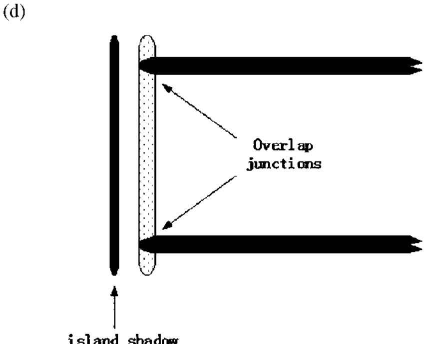


Figure 17. (a) Cross-sectional view along one of the lead channels showing where metal is deposited in the first evaporation. Because of the suspended resist bridge (RB), the first layer of the lead does not connect to the island. (b) Metal pattern after the first evaporation. (c) Cross-sectional view of the second evaporation at an angle 45° toward the island. (d) Two junctions are formed after the second evaporation.

4. APPLICATION OF SINGLE-ELECTRON TRANSISTORS

In the semiconducting industry, it is now possible to produce MOSFETs and integrated circuits with gate oxides less than 10 atoms across. Such thin films are required to maintain the current response of the transistor to lower voltages at the gate electrode. Manufacturers need to lower the power supply to individual components in order to pack more devices onto a chip. The roadmap projects that, by 2012, the gate oxide must be reduced to a thickness of five Si atomic layers as the device size miniaturizes [42]. Such scaling down of the gate oxide will eventually reach its fundamental physical limit, as the quantum-mechanical tunneling of electrons will cause the breakdown of electrical insulation. There are also other limitations associated with device miniaturization. For example, the power dissipated will be too high to allow the realization of integrated circuits with about 10^{11} devices per cm^2, high electric fields will cause device breakdown, and the shrinkage of depletion regions causes leakage. As a result, nanoscale devices become an attractive option for the development of integrated circuits with dimensions and performances beyond the ultimate roadmap. Among them, SET presents promising features such as reduced dimension, extremely low power consumption, and high charge sensitivity.

Many measurements have been made studying charge transport in the SETs. As we have discussed, by adding electrons to the island of an SET transistor, one can study the interaction of electrons on the island, the coupling of states in the leads to states in the island, and how superconductivity changes the system behavior. The remarkable capability to control the current allows SETs to be used in metrological applications such as ultrasensitive electrometer and current standard. The low power dissipation makes them potentially useful in high-density memory and logic circuits. SETs are also very sensitive to applied radiation. Photon-assisted tunneling has been observed, and the absorption of individual photons of microwave radiation can be detected. Furthermore, SETs have potential applications in the magnetic recording industry and developing new spintronics devices. In this section, we will discuss some possible applications of SETs.

4.1. Metrological Applications

4.1.1. Precision Measurements

Because of their charge sensitivity, SETs are excellent for making precision charge measurements. SET can be used to measure charge either in the normal state or in the superconducting state. Typically, it is voltage biased at a point where there is a large modulation of the current as a function of the gate charge Q_o. The charge that is to be measured is coupled to the gate, and the current modulated by the gate is monitored. Thus, charges much smaller than the electron charge e can be measured. The charge resolution that can be achieved is $8 \times 10^{-6} e/\sqrt{Hz}$ at 10 Hz [44]. SETs offer by far the best charge resolution among the available charge measurement devices.

SETs have also been capacitively coupled to a variety of systems so that the charge motion of those systems could

be monitored. Metallic SETs have been coupled to semiconductor quantum dots to observe the charge fluctuations in the quantum dot [45]. They have been coupled to superconducting particles where it is possible to observe whether the particle has an even number or an odd number of electrons on it [46]. SETs have also been scanned over semiconductors to measure fluctuations in the dopant distribution [47].

4.1.2. Current Standard

One of the applications of SET devices is a fundamental current standard. In such a device, a known current is established by transferring individual electrons through the device with a frequency f. This results in a current $I = ef$. A number of different designs have been made. They include modulating the gates coupled to the islands in an array of tunnel junctions, [48, 49], modulating the tunnel barriers in a semiconducting quantum dot [50], and transferring Cooper pairs in a superconducting circuit [51]. The most intensively studied current standard is called an electron pump. It consists of a number of tunnel junctions in series with a gate connected to each island between the junctions. By modulating the gates successively, one can draw a single electron through the array of tunnel junctions. The accuracy that has been achieved with this current standard is 15 parts per billion [52].

4.2. Information Technology

4.2.1. SET Memories

The small size and low power consumption of SETs make them a promising candidate for the information technology industry. In comparison with single-electron logic, single-electron memory is much easier for implementation, and therefore more promising from the practical point of view. All major problems of single-electron logic (operation temperature, background charge fluctuations, and power dissipation.) can be solved much easier for single-electron memories [53]. Here, we will discuss two memory schemes. The first is an offset-charge-independent DRAM cell which was described by Likharev and Korotkov [54]. In their circuit, a bit is represented by the presence or absence of the charge of a few electrons which are stored on an island. The charge on the island is monitored by an SET. When the memory cell is read, if there are charges on the island, the current through the SET undergoes oscillations as each electron tunnels off the island. The current oscillations through the device occur for any value of the offset charge. If there are no charges stored on the island, there will be no current oscillations.

Another SET memory type, called a single-electron MOS memory (SEMM), is also based on the motion of individual electrons [55]. This device is very similar to a conventional floating-gate MOS memory, which has a floating gate between a channel and a control gate. Information is represented by storing charges on the floating gate, which modulates the conduction through the channel. The ultimate limit is to use a single electron to represent a bit.

In the silicon SEMM design by Chou's group [56], there are two key features: (1) the silicon channel width (~10 nm) is narrower than the Debye screening length of a single electron, and (2) the floating gate is a nanoscale polysilicon dot (~7 nm × 7 nm) embedded between the control gate and the channel. The narrow channel ensures that the storage of one electron on the floating gate is sufficient to screen the entire channel width from the potential on the control gate. The tiny floating gate significantly increases the electron quantum energy and electron charging energy, so that the threshold voltage shift and the charging voltage become discrete and well separated at room temperature. When a single electron is added to the floating gate, the conduction through the channel significantly changes.

In the fabrication of this Si SEMM [66], a polysilicon film 11 nm thick was deposited on an Si wafer that had a 35 nm thick top layer of crystalline Si. The polysilicon film and the Si layer were separated by a layer of 1 nm thick native oxide. The first procedure of e-beam lithography and chlorine-based reactive ion etching (RIE) patterned the width of the floating gate and the narrow Si channel. The initial channel width ranges from 25 to 120 nm. Then, a second procedure of e-beam lithography and RIE patterned the length of the floating gate. An 18 nm thick layer of oxide was then thermally grown, reducing the thickness of the polysilicon dot by ~9 nm, and the lateral size of the dot and the width of the Si channel by 18 nm. The polysilicon dot functions as the floating gate. Then a thick layer of oxide was deposited by plasma-enhanced chemical vapor deposition. Polysilicon was then deposited, and the control gate was patterned to a length of 3 mm, which covered the floating gate and part of the narrow channel. In order to allow fast charging and to minimize the potential difference between the channel and the floating gate during the charging process, no tunnel oxide was intentionally grown between the channel and polysilicon floating gate. However, a potential barrier still exists between the channel and the floating gate because of the grain boundary in polysilicon and the thin native oxide. The devices were characterized at room temperature. With the drain voltage fixed at 50 mV, a positive voltage pulse was first applied to the control gate, causing the electrons to tunnel from the channel to the floating gate. The drain current was then measured as a function of the gate voltage. It was found that the charge stored at the floating gate could be held for ~5 s after the control gate potential was set back to the ground.

4.2.2. SET Logic Circuits

In the application of logic circuits, quite a number of logic designs based on SET circuits have been proposed. Some of the designs are very similar to CMOS, where bits are represented by voltage levels [57, 58]. Some logic designs resemble superconducting single-flux quantum logic [8]. In this case, bits are represented by the presence or the absence of individual electrons. Other logic designs contain elements that act like electron pumps for transferring charges [59].

Hadley's group at the Delft University of Technology fabricated a CMOS-like SET inverter using two capacitively coupled single-electron transistors [60]. This inverter has a voltage gain of 2.6 at 25 mK, and remained larger than 1 for temperatures up to 140 mK. Inverters can be used as a fundamental building block of SET logic circuits and memory

elements. NAND and NOR gates are realized by making slight variations of the inverter. With two inverters, a static RAM memory cell can also be constructed.

4.3. Microwave Detector

It has been shown that, when the island is superconducting, the effects of photon-assisted tunneling (PAT) in the SET greatly enhances the rates of some energetically forbidden tunneling processes. The most dramatic effect, the emerging of a secondary peark near $Q_o = 0$, is directly linked to superconductivity in the island. It results when PAT introduces a single quasiparticle to the island, a process which is otherwise forbidden at low voltage and temperature. This switches on the two-electron tunneling process, which continues until the single quasiparticle escapes from the island after a relatively long time, $\sim$1 μs. Consequently, the effect of PAT is magnified because many electrons tunnel through the system for each absorbed photon. It has been demonstrated that, in an all-superconducting SET sample, approximately 100 electrons can tunnel through the system per absorbed photon, giving rise to a secondary peak up to $\sim$1 nA [43]. In the presence of very small amounts of microwave radiation, the secondary peak provides a direct measure of the absorbed power. Therefore, the superconducting SET can be used as an extremely sensitive microwave detector.

4.4. Magnetoelectronics

There has been much interest in the past few years to study the spin transport and dynamics in ferromganetic SETs with ferromagnetic electrodes. [61–63]. Here, let us consider a double-junction system where the two electrodes are ferromagnets and the island is a normal metal. Such a tunnel junction can be used in ferromagnetic alignment (the magnetizations of the two electrodes are parallel) and in antiferromagnetic alignment (the magnetizations of the two electrodes are antiparallel, as shown in Fig. 18). Already for relatively large junctions (where the Coulomb blockade can be neglected), the anti-ferromagnetic alignment is interesting because a tunnel current generates a spin accumulation on the island. For nanoscale double-junction systems with ferromagnetic leads, it has been calculated that spin accumulation caused by cotunneling squeezes the Coulomb blockade in the antiferromagnetic configuration. And the tunnel magnetoresistance (TMR) is strongly enhanced by the cotunneling process in the Coulomb blockade region [64–66] in which the sequential tunneling is suppressed.

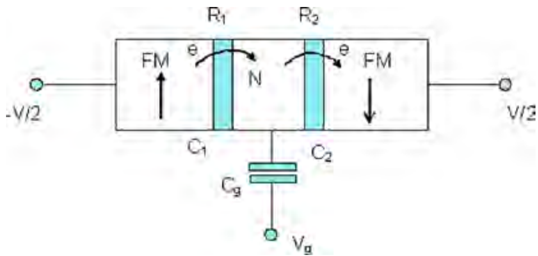


Figure 18. A ferromagnetic SET in the antiferromagnetic configuration. Insulating barriers are shown by the shaded areas.

As discussed earlier, cotunneling is a higher order process of tunneling through both junctions via a virtual intermediate state with increased electrostatic energy. The cotunneling process is energetically favorable since the cotunneling lowers the electrostatic energy by eV. One can calculate the TMR using the perturbation theory, and obtain the TMR ratio in the case of a symmetric junction:

$$\frac{R_A - R_F}{R_F} = \frac{2P^2}{1 - P^2} \qquad (k_B T \gg E_c) \tag{78}$$

$$\frac{R_A - R_F}{R_F} = \frac{4P^2}{(1 - P^2)^2} \qquad (k_B T \ll E_c) \tag{79}$$

where R_A and R_F are the resistances when the electrodes are in antiparallel and parallel alignments, respectively, and P is the polarization of the magnetic electrodes. Therefore, the TMR for $k_B T \ll E_c$ in the cotunneling regime is enhanced by $2/(1 - P^2)$ compared with TMR for $k_B T \gg E_c$ in the absence of the Coulomb blockade effect [67].

Extending the result to a superconducting island, it is found that the spin accumulation signal is strongly enhanced since a superconductor is a low carrier system for spin transport, but not for charge [68]. These aspects lead SET into new spin electronic devices useful for the magnetic recording industry as read head sensors and magnetic memory elements.

5. SUMMARY

We have discussed the charge-transport mechanisms in single-electron transistors, with primary concentration on the normal state SETs and superconducting state SETs. We have shown that superconductivity in the island leads to very interesting even–odd electron number effects, even if there are as many as one billion conduction electrons in the island. Single-charge-tunneling effects play an important role in devices with very small dimensions as device miniaturization continues for dense integrated circuits. Because the phenomena of single electronics that most groups have been studying manifest themselves at low temperatures, the main drawback of SET for applications is the necessity for low-temperature operation. However, the demonstration of a silicon-based SET operating at room temperature, the recent demonstration of carbon nanotube- and nanowire-based logic circuits operating at room temperature, and ongoing advances in nanofabrication to make even smaller feature sizes indicate that SET circuits can operate at room temperatures. Therefore, it is highly promising for single-electron devices to realize their practical industrial applications.

GLOSSARY

Coulomb blockade Electron transfer is energetically forbidden for voltages below a threshold voltage, showing a zero tunnel current region in the I–V characteristics.

Coulomb charging energy Energy required for adding a single electron onto an island.

Even–odd free energy difference Difference in free energy between two parity cases.

Inelastic cotunneling In the Coulomb blockade region, where sequential tunneling of an electron is energetically unfavorable, electrons can tunnel simultaneously across two junctions via a virtual intermediate state, leaving an electron-hole excitation on the island.

Island A region (for example a piece of metal, superconductor, or semiconductor), that is weakly coupled to its surrouding high resistance tunnel junctions. Charges can be transferred only through tunnel junctions.

Orthodox theory Formalism to calculate the current through the single electron transistor by determining the island charge distribution function and by considering the transitions across a single junction for any possible island charge.

Parity effect In a superconductor island, due to electron pairing, the ground state energy of the system is higher when the total number of electrons on the island, N is odd compared to when N is even. As a result, the states of even N become energetically favorable over those of odd N. This effect is also called even–odd electron number effect.

Single-electron charging effects When the capacitance of the island to the surrounding is very small, the charging energy required to add a single electron becomes the dominant energy, and current transport is significantly influenced by the discreteness of the electronic charge.

Single-electron transistor A double tunnel junction system in which the island is weakly coupled to the bias leads through small capacitance and high resistance tunnel junctions, and capacitively coupled to a gate electrode.

Total capacitance of the island Sum of the capacitances of the island to the three electrodes in a double junction system.

REFERENCES

1. Review by G. Bergmann, *Phys. Rep.* 107, 1 (1984).
2. S. Washburn and R. A. Webb, *Adv. Phys.* 35, 375 (1986).
3. W. F. Smith, T. S. Tighe, G. C. Spalding, M. Tinkham, and C. J. Lobb, *Phys. Rev. B* 43, 12267 (1991).
4. R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, *Phys. Rev. Lett.* 54, 2696 (1985).
5. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, *Phys. Rev. Lett.* 60, 848 (1988).
6. P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir, N. S. Wingreen, and S. J. Wind, *Phys. Rev. Lett.* 66, 1926 (1991).
7. D. C. Ralph, C. T. Black, and M. Tinkham, *Phys. Rev. Lett.* 74, 3241 (1995).
8. K. K. Likharev and A. B. Zorin, *J. Low Temp. Phys.* 59, 347 (1985).
9. D. V. Averin and K. K. Likharev, *J. Low Temp. Phys.* 62, 345 (1986).
10. D. V. Averin and K. K. Likharev, in "Mesoscopic Phenomena in Solids" (B. L. Altshuler, P. A. Lee, and R. A. Webb, Eds.). Elsevier, Amsterdam, 1991.
11. G.-L. Ingold and Yu.V. Nazarov, in "Single Charge Tunneling" (M. H. Devoret and H. Grabert, Eds.). Plenum, New York, 1992.
12. G. Schön, in "Quantum Transport and Dissipation" (T. Dittrich, P. Hänggie, G.-L. Ingold, B. Kramer, G. Schön, and W. Zwerger, Eds.). Wiley-VCH, 1997.
13. M. H. Cohen, L. M. Falicov, and J. C. Phillips, *Phys. Rev. Lett.* 8, 316 (1962).
14. E. Merzbacher, "Quantum Mechanics," John Wiley & Sons, New York, 1970.
15. M. Tinkham, "Introduction to Superconductivity," 2nd ed. McGraw-Hill, New York, 1995.
16. D. V. Averin and A. A. Odintsov, *Phys. Lett. A* 140, 251 (1989).
17. D. V. Averin and Yu.V. Nazarov, in "Single Charge Tunneling" (M. H. Devoret and H. Grabert, Eds.). Plenum, New York, 1992.
18. D. V. Averin and Yu.V. Nazarov, *Phys. Rev. Lett.* 65, 2446 (1990).
19. L. J. Geerligs, D. V. Averin, and J. E. Mooij, *Phys. Rev. Lett.* 65, 3037 (1990).
20. A. F. Andreev, *Sov. Phys. JETP* 19, 1228 (1964).
21. F. W. J. Hekking, L. I. Glazman, K. A. Matveev, and R. I. Shekhter, *Phys. Rev. Lett.* 70, 4138 (1993).
22. C. W. J. Beenakker, *Phys. Rev. B* 46, 12841 (1992).
23. F. W. J. Hekking and Yu.V. Nazarov, *Phys. Rev. B* 49, 6847 (1994).
24. B. J. van Wees, P. de Vries, P. Magnee, and T. M. Klapwijk, *Phys. Rev. Lett.* 69, 510 (1992).
25. A. D. Zaikin, *Physica B* 203, 255 (1994).
26. J. G. Lu, J. M. Hergenrother, and M. Tinkham, *Phys. Rev. B* 53, 3543 (1996).
27. T. A. Fulton, P. L. Gammel, D. J. Bishop, and L. N. Dunkleberger, *Phys. Rev. Lett.* 63, 1307 (1989).
28. A. Maassen van den Brink, G. Schön, and L. J. Geerligs, *Phys. Rev. Lett.* 67, 3030 (1991).
29. M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, *IEEE Trans. Appl. Supercond.* 3, 1972 (1993).
30. P. J. Siemens and A. S. Jensen, "Elements of Nuclei," Chap. 6. Addison-Wesley, 1987.
31. D. H. Cobden, M. Bockrath, P. L. McEuen, A. G. Rinzler and R. E. Smalley, *Phys. Rev. Lett.* 81, 681 (1998).
32. W. P. Halperin, *Rev. Mod. Phys.* 58, 533 (1986).
33. D. C. Ralph, C. T. Black, J. M. Hergenrother, J. G. Lu, and M. Tinkham, in "Mesoscopic Electron Transport" (L. L. Sohn, L. P. Kouwenhoven, and G. Schön, Eds.). Kluwer Academic, Dordrecht, 1997.
34. D. V. Averin and Yu. V. Nazarov, *Phys. Rev. Lett.* 69, 1993 (1992).
35. M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, *Phys. Rev. Lett.* 69, 1997 (1992).
36. M. Tinkham, J. M. Hergenrother, and J. G. Lu, *Phys. Rev. B* 51, 12649 (1995).
37. G. Schön and A. D. Zaikin, *Europhys. Lett.* 26, 695 (1994).
38. J. M. Hergenrother, M. T. Tuominen, J. G. Lu, D. C. Ralph, and M. Tinkham, *Physica B* 203, 327 (1994).
39. M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, *Phys. Rev. B* 47, 11599 (1993).
40. G. J. Dolan, *Appl. Phys. Lett.* 31, 337 (1997).
41. G. Johansson, P. Delsing, K. Bladh, D. Gunnarsson, T. Duty, A. Kack, G. Wendin, and A. Aassime, arXiv:cond-mat/0210163, submitted to "Proceedings of NATO ARW, Quantum Noise in Mesoscopic Physics," Oct. 8, 2002.
42. M. Schulz, *Nature* 399, 729 (1999).
43. J. M. Hergenrother, J. G. Lu, M. T. Tuominen, D. C. Ralph, and M. Tinkham, *Phys. Rev. B* 51, 9407 (1995).
44. P. Hadley, Lectures on superconductivity in networks and mesoscopic systems, (C. Giovannella and C. J. Lambert, Eds.), AIP conference proceedings 427, Woodbury, New York (1998); V. A. Krupenin, D. E. Presnov, A. B. Zorin, and J. Niemeyer, *J. Low Temp. Phys.* 118, 287 (2000).
45. D. Berman, N. B. Zhitenev, R. C. Ashoori, H. I. Smith, and M. R. Melloch, *J. Vac. Sci. Technol. B* 15, 2844 (1997).
46. C. T. Black, D. C. Ralph, and M .Tinkham, *Phys. Rev. Lett.* 76, 688 (1996).
47. M. J. Yoo, T. A. Fulton, J. F. Hess, R. L. Willett, L. N. Dunkleberger, R. J. Chichester, L. N. Pfeiffer, and K. W. West, *Science* 276, 579 (1997).

48. L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 64, 2691 (1990).
49. H. Pothier, P. Lafarge, C. Urbina, D. Esteve, and M. H. Devoret, *Europhys. Lett.* 17, 249 (1992).
50. L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, and C. J. P. M. Harmans, *Phys. Rev. Lett.* 67, 1626 (1991).
51. L. J. Geerligs, S. M. Verbrugh, P. Hadley, J. E. Mooij, H. Pothier, P. Lafarge, C. Urbina, D. Esteve, and M. H. Devoret, *Z. Phys. B* 85, 349 (1991).
52. M. W. Keller, J. M. Martinis, N. M. Zimmerman, and A. H. Steinbach, *Appl. Phys. Lett.* 69, 1804 (1996).
53. A. N. Korotkov, *J. Appl. Phys.* 92, 7291 (2002).
54. K. K. Likharev and A. N. Korotkov, *VLSI Design* 6, 341 (1998).
55. A. N. Korotkov, Coulomb blockade and digital single-electron devices, in "Molecular Electronics" (J. Jortner and M. A. Ratner, Eds.), pp. 157–189. Blackwell, Oxford, 1997; A. N. Korotkov, *Int. J. Electron.* 86, 511 (1999).
56. L. Guo, E. Leobandung, and S. Y. Chou, *Science* 275, 649 (1997).
57. N. Yoshikawa, Y. Jinguu, J. Ishibashi, and M. Sugahara, *Jpn. J. Appl. Phys.* 35, 1140 (1996).
58. A. N. Korotkov, R. H. Chen, and K. K. Likharev, *J. Appl. Phys.* 78, 2520 (1995).
59. M. G. Ancona, *J. Appl. Phys.* 79, 526 (1996).
60. C. P. Heij, P. Hadley, and J. E. Mooij, *Appl. Phys. Lett.* 78, 1140 (2001).
61. Y. Ootuka, R. Matsuda, K. Ono, and J. Shimada, *Physica B* 280, 394 (2000).
62. J. Barnaś, J. Martinek, G. Machalek, B. R. Bulka, and A. Fert, *Phys. Rev. B* 62, 12363 (2000).
63. H. Imamura, Y. Utsumi, and H. Ebisawa, *Phys. Rev. B* 66, 054503 (2002).
64. H. Imamura, S. Takahashi, and S. Maekawa, *Phys. Rev. B* 59, 6017 (1999).
65. A. Brataas, Yu.V. Nazarov, J. Inoue, and G. E. W. Bauer, *Phys. Rev. B* 59, 93 (1999).
66. J. Martinek, J. Barnaś, S. Maekawa, H. Schoeller, and G. Schön, *Phys. Rev. B* 66, 014402 (2002).
67. S. Maekawa, Theory of tunnel magnetoresistance, in "Spin Dependent Transport in Magnetic Nanostructures" (S. Maekawa and T. Shinjo, Eds.). Taylor & Francis, London, 2002.
68. S. Takahashi, H. Imamura, and S. Maekawa, *Phys. Rev. Lett.* 82, 3911 (1999).

www.aspbs.com/enn

Single Wall Carbon Nanotubes

Ákos Kukovecz
University of Szeged, Szeged, Hungary and University of Vienna, Vienna, Austria

Zoltán Kónya, Imre Kiricsi
University of Szeged, Szeged, Hungary

CONTENTS

1. Introduction
2. Synthesis of Single Wall Carbon Nanotubes
3. Functionalization of Single Wall Carbon Nanotubes
4. Characterization of Single Wall Carbon Nanotubes
5. Properties of Single Wall Carbon Nanotubes
6. Applications of Single Wall Carbon Nanotubes
7. Conclusions
Glossary
References

1. INTRODUCTION

Single wall carbon nanotubes (SWCNTs) are very thin (0.7–2 nm in diameter) and very long (lengths exceeding 10 μm are not uncommon) hollow tubules made up of a covalently bonded network of sp^2 carbon atoms. The main difference between SWCNTs and carbon fibers is that in principle one SWCNT is one giant molecule while fibers consist of entities significantly smaller than the fiber itself, held together by secondary bonding forces. Their exceptionally high aspect ratio makes SWCNTs the perfect one-dimensional model compound and results in many special physical properties. Besides being fascinating materials in their own right, there is considerable interest in nanotubes in the plastic, automotive, and nanoelectronics industries.

Advances in carbon nanotubes were reviewed by several authors since their discovery in 1991 [1–12]. In this chapter our goal is to give an overview of the SWCNT field for nonspecialists and to summarize the new results of the last few years for specialists.

1.1. Single Wall Carbon Nanotubes Timeline

Carbon soot containing a host of pyrolysis products has been around ever since mankind started using fire. Industrial-scale interest in artificial carbon forms was raised at the end of the 19th century by paint, adsorbent, and especially rubber manufacturers and was boosted by the spreading of cars running on black tires. Carbon fibers have been researched intensively since the 1950s. In the 1970s and 1980s vapor grown fibers with diameters below 10 nm were grown catalytically [13, 14] and in 1985 fullerenes were discovered [15]. Even though it is rather likely that carbon nanotubes were also produced in these experiments, the nanotube field started only in 1991 with Iijima's discovery of multiwall carbon nanotubes in the product soot of a dc arc discharge experiment utilizing graphite electrodes [16].

SWCNTs were discovered in 1993 simultaneously by Iijima and Ichihashi [17] and Bethune et al. [18]. SWCNT research increased in the second half of the 1990s as substantial amounts of nanotubes became available thanks to the optimization of the dc arc discharge by Journet et al. [19] and to development of the laser ablation process by Thess and co-workers [20]. Later, the chemical vapor deposition (CVD) SWCNT synthesis method was developed [21–24] and the first method offering feasible up-scaling of production was described [25, 26].

The end of the 1990s also brought the functionalization of carbon nanotubes [27–29], nanotube field effect transistors [30], and nanotube based chemical force microscopy tips [31] as well as nanotweezers [32]. Focal points of contemporary research are (i) achieving full control over the diameter and purity in nanotube synthesis, (ii) utilizing nanotubes in nanoelectronic devices like transistors, sensors, and actuators, (iii) bringing nanotube-based field emission devices (e.g., flat panel displays, lamps) and composites to the commercial market, and (iv) using nanotubes in the energy sector to improve batteries, fuel cells, and H_2 storage devices.

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

Encyclopedia of Nanoscience and Nanotechnology
Edited by H. S. Nalwa
Volume 9: Pages (923–946)

1.2. Single Wall Carbon Nanotubes Structure

This chapter introduces the basic concepts and vocabulary necessary for newcomers to understand the language of current SWCNT research papers. The formalism used for the presentation of the material here is adopted from the work of Saito et al. [2] and Dresselhaus and Eklund [33].

1.2.1. Geometry

Graphite is a three-dimensional (3D) material consisting of several 2D layers (graphene sheets) of hexagons of sp^2 carbon atoms arranged in a honeycomb lattice. SWCNTs can be derived by taking a long, thin strip of hexagons out of one graphene sheet and rolling it up to make the opposite long edges meet. As-synthesized tubes are always capped by fullerene hemispheres on both ends which can be opened later by chemical reactions. The present lower limit for SWCNT diameter is 0.4 nm for tubes confined into the channels of a zeolite host and 0.7 nm for freestanding tubes [25]. Although there is no theoretical upper limit for tube diameter or length, most SWCNT samples produced today fall into the range of $d =$ 0.9–2.5 nm and $l =$ 0.2–10 μm.

The beauty and the complexity of nanotube science both originate from the same ground: that one has nearly complete freedom in choosing the strip width and cut direction when slicing the graphene sheet. Each choice results in a nanotube with a different orientation of hexagons in the cylindrical wall (different *helicity*) and, consequently, with different electronic properties. It is therefore mandatory to understand Figure 1, which summarizes the geometrical quantities characterizing a nanotube. By connecting the crystallographically equivalent points O and A this strip can be rolled up into a (4, 2) SWCNT. Relocating point A to A* or A** would result in different armchair (4, 4) and zigzag (4, 0) nanotubes, respectively. Each tube is unambiguously identified by its *chiral vector* $\mathbf{C_h}$ (also called the *Hamada* vector [34]) which is perpendicular to the tube axis and is connecting O and A. $\mathbf{C_h}$ can be expressed as the sum of integer multiples of the real space unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of the unrolled honeycomb sheet,

$$\vec{a}_1 = \left(\frac{\sqrt{3}}{2}a, \frac{a}{2}\right) \qquad \vec{a}_2 = \left(\frac{\sqrt{3}}{2}a, -\frac{a}{2}\right) \tag{1}$$

where $a = 1.44$ Å $\times\sqrt{3} = 2.49$ Å is the lattice constant of the unrolled honeycomb lattice [35]:

$$C_h = n\vec{a}_1 + m\vec{a}_2 \equiv (n, m) \quad (n, m \in \mathbf{Z} \quad \text{and} \quad 0 \leq m \leq n) \tag{2}$$

Tubes possessing identical mirror images are called *achiral*; these are the so-called *armchair* ($n = m$) and *zigzag* ($m = 0$) SWCNTs while the general (n, m) tube is always *chiral*. The *diameter* d of a SWCNT is given by

$$d = \frac{a}{\pi}\sqrt{n^2 + m^2 + nm} \tag{3}$$

The *chiral angle* Θ denotes the tilt angle of the hexagons with respect to the direction of the SWCNT axis and can be expressed from

$$\cos\Theta = \frac{2n + m}{2\sqrt{n^2 + m^2 + nm}} \tag{4}$$

The *translational vector* $\mathbf{T}$ points from O to B which is the first lattice point through which a line normal to $\mathbf{C_h}$ passes from O:

$$\vec{T} = t_1\vec{a}_1 + t_2\vec{a}_2 \equiv (t_1, t_2) \qquad (t_1, t_2 \in \mathbf{Z}) \tag{5}$$

The *unit cell* of the nanotube is the OAB'B rectangle. Let d_H denote the highest common divisor of n and m and let us define d_R as

$$d_R = \begin{cases} d_H & \text{if } (n - m) \text{ is not a multiple of } 3d_H \\ 3d_H & \text{if } (n - m) \text{ is a multiple of } 3d_H \end{cases} \tag{6}$$

The N number of hexagons in the unit cell is then obtained as

$$N = \frac{2(m^2 + n^2 + nm)}{d_R} \tag{7}$$

The unit cell contains $2N$ carbon atoms.

The Brillouin zone (the reciprocal space counterpart of the unit cell) of 2D graphite is a regular hexagon. There are three highly symmetric points in the zone: the $\Gamma = (0, 0)$ point in the center, the $\mathbf{K} = (0, \frac{4\pi}{3a})$ point in the corner, and the $\mathbf{M} = (\frac{2\pi}{\sqrt{3}a}, 0)$ at the center of the edge. The reciprocal unit vectors are defined as

$$\vec{b}_1 = \left(\frac{1}{\sqrt{3}}, 1\right)\frac{2\pi}{a} \qquad \vec{b}_2 = \left(\frac{1}{\sqrt{3}}, -1\right)\frac{2\pi}{a} \tag{8}$$

where a is the lattice constant of the graphene sheet. The Brillouin zone of SWCNTs is derived from that of 2D

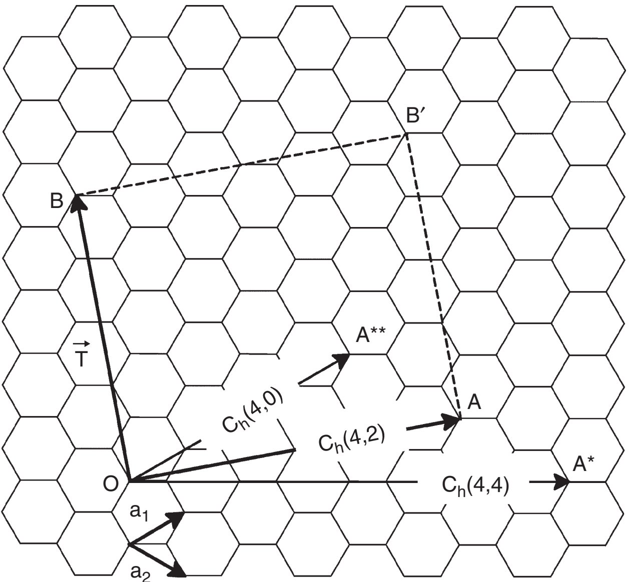


Figure 1. A graphene sheet can be folded into a SWCNT by connecting points O and point A. The resulting (4, 2) tube will have the translation vector **T** and unit cell OAB′B. Relocating point A to A* or A** yields the armchair (4, 4) and zigzag (4, 0) SWCNTs, respectively. The real space unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are also shown.

graphite by calculating the nanotube reciprocal lattice vectors $\mathbf{K}_1$ and $\mathbf{K}_2$ corresponding to $\mathbf{C_h}$ and $\mathbf{T}$, respectively as

$$\vec{K}_1 = \frac{1}{N}(-t_2\vec{b}_1 + t_1\vec{b}_2) \qquad \vec{K}_2 = \frac{1}{N}(m\vec{b}_1 - n\vec{b}_2) \quad (9)$$

Since the single wall carbon nanotube is a one-dimensional system, only $\mathbf{K}_2$ is a true reciprocal unit vector. The Brillouin zone is a $|\vec{K}_2|$ long line segment passing through the Γ point parallel with $\mathbf{K}_2$. $\mathbf{K}_1$ gives discrete values for the wave vector $\mathbf{k}$ in the direction of the Hamada vector.

The symmetry of SWCNTs is determined by the geometry of the unit cell. The achiral (n, n) armchair and $(n, 0)$ zigzag nanotubes possess one inversion center (i) and one n-fold (C_n) axis and one $n \times C_2$ rotational axis in the directions parallel and perpendicular to the nanotube axis, respectively. The $n = 2k$ $(k \in \mathbf{Z})$ achiral tubes have an additional horizontal mirror plane (σ_h) and thus belong into the point group D_{nh} while the $n = 2k + 1$ $(k \in \mathbf{Z})$ tubes belong into D_{nd}. The basic symmetry operation of chiral (n, m) SWCNTs is a rotation around the tube axis followed by a translation along the axis. This operation corresponds to the $\mathbf{R} = p\mathbf{a}_1 + q\mathbf{a}_2$ vector where p and q are calculated from

$$d_H = mp - nq \quad \text{with the conditions } p < \frac{n}{d_H} \quad q < \frac{m}{d_H} \quad (10)$$

Let us define Ω as

$$\Omega = \frac{p(m+2n) + q(n+2m)}{d_H/d_R} \quad (11)$$

The point group of the chiral (n, m) tube is $C_{N/\Omega}$ if $d_H = 1$ and is obtained as a direct product $C_{d_H} \otimes C'_{N/\Omega}$ if $d_H \neq 1$.

SWCNTs are seldom found alone; rather, they stack together and form bundles which may contain from a few up to a few hundred tubes. Thess et al. provided experimental [transmission electron microscopy (TEM) and X-ray diffraction (XRD)] evidence [20] that the bundles are quasi-crystalline arrays of nanotubes arranged into a trigonal lattice of space group *P6/mcc* [36]. It is yet uncertain if the arrays are made up of all identical (n, m) tubes or of SWCNTs with similar diameters but different helicities. The number of tubes in a bundle can be increased by functionalization [37] or decreased by ultrasonication in aqueous ionic [38, 39] or nonionic [40] surfactant solutions. However, it is quite difficult to separate a macroscopic SWCNT sample into truly individual tubes. Therefore, most of the macroscopic nanotube-related studies reported so far in the literature refer to work on SWCNT bundles even if not stated explicitly.

1.2.2. Electronic Structure

The electronic structure of SWCNTs is obtained simply by applying periodic boundary conditions in the $\mathbf{C_h}$ direction to 2D graphite. Thus the wave vector perpendicular to the nanotube axis becomes quantized while the parallel one remains continuous [41]. Following Saito et al., for a SWCNT with N hexagons in the unit cell we can think of the quantization as making N cross-sections in the energy dispersion surface of 2D graphite, resulting in N pairs of one-dimensional energy dispersion curves of the nanotube:

$$E^{\pm}_{\mu}(k) = E^{\pm}_{g2D}\left(k\frac{\vec{K}_2}{|\vec{K}_2|} + \mu\vec{K}_1\right) \qquad \mu = 0, \ldots, N-1$$
$$\text{and} \qquad -\frac{\pi}{|\vec{T}|} < k < \frac{\pi}{|\vec{T}|} \quad (12)$$

Here, $k \equiv |\vec{k}|$ is the magnitude of the wave vector, $E^{\pm}_{g2D}(k)$ denotes the energy dispersion relation of 2D graphite, and the "+" and "−" indices correspond to the π and π^* bands, respectively. The cross-sections are made along $\mathbf{K}_1$ spaced, $|\vec{K}_2|$ long line segments in the hexagonal Brillouin zone. If for a particular nanotube a segment happens to go through a $\mathbf{K}$ point where the π and π^* bands of graphene are degenerate, than there will be a zero gap between the valence (+) and conduction (−) branches of the corresponding one-dimensional energy dispersion curve. Since the degenerate point matches the Fermi energy where in such cases the electronic density of states has a nonzero value, these SWCNTs are *metallic*. On the other hand, if none of the N segments passes through a $\mathbf{K}$ point then the corresponding SWCNT will possess a finite gap, its electronic density of states will be zero at the Fermi energy, and therefore the tube will be *semiconducting*. It can be shown that tubes with $(n - m) = 3k$, $k \in \mathbf{Z}$ are the metallic ones and all others (the remaining 2/3 of all possible helicities) are semiconducting. It is useful to remember that armchair SWCNTs are always metallic and zigzags are metallic only if n is a multiple of 3.

The electronic density of states (DOS) of SWCNTs is calculated as

$$D(E) = \frac{|\vec{T}|}{2\pi N}\sum_{\pm}\sum_{\mu=0}^{N-1}\int \frac{1}{\left|\frac{dE^{\pm}_{\mu}(k)}{dk}\right|}\delta(E^{\pm}_{\mu}(k) - E)\,dE$$
$$[\text{States/C atom/eV}] \quad (13)$$

It is clear from this equation that very high $D(E)$ values can be expected close to the minima and maxima of the 1D energy dispersion curves as defined in Eq. (12). These peaks in the DOS are called *van Hove singularities* (vHs) and they play a crucial role in several SWCNT spectroscopies. The singularities are indexed as $i = 1, 2, \ldots$ from the Fermi energy to both directions (valence vs conduction band) and optical transitions between the ith valence and jth conductance band vHs are denoted as E_{ij}. The energy separation between the highest valence band vHs and the lowest conduction band vHs (the gap size of semiconducting SWCNTs) is independent of the chiral angle as shown in the formulae

$$E^{m}_{11} = \frac{6a_{\mathrm{C-C}}\gamma_0}{d} \qquad \text{and} \qquad E^{sc}_{11} = \frac{2a_{\mathrm{C-C}}\gamma_0}{d} \quad (14)$$

for metallic and semiconducting SWCNTs, respectively. In Eq. (14) $a_{\mathrm{C-C}} = a/\sqrt{3}$ is the nearest neighbor C–C distance in 2D graphite, $\gamma_0 \cong 2.5$ eV is the C–C transfer energy, and d is the diameter of the nanotube. These equations are valid strictly only within the so-called "linear k approximation" [42] which fails for small SWCNT diameters. In thin tubes the equienergy contour lines of the energy dispersion surface of 2D graphite are distorted from circles into trigonal shapes

for large k values. This phenomenon is called the *"trigonal warping effect"* [43] and is responsible for the splitting of DOS peaks in zigzag and in metallic chiral nanotubes. Armchair and semiconducting chiral nanotubes do not suffer any splitting due to the warping effect.

2. SYNTHESIS OF SINGLE WALL CARBON NANOTUBES

With nanotube research being a very young science, the issue of nanotube synthesis has not yet settled down so well as, for example, in the case of the six years older fullerenes. Authors of all research papers are still expected to provide details about their nanotube source and purification procedure because these have a considerable effect on the properties of the material. Several extensive reviews are available on the topic of SWCNT synthesis [44–46].

2.1. Synthesis Using Arc discharge

The temperature in an electric discharge may reach as high as 4000 °C. Under its operation material transfer takes place between the electrodes. In 1990, successful fullerene [47] and in 1991 successful carbon nanotube syntheses [16] were performed using this method.

Lozovik and Popov [48] divided the operation mode of electric discharge systems into noisy and quiet ones. Continuous transition between the two modes can be regulated simply by changing the current density. It was found that high current density and low inert gas pressure result in the formation of fullerenes, while for low current density and high pressure nanoparticles and nanotubes are predominately formed. Ebbessen and Ajayan [49] developed a technique suitable for large scale production. He gas was run through the reactor with controlled rate and pressure. Generally the anode graphite rod contained a hole filled with the metal catalyst. The product formed on the cathode was composed of various matter including nanotubes, fullerenes, carbonaceous nanoparticles, and catalyst. After purification only 1% of the initial deposit remained as final product. Journet et al. [19] reported a novel technology leading to yields of nanotube production as 70–90%. Bundles of a few tens of tubes and only a few isolated tubes were formed. Bethune et al. [18] reported that co-vaporizing carbon and Co under He atmosphere resulted in the formation of SWCNTs of very small diameter (around 1.2 nm). They proved that no nanotube formation was found when using Fe, Ni, or a mixture of Ni–Cu. Seraphin and Zhou [50] have reported an efficient synthesis of SWCNTs in the presence of bimetallic catalysts such as Fe–Ni, Fe–Co in argon atmosphere. They obtained gram quantities of SWCNTs. Iijima and Ichihashi [17] prepared SWCNT by co-vaporization of graphite and Fe in an Ar–methane atmosphere. Here, the atmosphere contained a hydrocarbon component reactive at the temperature of arc discharge. The average diameter of the tubes centered at around 1 nm. No tube was formed in absence of iron catalyst. Kiang and Goddard [51] proved that various additives such as S, Bi, and Pb increased the yield of nanotube production and substantially modified the diameter distribution of SWCNTs.

There is no a well defined mechanism describing the formation of SWCNTs in the arc discharge process. *Ab ovo* there are two main ways of formation: the noncatalytic and the catalytic. For the noncatalytic process one of the most probable methods of generation suggested by several authors [52] is as follows. The tube grows from an empty fullerene cage by repeated additions of C_2 units to the open ends. The growth by this mechanism lengthens a closed tube. According to other authors [53–55] the tubes are open during the growth and the carbon atoms bond to the active dangling bond edge sites. Iijima [54] assumed that the axial growth predominated over the shell growth for SWCNTs.

More and even complex mechanism assumptions are reported on the growth of SWCNTs in the catalytic version of arc discharge. Here, the metal atoms generated by evaporation of catalyst parallel with the carbon electrode should agglomerate to small clusters, and simultaneously (while growing the metal particles) or after forming agglomerates of a suitable size the carbon nanotubes should be generated. Alternatively, SWCNT formation takes place on the metal nanoparticles deposited on the cool reactor parts and counterelectrode surface.

Iijima and Ichihashi [17] assumed that Fe particles act as a homogeneous catalyst in the gas phase assisting in some way the formation of SWCNTs. Kiang et al. [56] described a model in which small catalytic particles rapidly assemble in a region of high carbon density. SWCNTs nucleate and grow very rapidly on these particles as soon as they reach a critical size. By other authors [51] planar polyyne rings serve as nuclei for the formation of SWCNTs. They assumed that C_{32} and C_{38} monocyclic rings lead to SWCNTs with diameters of 1.25 and 1.5 nm, respectively. Co_mC_n type compounds are assumed to act as catalysts. Seraphin and Zhou [50] found that there is no strict correlation between the diameter of SWCNTs and that of the catalyst particle and that the tubes do not grow out from the metal particles. They concluded that the metal acts at the atomic level rather than acting as a heterogeneous catalyst in the growth of SWCNTs. Summarizing the described mechanisms, the following generalized picture can be drawn. The plasma generated by arc discharge of catalyst containing carbon sources, electrons, and other charged species, together with the atoms of catalyst and carbon, is a rather complex reacting system. When a metal catalyst is evaporated with carbon, the metal and carbon atoms may condensate and form alloy particles. As the particles are cooled, carbon dissolved in the particles starts to segregate on the surface of particles since the solubility of carbon in the metal decreases with decreasing temperature. In general the Baker et al. [57, 58] mechanism may be regarded as a realistic description even for the formation of carbon nanotubes. In short, they suggested that the first step is the decomposition of hydrocarbons on the metal surfaces and dissolution of carbon in the metal. Upon this generally exothermic procedure the metal warms up. The dissolved carbon moves to the cooler part of the metal particle and leaves the metal there. The carbon moved from the metal initiates the growth of carbon fibers and/or nanotubes. In this process the diffusion of carbon in the metal is the rate determining step. Opposite to this, Oberlin et al. [59] concluded that the diffusion of carbon in the metal is negligible and diffusion takes place dominantly on the surface of metal

particles. However, Baker et al. [57] have found a correlation between the carbon diffusion in the metal particle and the rate of the growth of the carbon fibers. Baker et al. [60] answered an additional important question concerning the interaction of metal particles to the support material. They classified this interaction into two groups. For the weak interaction the growing tube pushes the catalytic metal particles from the support surface into the gas phase. This way carbon fibers or nanotubes with metal particles in their tips are formed. When the interaction is strong, metal particles remain bonded to the surface and nanotubes are growing on these metal islands. In both cases there are free surfaces toward the gas phase, not covered by carbon fibers or tubes, where the carbon transport may occur from the gas phase. Surface species formed by nucleation may catalyze the growth of SWCNTs. While the tube is growing carbon is supplied from the metal that dissolves carbon coming from the adsorbing carbon atoms. Theoretical calculations [61, 62] have also been performed to clarify the formation mechanisms of SWCNTs. A more complex mechanism has been proposed by Kanzow and Ding [63]. They divided the procedures into three main streams depending on the correlation between the carbon supply and the kinetic energy of the carbon source. These procedures take place at different temperature ranges. The condition with rich carbon supply and insufficient kinetic energy can be represented by significantly lower temperature than for insufficient carbon supply and high kinetic energy. For the former case MWNTs dominate, while under the latter conditions SWCNT formation dominates.

2.2. Synthesis of Single Wall Carbon Nanotubes with Laser Ablation

When a laser beam is used for evaporation of a carbon target generally containing transition metals as catalyst components, procedures similar to those for arc discharge take place [20]. In a typical laser ablation experiment, the second harmonic of a Nd:YAG laser pulse is focused on a graphite target containing 1–2 at% catalyst metal [64–66]. The target is kept in inert atmosphere and the system is heated to 800–1400 °C by an electric furnace. The result of laser ablation is also plasma. Dal et al. [67] concluded that by the application of this direct evaporization method better control over the growth of tubes can be achieved and also the quality of SWCNTs is better. Furthermore, this method allows quasi-continuous production and consequently the yield is much higher as well. Bimetallic catalysts such as Co/Pt, Co/Ni, etc. also gave nice nanotubes. They assumed that SWCNTs form on such a metal particle whose diameter is too small for generation of second and further walls. Kataura et al. showed that varying the type of catalyst metals and the furnace temperature can control the diameter of the grown nanotubes rather well [68]. Since SWCNT formation takes place in the plasma generated by the laser pulse [69], it is thought that nanotube growth in a laser ablation experiment follows essentially the same mechanisms as those detailed for the dc arc discharge [70]. The laser ablation method is often referred to as pulsed laser vaporization (PLV) method as well.

2.3. Chemical Vapor Deposition Synthesis of Single Wall Carbon Nanotubes

There are several studies focused on the decrease of the temperature of CVD. In these works either novel catalyst systems or new carbon sources are applied in the synthesis. Harutyunyan et al. [71] found that addition of 20% of molybdenum to the alumina supported Fe catalyst is beneficial to the SWCNT production, since it lowered the reaction temperature and eliminated the reduction of catalyst prior to the synthesis. The catalyst was active as low as 680 °C. No multiwall nanotube formation was observed.

The generation of SWCNTs is one of the most intriguing questions. Several attempts aimed at the particle size of the metal catalysts, turbulent flow of the carbon source around the catalyst, the role of metal carbide generally detected by instrumental methods, etc. have been made. The role of the size of catalyst metal particles was investigated by Li et al. [72] applying 1–2 or 3–5 nm particles in apoferritin. The instrumental characterization involving atomic force microscopy (AFM), TEM, and Raman spectroscopy proved that the diameter of SWCNTs is strictly correlated to the size of metal particles. Furthermore they proposed a mechanism for the formation of tubes from methane. According to their TEM measurements the metal particles are sitting on the substrate and the isolated carbon nanotubes are growing from them toward the gas phase.

Colomer et al. [24] reported on the production of single wall carbon nanotubes by CCVD of ethylene over Co, Ni, and Fe or their mixture on various supports. They assumed that the support helps to disperse the metal particles. The reaction carried out at 1080 °C gave a significant amount of SWCNT bundles on the Co–Fe, Fe–Co–Ni catalysts.

A real breakthrough was achieved in CVD nanotube production when the silica, zeolite, and alumina types of catalyst supports were replaced by alkaline earth oxides or carbonates [73–75]. In this latter case the removal of support became much easier.

2.4. Gas Phase Catalytic Single Wall Carbon Nanotubes Growth

Industrial scale carbon nanotube applications can only be expected if economic industrial scale SWCNT production methods become available. Arc discharge and laser ablation are extremely power hungry batch methods. It is unlikely that they can ever be scaled up to deliver large quantities of SWCNTs at a reasonable price [e.g., in kg/(unit × day) magnitude]. CVD methods use continuous carbon feedstock but require an additional purification step to separate the synthesized carbonaceous deposit from the catalyst support (SiO_2, Al_2O_3, MgO, etc.). A promising alternative is to eliminate the support and generate the metal catalyst particles *in-situ* in a continuous flow gas phase reactor. This approach has been shown to work utilizing metallocenes and $Fe(CO)_5$ as catalyst precursors and hexane, benzene, C_2H_2, CH_4, and CO as carbon sources. Since most hydrocarbons pyrolize at 6–700 °C which is in all cases below the required operating temperature of the reactor, nanotubes grown from hydrocarbon sources often show an undesired amorphous carbon coating. The most successful gas phase

catalytic SWCNT production method uses carbon monoxide as the carbon source and $Fe(CO)_5$ as the catalyst precursor. It was developed by Bronikowski et al. at Rice University and is called the HiPco® (high pressure CO decomposition) process [25, 26].

In a typical HiPco® synthesis a flow of CO (1–2 standard dm^3/min) dosed with 4–8 ppm $Fe(CO)_5$ is led through a quartz tube kept at 800–1200 °C. Iron nanoclusters formed by Fe atoms from the thermal decomposition of $Fe(CO)_5$ serve as catalyst particles on which SWCNTs can grow from the carbon produced by the Boudouard reaction from CO. Carbon nanotubes are collected from the reactor walls and from a filter placed into the flow at the reactor outlet. SWCNT yield can be increased by raising the CO pressure (30–50 atm) and the temperature (1100–1200 °C) and by introducing a little (<1 vol%) CH_4 into the feedstock. Bronikowski et al. [26] optimized the process parameters and obtained SWCNTs of up to 97 mol% purity at 450 mg/h yield. An extensive analysis of HiPco® kinetics was performed by Dateo et al. [76, 77]. The end product of the HiPco® process contains remarkably little non-SWCNT carbon material and almost no nanotubes exhibit an amorphous carbon coating. The dominant contamination (4–5 at% Fe in the raw product) comes from the iron catalyst particles which are encapsulated in the SWCNT cap in the closing step of the growth process. Chiang et al. [78] described a purification protocol consisting of two wet air oxidation and two HCl sonication steps which could decrease the iron content of a raw HiPco® sample from 5.06 to 0.05 at%. High temperature (800 °C) annealing of the purified material in Ar results in better quality SWCNTs because defects introduced during the synthesis and oxidation steps are partially healed. It is still debated if the annealing improves also the crystallinity of HiPco® samples [78] or not [79].

Although the diameter distribution of HiPco® SWCNT samples can be controlled to some extent by changing the CO pressure, the currently available typical HiPco® materials are characterized by a broader distribution peaking at a lower mean diameter than those obtained by, for example, arc discharge. Diameters of HiPco® tubes were found to fall between 0.79–1.2 and 0.8–1.4 nm by Yudasaka et al. [80] and Zhou et al. [79], respectively. In a recent study [81] Kukovecz et al. reported the values $d = 1.05$ nm, $\sigma = 0.15$ nm for Gaussian distribution mean diameter and variance, respectively. The observed discrepancies between measured and simulated HiPco® Raman spectra suggested that the diameter distribution of HiPco® materials could also be non-Gaussian. HiPco® nanotubes have proven to be excellent subjects for diameter selective observations because of their broad diameter distribution. The oxidation rate of SWCNTs scales roughly with $1/d$ allowing the selective oxidation of thin tubes while leaving the thicker ones intact. Reported selective oxidation conditions are: (i) 425 °C/1 h in flowing wet air [78], (ii) 460–650 °C/10 min in stagnant dry air [79], (iii) oxidation in a reduced O_2 atmosphere [82], and (iv) focused laser pulse at ambient conditions [83]. Similar to the coalescent diameter doubling phenomenon [84] the mean diameter of a HiPco® SWCNT sample could also be increased by heating it to 1600–1800 °C for 5 hours in high vacuum [80]. The diameter enhancement itself was evidenced by Raman and TEM measurements as well as by the increased burn-off temperature of the sample, but the mechanism for this remarkable reaction is yet unclear.

2.5. Other Single Wall Carbon Nanotubes Synthesis Methods

Two additional catalytic SWCNT synthesis methods should also be mentioned here, even though they are not as widespread as the ones detailed previously at present. The solar route [85, 86] focuses the beams of the sun on a catalyst-containing graphite target. The growth dynamics inside the solar furnace are thought to be similar to those in laser ablation [87] and the product obtained is high quality nanotubes containing little amorphous carbon [88]. SWCNTs can also be prepared in low pressure hydrocarbon (e.g., C_2H_2, C_2H_4) flames by mixing catalyst precursors [e.g., $Fe(CO)_5$ or ferrocene] into the feedstock [89, 90]. The principle of this method is quite similar to the gas phase catalytic decomposition detailed previously and, similarly, the best results for SWCNTs are obtained when using CO as carbon source [91, 92]. However, the chemical engineering side of flame reactors is well known and instruments from the laboratory to the production plant scale are already available commercially. Therefore, once the optimal conditions are found [93], flame pyrolysis could rapidly become the dominant large scale supplier of carbon nanotubes in the future.

Under special conditions SWCNTs can also be synthesized without metal catalysts. The thermal coalescence of C_{60} molecules within a SWCNT (peapod system; see Section 3.1) produces a thin SWCNT inside the larger diameter outer tube which acts as a nanoreactor in this case. Another example of SWCNT growth in a confined environment is the pyrolysis of organic template molecules in freshly synthesized $AlPO_4$-5 zeolite pores [94, 95] which gives uniform 0.4 nm diameter SWCNTs superconducting below 20 K [96]. Interestingly, the lowest diameter (0.33 nm) SWCNT reported so far was not produced under steric restrictions. This (4,0) SWCNT was grown from a larger nanotube under electron beam irradiation from a TEM field emission gun [97].

2.6. Single Wall Carbon Nanotubes Purification

The raw product of a single wall nanotube synthesis contains various forms of carbon, that is, amorphous, active carbonlike matter, carbon nanoparticles including fullerene-like material, SWCNTs, as desired product and the catalysts particles. Depending on the preparation method the catalyst component may be the simple metal nanoparticle itself or more complex, supported metal nanoparticles. The aim of the SWCNT purification is the removal of the amorphous carbon forms and the catalyst as completely as possible. A number of purification methods have been suggested; however, they can be classified into four main groups: acidic oxidation, gas phase oxidation, filtration and chromatography, and their combination.

In the acid treatment, first described by Rinzler's group [98], the raw product is refluxed in nitric acid in order to oxidize the metal and the amorphous carbon. This base method has been modified several times suggesting alternative acid

concentrations and reflux times. However, the characteristics of the resulting, purified materials were very similar. Generally upon acid treatment the ends of nanotubes have been decapped and the open ends were terminated by carboxyl groups. Furthermore, structural imperfections closed by hydroxyl and/or carboxyl groups must have been generated as well. The appearance of these functional groups is advantageous for those samples that are to be applied to polymers as a strengthening material; however, they are disadvantageous when the nanotubes are to be utilized in nanoelectronic devices. Zimmerman et al. [99] discussed an oxidation technology for purification of SWCNTs using HCl, Cl_2, and H_2O. A multistep purification procedure was discussed in the paper by Li et al. [100]. The method was developed for purification of SWCNT samples produced by catalytic decomposition of hydrocarbons. The yield of the treatment is 40% with a purity of about 95%. The first step of the procedure is an extraction. Using benzene as solvent the soluble material is removed from the sample. The next step is the acid leaching to remove the iron catalyst particles. The procedure is enhanced by ultrasonication. After filtering the sample was frozen at −196 °C for 10 h followed by further ultrasonication. Finally the sample was washed with deionized water and dried. Starting from 200 mg of raw product 75 mg of purified sample was obtained.

Gas phase oxidation is the most commonly used method for purification of SWCNTs [101–103]. Dillon's group [104] described a method that produces >98 wt% pure SWCNTs. First they subjected the carbonaceous matter to a nitric acid treatment at reflux temperature followed by the gas phase oxidation at 550 °C, for 30 min. They proved that SWCNTs purified in this way are stable as high as 600 °C in oxygen. The acid treatment prior to the gaseous oxidation procedure is obligatory in order to oxidize or dissolve the metal content of the raw product since several metals used for nanotube production by catalytic methods catalyze also the oxidation of carbon nanotubes. Chiang et al. [105] reported a high efficiency purification method for SWCNTs produced by laser oven growth. They investigated the influence of the metal content of the tubes as well and found that with increasing metal content (the origin of which is the catalyst) the oxidation rate of carbon nanotube is enhanced. By the method proposed by Moon et al. [106] high-yield purification of SWCNTs can be achieved. They designed a rotating oven for annealing the nanotube samples in the first step of the procedure. The second step consists of acid leaching for 24 h in order to dissolve the metal impurities. They claim a product with very low, less than 1 wt%, metal content.

Microfiltering proved to be an advantageous procedure for purification of carbon nanotubes. Shelimov et al. [107] developed an ultrasound assisted ultrafiltration method resulting in as high as >90% pure SWCNTs.

Nanotubes of high purity and rather uniform length distribution were obtained by Duesberg et al. [108] using column chromatography for separation of aqueous solution of SWCNTs. Niyogi et al. [109] reported a gel permeation chromatographic method for purification of soluble single wall carbon nanotubes. Two columns, Styragel HMW7 and PLgelMIXED-A, were used and the effluents were analyzed by AFM, ultraviolet-visible (UV-vis), near-infrared (NIR), fluorescent, and Raman spectroscopies. The first fraction of effluent contained 74% of soluble SWCNTs. The authors have claimed that the refinement of this chromatographic technique gives a promise of SWCNT products separated by their length, diameter, and even chirality [109].

3. FUNCTIONALIZATION OF SINGLE WALL CARBON NANOTUBES

Adapting the notation of fullerene chemistry we may classify SWCNT functionalization methods into three groups: (i) endohedral functionalization takes place when materials are admitted into the hollow interior of the nanotube, (ii) exohedral functionalization refers to sidechains chemically bound to the outer surface of the tube, and (iii) lattice functionalization occurs if the SWCNT properties are changed because agents are introduced into the intertube channels of the nanotube lattice. As a limiting case of exohedral functionalization one must also mention the in-wall chemistry of SWCNTs when a number of carbon atoms are substituted by boron and/or nitrogen.

3.1. Endohedral Functionalization

Several groups reported considerable success in forming 1D single crystals of metals, metal carbides, metal oxides, and metal halides within the nanospace inside the SWCNT. These materials are expected to show interesting electrical and magnetic properties like quantum wire effects. Utilized encapsulation methods are based on (i) the capillary effects, (ii) wet chemical techniques, (iii) one-step synthesis of the filled tube by arc dischare, (iv) condensed phase electrolysis, and (v) pyrolysis. Reportedly introduced materials are various metals [110], carbides [111], oxides [112], and halides [113–116]. The reader is referred to some recent extensive reviews for more details on the topic [117, 118].

Fullerenes can be encapsulated within SWCNTs if the tube diameter is adequate [119]. The resulting materials form a subclass of endohedrally functionalized SWCNTs called "peapods" (e.g., C_{60}@SWCNT). A spectacular high resolution TEM (HRTEM) image of a C_{60} peapod is presented in Figure 2a. Peapods were first described in 1998 as spontaneously formed by-products of the purification of PLV derived SWCNTs [120–122]. They quickly gained popularity within the nanotube community and their large scale synthesis was worked out by Smith et al. at the University of Pennsylvania [123, 124] and Kataura et al. at Tokyo Metropolitan University [125]. Since C_{60} favors the inner space of wider nanotubes [126, 127] the synthesis principle is simple: the purified fullerene and a buckypaper of opened SWCNTs are heated together in an evacuated quartz ampoule at 600 °C for 2 hours and then the excess C_{60} is removed by washing the formed peapod buckypaper with toluene. The method is sensitive to the purity of the starting materials and especially to the opening of the nanotubes which can be achieved by selective oxidation of the caps with, for example, H_2O_2, HNO_3, or hot wet air. In the ideal case 90–100% of all available SWCNTs can be filled with fullerenes. Besides direct TEM observation the filling factor can also be determined by electron energy loss spectroscopy (EELS) [128] and Raman spectroscopy [129]. To now peapods have been characterized

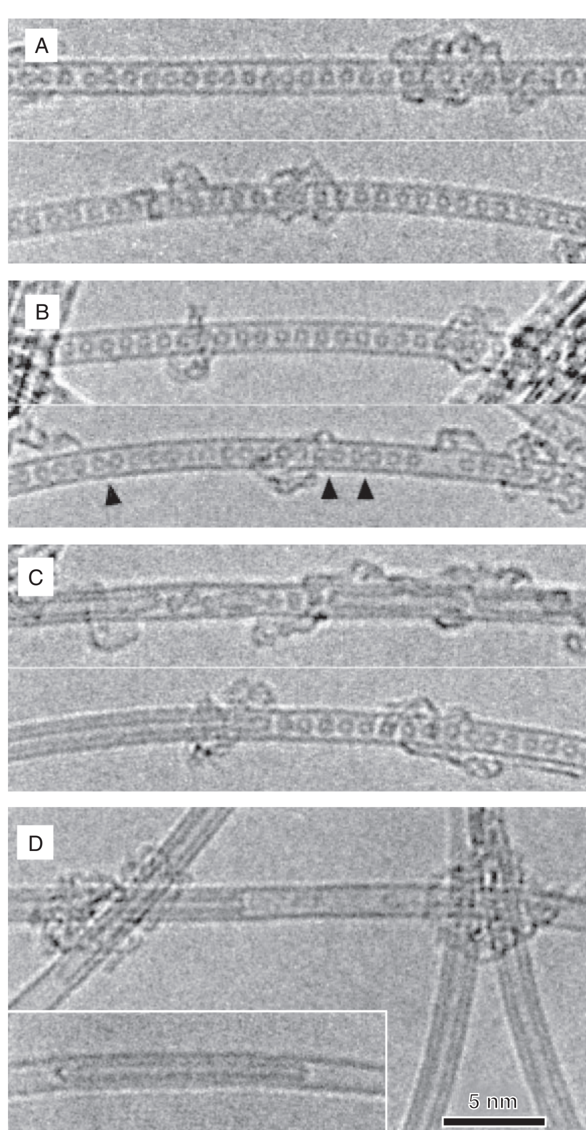


Figure 2. HRTEM images showing a C_{60}@SWCNT peapod (a) and the formation of a double wall carbon nanotube by the coalescence of fullerenes within the nanotube as the peapod is heated to 800, 1000, and 1200 °C (b, c, and d) respectively. Black arrows in (b) indicate locations where the first step of coalescence is observable. Reprinted with permission from [135], S. Bandow et al., *Chem. Phys. Lett.* 337, 48 (2001). © 2001, Elsevier Science.

by electron diffraction [130], EELS [131], optical absorption, and Raman spectroscopy [132, 133] as well as TEM and scanning tunneling microscopy (STM) [134].

Peapods are expected to outperform both fullerenes and empty SWCNTs in applications demanding mechanical strength, H_2 storage [136], and tunable electronic properties [137]. Especially promising studies were reported on metallofullerene peapods [138], for example, on the Gd@C_{82}@SWCNT system [139, 140] which can perform as an ambipolar transistor [141]. The fullerene molecules polymerize inside the SWCNT upon doping [142] and coalesce to yield a thin inner nanotube within the parent SWCNT when heated [135, 143]. This latter process is very visible in Figure 2, where (b)–(d) show the formation of a double-wall carbon nanotube as the peapod depicted in (a) is heated to 800, 1000, and 1200 °C in vacuum, respectively.

3.2. Exohedral Functionalization

Just like fullerenes, the perfect SWCNT is without functional groups and is therefore chemically quite inert. On the other hand, doing addition-based covalent chemistry on nanotubes is considerably easier than functionalizing graphite. The two main sources of reactivity in SWCNTs are (i) the curvature-induced strain arising from the nonplanar geometry of sp^2 carbons and (ii) the misalignment of the π orbitals. The former is more pronounced (i.e., the pyramidalization angle is larger) at the capping fullerene hemispheres while the latter affects mainly the sidewalls. All in all the most reactive places in any nanotube sample are found in the cap of the thinnest tube and least reactive are the bonds running perpendicular to the axis of the largest diameter SWCNT. Recent reviews on the covalent chemistry of SWCNTs are available [12].

Derivatization reactions can be roughly divided into two categories: (i) direct attachment of the required functional groups onto the nanotube and (ii) building bridgeheads by oxidizing some atoms in the tube wall and then proceeding with the functionalization using substitution reactions on the simple (-F, -OH, -COOH) groups formed. Reactions falling into the first category are the 1,3-dipolar cycloaddition [144], the Birch reduction [145], and reactions with nitrenes, radicals, and carbenes [146]. In the second case the first reaction step is usually an aggressive treatment, for example, (i) oxidation in refluxing cc. HNO_3 (yields carboxyl groups) [147–149], (ii) ozone treatment [150, 151], (iii) ball milling in reactive atmosphere [152] or together with solid KOH (hydroxyl groups) [153], and (iv) HF reaction (-F groups) [154, 155]. Then in the second step these bridgeheads are converted to more reactive groups (e.g., carboxyl groups are replaced by acyl chlorides after $SOCl_2$ treatment) and finally, conventional organic synthesis reactions like the Grignard reaction [156] are applied to build the desired functionality onto the anchor sp^3 carbon in the nanotube wall.

Covalent functionalization opens new routes for SWCNT applications. Most important of all, functionalized tubes are often soluble in organic solvents [157, 158] or even water [159, 160]. Georgakilas et al. suggest that a nanotube purification protocol could be based on the solubility difference between functionalized tubes and unharmed carbon/metallic species [161]. Sidegroups may also serve as covalent anchors for fixing SWCNTs onto surfaces [162] or within a polymer matrix [163]. Nanotube–polymer composites have superb mechanical properties and are expected to become a major SWCNT application field in the near future [164, 165].

By heating a SWCNT-B_2O_3 mixture in nitrogen containing atmosphere to ~1600 K it is possible to substitute some of the carbon atoms forming the nanotube wall itself by boron and nitrogen [166, 167]. This reaction—a limiting case of exohedral functionalization—yields $C_xB_yN_z$ tubes. Finally, we should note that investigations on the stacking-based, noncovalent functionalization of SWCNTs are also in progress [168–170].

3.3. Lattice Functionalization (Doping)

The properties of a SWCNT sample can be modified by introducing electron donors or acceptors into the trigonal nanotube lattice [171, 172] to get the nanotube analogs of graphite intercalation compounds. In the former, n-type doping can be achieved by, for example, alkali metals and in the latter, p-type doping can be achieved by, for example, Br_2 or $FeCl_3$. The conductance of doped samples increases [173] and their optically allowed transitions are quenched which causes characteristic changes in their optical [174] and Raman spectra [175]. The extent of intercalation depends on the type of doping species and on the diameter distribution of the nanotube sample [176]. It is possible to dope one-half of a single SWCNT n-type and the other half p-type which leads to an intramolecular nanoelectronics device [177]. The occupancy of the valence/conduction bands can also be modified without doping agents by electrochemical means [178, 179]. A recent review of doping is available from Fischer [180].

4. CHARACTERIZATION OF SINGLE WALL CARBON NANOTUBES

In spite of the significant research effort invested [68, 82, 181–183] it is still impossible to synthesize one selected (n, m) SWCNT in macroscopic quantities. Synthesis products are always obtained as mixtures of different helicity nanotubes. These mixtures cannot be separated into SWCNT fractions using the currently available techniques. Therefore, the reader should be aware that the macroscopical SWCNT studies published so far all investigate a set of different (n, m) nanotubes. It is common to characterize SWCNT samples by estimating the parameters of their diameter distribution from TEM, AFM, or Raman measurements. The distribution function is usually close to a Gaussian, though notable exceptions are also found.

Even though idealized nanotube geometries and detailed electronic models are readily available, experimental SWCNT characterization is no simple task. The main reasons for this are: (i) no well-defined reference tube samples are available, (ii) the sample to be characterized is always a mixture of various (n, m) chiralities, (iii) the sample is often contaminated by non-SWCNT carbon and leftover catalyst particles, and (iv) because of its low atomic number ($Z = 6$) carbon exhibits small cross-section for electron and X-ray diffraction. In this section we discuss the various techniques applied in nanotube research.

4.1. Calculations on Nanotubes

The relationship between SWCNTs and calculations performed on them is a unique one. Since the ideal nanotube structure is well defined but pure and monodisperse samples are scarce, experiments in nanotube science were not *initiating* first principles calculations but quite often *verifying* them. The special physics caused by the one-dimensional character of SWCNTs was well understood by 1995 [184]. The Dresselhaus group at MIT has systematically studied the electronic [185], magnetic [186], and phonon [187] band structure of SWCNTs. Nanotubes were shown to be thermodynamically stable [188] and to favor hexagonal packing [36] with only weak intertube coupling [189].

Currently structural *ab initio* calculations are mostly used to explore the effects of deviations from the ideal cylindrical sp^2 carbon network due to sp^3 defects [190, 191], B and/or N substitution [192], and doping [193–196]. Molecular dynamics simulations are used continuously to suggest nanotube formation mechanisms [62, 197–202]. Some other fields of nanotube research where calculations are making important contributions are: (i) the refinement of our understanding of the special quantized nature of nanotubes [203–208], (ii) H_2 storage related simulations [209–215], and (iii) studies on functionalized SWCNTs [127, 216–218].

4.2. Nanotube Microscopy

Single wall carbon nanotubes were first observed by TEM. Low resolution TEM is regularly used to detect non-nanotube carbon and leftover catalyst contamination in SWCNT samples. Images of nanotube bundles curving through the TEM focal plane provided evidence for the trigonal lattice of SWCNT quasi-crystals and several interesting carbon nanostructures like nanotube coils [219], nano-onions [220], nanocones [221], bamboo-shaped rods [222, 223], etc. were also found as side products of SWCNT electron microscopy. HRTEM can be used to directly measure tube diameters and observe various cap shapes. Iijima et al. [53] published HRTEM images of the continuous junction between tubes of two different diameters proving that heptagon–pentagon defect pairs may occur even in otherwise undamaged nanotube walls. Young's modulus of multiwall [224] and single wall carbon nanotubes [225] was estimated for the first time from intuitive TEM experiments by Treacy et al.

The scanning probe microscopy techniques complement TEM in SWCNT research well [226–229]. Though it was shown theoretically that AFM is capable of reaching atomic resolution on nanotubes [230], AFM is typically used at low magnification levels as a fast and economical means to acquire nanotube length, bundle diameter, and sample purity data. STM, on the other hand, has been applied with considerable success to SWCNT atomic structure elucidation. The first low temperature measurements by Wildoer et al. [231] and Odom et al. [232] showed that the nanotube walls are made of hexagonally arranged carbon atoms and that metallic and semiconducting tubes coexist in SWCNT samples. This was further confirmed by the room temperature results of Hassanien et al. [233]. The chiral angle Θ can often be precisely measured on atomically resolved STM images. Several different angles were observed and this put an end to the early TEM based assumptions that SWCNT samples are dominated by achiral (10,10) tubes. However, neighboring nanotubes of the same apparent helicity (zigzag) were observed in a SWCNT bundle by Hassanien et al. which argues in favor of the "one rope—one helicity theory." Nevertheless, care must be exercised when extracting helicity information from STM images since (i) distinguishing between the intrinsic chirality of the SWCNT and a large scale axial twisting of the tube is not

straightforward [234] and (ii) the geometry of the interaction of the STM tip with a thin nanotube cylinder is rather complex [229, 235].

The electrical conducting properties of nanotubes can be directly measured by scanning tunneling spectroscopy (STS), an STM modus where the tip is held fixed above a chosen sample point and the tunneling current is measured as a function of the bias voltage [236, 237]. The distinction between metallic and semiconducting tubes is then readily made on the basis of the measured I–V characteristics. It is even more useful to study the $\frac{\partial I}{\partial V}/\frac{I}{V}$ vs V curve which is the measure of the electronic DOS. STS experiments on isolated SWCNTs proved the existence of van Hove singularities [238] and made it possible to measure the bandgap of nanotube semiconductors. It was possible to probe the local density of states near the tip of SWCNTs by STS and find sharp resonant valence band states filling the bandgap [239, 240]. While most of the available STS data support the predictions of the simple model described in Section 1.2.2, in a recent report Ouyang et al. [241] argued that only isolated armchair SWCNTs are true metals. Bundled armchair tubes have a pseudogap whose width scales roughly with the inverse tube diameter, and "metallic" zigzag SWCNTs in fact do possess real energy gaps scaling with the inverse square root of tube radius [242].

Scanning probe microscopy is suitable for measuring the mechanical properties (Young's modulus, toughness, strength) of SWCNTs [243–245]. It can also be used for manipulating individual SWCNTs to create, for example, nanotube junctions [246] or field effect transistors (FETs) and to cut nanotubes [247]. The laborious pushing–rolling cycles of moving a nanotube around a surface can be avoided by using the new technique described by Cheung et al. [248]. Nanotubes themselves make superior SPM tips for high resolution topography imaging [249, 250]. Lieber's group pioneered the use of nanotubes as probes for individual biomolecules [31, 251–253]. The best results are achieved when the tubes are grown directly on the SPM cantilever [254–258]. An exciting new field is the utilization of covalently functionalized nanotubes as tips for chemical force microscopy [259, 260].

Some recent advances in SWCNT microscopy include the introduction of Fresnel projection microscopy [261, 262] and near field spectroscopy.

4.3. Optical Spectroscopy

Optically allowed transitions (OATs) from the ith valence to the jth conduction band follow the selection rules $(j - i) = 0$ and $|j - i| = 1$ for electric fields polarized parallel and perpendicular to the nanotube axis, respectively. Since in the latter case the OAT is suppressed because of the depolarization effect, in practice we only have to consider the symmetric E_{ii} transitions between the equidistant van Hove singularities below and above the Fermi energy. Each transition gives rise to a peak in the nanotube UV-vis–NIR absorption spectrum. Peaks at around 0.7, 1.2, 1.8, and 2.5 eV come from the E_{11}^{sc}, E_{22}^{sc}, E_{11}^{m}, and E_{33}^{sc} OATs where "*sc*" and "*m*" denote semiconducting and metallic tubes, respectively. The exact position, shape, and width of the peaks are determined by the diameter distribution of the sample and the DOSs of the individual tubes. Bachilo et al. have suggested a method for rapidly surveying the (n, m) composition of bulk SWCNT samples using bandgap fluorescence spectroscopy [39, 263].

Optical absorption measurements have provided the data to estimate the diameter distribution [264, 265], the bandgap size, the γ_0 transfer energy, and the dielectric function of SWCNTs. Kazaoui et al. [266, 267] conducted an optical absorption study on the n- and p-type doping of a SWCNT sample using Cs as electron donor and Br_2 as electron acceptor partners. In both cases, the intensity of the first few (low energy) absorption peaks decreased with increasing dopant concentration. This could be explained by the gradual filling of the conduction band or the valence band with electrons/holes in n- and p-doping, respectively. Both effects extinguish van Hove singularities gradually, decreasing the possibility of an OAT involving that particular vHs which manifests as a loss of absorption band intensity in the spectrum. Close to saturation the authors observed new peaks at 1.07 and 1.30 eV for Br- and Cs-doping, respectively. They tentatively assigned these to intersubband transitions in the valence (for Br) and the conduction (for Cs) bands. Finally, the presence of the characteristic peaks in the UV-vis–NIR spectrum can be used to demonstrate that the sample retained its special one-dimensional electronic structure even after performing purification and/or functionalization chemistry on it.

4.4. Raman Spectroscopy

Raman spectroscopy is a very potent SWCNT characterization technique. High resolution spectra can be collected from practically all sample forms, at any sample temperature, even *in-situ* while, for example, monitoring the progress of a chemical reaction. Moreover, the SWCNT Raman spectrum is relatively simple to interpret yet informs us about the diameter distribution, electronic properties, and purity of the sample. This section discusses the major components of the spectrum and refers the reader to the available reviews [33, 268, 269] for more details.

4.4.1. Basics of Single Wall Carbon Nanotubes Raman spectroscopy

Phonon dispersion relations and phonon densities of states of SWCNTs are calculated by zone folding of a graphene sheet with special attention to certain nanotube specific differences [187]. A SWCNT with N hexagons in its unit cell [Eq. (7)] has altogether $6N$ possible phonon modes. The number of actual distinct phonon branches is usually much smaller because of mode degeneracies, for example, 66 (54 doubly degenerate plus 12 nondegenerate) instead of 120 in the case of the $(10, 10)$ tube. Four phonons are always acoustic: two degenerate transverse acoustic modes corresponding to x and y displacements perpendicular to the nanotube axis, a twisting acoustic mode for the rigid rotation around the tube axis, and a longitudinal acoustic for displacement along the axis. These four give rise to a constant nonzero phonon DOS at very low energies, which increases stepwise with the energy as the optical phonons enter. Each new subband introduces a 1D van Hove singularity into the phonon DOS.

Only a small portion of the $6N - 4$ optical phonon modes is Raman active because of symmetry reasons. As only k vectors very close to $k = 0$ can couple to the incident light, one has to consider only the symmetry of the vibrations at the center of the Brillouin zone (Γ point). Therefore, the point group of the unit cell can be used to find the Raman active vibrations which are $\Gamma^{\text{vib, Raman}}_{(n,n)|n=2k} = 4A_{1g} + 4E_{1g} + 8E_{2g}$, $\Gamma^{\text{vib, Raman}}_{(n,m)|n\neq m\neq 0} = 4A + 5E_1 + 6E_2$, and $\Gamma^{\text{vib, Raman}}_{\text{others}} = 3A_{1g} + 6E_{1g} + 6E_{2g}$. It is important to note that the number of Raman active modes (16 for even n armchair tubes, 15 for all others) is independent of the tube diameter and chirality, even though the number of total optical phonon modes ($6N - 4$) varies a lot. The *frequency* of a particular mode, however, may or may not be sensitive to the tube diameter. For example, the low frequency (below 500 cm^{-1}) modes soften quasi-linearly with $1/d$ [270] while other modes between 1000 and 1200 cm^{-1} show an opposite behavior and the highest frequency modes around 1580 cm^{-1} are not very diameter sensitive [271–273]. Some Raman active modes have only a very small Raman cross-section and therefore, only 6–7 modes are typically observable experimentally. On the other hand, SWCNT samples are always [274] mixtures of different helicities. Consequently, it is rather common to find several (even more than 16) peaks in a measured spectrum since the same symmetry mode could appear at j different positions for j different tubes.

Perhaps the most striking feature of SWCNT Raman spectroscopy is the enormous *resonance enhancement* of the signal at selected laser wavelengths. The scattering intensity is enlarged by orders of magnitude when the energy of the incident (laser photon) or the scattered (laser photon–absorbed phonon) light matches the energy of an OAT between a vHs pair in the electronic DOS. The enhancement is so strong that even fifth order Raman spectra could be observed with its help [275]. Since the position of the vHs's and, accordingly, the E_{ii} OAT energies are helicity dependent, it is possible to measure only a few desired (n, m) tubes selectively even in heterogeneous samples by selecting the excitation laser to match the corresponding OAT frequency [276]. The large enhancement factor will then ensure that most of the Raman signal comes from the chosen tubes and the others remain "silent." In an inverse experiment Jorio et al. succeeded in focusing the laser beam on one SWCNT and fine-tuned the excitation laser around an OAT frequency [277]. They plotted the intensity of a selected Raman peak as a function of laser energy and thus were able to directly map the singularities in the SWCNT joint DOS. Their results indicate that a vHs may be significantly narrower ($\sim$1 meV) than previously estimated from STS data ($\sim$100 meV).

In Figure 3, the Fourier transform (FT)-Raman spectrum of an arc discharge derived SWCNT sample is depicted. It can be seen that all the theoretical considerations mentioned previously boil down to an admirably simple spectrum which has only three main first order components: the so-called radial breathing mode (RBM) at around 180 cm^{-1}, the D-line at 1330 cm^{-1}, and the G-line between 1560 and 1600 cm^{-1}. Typical second order and combination bands are also depicted. In the following sections we will discuss the

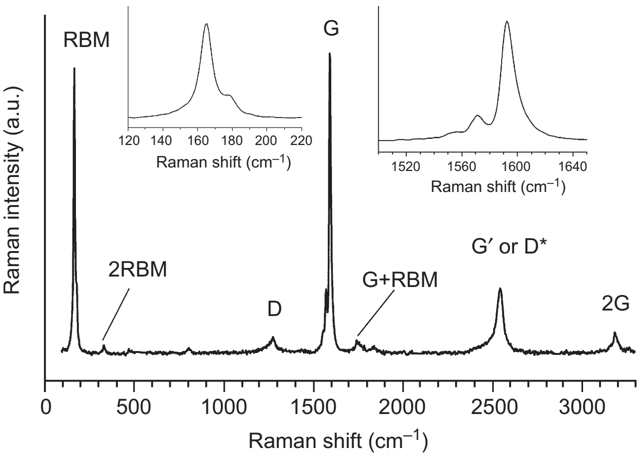


Figure 3. FT-Raman spectrum of a SWCNT sample showing the characteristic first and second order bands as well as the G+RBM combination band. The insets reveal the typical detailed shapes of the RBM and the G-line.

information one can get from each of these bands. Unless indicated otherwise, all experiments refer to Stokes Raman spectra ($E_{\text{scattered}} = E_{\text{laser}} - E_{\text{phonon}}$).

4.4.2. Radial Breathing Mode

The radial breathing mode is a low frequency (150–300 cm^{-1}) A_1 mode corresponding to the total symmetric in-phase motion of all carbon atoms in the SWCNT wall perpendicular to the tube axis. It has a large Raman cross-section and its position is sensitive to the tube diameter (d). Therefore, it is a valuable tool for SWCNT research. The RBM Raman shift ω_{RBM} scales linearly with $1/d$ as

$$\omega_{\text{RBM}} = \frac{C_1}{d} + C_2 \qquad C_1, C_2 \in \mathbf{R}$$
$$[C_1] = \text{cm}^{-1}\,\text{nm} \qquad [C_2] = \text{cm}^{-1} \tag{15}$$

It should be noted that although the RBM frequency is a reliable means of investigating SWCNT diameters in the d =0.7–2.0 nm range, its cross-section is itself a function of d and decreases rapidly with increasing tube diameter, complicating studies on thick tubes. In Eq. (15) C_1 is the proportionality constant and C_2 compensates for intertube interactions. Choosing the right values for these two is the key to get accurate tube diameter estimates and has been frequently addressed in the literature. Suggested (C_1, C_2) pairs based on fitting experimental data are (224, 14) [278] (232, 6.5) [279] (214, 6) [280], and (239, 8.5) [81]. Values based on theoretical calculations were reported as (234, 0) [203] and (223.75, 0) [281]. Jorio et al. have found (248, 0) to describe their single tube experiments the best [282].

A remarkable feature of the RBM is that its position varies with the energy of the excitation laser in an oscillatory fashion even when measuring the very same SWCNT sample [283, 284]. The phenomenon can be explained by considering that (i) SWCNT samples are always made of several different diameter tubes and (ii) the position of the van Hove singularities in the DOS and, consequently, the E_{ii} energies of the OATs responsible for the resonance Raman

enhancement are dependent on the diameter (helicity) of the SWCNT. Each laser line brings different (n, m) nanotubes into resonance. The contribution of these few tubes to the total RBM will be dominant and the spectrum will show a peak at the $\omega_{\mathrm{RBM}}(d)$ frequency determined by Eqs. (3) and (15). Tuning to another laser wavelength brings a new set of tubes into resonance, and therefore the dominant RBM peak will appear at a different position. It is quite useful to study the so-called *Kataura plot* depicted in Figure 4 to visualize the events for ourselves. Each point in this plot corresponds to an OAT energy (ordinate) for a chosen nanotube diameter (abscissa). Let us now consider a sample containing SWCNTs in the $d = 1.0$–1.5 nm diameter range and excite it with a laser emitting $E_{\mathrm{laser}} = 2.0$ eV photons. This energy matches the first optical transition of $d \sim 1.3$ nm metallic nanotubes found in the sample, and therefore the dominant RBM peak will appear at $\omega \sim 190$ cm^{-1} (top x axis). By increasing E_{laser} slowly we hit the $E_{11}^{\mathrm{metallic}}$ OAT of thinner and thinner metallic tubes so the RBM peak will wander to higher and higher frequencies. At a certain point the E_{laser} energy will be high enough to excite the $E_{33}^{\mathrm{semicond}}$ transition of the thickest semiconducting SWCNTs in our sample; thus there will be two major RBM peaks: one corresponding to the diameter of the thinnest metallic (at $\omega \sim 248$ cm^{-1}) and one to the thickest semiconducting tube ($\omega \sim 190$ cm^{-1}). The former will disappear when increasing E_{laser} any further, the latter will start wandering upward, and the process will repeat itself again when the excitation energy reaches the $E_{44}^{\mathrm{semicond}}$ OAT and so on. Even though the overall result is apparently a RBM peak moving between 190 and 248 cm^{-1} the reader should remember that this is a resonance effect and that the position of the RBM peak of any given SWCNT depends neither on E_{laser} nor on tube conductivity. Kataura plots can be used to do approximate (n, m) assignments [284, 285]; however, the accuracy of such calculations is limited by the accuracy of the underlying DOS

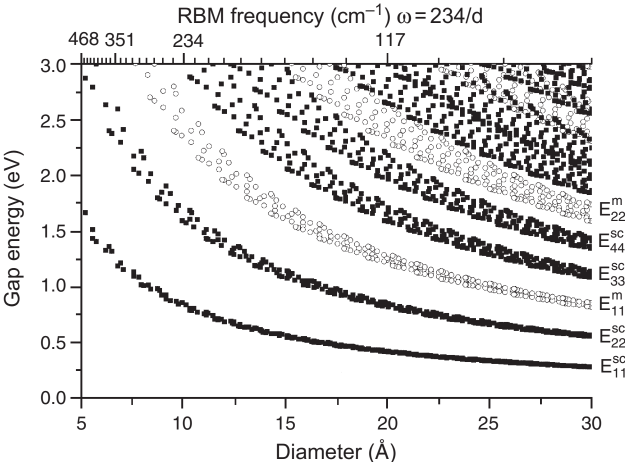


Figure 4. A plot of the bandgap energy vs nanotube diameter for all possible nanotubes within the diameter range 0.5–3.0 nm ("Kataura plot"). Black squares and open circles denote semiconducting and metallic tubes, respectively. The calculated position of the Raman RBM peak for each diameter is given on the top x axis. Public domain data from the homepage of Dr. S. Maruyama (University of Tokyo, Tokyo, Japan) were used for creating this figure.

calculation as well as by experimental issues like intertube interactions and spectral resolution.

The positions of the RBM peaks inform us about individual tubes only. In order to obtain the more useful *diameter distribution* parameters it is advantageous to analyze the oscillations of the first and second spectral moments [286] of the whole radial breathing mode band measured at several different laser energies to avoid biasing the distribution toward a certain diameter via the resonance effect. The required formalism has been recently reviewed by Kuzmany et al. [287].

4.4.3. Defect Induced Mode (D-Line)

The so-called D-line is a weak to medium intensity peak observed between 1300 and 1350 cm^{-1} in visible excitation SWCNT Raman spectra. Its most striking feature is that it upshifts approximately 50 cm^{-1}/eV when increasing the laser excitation frequency. The position of the D-line is not particularly sensitive to the diameter distribution of the sample. In the early SWCNT Raman investigations the D-line was usually assigned to the unidentified graphitic/amorphous carbon contamination accompanying the nanotubes. The SWCNT D-line has a similar counterpart in graphite which itself has puzzled researchers for decades and initiated several incorrect interpretations. Finally, Thomsen and Reich succeeded in explaining the graphitic D-line by developing a double resonance theory for it [288]. The authors have shown that the double resonance model is applicable to SWCNTs as well and that only the $(n - m)/3d_H \in \mathbf{Z}$ [see Eq. (6)] nanotubes contribute to the D-line [289]. Single tube measurements by Souza et al. verified that the position of the D-line is sensitive to the chirality of the nanotube and explained this with a resonance effect between the van Hove singularities in the SWCNT DOS [290]. The most comprehensive theory of the SWCNT Raman D-band is that of Kürti et al. who take both the double resonance process and the van Hove singularities into account [207]. This theory is able to predict the small deviations from linearity in the dispersion of the SWCNT D-line reported by Grüneis et al. Kürti et al. suggest that the fine structure of the D-line dispersion profile could be sensitive to the diameter distribution.

The message of these theoretical advances is that it is not necessary to have macroscopic impurities in a nanotube sample to observe the D-line. However, the presence of a defect site is certainly required by all double resonance models. It is therefore correct to regard the D-line as an overall indicator of defects: sp^3 carbon atoms, holes in the SWCNT walls, attached functional groups, and, indeed, non-nanotube graphitic domains may all contribute to the D-line. It is customary in the nanotube literature to report the intensity ratio of the D-line to the G-line as an indicator of sample purity ("pure" meaning both "amorphous carbon-free" and "defect-free" as discussed previously). As a rule of thumb we can say that the larger the G/D ratio, the higher the sample quality.

An interesting feature in the second order SWCNT Raman spectrum is the G′-line (also called the D*-line by some authors) found between 2500 and 2900 cm^{-1} [291]. The G′-line shows a dispersion of ~100 cm^{-1}/eV with E_{laser}. This first overtone of the D-line is generally more intense

than the D-line itself because the G′ is symmetry allowed. The recent work of Souza et al. indicates that the fine structure of the G′-line is rather informative and should be studied in detail [292].

4.4.4. Tangential Modes (G Band)

The most intense lines of the SWCNT Raman spectrum are found between 1560 and 1600 cm^{-1} in the so-called the G-band which corresponds to the tangential C–C stretching vibrations. The band consists of non- or weakly diameter dispersive phonons of A_1, E_1, and E_2 symmetries. It is rather insensitive to E_{laser} as long as the character of the dominant, resonantly enhanced nanotubes is the same (i.e., either semiconducting or metallic). When the main spectral contribution comes from semiconducting SWCNTs (e.g., 2.1 eV < E_{laser} < 2.7 eV for a typical sample with Gaussian diameter distribution $d = 1.35$ nm, $\sigma = 0.15$ nm) then the G-band is dominated by a sharp peak at ~1592 cm^{-1} accompanied by a sideband at ~1560 cm^{-1}. G-bands of semiconducting SWCNTs can be fitted very well with six Lorenztian oscillators [one longitudinal (LO) and one transversal (TO) mode for each of the three allowed symmetries] [293].

Although peaks from semiconducting tubes are always present in the G-band to some extent, the characteristic features of G-bands dominated by metallic nanotubes (1.7 eV < E_{laser} < 2 eV for the aforementioned distribution) are three new, broader peaks at ~1515, 1540, and 1580 cm^{-1}. The 1540 cm^{-1} peak is intense, broad, and asymmetrically elongated toward the lower frequencies. Unlike the other peaks which are well fitted with Lorentzian oscillators, the 1540 cm^{-1} one has a Breit–Wigner–Fano (BWF) lineshape [264],

$$I_\omega = I_0 \frac{[1 + (\omega - \omega_{BWF})/q\Gamma]^2}{1 + [(\omega - \omega_{BWF})/\Gamma]^2} \tag{16}$$

where I_0, ω_{BWF}, $1/q$, and Γ denote the maximum intensity, the renormalized frequency, a coupling parameter, and the full width at half maximum, respectively. BWF lineshapes have been reported previously for several metallic carbon species like alkali metal intercalated graphite [294] and C_{60} [295]. They arise from the coupling of a phonon mode (in the case of SWCNTs an A_1(LO) [208] or A_1(TO) [273, 296] phonon) to the electron continuum and can be regarded as a telltale signal of the presence of metallic nanotubes. The intensity ratio I_{1540}/I_{1592} is routinely used as a semiquantitative indicator of the relative contribution of metallic SWCNTs to a given Raman spectrum. The position of the BWF line has been found to be more sensitive to the diameter of the investigated nanotubes than other components of the G-band [208].

The main semiconducting G-line peak found at 1592 cm^{-1} at room temperature shifts about –0.012 cm^{-1}/K with increasing sample temperature [297, 298]. This effect can be used to estimate the local temperature experienced by the sample hit by the Raman laser and to characterize sample purity [299]. The former effect is particularly useful since the usual temperature sensor of Raman spectroscopy, the Stokes/anti-Stokes intensity ratio, fails for SWCNTs because of the resonance effect. Charge transfer to and from the nanotube (n- and p-type doping) also shifts the G-band to lower and higher frequencies, respectively [172, 175]. Maultzsch et al. examined the dispersion of the various G-band components with E_{laser} and suggested that the G-band could also result from a double resonance process similar to the D-line [206].

4.4.5. Other Raman Features of Interest

Although understanding the RBM, the D-line, and the G-band should enable the reader to benefit from the majority of the SWCNT Raman research papers available, one should not forget about developments that took place in more specialized fields of nanotube Raman spectroscopy. Peaks showing strong energy dispersion have been found in the intermediate (600–1200 cm^{-1}) frequency range of the spectrum [300]. Differences between semiconducting and metallic tubes were found when studying the anti-Stokes Raman spectra [301–303] and when performing surface enhanced resonant Raman experiments [304–306]. Polarized Raman experiments helped in identifying the symmetry of the components of the G-band [307–310]. UV Raman studies uncovered notable differences between the RBM and D-line intensities in UV ($E_{laser} = 4.8$ eV) and visible SWCNT Raman spectra [311]. The possibility of helicity assignment based on the splitting of the van Hove singularities in the DOS due to the trigonal warping effect was suggested by Saito et al. [43].

4.5. Other Characterization Methods

Even though SWCNTs are very thin and beam-sensitive, some groups have managed to calculate [312] and measure nanotube electron diffraction (ED) patterns [313, 314] and in some cases even succeeded in ED (n, m) assignment [315, 316]. The parameters of the trigonal SWCNT lattice in bundles have been obtained from X-ray diffraction studies [20, 281]. Rols et al. have recently shown that diameter distribution estimates based on XRD agree sufficiently with, for example, TEM results if the finite size of the nanotube bundle is taken into account upon XRD evaluation [317, 318]. Inelastic neutron scattering is a powerful alternative to Raman spectroscopy for studying the SWCNT phonon structure because in contrast to optical probes, all phonons are active to neutrons [319–321]. Rayleigh scattering was utilized by Yu and Brus [322] to study the interband transitions of SWCNTs and by Puretzky et al. to investigate the dynamics of the PLV synthesis method [323].

Investigations based on magnetic interactions [electron paramagnetic and nuclear magnetic resonace (EPR, NMR)] in SWCNTs are generally not as informative as for other materials. Pure SWCNTs do not show any intrinsic EPR features; the formerly reported $g = 2.002$ asymmetric line [324] is now thought to be due to charge transfer effects. On the other hand, SWCNTs do have their own ^{13}C MAS NMR signature: a broad isotropic line at 126 ppm (relative to TMS) which can be deconvoluted into Lorentzian semiconducting and metallic contributions [325, 326]. The recent increased interest in SWCNT functionalization, however, has opened up new fields of application for ^{13}C NMR in nanotube research. Functional groups are anchored to the nanotube via sp^3 carbons giving a signal at 29 ppm and also show their own chemical shifts (e.g., 182 ppm for –COOH groups) [327].

Although SWCNTs do have infrared (IR) active phonons that can be described as $\Gamma^{\text{vib, IR}}_{(n,n)|n=2k} = A_{2u} + 7E_{1u}$, $\Gamma^{\text{vib, IR}}_{(n,m)|n\neq m\neq 0} = 4A + 5E_1$, and $\Gamma^{\text{vib, IR}}_{\text{others}} = 2A_{2u} + 5E_{1u}$, infrared spectroscopy is seldom used for characterizing pure SWCNT samples because all IR active modes are very weak. IR reflectance studies on untreated SWCNT samples are available in the mid-IR regime from Kuhlmann et al. [328, 329] and in the far-IR from Ugawa et al. [330, 331]. In a mid-IR transmission experiment Kastner et al. found one broad and asymmetric line at 1575 cm^{-1} and a weaker line at 868 cm^{-1} [332]. On the other hand, once the symmetry of the SWCNT structure is broken, IR spectroscopy becomes a rather helpful tool. Several authors used the IR spectrum of sidewall functionalized nanotubes to prove the presence of, for example, carboxylic acid [37, 148, 157, 333], ester [334], amide [335, 336], quinone [333], and fluoride [337] groups as well as the formation of more complex heterostructures [338] and physisorbed complexes [339]. Substituting carbon atoms in the SWCNT wall can also be studied by IR spectroscopy as it was shown for boron [340] or nitrogen [341, 342] substitution recently.

The EELS of SWCNTs was measured in 1994 by Kuzuo et al. [343] to be different from that of multiwall carbon nanotubes [344] and of graphite. According to Knupfer et al. [345] the two strongest peaks in the loss function arise from the π plasmon (the collective excitation of the π system) and the $\pi + \sigma$ plasmon (the excitation of all valence electrons) at 5.8 and 21.5 eV for $q = 0.15$ Å^{-1}, respectively. Another set of low intensity, low energy (0.6, 1.2 eV) nondispersive peaks could be assigned to interband transitions between the van Hove singularities in the SWCNT electronic density of states. These peaks can be correlated with optical absorption peaks after performing Kramers–Kronig analysis and could be used for studying alkali metal intercalation into SWCNT bundles [346–349]. EELS was successfully applied for detecting chemical bonds between SWCNTs and Pt nanoclusters [350] and for identifying endohedral metallic oxide functionalization [351]. Finally, EELS has become an indispensable tool in B- and N-doped SWCNT research since it allows the simultaneous determination of atomic composition and heteroatom hybridization state [167, 352–354].

The gas adsorption properties of SWCNTs received much attention because of their potential application for H_2 storage [104]. Isotherms are available for, for example, N_2 [355–359], O_2 [356], Ar [359, 360], Kr [361], Xe [333, 362–365], CH_4 [361, 366, 367], CF_4 [364], SF_6 [364], C_6H_6 [355], and CH_3OH [355]. Small molecules are predicted to adsorb both within open tubes and in the interstitial space while large molecules clearly prefer adsorption inside the nanotube [368]. As expected, single wall carbon nanotubes are essentially microporous adsorbents with BET surface areas exceeding 400 m^2/g [369]. However, it is important to realize that closed (as synthesized) and opened (oxidized) SWCNTs have different sorption properties. In the latter adsorbates enjoy a favorable potential within the tubes and thus the tube interior is filled up completely before adsorption starts in the interstitial channels [356]. Ohba and Kaneko argue that the BET analysis is inapplicable for nanotube surface area determination [370]. According to Du et al. the diameter of the open pores can be reliably analyzed using the Horvath–Kawazoe method [359].

5. PROPERTIES OF SINGLE WALL CARBON NANOTUBES

One should always be aware of the heterogeneity of the SWCNT samples used for determining the physical properties of the material. While the SWCNT as a molecule is a well-defined entity, bulk nanotube samples available currently are mixtures of various different helicities at best. Samples reported in the literature come from different sources, may contain amorphous carbon impurities and/or leftover catalyst particles, and may have been subjected to different postsynthetic treatments. In this section we report on what appear to be the basic common characteristics of SWCNTs. The most important parameters are summarized in Table 1.

5.1. Macroscopic Properties and Handling

A SWCNT sample usually appears as a fluffy black powder with very low apparent density. It cannot be dissolved in any solvent and does not melt or sublimate in vacuum up to 1400 °C. Apart from irritating the respiratory system if inhaled, it is a nontoxic material. However, samples may contain harmful contaminants like carcinogen polyaromatic hydrocarbons and also the intrinsic carcinogenety of nanotubes cannot be ruled out at the present state of knowledge. For ease of handling the powder is often transformed into a randomly oriented mat of entangled SWCNTs (*buckypaper*) by dispersing the nanotubes in a solvent and filtering the dispersion through a ~0.4 μm pore filter. The dried buckypaper can be investigated by Raman spectroscopy and scanning

Table 1. A summary of the physical properties of single wall carbon nanotubes.

Property	Value	Ref.
Length	50 nm–1 mm	[371, 372]
Diameter	0.4–5.6 nm	[94, 373]
Density	0.040–2 g/cm^3 (buckypaper—individual tube)	[374, 375]
Lattice	trigonal *P6/mcc* space group	[20, 36]
Melting point	>1400 °C in vacuum	[2]
Specific surface area	151–1315 m^2/g	[376]
Specific heat	~610 mJ/(gK)	[377]
Thermal conductivity	35 W/mK (disordered) 200 W/mK (aligned)	[378, 379]
Debye temperature	960 K (in-tube) 13 K (intertube)	[377]
Dc conductivity at 300 K	6.9×10^4 S/m	[375, 380]
Dielectric constant at 10 kHz	$\sim -1.8 \times 10^4$	[375]
Susceptibility	$\sim -10 \times 10^{-6}$ emu/g (anisotropic)	[381]
Maximum electric current density	$>10^9$ A/cm^2	[382, 383]
Bandgap	~0.7 eV (semiconducting)	[384]
Young's modulus	1.25–0.35/+0.45 TPa	[225]
Tensile strength	~45 GPa	[243]

electron microscopy or used in electrochemical experiments without further treatment.

Though not soluble in the classical sense, SWCNTs can be finely dispersed in liquid phase by ultrasonication. Care must be exercised when setting the parameters as tubes may break during the process [107, 385]. The right solvent is chosen by trial-and-error because the dispersability of a certain nanotube sample in a given solvent depends largely on the type and parameters of the SWCNT synthesis method, the sample morphology, and also the postsynthetic treatments (e.g., purification) applied. In general 1,2-dichlorobenzene, CS_2, N,N-dimethyl-formamide, and 2-propanol are worth trying before others. Recently Walters et al. [386] succeeded in preparing aligned nanotube membranes by filtering a SWCNT solution in a strong magnetic field. Dispersions containing individually separated SWCNTs can be obtained by ultrasonicating nanotubes in an aqueous detergent solution. Ionic (e.g., alkali-dodecylsulphates) [39, 387] and nonionic [388] surfactants were both shown to perform adequately. Surfactant assisted suspensions are the only systems where successful high pressure liquid chromatography separation of unfunctionalized nanotubes was reported [38, 108, 389]. Nanotube dispersions can be directly studied by techniques like UV-vis–NIR spectroscopy or light scattering. When a SWCNT film is more appropriate for the measurement it can be obtained by drop-, spin-, or spray-coating substrate. Such samples are used in optical absorption spectroscopy and IR spectroscopy. In the former case the substrate is quartz, and in the latter it is a gold mirror which is measured in reflection–absorption geometry. Samples for nanotube microscopy are generally produced by drop-coating an adequate substrate (e.g, highly oriented pyrolitic graphite, gold, Si) with a highly diluted nanotube suspension.

5.2. Mechanical Properties

Nanotubes have exceptional mechanical properties [390–392]. Walters et al. [243] measured the tensile strength of SWCNTs and found a maximum elastic strain of $5.8 \pm 0.9\%$. Taking 1.25 TPa as the Young's modulus of a SWCNT [225] they calculate the yield strength of a SWCNT rope to be 45 ± 7 GPa—an order of magnitude better than that of high quality steels. Chesnokov et al. found that the reversible work done in compressing a nanotube sample to 29 kbar is 0.18 eV/C atom and suggested that nanotubes could be applied as strings for mechanical energy storage [393].

5.3. Electronic and Heat Conductance

SWCNTs are one-dimensional conductors with only two open conduction channels in the case of armchair tubes, and therefore it was expected that their conduction electrons should localize with increasing length. However, White and Todorov showed that SWCNTs could in fact be ballistic conductors with localization lengths over 10 mm [394]. Experimental evidence for the ballistic conductance of MWCNTs [395] and SWCNTs [396] was also found. Very thin SWCNTs synthesized within the pores of aluminosilicates were reported as superconductive [96] with a transition temperature of 15 K. Another interesting feature of SWCNTs is that they exhibit strong electron–electron correlation effects (Luttinger liquid behavior) [397, 398]. For some recent work on nanotube transport properties, see, for example, [399–402].

The thermal properties of SWCNTs are determined by the 1D quantized phonon band structure [403–406]. The specific heat of SWCNTs is very close to a graphene sheet above $T = 50$ K but is significantly smaller at lower temperatures [377]. Macroscopic thermal properties were found to be rather sensitive to the alignment of SWCNTs in the sample: the room temperature thermal conductivity of a random nanotube mat is 35 W/mK while that of an aligned sample is ~200 W/mK [379].

5.4. Optical Properties

The linear polarizability α and third-order nonlinear polarizability γ of SWCNTs was calculated for several finite length SWCNTs by Wan et al. [407, 408]. They found both parameters to be heavily chirality and size dependent. Narrower tubes and tubes with a higher helical degree have larger polarizability [409]. Semiconducting SWCNTs have smaller α and much smaller γ values than metallic ones. The calculations indicate that SWCNTs are promising nonlinear optical materials. Vivien et al. recently published a study on the optical limiting properties of SWCNT suspensions [410]. Ruzicka et al. measured the optical reflectivity and dc resistivity of pristine and potassium doped SWCNT mats [374].

6. APPLICATIONS OF SINGLE WALL CARBON NANOTUBES

In the first decade of SWCNT research efforts were mainly focused on synthesizing nanotubes in sufficient quantity and quality and on understanding their unique physical properties. While there is certainly a lot left to do in these fields as well, it is expected that in the second decade the spotlight will move to nanotube-based applications, devices, and marketable items. In this section we present a brief overview of the current mainstream SWCNT application themes [411].

6.1. Nanoelectronic Devices

The traditional silicon-based electronics industry is quickly approaching the limits of miniaturization where size reduction is no longer possible because of inevitable quantum effects. Nanoelectronics or "single-molecule electronics" therefore became an extensively researched field in the 1990s. Because of their unique properties SWCNTs are regarded by many as potential building blocks for nanoelectronic devices. The three major research directions are currently nanotube semiconductors, sensors, and field emission displays.

FETs utilizing a single carbon nanotube as the active element were first reported in 1998 [30, 412] and have been extensively researched ever since [413–415]. Other elementary electronics units like diodes [177, 398], logical gates [416, 417], and random access memory cells [418] were also prepared from carbon nanotubes. Establishing good contacts between the nanotube and other parts of the circuitry is crucial for the performance of such devices. In fact it was

recently shown by Heinze et al. [419] that transistor action may occur primarily on the contact resistances (Schottky barrier transistors). For a recent review on nanotube semiconductor applications see, for example, Avouris [420].

SWCNT based chemical sensors were first reported in 2000. Collins et al. observed the extreme oxygen sensitivity of the electrical resistance of SWCNTs [421] and Kong et al. used a similar setup to detect small amounts of NO_2 and NH_3 [422]. The detection limit of nanotube based gas sensors already matches or exceeds that of conventional solid state sensors. However, recovery time and the number of duty cycles are still to be improved. The selectivity of nanotube sensors can be enhanced by using a two-layered construction where the outer layer interacts specifically with a target molecule and the inner layer transmits the changes induced by this interaction to the SWCNT in the core of the sensor. Nanotubes were also successfully tested as biosensors [423–425]. Mechanical and thermal strain experienced by a SWCNT containing polymer can also be measured on the basis of specific Raman shifts [426, 427]. An up-to-date review on nanotube based sensors is available from Dai et al. [428].

The market of flat panel displays (FPDs) is rapidly expanding. Active TFT screens can be manufactured cost-effectively only up to approximately 20 inch diagonal and plasma or projection displays are built with diagonals starting at 40 inches. Therefore, the demand for economical FPDs in the size of a normal home TV set is not yet satisfied. One possible solution is the field emission display (FED) which keeps the phosphorescent coating of the old cathode ray tube but uses a separate electron emitter for each pixel. The excellent field emission properties of carbon nanotubes were realized as early as 1995 [429, 430]. The first working display was reported in 1998 [431, 432] and commercial devices are expected to be introduced in 2004. Nanotube FEDs work at quite low turn-on voltages (0.7–4 V/μm) where nanotubes still follow the conventional Fowler–Nordheim model (up to a critical current density [433]). Nanotube FEDs generally utilize aligned nanotube arrays [434–436] grown by the CVD technique directly on the back plate of the display [437, 438].

6.2. Hydrogen Storage

Due to the environmental hazard caused by car exhaust gases an enormous effort is focused to the production of car engines working by combustion of hydrogen with oxygen. If hydrogen fuel cells were used to power automobiles and other vehicles, air pollution would be reduced. Such a fuel cell would be practical only if the hydrogen could be stored on board the vehicle in a safe, efficient, compact, and economical manner. For this purpose, metal hydrides and various carbon materials were investigated [439]. After the discovery of carbon nanotubes they became the most attractive materials for hydrogen storage since they were tubular, the interior of the tubes might be filled with hydrogen, and they are lightweight materials more suitable for preparing adsorbents than metals or metal oxides and hydrides [104, 440].

Optimistic calculations predict that SWCNTs are capable of exceeding the U.S. Department of Energy (DOE) target for H_2 storage; for example, Lee et al. calculated 14 wt% for a (10, 10) tube [441, 442]. While certain studies warn that gases cannot be adsorbed in the interstitial channels of SWCNT bundles [443] Ren and Price interpret their neutron scattering data as evidence for H_2 adsorption in the interstitial space [444]. Wang and Johnson argue on the basis of geometric considerations that it is impossible the meet the DOE target (6 wt%) using SWCNTs at ambient temperature [445, 446]. Recent calculations indicate that the hydrogen storage capacity could be a function of nanotube diameter and chirality [209, 447]. Some published experimental H_2 uptake values for SWCNTs are: (i) 4.2 wt% at room temperature (RT) [448], (ii) 0.932 wt% at RT [449], (iii) 4 wt% at RT [450], and 8 wt% at 77 K [451]. Typically 70–78% of the total hydrogen adsorbed can be released under ambient pressure at RT. Alkali doped SWCNTs appear to perform considerably better: Chen et al. report 14–20 wt% H_2 uptake [452]. Froudakis explains this observation by suggesting that charge transfer from the alkali metal to the H_2 polarizes the latter and the charge-induced dipole interaction makes higher hydrogen storage capacities possible [453]. On the other hand, some doubt has been raised recently that some of the weight increase in the earlier experiments was due to H_2O and not to H_2 uptake [454].

Summarizing, H_2 storage in SWCNTs is at present more of a controversial scientific challenge than a mature technology. Possible future research directions were recently sketched by the experts of the field [439, 455–458].

6.3. Other Applications

Incorporating 0.2–5% nanotubes into plastics [459] offers considerable benefits in two major areas: (i) conducting polymers and (ii) enhanced durability plastics. Nanotube–polymer composites are available with conductances in the 0.01–0.1 S/cm range which is more than enough to dissipate electrostatic charge in, for example, pipelines and filters threatened by static discharge induced explosions. Qian et al. reported that 1% multiwall nanotube incorporation can increase the elastic modulus and the break stress of a polystyrene matrix by 36–42% and ~25%, respectively [460]. Biercuk et al. observed a monotonic increase up to 350% (at 2 wt% SWCNT content) in the Vickers hardness of SWCNT–epoxy composites as a function of nanotube loading as well as a 125% increase (at 1 wt% SWCNT) in the thermal conductivity of the composite [461]. The performance of composites is governed by the quality of the nanotube–polymer interface [218, 462]. It is expected that even better results could be obtained by anchoring functionalized tubes into the matrix through covalent bonds [463, 464].

Some other fields with large potential for development in carbon nanotube applications are nonlinear optics [410, 465–467], heterogeneous catalysis [468–470], and biotechnology [471–474].

7. CONCLUSIONS

Since Iijima's discovery in 1991 carbon nanotube science has developed from a peculiar side branch of fullerene research into a discipline of its own with well over 100 genuine papers published monthly. Though a good deal of the basic physical

and chemical properties is known already, there is still a lot of exciting science left waiting to be explored. It is expected that SWCNT-based devices will make it from the laboratories to the consumer market in the next 5–7 years. The most promising candidates for this seem to be nanotube-based gas sensors, field emission displays, lamps, and nanotube-enforced composites at the moment. The most important challenge today is preparing monodisperse (n, m) samples in macroscopic quantities as this would open the road for manufacturing single-molecule nanoelectronic devices with SWCNTs as active elements. In the long run, the hollow inner space of nanotubes might also be utilized at an industrial scale as a "nano cleanroom" for a whole new chemistry. One thing is certain: this coming second decade of SWCNT research is bound to be lots of fun so be sure to follow it as closely as you can!

GLOSSARY

Buckypaper A thin mat of randomly oriented carbon nanotubes obtained by filtering a carbon nanotube suspension, much like normal paper is obtained by filtering cellulose fiber suspension.

BWF peak A peak with a characteristic Breit-Wigner-Fano lineshape occuring in the low wavenumber branch of the tangential Raman band (G-band) of metallic carbon nanotubes. The asymmetric BWF lineshape is the result of the coupling of a phonon to the electronic continuum and is reduced to a symmetric Lorentzian if there is no such coupling.

Catalytic chemical vapor deposition (CCVD) A gas-solid heterogeneous catalytic synthesis reaction. The carbon source molecule mixed into the gas feed decomposes on the heated solid catalyst particles and forms carbon nanotubes. Sometimes also referred to as CVD.

Density of states (DOS) A characteristic function of a solid which shows the number of available electronic states at a certain energy relative to the Fermi level.

Hamada vector The folding vector along which a hypothetical graphene sheet can be rolled up into itself to form a particular SWCNT. Also referred to as "chiral vector", the Hamada vector determines all of the properties of the nanotube.

Kataura plot A plot of the energies of all optically allowed transitions between the valence and the conduction band of SWCNTs as a function of nanotube diameter. The Kataura plot is very helpful in understanding the peculiar resonance behavior of SWCNTs.

Peapod A nanostructure consisting of fullerenes ("peas") filled into the hollow core of a SWCNT ("pod").

Pulsed laser vaporization (PLV) A synthesis method for preparing carbon nanostructures by evaporating a part of a graphite target using a laser pulse. The nanostructures are formed in the plasma during cooling.

Radial breathing model (RBM) A totally symmetric vibrational modus of SWCNTs when all carbon atoms move in-phase perpendicular to the tube axis.

Single wall carbon nanotube (SWCNT) A quasi one dimensional form of carbon which can be visualized as a very long (above 1 μm) and thin (1–2 nm diameter) rolled-up graphene sheet.

ACKNOWLEDGMENTS

A.K. acknowledges funding by the EU RTN FUNCARS (HPRN-CT-1999-00011) and thanks Professor Hans Kuzmany at the University of Vienna, Austria for his support. A. K. and Z. K. acknowledge support from Zoltan magyary postdoctoral fellowships.

REFERENCES

1. M. S. Dresselhaus and G. Dresselhaus, "Science of Fullerenes and Carbon Nanotubes," 1996.
2. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, "Physical Properties of Carbon Nanotubes." Imperial College Press, London, 1999.
3. *Appl. Phys. A* 67 (1995).
4. *Carbon* 33 (1995).
5. *Carbon* 38 (2000).
6. *Carbon* 40 (2002).
7. *Physica B* 323 (2002).
8. M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, in "Topics in Applied Physics," Vol. 80. Springer-Verlag, Berlin/Heidelberg, 2001.
9. P. M. Ajayan, *Chem. Rev.* 99, 1787 (1999).
10. H. J. Dai, *Surf. Sci.* 500, 218 (2002).
11. C. N. R. Rao, B. C. Satishkumar, A. Govindaraj, and M. Nath, *Chem. Phys. Chem.* 2, 78 (2001).
12. *Acc. Chem. Res.* 35 (2002).
13. A. Q. Oberlin, M. Endo, and T. Koyama, *Carbon* 14, 133 (1976).
14. S. Iijima, *J. Cryst. Growth* 55, 675 (1980).
15. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
16. S. Iijima, *Nature* 354, 56 (1991).
17. S. Iijima and T. Ichihashi, *Nature* 363, 603 (1993).
18. D. S. Bethune, C. H. Kiang, M. S. Devries, G. Gorman, R. Savoy, J. Vazquez, and R. Beyers, *Nature* 363, 605 (1993).
19. C. Journet, W. K. Maser, P. Bernier, A. Loiseau, M. L. delaChapelle, S. Lefrant, P. Deniard, R. Lee, and J. E. Fischer, *Nature* 388, 756 (1997).
20. A. Thess, R. Lee, P. Nikolaev, H. J. Dai, P. Petit, J. Robert, C. H. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tomanek, J. E. Fischer, and R. E. Smalley, *Science* 273, 483 (1996).
21. E. F. Kukovitskii, L. A. Chernozatonskii, N. A. Kiselev, O. I. Lebedev, A. B. Ormont, N. N. Melnik, O. E. Omeljanovskii, and V. I. Tsebro, *Mol. Cryst. Liq. Cryst. C* 8, 17 (1996).
22. J. Kong, H. T. Soh, A. M. Cassell, C. F. Quate, and H. J. Dai, *Nature* 395, 878 (1998).
23. J. Kong, A. M. Cassell, and H. J. Dai, *Chem. Phys. Lett.* 292, 567 (1998).
24. J. F. Colomer, G. Bister, I. Willems, Z. Konya, A. Fonseca, G. Van Tendeloo, and J. B. Nagy, *Chem. Commun.* 1343 (1999).
25. P. Nikolaev, M. J. Bronikowski, R. K. Bradley, F. Rohmund, D. T. Colbert, K. A. Smith, and R. E. Smalley, *Chem. Phys. Lett.* 313, 91 (1999).
26. M. J. Bronikowski, P. A. Willis, D. T. Colbert, K. A. Smith, and R. E. Smalley, *J. Vac. Sci. Technol. A* 19, 1800 (2001).
27. T. W. Ebbesen, H. Hiura, M. E. Bisher, M. M. J. Treacy, J. L. Shreeve Keyer, and R. C. Haushalter, *Adv. Mater.* 8, 155 (1996).
28. J. Chen, M. A. Hamon, H. Hu, Y. S. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon, *Science* 282, 95 (1998).
29. Y. Chen, R. C. Haddon, S. Fang, A. M. Rao, W. H. Lee, E. C. Dickey, E. A. Grulke, J. C. Pendergrass, A. Chavan, B. E. Haley, and R. E. Smalley, *J. Mater. Res.* 13, 2423 (1998).

30. S. J. Tans, A. R. M. Verschueren, and C. Dekker, *Nature* 393, 49 (1998).
31. S. S. Wong, E. Joselevich, A. T. Woolley, C. L. Cheung, and C. M. Lieber, *Nature* 394, 52 (1998).
32. P. Kim and C. M. Lieber, *Science* 286, 2148 (1999).
33. M. S. Dresselhaus and P. C. Eklund, *Adv. Phys.* 49, 705 (2000).
34. N. Hamada, *Mater. Sci. Eng. B* 19, 181 (1993).
35. The $a = 2.49$ Å lattice constant of the unrolled nanotube honeycomb lattice differs from the 2.46 Å value of graphite because the C–C bond distance is 1.44 Å in the former and 1.42 Å in the latter.
36. J. C. Charlier, X. Gonze, and J. P. Michenaud, *Europhys. Lett.* 29, 43 (1995).
37. A. Kukovecz, C. Kramberger, M. Holzinger, H. Kuzmany, J. Schalko, M. Mannsberger, and A. Hirsch, *J. Phys. Chem. B* 106, 6374 (2002).
38. G. S. Duesberg, J. Muster, V. Krstic, M. Burghard, and S. Roth, *Appl. Phys. A* 67, 117 (1998).
39. M. J. O'Connell, S. M. Bachilo, C. B. Huffman, V. C. Moore, M. S. Strano, E. H. Haroz, K. L. Rialon, P. J. Boul, W. H. Noon, C. Kittrell, J. P. Ma, R. H. Hauge, R. B. Weisman, and R. E. Smalley, *Science* 297, 593 (2002).
40. X. Y. Gong, J. Liu, S. Baskaran, R. D. Voise, and J. S. Young, *Chem. Mater.* 12, 1049 (2000).
41. Actually, the wave vector along the nanotube axis is also quantized for finite length SWCNTs. However, because of the large aspect ratio of carbon nanotubes this quantization becomes important only when examining transport properties at low temperatures.
42. The so-called "small (or linear) k" approximation holds near the K point of the graphene Brillouin zone. In this approximation the valence and the conduction bands are symmetric and the energy dispersion relation is a linear function of k.
43. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 61, 2981 (2000).
44. C. Journet and P. Bernier, *Appl. Phys. A* 67, 1 (1998).
45. H. J. Dai, in *Carbon Nanotubes*, 80, 29 (2001).
46. A. Huczko, *Appl. Phys. A* 74, 617 (2002).
47. W. Kratschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, *Nature* 347, 354 (1990).
48. Y. E. Lozovik and A. M. Popov, *Uspekhi Fiz. Nauk* 167, 751 (1997).
49. T. W. Ebbesen and P. M. Ajayan, *Nature* 358, 220 (1992).
50. S. Seraphin and D. Zhou, *Appl. Phys. Lett.* 64, 2087 (1994).
51. C. H. Kiang and W. A. Goddard, *Phys. Rev. Lett.* 76, 2515 (1996).
52. M. Endo, K. Takeuchi, S. Igarashi, K. Kobori, M. Shiraishi, and H. W. Kroto, *J. Phys. Chem. Solids* 54, 1841 (1993).
53. S. Iijima, P. M. Ajayan, and T. Ichihashi, *Phys. Rev. Lett.* 69, 3100 (1992).
54. S. Iijima, *Mater. Sci. Eng. B* 19, 172 (1993).
55. R. E. Smalley, *Mater. Sci. Eng. B* 19, 1 (1993).
56. C. H. Kiang, W. A. Goddard, R. Beyers, J. R. Salem, and D. S. Bethune, *J. Phys. Chem. Solids* 57, 35 (1996).
57. R. T. K. Baker, M. A. Baker, P. S. Barber, P. S. Harris, F. S. Feates, and R. J. Waite, *J. Catal.* 26, 51 (1972).
58. R. T. K. Baker and P. S. Harris, in "Chemistry and Physics of Carbon" (P. L. Walker and P. A. Thrower, Eds.). Dekker, New York, 1978.
59. A. Q. Oberlin, M. Endo, and T. Koyama, *J. Cryst. Growth* 32, 335 (1978).
60. R. T. K. Baker, P. S. Harris, F. Henderson, and R. B. Thomas, *Carbon* 13, 17 (1975).
61. A. Maiti, C. J. Brabec, and J. Bernholc, *Phys. Rev. B* 55, R6097 (1997).
62. J. Gavillet, A. Loiseau, C. Journet, F. Willaime, F. Ducastelle, and J. C. Charlier, *Phys. Rev. Lett.* 8727, 275504 (2001).
63. H. Kanzow and A. Ding, *Phys. Rev. B* 60, 11180 (1999).
64. T. Guo, P. Nikolaev, A. Thess, D. T. Colbert, and R. E. Smalley, *Chem. Phys. Lett.* 243, 49 (1995).
65. S. Iijima, T. Wakabayashi, and Y. Achiba, *J. Phys. Chem.* 100, 5839 (1996).
66. T. Wakabayashi, D. Kasuya, H. Shiromaru, S. Suzuki, K. Kikuchi, and Y. Achiba, *Z. Phys. D* 40, 414 (1997).
67. H. J. Dai, A. G. Rinzler, P. Nikolaev, A. Thess, D. T. Colbert, and R. E. Smalley, *Chem. Phys. Lett.* 260, 471 (1996).
68. H. Kataura, Y. Kumazawa, Y. Maniwa, Y. Ohtsuka, R. Sen, S. Suzuki, and Y. Achiba, *Carbon* 38, 1691 (2000).
69. A. A. Puretzky, D. B. Geohegan, X. Fan, and S. J. Pennycook, *Appl. Phys. A* 70, 153 (2000).
70. F. Kokai, K. Takahashi, M. Yudasaka, R. Yamada, T. Ichihashi, and S. Iijima, *J. Phys. Chem. B* 103, 4346 (1999).
71. A. R. Harutyunyan, B. K. Pradhan, U. J. Kim, G. G. Chen, and P. C. Eklund, *Nano Lett.* 2, 525 (2002).
72. Y. M. Li, W. Kim, Y. G. Zhang, M. Rolandi, D. W. Wang, and H. J. Dai, *J. Phys. Chem. B* 105, 11424 (2001).
73. J. F. Colomer, C. Stephan, S. Lefrant, G. Van Tendeloo, I. Willems, Z. Konya, A. Fonseca, C. Laurent, and J. B. Nagy, *Chem. Phys. Lett.* 317, 83 (2000).
74. H. Yan, Q. W. Li, J. Zhang, and Z. F. Liu, *Carbon* 40, 2693 (2002).
75. Q. W. Li, H. Yan, Y. Cheng, J. Zhang, and Z. F. Liu, *J. Mater. Chem.* 12, 1179 (2002).
76. C. E. Dateo, T. Gokcen, and M. Meyyappan, *J. Nanosci. Nanotechnol.* 2, 523 (2002).
77. T. Gokcen, C. E. Dateo, and M. Meyyappan, *J. Nanosci. Nanotechnol.* 2, 535 (2002).
78. I. W. Chiang, B. E. Brinson, A. Y. Huang, P. A. Willis, M. J. Bronikowski, J. L. Margrave, R. E. Smalley, and R. H. Hauge, *J. Phys. Chem. B* 105, 8297 (2001).
79. W. Zhou, Y. H. Ooi, R. Russo, P. Papanek, D. E. Luzzi, J. E. Fischer, M. J. Bronikowski, P. A. Willis, and R. E. Smalley, *Chem. Phys. Lett.* 350, 6 (2001).
80. M. Yudasaka, H. Kataura, T. Ichihashi, L. C. Qin, S. Kar, and S. Iijima, *Nano Lett.* 1, 487 (2001).
81. A. Kukovecz, C. Kramberger, V. Georgakilas, M. Prato, and H. Kuzmany, *Eur. Phys. J. B* 28, 223 (2002).
82. E. Borowiak-Palen, T. Pichler, X. Liu, M. Knupfer, A. Graff, O. Jost, W. Pompe, R. J. Kalenczuk, and J. Fink, *Chem. Phys. Lett.* 363, 567 (2002).
83. S. L. Bokova, personal communication, 2002.
84. S. L. Fang, A. M. Rao, P. C. Eklund, P. Nikolaev, A. G. Rinzler, and R. E. Smalley, *J. Mater. Res.* 13, 2405 (1998).
85. D. Laplaze, P. Bernier, W. K. Maser, G. Flamant, T. Guillard, and A. Loiseau, *Carbon* 36, 685 (1998).
86. T. Guillard, S. Cetout, L. Alvarez, J. L. Sauvajol, E. Anglaret, P. Bernier, G. Flamant, and D. Laplaze, *Eur. Phys. J.* 5, 251 (1999).
87. L. Alvarez, T. Guillard, J. L. Sauvajol, G. Flamant, and D. Laplaze, *Appl. Phys. A* 70, 169 (2000).
88. E. Anglaret, N. Bendiab, T. Guillard, C. Journet, G. Flamant, D. Laplaze, P. Bernier, and J. L. Sauvajol, *Carbon* 36, 1815 (1998).
89. M. D. Diener, N. Nichelson, and J. M. Alford, *J. Phys. Chem. B* 104, 9615 (2000).
90. R. L. Vander Wal, T. M. Ticich, and V. E. Curtis, *J. Phys. Chem. A* 104, 7209 (2000).
91. R. L. Vander Wal and T. M. Ticich, *Chem. Phys. Lett.* 336, 24 (2001).
92. R. L. V. Wal and T. M. Ticich, *J. Phys. Chem. B* 105, 10249 (2001).
93. R. L. V. Wal, G. M. Berger, and L. J. Hall, *J. Phys. Chem. B* 106, 3564 (2002).
94. Z. K. Tang, H. D. Sun, J. Wang, J. Chen, and G. Li, *Appl. Phys. Lett.* 73, 2287 (1998).
95. H. D. Sun, Z. K. Tang, J. Chen, and G. Li, *Appl. Phys. A* 69, 381 (1999).
96. Z. K. Tang, L. Y. Zhang, N. Wang, X. X. Zhang, G. H. Wen, G. D. Li, J. N. Wang, C. T. Chan, and P. Sheng, *Science* 292, 2462 (2001).
97. L. M. Peng, Z. L. Zhang, Z. Q. Xue, Q. D. Wu, Z. N. Gu, and D. G. Pettifor, *Phys. Rev. Lett.* 85, 3249 (2000).

98. A. G. Rinzler, J. Liu, H. Dai, P. Nikolaev, C. B. Huffman, F. J. Rodriguez-Macias, P. J. Boul, A. H. Lu, D. Heymann, D. T. Colbert, R. S. Lee, J. E. Fischer, A. M. Rao, P. C. Eklund, and R. E. Smalley, *Appl. Phys. A* 67, 29 (1998).
99. J. L. Zimmerman, R. K. Bradley, C. B. Huffman, R. H. Hauge, and J. L. Margrave, *Chem. Mater.* 12, 1361 (2000).
100. F. Li, H. M. Cheng, Y. T. Xing, P. H. Tan, and G. Su, *Carbon* 38, 2041 (2000).
101. E. Dujardin, T. W. Ebbesen, A. Krishnan, and M. M. J. Treacy, *Adv. Mater.* 10, 611 (1998).
102. T. W. Ebbesen, P. M. Ajayan, H. Hiura, and K. Tanigaki, *Nature* 367, 519 (1994).
103. H. Hiura, T. W. Ebbesen, and K. Tanigaki, *Adv. Mater.* 7, 275 (1995).
104. A. C. Dillon, K. M. Jones, T. A. Bekkedahl, C. H. Kiang, D. S. Bethune, and M. J. Heben, *Nature* 386, 377 (1997).
105. I. W. Chiang, B. E. Brinson, R. E. Smalley, J. L. Margrave, and R. H. Hauge, *J. Phys. Chem. B* 105, 1157 (2001).
106. J. M. Moon, K. H. An, Y. H. Lee, Y. S. Park, D. J. Bae, and G. S. Park, *J. Phys. Chem. B* 105, 5677 (2001).
107. K. B. Shelimov, R. O. Esenaliev, A. G. Rinzler, C. B. Huffman, and R. E. Smalley, *Chem. Phys. Lett.* 282, 429 (1998).
108. G. S. Duesberg, M. Burghard, J. Muster, G. Philipp, and S. Roth, *Chem. Commun.* 435 (1998).
109. S. Niyogi, H. Hu, M. A. Hamon, P. Bhowmik, B. Zhao, S. M. Rozenzhak, J. Chen, M. E. Itkis, M. S. Meier, and R. C. Haddon, *J. Am. Chem. Soc.* 123, 733 (2001).
110. S. Seraphin, *J. Electrochem. Soc.* 142, 290 (1995).
111. S. Seraphin, D. Zhou, J. Jiao, J. C. Withers, and R. Loutfy, *Nature* 362, 503 (1993).
112. S. Friedrichs, R. R. Meyer, J. Sloan, A. I. Kirkland, J. L. Hutchison, and M. L. H. Green, *Chem. Commun.* 929 (2001).
113. J. Sloan, D. M. Wright, H. G. Woo, S. Bailey, G. Brown, A. P. E. York, K. S. Coleman, J. L. Hutchison, and M. L. H. Green, *Chem. Commun.* 699 (1999).
114. C. G. Xu, J. Sloan, G. Brown, S. Bailey, V. C. Williams, S. Friedrichs, K. S. Coleman, E. Flahaut, J. L. Hutchison, R. E. Dunin-Borkowski, and M. L. H. Green, *Chem. Commun.* 2427 (2000).
115. S. Friedrichs, R. R. Meyer, J. Sloan, A. I. Kirkland, J. L. Hutchison, and M. L. Green, in "Electron Microscopy and Analysis 2001," pp. 279–282.
116. J. Sloan, S. J. Grosvenor, S. Friedrichs, A. I. Kirkland, J. L. Hutchison, and M. L. H. Green, *Angew. Chem. Int. Ed.* 41, 1156 (2002).
117. F. Banhart, N. Grobert, M. Terrones, J. C. Charlier, and P. M. Ajayan, *Int. J. Mod. Phys. B* 15, 4037 (2001).
118. M. Monthioux, *Carbon* 40, 1809 (2002).
119. S. Bandow, M. Takizawa, H. Kato, T. Okazaki, H. Shinohara, and S. Iijima, *Chem. Phys. Lett.* 347, 23 (2001).
120. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Nature* 396, 323 (1998).
121. B. Burteaux, A. Claye, B. W. Smith, M. Monthioux, D. E. Luzzi, and J. E. Fischer, *Chem. Phys. Lett.* 310, 21 (1999).
122. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Chem. Phys. Lett.* 315, 31 (1999).
123. B. W. Smith and D. E. Luzzi, *Chem. Phys. Lett.* 321, 169 (2000).
124. B. W. Smith, R. M. Russo, S. B. Chikkannanavar, and D. E. Luzzi, *J. Appl. Phys.* 91, 9333 (2002).
125. H. Kataura, Y. Maniwa, T. Kodama, K. Kikuchi, K. Hirahara, K. Suenaga, S. Iijima, S. Suzuki, Y. Achiba, and W. Kratschmer, *Synth. Met.* 121, 1195 (2001).
126. S. Okada, S. Saito, and A. Oshiyama, *Phys. Rev. Lett.* 86, 3835 (2001).
127. S. Berber, Y. K. Kwon, and D. Tomanek, *Phys. Rev. Lett.* 88, 185502 (2002).
128. X. Liu, T. Pichler, A. Knupfer, M. S. Golden, J. Fink, H. Kataura, Y. Achiba, K. Hirahara, and S. Iijima, *Phys. Rev. B* 6504, 045419 (2002).
129. R. Pfeiffer, H. Kuzmany, W. Plank, T. Pichler, H. Kataura, and Y. Achiba, *Diamond Relat. Mater.* 11, 957 (2002).
130. K. Hirahara, S. Bandow, K. Suenaga, H. Kato, T. Okazaki, H. Shinohara, and S. Iijima, *Phys. Rev. B* 6411, 115420 (2001).
131. T. Pichler, *New Diamond Front. Carbon Technol.* 11, 375 (2001).
132. H. Kataura, Y. Maniwa, M. Abe, A. Fujiwara, T. Kodama, K. Kikuchi, H. Imahori, Y. Misaki, S. Suzuki, and Y. Achiba, *Appl. Phys. A* 74, 349 (2002).
133. L. Kavan, L. Dunsch, and H. Kataura, *Chem. Phys. Lett.* 361, 79 (2002).
134. D. J. Hornbaker, S. J. Kahng, S. Misra, B. W. Smith, A. T. Johnson, E. J. Mele, D. E. Luzzi, and A. Yazdani, *Science* 295, 828 (2002).
135. S. Bandow, M. Takizawa, K. Hirahara, M. Yudasaka, and S. Iijima, *Chem. Phys. Lett.* 337, 48 (2001).
136. M. Shiraishi, T. Takenobu, A. Yamada, M. Ata, and H. Kataura, *Chem. Phys. Lett.* 358, 213 (2002).
137. A. Oshiyama, S. Okada, and S. Saito, *Physica B* 323, 21 (2002).
138. T. Okazaki, K. Suenaga, K. Hirahara, S. Bandow, S. Iijima, and H. Shinohara, *Physica B* 323, 97 (2002).
139. K. Suenaga, T. Tence, C. Mory, C. Colliex, H. Kato, T. Okazaki, H. Shinohara, K. Hirahara, S. Bandow, and S. Iijima, *Science* 290, 2280 (2000).
140. K. Hirahara, K. Suenaga, S. Bandow, H. Kato, T. Okazaki, H. Shinohara, and S. Iijima, *Phys. Rev. Lett.* 85, 5384 (2000).
141. T. Shimada, T. Okazaki, R. Taniguchi, T. Sugai, H. Shinohara, K. Suenaga, Y. Ohno, S. Mizuno, S. Kishimoto, and T. Mizutani, *Appl. Phys. Lett.* 81, 4067 (2002).
142. T. Pichler, H. Kuzmany, H. Kataura, and Y. Achiba, *Phys. Rev. Lett.* 8726, 267401 (2001).
143. S. Bandow, G. Chen, G. U. Sumanasekera, R. Gupta, M. Yudasaka, S. Iijima, and P. C. Eklund, *Phys. Rev. B* 66, 075416 (2002).
144. V. Georgakilas, K. Kordatos, M. Prato, D. M. Guldi, M. Holzinger, and A. Hirsch, *J. Am. Chem. Soc.* 124, 760 (2002).
145. S. Pekker, J. P. Salvetat, E. Jakab, J. M. Bonard, and L. Forro, *J. Phys. Chem. B* 105, 7938 (2001).
146. M. Holzinger, O. Vostrowsky, A. Hirsch, F. Hennrich, M. Kappes, R. Weiss, and F. Jellen, *Angew. Chem. Int. Ed.* 40, 4002 (2001).
147. C. N. R. Rao, A. Govindaraj, and B. C. Satishkumar, *Chem. Commun.* 1525 (1996).
148. A. Kuznetsova, I. Popova, J. T. Yates, M. J. Bronikowski, C. B. Huffman, J. Liu, R. E. Smalley, H. H. Hwu, and J. G. G. Chen, *J. Am. Chem. Soc.* 123, 10699 (2001).
149. Z. H. Yu and L. E. Brus, *J. Phys. Chem. A* 104, 10995 (2000).
150. D. B. Mawhinney, V. Naumenko, A. Kuznetsova, J. T. Yates, J. Liu, and R. E. Smalley, *J. Am. Chem. Soc.* 122, 2383 (2000).
151. L. T. Cai, J. L. Bahr, Y. X. Yao, and J. M. Tour, *Chem. Mater.* 14, 4235 (2002).
152. Z. Konya, I. Vesselenyi, K. Niesz, A. Kukovecz, A. Demortier, A. Fonseca, J. Delhalle, Z. Mekhalif, J. B. Nagy, A. A. Koos, Z. Osvath, A. Kocsonya, L. P. Biro, and I. Kiricsi, *Chem. Phys. Lett.* 360, 429 (2002).
153. H. Pan, L. Liu, Z.-X. Guo, L. Dai, F. Zhang, D. B. Zhu, R. Czerw, and D. Carroll, *Nano Lett.* 3, 29 (2003).
154. E. T. Mickelson, C. B. Huffman, A. G. Rinzler, R. E. Smalley, R. H. Hauge, and J. L. Margrave, *Chem. Phys. Lett.* 296, 188 (1998).
155. K. F. Kelly, I. W. Chiang, E. T. Mickelson, R. H. Hauge, J. L. Margrave, X. Wang, G. E. Scuseria, C. Radloff, and N. J. Halas, *Chem. Phys. Lett.* 313, 445 (1999).
156. P. J. Boul, J. Liu, E. T. Mickelson, C. B. Huffman, L. M. Ericson, I. W. Chiang, K. A. Smith, D. T. Colbert, R. H. Hauge, J. L. Margrave, and R. E. Smalley, *Chem. Phys. Lett.* 310, 367 (1999).

157. J. Chen, A. M. Rao, S. Lyuksyutov, M. E. Itkis, M. A. Hamon, H. Hu, R. W. Cohn, P. C. Eklund, D. T. Colbert, R. E. Smalley, and R. C. Haddon, *J. Phys. Chem. B* 105, 2525 (2001).
158. Y. P. Sun, W. J. Huang, Y. Lin, K. F. Fu, A. Kitaygorodskiy, L. A. Riddle, Y. J. Yu, and D. L. Carroll, *Chem. Mater.* 13, 2864 (2001).
159. B. Li, Z. J. Shi, Y. F. Lian, and Z. N. Gu, *Chem. Lett.* 598 (2001).
160. V. Georgakilas, N. Tagmatarchis, D. Pantarotto, A. Bianco, J. P. Briand, and M. Prato, *Chem. Commun.* 3050 (2002).
161. V. Georgakilas, D. Voulgaris, E. Vazquez, M. Prato, D. M. Guldi, A. Kukovecz, and H. Kuzmany, *J. Am. Chem. Soc.* 124, 14318 (2002).
162. B. Wu, J. Zhang, Z. Wei, S. M. Cai, and Z. F. Liu, *J. Phys. Chem. B* 105, 5075 (2001).
163. M. Sano, A. Kamino, J. Okamura, and S. Shinkai, *Langmuir* 17, 5125 (2001).
164. L. M. Dai and A. W. H. Mau, *Adv. Mater.* 13, 899 (2001).
165. K. T. Lau and D. Hui, *Compos. B* 33, 263 (2002).
166. W. Q. Han, Y. Bando, K. Kurashima, and T. Sato, *Appl. Phys. Lett.* 73, 3085 (1998).
167. W. Q. Han, J. Cumings, X. S. Huang, K. Bradley, and A. Zettl, *Chem. Phys. Lett.* 346, 368 (2001).
168. R. J. Chen, Y. G. Zhan, D. W. Wang, and H. J. Dai, *J. Am. Chem. Soc.* 123, 3838 (2001).
169. J. Chen, H. Y. Liu, W. A. Weimer, M. D. Halls, D. H. Waldeck, and G. C. Walker, *J. Am. Chem. Soc.* 124, 9034 (2002).
170. N. Nakashima, Y. Tomonari, and H. Murakami, *Chem. Lett.* 638 (2002).
171. R. S. Lee, H. J. Kim, J. E. Fischer, A. Thess, and R. E. Smalley, *Nature* 388, 255 (1997).
172. A. M. Rao, P. C. Eklund, S. Bandow, A. Thess, and R. E. Smalley, *Nature* 388, 257 (1997).
173. A. S. Claye, N. M. Nemes, A. Janossy, and J. E. Fischer, *Phys. Rev. B* 62, R4845 (2000).
174. S. Kazaoui, N. Minami, H. Kataura, and Y. Achiba, *Synth. Met.* 121, 1201 (2001).
175. A. Claye, S. Rahman, J. E. Fischer, A. Sirenko, G. U. Sumanasekera, and P. C. Eklund, *Chem. Phys. Lett.* 333, 16 (2001).
176. A. Kukovecz, T. Pichler, R. Pfeiffer, and H. Kuzmany, *Chem. Commun.* 1730 (2002).
177. C. W. Zhou, J. Kong, E. Yenilmez, and H. J. Dai, *Science* 290, 1552 (2000).
178. L. Kavan, P. Rapta, and L. Dunsch, *Chem. Phys. Lett.* 328, 363 (2000).
179. L. Kavan, P. Rapta, L. Dunsch, M. J. Bronikowski, P. Willis, and R. E. Smalley, *J. Phys. Chem. B* 105, 10764 (2001).
180. J. E. Fischer, *Acc. Chem. Res.* 35, 1079 (2002).
181. M. Takizawa, S. Bandow, M. Yudasaka, Y. Ando, H. Shimoyama, and S. Iijima, *Chem. Phys. Lett.* 326, 351 (2000).
182. M. Kanai, A. Koshio, H. Shinohara, T. Mieno, A. Kasuya, Y. Ando, and X. Zhao, *Appl. Phys. Lett.* 79, 2967 (2001).
183. C. L. Cheung, A. Kurtz, H. Park, and C. M. Lieber, *J. Phys. Chem. B* 106, 2429 (2002).
184. M. S. Dresselhaus, G. Dresselhaus, and R. Saito, *Carbon* 33, 883 (1995).
185. R. A. Jishi, D. Inomata, K. Nakao, M. S. Dresselhaus, and G. Dresselhaus, *J. Phys. Soc. Jpn.* 63, 2252 (1994).
186. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 50, 14698 (1994).
187. R. A. Jishi, L. Venkataraman, M. S. Dresselhaus, and G. Dresselhaus, *Chem. Phys. Lett.* 209, 77 (1993).
188. C. J. Mei and V. H. Smith, *Physica C* 213, 157 (1993).
189. Y. K. Kwon, D. Tomanek, Y. H. Lee, K. H. Lee, and S. Saito, *J. Mater. Res.* 13, 2363 (1998).
190. H. Hiura, T. W. Ebbesen, J. Fujita, K. Tanigaki, and T. Takada, *Nature* 367, 148 (1994).
191. J. C. Charlier, P. Lambin, and T. W. Ebbesen, *Phys. Rev. B* 54, R8377 (1996).
192. X. Blase, J. C. Charlier, A. DeVita, and R. Car, *Appl. Phys. Lett.* 70, 197 (1997).
193. A. A. Farajian, K. Ohno, K. Esfarjani, Y. Maruyama, and Y. Kawazoe, *J. Chem. Phys.* 111, 2164 (1999).
194. T. Kar, J. Pattanayak, and S. Scheiner, *J. Phys. Chem. A* 105, 10397 (2001).
195. C. Jo, C. Kim, and Y. H. Lee, *Phys. Rev. B* 6503, 035420 (2002).
196. S. H. Jhi, S. G. Louie, and M. L. Cohen, *Solid State Commun.* 123, 495 (2002).
197. A. Maiti, C. J. Brabec, C. Roland, and J. Bernholc, *Phys. Rev. B* 52, 14850 (1995).
198. J. C. Charlier, A. DeVita, X. Blase, and R. Car, *Science* 275, 646 (1997).
199. Y. H. Lee, S. G. Kim, and D. Tomanek, *Phys. Rev. Lett.* 78, 2393 (1997).
200. J. Bernholc, C. Brabec, M. B. Nardelli, A. Maiti, C. Roland, and B. I. Yakobson, *Appl. Phys. A* 67, 39 (1998).
201. C. Roland, J. Bernholc, C. Brabec, M. B. Nardelli, and A. Maiti, *Mol. Simul.* 25, 1 (2000).
202. Y. Kumeda, Y. Fukuhiro, T. Taketsugu, and T. Hirano, *Chem. Phys. Lett.* 333, 29 (2001).
203. J. Kurti, G. Kresse, and H. Kuzmany, *Phys. Rev. B* 58, R8869 (1998).
204. S. Reich, C. Thomsen, and P. Ordejon, *Phys. Rev. B* 6419, 195416 (2001).
205. M. Machon, S. Reich, C. Thomsen, D. Sanchez-Portal, and P. Ordejon, *Phys. Rev. B* 66, 155410 (2002).
206. J. Maultzsch, S. Reich, and C. Thomsen, *Phys. Rev. B* 65, 233402 (2002).
207. J. Kurti, V. Zolyomi, A. Gruneis, and H. Kuzmany, *Phys. Rev. B* 65, 165433 (2002).
208. O. Dubay, G. Kresse, and H. Kuzmany, *Phys. Rev. Lett.* 88, 235506 (2002).
209. K. Tada, S. Furuya, and K. Watanabe, *Phys. Rev. B* 6315, 155405 (2001).
210. G. E. Froudakis, *J. Phys. Condens. Matter* 14, R453 (2002).
211. S. P. Chan, G. Chen, X. G. Gong, and Z. F. Liu, *Phys. Rev. Lett.* 8720, 205502 (2001).
212. T. Yildirim, O. Gulseren, and S. Ciraci, *Phys. Rev. B* 6407, 075404 (2001).
213. O. Gulseren, T. Yildirim, and S. Ciraci, *Phys. Rev. B* 66, 121401 (2002).
214. E. C. Lee, Y. S. Kim, Y. G. Jin, and K. J. Chang, *Phys. Rev. B* 66, 073415 (2002).
215. J. J. Zhao, A. Buldum, J. Han, and J. P. Lu, *Nanotechnology* 13, 195 (2002).
216. F. Uhlik, Z. Slanina, and E. Osawa, *Mol. Mater.* 13, 231 (2000).
217. Y. K. Kwon, D. Tomanek, and S. Iijima, *Phys. Rev. Lett.* 82, 1470 (1999).
218. S. J. V. Frankland, A. Caglar, D. W. Brenner, and M. Griebel, *J. Phys. Chem. B* 106, 3046 (2002).
219. P. Piedigrosso, Z. Konya, J. F. Colomer, A. Fonseca, G. Van Tendeloo, and J. B. Nagy, *PCCP Phys. Chem. Chem. Phys.* 2, 163 (2000).
220. S. Subramoney, *Adv. Mater.* 10, 1157 (1998).
221. N. Koprinarov, M. Marinov, G. Pchelarov, M. Konstantinova, and R. Stefanov, *J. Phys. Chem.* 99, 2042 (1995).
222. Y. D. Li, J. L. Chen, Y. M. Ma, J. B. Zhao, Y. N. Qin, and L. Chang, *Chem. Commun.* 1141 (1999).
223. X. X. Zhang, Z. Q. Li, G. H. Wen, K. K. Fung, J. L. Chen, and Y. D. Li, *Chem. Phys. Lett.* 333, 509 (2001).
224. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996).
225. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14013 (1998).
226. T. W. Odom, J. H. Hafner, and C. M. Lieber, in "Topics in Applied Physics," 2001, Vol. 80, pp. 173–211.
227. Z. Zhang and C. M. Lieber, *Appl. Phys. Lett.* 62, 2792 (1993).

228. M. H. Ge and K. Sattler, *Appl. Phys. Lett.* 65, 2284 (1994).
229. L. P. Biro, S. Lazarescu, P. Lambin, P. A. Thiry, A. Fonseca, J. B. Nagy, and A. A. Lucas, *Phys. Rev. B* 56, 12490 (1997).
230. G. V. Dedkov and S. S. Rekhviashvili, *Tech. Phys.* 44, 982 (1999).
231. J. W. G. Wildoer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, *Nature* 391, 59 (1998).
232. T. W. Odom, J. L. Huang, P. Kim, and C. M. Lieber, *Nature* 391, 62 (1998).
233. A. Hassanien, M. Tokumoto, Y. Kumazawa, H. Kataura, Y. Maniwa, S. Suzuki, and Y. Achiba, *Appl. Phys. Lett.* 73, 3839 (1998).
234. W. Clauss, D. J. Bergeron, and A. T. Johnson, *Phys. Rev. B* 58, R4266 (1998).
235. V. Meunier and P. Lambin, *Phys. Rev. Lett.* 81, 5588 (1998).
236. C. H. Olk and J. P. Heremans, *J. Mater. Res.* 9, 259 (1994).
237. T. W. Odom, J. L. Huang, P. Kim, M. Ouyang, and C. M. Lieber, *J. Mater. Res.* 13, 2380 (1998).
238. J. P. Issi, L. Langer, J. Heremans, and C. H. Olk, *Carbon* 33, 941 (1995).
239. D. L. Carroll, P. Redlich, P. M. Ajayan, J. C. Charlier, X. Blase, A. DeVita, and R. Car, *Phys. Rev. Lett.* 78, 2811 (1997).
240. A. De Vita, J. C. Charlier, X. Blase, and R. Car, *Appl. Phys. A* 68, 283 (1999).
241. M. Ouyang, J. L. Huang, C. L. Cheung, and C. M. Lieber, *Science* 291, 97 (2001).
242. A. A. Maarouf, C. L. Kane, and E. J. Mele, *Phys. Rev. B* 61, 11156 (2000).
243. D. A. Walters, L. M. Ericson, M. J. Casavant, J. Liu, D. T. Colbert, K. A. Smith, and R. E. Smalley, *Appl. Phys. Lett.* 74, 3803 (1999).
244. J. P. Salvetat, J. M. Bonard, N. H. Thomson, A. J. Kulik, L. Forro, W. Benoit, and L. Zuppiroli, *Appl. Phys. A* 69, 255 (1999).
245. J. P. Salvetat, G. A. D. Briggs, J. M. Bonard, R. R. Bacsa, A. J. Kulik, T. Stockli, N. A. Burnham, and L. Forro, *Phys. Rev. Lett.* 82, 944 (1999).
246. J. Lefebvre, J. F. Lynch, M. Llaguno, M. Radosavljevic, and A. T. Johnson, *Appl. Phys. Lett.* 75, 3014 (1999).
247. L. C. Venema, J. W. G. Wildoer, H. Tuinstra, C. Dekker, A. G. Rinzler, and R. E. Smalley, *Appl. Phys. Lett.* 71, 2629 (1997).
248. C. L. Cheung, J. H. Hafner, T. W. Odom, K. Kim, and C. M. Lieber, *Appl. Phys. Lett.* 76, 3136 (2000).
249. H. J. Dai, J. H. Hafner, A. G. Rinzler, D. T. Colbert, and R. E. Smalley, *Nature* 384, 147 (1996).
250. J. A. Harrison, S. J. Stuart, D. H. Robertson, and C. T. White, *J. Phys. Chem. B* 101, 9682 (1997).
251. S. S. Wong, J. D. Harper, P. T. Lansbury, and C. M. Lieber, *J. Am. Chem. Soc.* 120, 603 (1998).
252. H. Nishijima, S. Kamo, S. Akita, Y. Nakayama, K. I. Hohmura, S. H. Yoshimura, and K. Takeyasu, *Appl. Phys. Lett.* 74, 4061 (1999).
253. T. Rhodin, K. Umemura, M. Gad, S. Jarvis, M. Ishikawa, and J. H. Fu, *Appl. Surf. Sci.* 188, 486 (2002).
254. E. B. Cooper, S. R. Manalis, H. Fang, H. Dai, K. Matsumoto, S. C. Minne, T. Hunt, and C. F. Quate, *Appl. Phys. Lett.* 75, 3566 (1999).
255. J. H. Hafner, C. L. Cheung, and C. M. Lieber, *J. Am. Chem. Soc.* 121, 9750 (1999).
256. C. L. Cheung, J. H. Hafner, and C. M. Lieber, *Proc. Nat. Acad. Sci. USA* 97, 3809 (2000).
257. T. Ono, H. Miyashita, and M. Esashi, *Nanotechnology* 13, 62 (2002).
258. Y. Shingaya, T. Nakayama, and M. Aono, *Physica B* 323, 153 (2002).
259. S. S. Wong, A. T. Wooley, E. Joselevich, and C. M. Lieber, *Chem. Phys. Lett.* 306, 219 (1999).
260. Y. L. Yang, J. Zhang, X. L. Nan, and Z. F. Liu, *J. Phys. Chem. B* 106, 4139 (2002).
261. V. T. Binh, P. Vincent, F. Feschet, and J. M. Bonard, *J. Appl. Phys.* 88, 3385 (2000).
262. D. Golberg, P. Dorozhkin, Y. Bando, M. Hasegawa, and Z. C. Dong, *Chem. Phys. Lett.* 359, 220 (2002).
263. S. M. Bachilo, M. S. Strano, C. Kittrell, R. Hauge, R. Smalley, and R. B. Weisman, *Science* 298, 2361 (2002).
264. H. Kataura, Y. Kumazawa, Y. Maniwa, I. Umezu, S. Suzuki, Y. Ohtsuka, and Y. Achiba, *Synth. Met.* 103, 2555 (1999).
265. X. Liu, T. Pichler, M. Knupfer, M. S. Golden, J. Fink, H. Kataura, and Y. Achiba, *Phys. Rev. B* 66, 045411 (2002).
266. S. Kazaoui, N. Minami, R. Jacquemin, H. Kataura, and Y. Achiba, *Phys. Rev. B* 60, 13339 (1999).
267. S. Kazaoui, N. Minami, N. Matsuda, H. Kataura, and Y. Achiba, *Appl. Phys. Lett.* 78, 3433 (2001).
268. P. C. Eklund, J. M. Holden, and R. A. Jishi, *Carbon* 33, 959 (1995).
269. R. Saito and H. Kataura, in "Topics in Applied Physics," 2001, Vol. 80, pp. 213–246.
270. R. Saito, T. Takeya, T. Kimura, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 57, 4145 (1998).
271. A. Kasuya, Y. Sasaki, Y. Saito, K. Tohji, and Y. Nishina, *Phys. Rev. Lett.* 78, 4434 (1997).
272. A. M. Rao, E. Richter, S. Bandow, B. Chase, P. C. Eklund, K. A. Williams, S. Fang, K. R. Subbaswamy, M. Menon, A. Thess, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *Science* 275, 187 (1997).
273. A. Jorio, A. G. Souza, G. Dresselhaus, M. S. Dresselhaus, A. K. Swan, M. S. Unlu, B. B. Goldberg, M. A. Pimenta, J. H. Hafner, C. M. Lieber, and R. Saito, *Phys. Rev. B* 65, 155412 (2002).
274. The Raman excitation laser is usually focused into a spot with diameter 1–2 mm (macroscopic measurements) or a few micrometers (microscopes). Even the latter is large enough to cover several SWCNTs in the sample. Thus, the measured Raman spectrum is really a weighted average of the various individual tube spectra. The averaging can only be avoided by sophisticated experiment design like depositing dilute SWCNT suspensions onto grids with microscopic markers or that used in the single tube measurements of Jorio et al.
275. P. H. Tan, Y. Tang, Y. M. Deng, F. Li, Y. L. Wei, and H. M. Cheng, *Appl. Phys. Lett.* 75, 1524 (1999).
276. L. Alvarez, A. Righi, S. Rols, E. Anglaret, J. L. Sauvajol, E. Munoz, W. K. Maser, A. M. Benito, M. T. Martinez, and G. F. de la Fuente, *Phys. Rev. B* 6315, 153401 (2001).
277. A. Jorio, A. G. Souza, G. Dresselhaus, M. S. Dresselhaus, R. Saito, J. H. Hafner, C. M. Lieber, F. M. Matinaga, M. S. S. Dantas, and M. A. Pimenta, *Phys. Rev. B* 6324, 245416 (2001).
278. U. D. Venkateswaran, A. M. Rao, E. Richter, M. Menon, A. Rinzler, R. E. Smalley, and P. C. Eklund, *Phys. Rev. B* 59, 10928 (1999).
279. L. Alvarez, A. Righi, T. Guillard, S. Rols, E. Anglaret, D. Laplaze, and J. L. Sauvajol, *Chem. Phys. Lett.* 316, 186 (2000).
280. D. Kahn and J. P. Lu, *Phys. Rev. B* 60, 6535 (1999).
281. S. Bandow, S. Asaka, Y. Saito, A. M. Rao, L. Grigorian, E. Richter, and P. C. Eklund, *Phys. Rev. Lett.* 80, 3779 (1998).
282. A. Jorio, R. Saito, J. H. Hafner, C. M. Lieber, M. Hunter, T. McClure, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 86, 1118 (2001).
283. M. Milnera, J. Kurti, M. Hulman, and H. Kuzmany, *Phys. Rev. Lett.* 84, 1324 (2000).
284. Z. H. Yu and L. E. Brus, *J. Phys. Chem. B* 105, 6831 (2001).
285. Y. F. Wang, X. W. Cao, S. F. Hu, Y. Y. Liu, and G. X. Lan, *Chem. Phys. Lett.* 336, 47 (2001).
286. M. Hulman, W. Plank, and H. Kuzmany, *Phys. Rev. B* 6308, 081406 (2001).
287. H. Kuzmany, W. Plank, M. Hulman, C. Kramberger, A. Gruneis, T. Pichler, H. Peterlik, H. Kataura, and Y. Achiba, *Eur. Phys. J. B* 22, 307 (2001).
288. C. Thomsen and S. Reich, *Phys. Rev. Lett.* 85, 5214 (2000).
289. J. Maultzsch, S. Reich, and C. Thomsen, *Phys. Rev. B* 6412, 121407 (2001).

290. A. G. Souza, A. Jorio, G. Dresselhaus, M. S. Dresselhaus, R. Saito, A. K. Swan, M. S. Unlu, B. B. Goldberg, J. H. Hafner, C. M. Lieber, and M. A. Pimenta, *Phys. Rev. B* 6503, 035404 (2002).
291. S. D. M. Brown, P. Corio, A. Marucci, M. A. Pimenta, M. S. Dresselhaus, and G. Dresselhaus, *Phys. Rev. B* 61, 7734 (2000).
292. A. G. Souza, A. Jorio, A. K. Swan, M. S. Unlu, B. B. Goldberg, R. Saito, J. H. Hafner, C. M. Lieber, M. A. Pimenta, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 65, 085417 (2002).
293. M. A. Pimenta, A. Marucci, S. D. M. Brown, M. J. Matthews, A. M. Rao, P. C. Eklund, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *J. Mater. Res.* 13, 2396 (1998).
294. P. Eklund, *Phys. Rev. B* 16, 3330 (1977).
295. P. Zhou, *Phys. Rev. B* 45, 10838 (1992).
296. S. D. M. Brown, A. Jorio, P. Corio, M. S. Dresselhaus, G. Dresselhaus, R. Saito, and K. Kneipp, *Phys. Rev. B* 6315, 155414 (2001).
297. E. D. Obraztsova, S. V. Terekhov, and A. V. Osadchy, in "XIVth International Winterschool on the Electronic Properties of Novel Materials" (H. Kuzmany, Ed.), Vol. 544, p. 276. Am. Inst. of Physics, Kirchberg, Austria, 2000.
298. H. D. Li, K. T. Yue, Z. L. Lian, Y. Zhan, L. X. Zhou, S. L. Zhang, Z. J. Shi, Z. N. Gu, B. B. Liu, R. S. Yang, H. B. Yang, G. T. Zou, Y. Zhang, and S. Iijima, *Appl. Phys. Lett.* 76, 2053 (2000).
299. S. V. Terekhov, E. D. Obraztsova, A. S. Lobach, and V. I. Konov, *Appl. Phys. A* 74, 393 (2002).
300. L. Alvarez, A. Righi, S. Rols, E. Anglaret, and J. L. Sauvajol, *Chem. Phys. Lett.* 320, 441 (2000).
301. S. D. M. Brown, P. Corio, A. Marucci, M. S. Dresselhaus, M. A. Pimenta, and K. Kneipp, *Phys. Rev. B* 61, R5137 (2000).
302. P. H. Tan, Y. Tang, C. Y. Hu, F. Li, Y. L. Wei, and H. M. Cheng, *Phys. Rev. B* 62, 5186 (2000).
303. E. Gregan, S. M. Keogh, T. G. Hedderman, B. McCarthy, G. Farrell, G. Chambers, and H. J. Byrne, in "XVIth Winterschool on the Electronic Properties of Novel Materials" (H. Kuzmany, Ed.), Vol. 633, p. 294. Am. Inst. of Physics, Kirchberg, Austria, 2002.
304. K. Kneipp, A. Jorio, H. Kneipp, S. D. M. Brown, K. Shafer, J. Motz, R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 6308, 081401 (2001).
305. K. Kneipp, L. T. Perelman, H. Kneipp, V. Backman, A. Jorio, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 6319, 193411 (2001).
306. H. Grebel, Z. Iqbal, and A. Lan, *Chem. Phys. Lett.* 348, 203 (2001).
307. J. Hwang, H. H. Gommans, A. Ugawa, H. Tashiro, R. Haggenmueller, K. I. Winey, J. E. Fischer, D. B. Tanner, and A. G. Rinzler, *Phys. Rev. B* 62, R13310 (2000).
308. A. Jorio, G. Dresselhaus, M. S. Dresselhaus, M. Souza, M. S. S. Dantas, M. A. Pimenta, A. M. Rao, R. Saito, C. Liu, and H. M. Cheng, *Phys. Rev. Lett.* 85, 2617 (2000).
309. A. Jorio, A. G. Souza, V. W. Brar, A. K. Swan, M. S. Unlu, B. B. Goldberg, A. Righi, J. H. Hafner, C. M. Lieber, R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 65, 121402 (2002).
310. J. Azoulay, A. Debarre, A. Richard, P. Tchenio, S. Bandow, and S. Iijima, *Europhys. Lett.* 53, 407 (2001).
311. T. R. Ravindran, B. R. Jackson, and J. V. Badding, *Chem. Mater.* 13, 4187 (2001).
312. P. Lambin and A. A. Lucas, *Phys. Rev. B* 56, 3571 (1997).
313. D. Bernaerts, A. Zettl, N. G. Chopra, A. Thess, and R. E. Smalley, *Solid State Commun.* 105, 145 (1998).
314. L. C. Qin, S. Iijima, H. Kataura, Y. Maniwa, S. Suzuki, and Y. Achiba, *Chem. Phys. Lett.* 268, 101 (1997).
315. J. F. Colomer, L. Henrard, P. Lambin, and G. Van Tendeloo, *Phys. Rev. B* 6412, 125425 (2001).
316. L. Henrard, A. Loiseau, C. Journet, and P. Bernier, *Eur. Phys. J. B* 13, 661 (2000).
317. E. Anglaret, S. Rols, and J. L. Sauvajol, *Phys. Rev. Lett.* 81, 4780 (1998).
318. S. Rols, R. Almairac, L. Henrard, E. Anglaret, and J. L. Sauvajol, *Eur. Phys. J. B* 10, 263 (1999).
319. S. Rols, E. Anglaret, J. L. Sauvajol, G. Coddens, and A. J. Dianoux, *Appl. Phys. A* 69, 591 (1999).
320. S. Rols, E. Anglaret, J. L. Sauvajol, G. Coddens, H. Schober, and A. J. Dianoux, *Physica B* 276, 276 (2000).
321. J. L. Sauvajol, E. Anglaret, S. Rols, and L. Alvarez, *Carbon* 40, 1697 (2002).
322. Z. H. Yu and L. Brus, *J. Phys. Chem. B* 105, 1123 (2001).
323. A. A. Puretzky, D. B. Geohegan, H. Schittenhelm, X. D. Fan, and M. A. Guillorn, *Appl. Surf. Sci.* 197, 552 (2002).
324. P. Petit, E. Jouguelet, J. E. Fischer, A. G. Rinzler, and R. E. Smalley, *Phys. Rev. B* 56, 9275 (1997).
325. X. P. Tang, A. Kleinhammes, H. Shimoda, L. Fleming, K. Y. Bennoune, S. Sinha, C. Bower, O. Zhou, and Y. Wu, *Science* 288, 492 (2000).
326. C. G. Bac, S. Latil, L. Vaccarini, P. Bernier, P. Gaveau, S. Tahir, V. Micholet, R. Aznar, A. Rubio, K. Metenier, and F. Beguin, *Phys. Rev. B* 6310, 100302 (2001).
327. C. Goze-Bac, S. Latil, P. Lauginie, V. Jourdain, J. Conard, L. Duclaux, A. Rubio, and P. Bernier, *Carbon* 40, 1825 (2002).
328. U. Kuhlmann, H. Jantoljak, N. Pfander, P. Bernier, C. Journet, and C. Thomsen, *Chem. Phys. Lett.* 294, 237 (1998).
329. U. Kuhlmann, H. Jantoljak, N. Pfander, C. Journet, P. Bernier, and C. Thomsen, *Synth. Met.* 103, 2506 (1999).
330. A. Ugawa, A. G. Rinzler, and D. B. Tanner, *Phys. Rev. B* 60, R11305 (1999).
331. A. Ugawa, A. G. Rinzler, and D. B. Tanner, *Ferroelectrics* 249, 145 (2001).
332. J. Kastner, T. Pichler, H. Kuzmany, S. Curran, W. Blau, D. N. Weldon, M. Delamesiere, S. Draper, and H. Zandbergen, *Chem. Phys. Lett.* 221, 53 (1994).
333. A. Kuznetsova, D. B. Mawhinney, V. Naumenko, J. T. Yates, J. Liu, and R. E. Smalley, *Chem. Phys. Lett.* 321, 292 (2000).
334. M. A. Hamon, H. Hui, P. Bhowmik, H. M. E. Itkis, and R. C. Haddon, *Appl. Phys. A* 74, 333 (2002).
335. M. A. Hamon, H. Hu, P. Bhowmik, S. Niyogi, B. Zhao, M. E. Itkis, and R. C. Haddon, *Chem. Phys. Lett.* 347, 8 (2001).
336. F. Pompeo and D. E. Resasco, *Nano Lett.* 2, 369 (2002).
337. W. Zhao, C. H. Song, B. Zheng, J. Liu, and T. Viswanathan, *J. Phys. Chem. B* 106, 293 (2002).
338. S. Banerjee and S. S. Wong, *Nano Lett.* 2, 49 (2002).
339. E. V. Basiuk, V. A. Basiuk, J. G. Banuelos, J. M. Saniger-Blesa, V. A. Pokrovskiy, T. Y. Gromovoy, A. V. Mischanchuk, and B. G. Mischanchuk, *J. Phys. Chem. B* 106, 1588 (2002).
340. O. E. Alon, *Phys. Rev. B* 6415, 153408 (2001).
341. S. L. Sung, S. H. Tsai, C. H. Tseng, F. K. Chiang, X. W. Liu, and H. C. Shih, *Appl. Phys. Lett.* 74, 197 (1999).
342. H. D. Klotz, R. Mach, F. Oleszak, H. E. Maneck, H. Goering, and K. W. Brzezinka, *Appl. Surf. Sci.* 179, 1 (2001).
343. R. Kuzuo, M. Terauchi, M. Tanaka, and Y. Saito, *Jpn. J. Appl. Phys. 2 Lett.* 33, L1316 (1994).
344. R. Kuzuo, M. Terauchi, and M. Tanaka, *Jpn. J. Appl. Phys. 2 Lett.* 31, L1484 (1992).
345. M. Knupfer, T. Pichler, M. S. Golden, J. Fink, A. Rinzler, and R. E. Smalley, *Carbon* 37, 733 (1999).
346. C. Bower, S. Suzuki, K. Tanigaki, and O. Zhou, *Appl. Phys. A* 67, 47 (1998).
347. S. Suzuki, C. Bower, and O. Zhou, *Chem. Phys. Lett.* 285, 230 (1998).
348. T. Pichler, M. Knupfer, M. S. Golden, J. Fink, A. Rinzler, and R. E. Smalley, *Synth. Met.* 103, 2515 (1999).
349. T. Pichler, M. Sing, M. Knupfer, M. S. Golden, and J. Fink, *Solid State Commun.* 109, 721 (1999).
350. V. Lordi, N. Yao, and J. Wei, *Chem. Mater.* 13, 733 (2001).
351. J. Mittal, M. Monthioux, H. Allouche, and O. Stephan, *Chem. Phys. Lett.* 339, 311 (2001).
352. D. Golberg, Y. Bando, W. Han, K. Kurashima, and T. Sato, *Chem. Phys. Lett.* 308, 337 (1999).

353. D. Golberg, Y. Bando, K. Kurashima, and T. Sato, *Chem. Phys. Lett.* 323, 185 (2000).
354. D. Golberg, Y. Bando, M. Mitome, K. Kurashima, N. Grobert, M. Reyes-Reyes, H. Terrones, and M. Terrones, *Chem. Phys. Lett.* 360, 1 (2002).
355. M. Eswaramoorthy, R. Sen, and C. N. R. Rao, *Chem. Phys. Lett.* 304, 207 (1999).
356. A. Fujiwara, K. Ishii, H. Suematsu, H. Kataura, Y. Maniwa, S. Suzuki, and Y. Achiba, *Chem. Phys. Lett.* 336, 205 (2001).
357. D. H. Yoo, G. H. Rue, Y. H. Hwang, and H. K. Kim, *J. Phys. Chem. B* 106, 3371 (2002).
358. C. M. Yang, H. Kanoh, K. Kaneko, M. Yudasaka, and S. Iijima, *J. Phys. Chem. B* 106, 8994 (2002).
359. W. F. Du, L. Wilson, J. Ripmeester, R. Dutrisac, B. Simard, and S. Denommee, *Nano Lett.* 2, 343 (2002).
360. D. H. Yoo, G. H. Rue, J. Y. Seo, Y. H. Hwang, M. H. W. Chan, and H. K. Kim, *J. Phys. Chem. B* 106, 9000 (2002).
361. M. Muris, N. Dufau, M. Bienfait, N. Dupont-Pavlovsky, Y. Grillet, and J. P. Palmari, *Langmuir* 16, 7019 (2000).
362. A. Kuznetsova, J. T. Yates, J. Liu, and R. E. Smalley, *J. Chem. Phys.* 112, 9590 (2000).
363. A. Kuznetsova, J. T. Yates, V. V. Simonyan, J. K. Johnson, C. B. Huffman, and R. E. Smalley, *J. Chem. Phys.* 115, 6691 (2001).
364. M. Muris, N. Dupont-Pavlovsky, M. Bienfait, and P. Zeppenfeld, *Surf. Sci.* 492, 67 (2001).
365. A. J. Zambano, S. Talapatra, and A. D. Migone, *Phys. Rev. B* 6407, 075415 (2001).
366. S. E. Weber, S. Talapatra, C. Journet, Z. Zambano, and A. D. Migone, *Phys. Rev. B* 61, 13150 (2000).
367. S. Talapatra and A. D. Migone, *Phys. Rev. B* 6504, 045416 (2002).
368. G. Stan, M. J. Bojan, S. Curtarolo, S. M. Gatica, and M. W. Cole, *Phys. Rev. B* 62, 2173 (2000).
369. R. R. Bacsa, C. Laurent, A. Peigney, W. S. Bacsa, T. Vaugien, and A. Rousset, *Chem. Phys. Lett.* 323, 566 (2000).
370. T. Ohba and K. Kaneko, *J. Phys. Chem. B* 106, 7171 (2002).
371. Z. Gu, H. Peng, R. H. Hauge, R. E. Smalley, and J. L. Margrave, *Nano Lett.* 2, 1009 (2002).
372. J. B. Nagy and A. Fonseca, *J. Phys. IV* 11, 411 (2001).
373. S. Lebedkin, P. Schweiss, B. Renker, S. Malik, F. Hennrich, M. Neumaier, C. Stoermer, and M. M. Kappes, *Carbon* 40, 417 (2002).
374. B. Ruzicka, L. Degiorgi, R. Gaal, L. Thien-Nga, R. Bacsa, J. P. Salvetat, and L. Forro, *Phys. Rev. B* 61, R2468 (2000).
375. O. Hilt, H. B. Brom, and M. Ahlskog, *Phys. Rev. B* 61, R5129 (2000).
376. A. Peigney, C. Laurent, E. Flahaut, R. R. Bacsa, and A. Rousset, *Carbon* 39, 507 (2001).
377. J. Hone, B. Batlogg, Z. Benes, A. T. Johnson, and J. E. Fischer, *Science* 289, 1730 (2000).
378. J. Hone, M. Whitney, and A. Zettl, *Synth. Met.* 103, 2498 (1999).
379. J. Hone, M. C. Llaguno, N. M. Nemes, A. T. Johnson, J. E. Fischer, D. A. Walters, M. J. Casavant, J. Schmidt, and R. E. Smalley, *Appl. Phys. Lett.* 77, 666 (2000).
380. H. C. F. Martens, O. Hilt, H. B. Brom, D. M. de Leeuw, and R. Menon, *Synth. Met.* 119, 423 (2001).
381. X. K. Wang, X. W. Lin, S. N. Song, V. P. Dravid, J. B. Ketterson, and R. P. H. Chang, *Carbon* 33, 949 (1995).
382. M. Radosavljevic, J. Lefebvre, and A. T. Johnson, *Phys. Rev. B* 6424, 241307 (2001).
383. Z. Yao, C. L. Kane, and C. Dekker, *Phys. Rev. Lett.* 84, 2941 (2000).
384. S. M. Bachilo, M. S. Strano, C. Kittrell, R. H. Hauge, R. E. Smalley, and R. B. Weisman, *Science* 298, 2361 (2002).
385. M. Yudasaka, M. Zhang, C. Jabs, and S. Iijima, *Appl. Phys. A* 71, 449 (2000).
386. D. A. Walters, M. J. Casavant, X. C. Qin, C. B. Huffman, P. J. Boul, L. M. Ericson, E. H. Haroz, M. J. O'Connell, K. Smith, D. T. Colbert, and R. E. Smalley, *Chem. Phys. Lett.* 338, 14 (2001).
387. V. Krstic, G. S. Duesberg, J. Muster, M. Burghard, and S. Roth, *Chem. Mater.* 10, 2338 (1998).
388. H. Ago, T. Kugler, F. Cacialli, K. Petritsch, R. H. Friend, W. R. Salaneck, Y. Ono, T. Yamabe, and K. Tanaka, *Synth. Met.* 103, 2494 (1999).
389. G. S. Duesberg, W. Blau, H. J. Byrne, J. Muster, M. Burghard, and S. Roth, *Synth. Met.* 103, 2484 (1999).
390. B. I. Yakobson and P. Avouris, in "Topics in Applied Physics," 2001, Vol. 80, pp. 287–327.
391. J. P. Salvetat-Delmotte and A. Rubio, *Carbon* 40, 1729 (2002).
392. M. Terrones, W. K. Hsu, H. W. Kroto, and D. R. M. Walton, in "Fullerenes and Related Structures," 1999, Vol. 199, pp. 189–234.
393. S. A. Chesnokov, V. A. Nalimova, A. G. Rinzler, R. E. Smalley, and J. E. Fischer, *Phys. Rev. Lett.* 82, 343 (1999).
394. C. T. White and T. N. Todorov, *Nature* 393, 240 (1998).
395. S. Frank, P. Poncharal, Z. L. Wang, and W. A. de Heer, *Science* 280, 1744 (1998).
396. J. Kong, E. Yenilmez, T. W. Tombler, W. Kim, H. J. Dai, R. B. Laughlin, L. Liu, C. S. Jayanthi, and S. Y. Wu, *Phys. Rev. Lett.* 8710, 106801 (2001).
397. M. Bockrath, D. H. Cobden, J. Lu, A. G. Rinzler, R. E. Smalley, T. Balents, and P. L. McEuen, *Nature* 397, 598 (1999).
398. Z. Yao, H. W. C. Postma, L. Balents, and C. Dekker, *Nature* 402, 273 (1999).
399. Z. Yao, C. Dekker, and P. Avouris, in "Topics in Applied Physics," 2001, Vol. 80, pp. 147–171.
400. S. Roth, V. Krstic, and G. Rikken, *Curr. Appl. Phys.* 2, 155 (2002).
401. M. Shiraishi and M. Ata, *Synth. Met.* 128, 235 (2002).
402. P. J. de Pablo, C. Gomez-Navarro, J. Colchero, P. A. Serena, J. Gomez-Herrero, and A. M. Baro, *Phys. Rev. Lett.* 8803, 036804 (2002).
403. J. Hone, in "Carbon Nanotubes," 2001, Vol. 80, pp. 273–286.
404. J. C. Lasjaunias, K. Biljakovic, Z. Benes, and J. E. Fischer, *Physica B* 316, 468 (2002).
405. J. Hone, M. C. Llaguno, M. J. Biercuk, A. T. Johnson, B. Batlogg, Z. Benes, and J. E. Fischer, *Appl. Phys. A* 74, 339 (2002).
406. J. C. Lasjaunias, K. Biljakovic, Z. Benes, J. E. Fischer, and P. Monceau, *Phys. Rev. B* 65, 113409 (2002).
407. X. G. Wan, J. M. Dong, J. Jiang, and D. Y. Xing, *Phys. Status Solidi B* 199, 571 (1997).
408. X. G. Wan, J. M. Dong, and D. Y. Xing, *Phys. Rev. B* 58, 6756 (1998).
409. J. Jiang, J. M. Dong, and D. Y. Xing, *Phys. Rev. B* 59, 9838 (1999).
410. L. Vivien, P. Lancon, D. Riehl, F. Hache, and E. Anglaret, *Carbon* 40, 1789 (2002).
411. R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, *Science* 297, 787 (2002).
412. R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, *Appl. Phys. Lett.* 73, 2447 (1998).
413. C. W. Zhou, J. Kong, and H. J. Dai, *Appl. Phys. Lett.* 76, 1597 (2000).
414. X. L. Liu, C. Lee, C. W. Zhou, and J. Han, *Appl. Phys. Lett.* 79, 3329 (2001).
415. F. Leonard and J. Tersoff, *Phys. Rev. Lett.* 88, 258302 (2002).
416. V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, *Nano Lett.* 1, 453 (2001).
417. A. Javey, Q. Wang, A. Ural, Y. M. Li, and H. J. Dai, *Nano Lett.* 2, 929 (2002).
418. T. Rueckes, K. Kim, E. Joselevich, G. Y. Tseng, C. L. Cheung, and C. M. Lieber, *Science* 289, 94 (2000).
419. S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and P. Avouris, *Phys. Rev. Lett.* 89, 106801 (2002).
420. P. Avouris, *Chem. Phys.* 281, 429 (2002).
421. P. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, *Science* 287, 1801 (2000).
422. J. Kong, N. R. Franklin, C. W. Zhou, M. G. Chapline, S. Peng, K. J. Cho, and H. J. Dai, *Science* 287, 622 (2000).

423. F. Balavoine, P. Schultz, C. Richard, V. Mallouh, T. W. Ebbesen, and C. Mioskowski, *Angew. Chem. Int. Ed.* 38, 1912 (1999).
424. A. Guiseppi-Elie, C. H. Lei, and R. H. Baughman, *Nanotechnology* 13, 559 (2002).
425. H. G. Xue, W. L. Sun, B. J. He, and Z. Q. Shen, *Chin. Chem. Lett.* 13, 799 (2002).
426. J. R. Wood and H. D. Wagner, *Appl. Phys. Lett.* 76, 2883 (2000).
427. J. R. Wood, Q. Zhao, M. D. Frogley, E. R. Meurs, A. D. Prins, T. Peijs, D. J. Dunstan, and H. D. Wagner, *Phys. Rev. B* 62, 7571 (2000).
428. L. M. Dai, P. Soundarrajan, and T. Kim, *Pure Appl. Chem.* 74, 1753 (2002).
429. W. A. Deheer, A. Chatelain, and D. Ugarte, *Science* 270, 1179 (1995).
430. A. G. Rinzler, J. H. Hafner, P. Nikolaev, L. Lou, S. G. Kim, D. Tomanek, P. Nordlander, D. T. Colbert, and R. E. Smalley, *Science* 269, 1550 (1995).
431. Q. H. Wang, A. A. Setlur, J. M. Lauerhaas, J. Y. Dai, E. W. Seelig, and R. P. H. Chang, *Appl. Phys. Lett.* 72, 2912 (1998).
432. W. B. Choi, D. S. Chung, J. H. Kang, H. Y. Kim, Y. W. Jin, I. T. Han, Y. H. Lee, J. E. Jung, N. S. Lee, G. S. Park, and J. M. Kim, *Appl. Phys. Lett.* 75, 3129 (1999).
433. P. G. Collins and A. Zettl, *Phys. Rev. B* 55, 9391 (1997).
434. W. A. Deheer, W. S. Bacsa, A. Chatelain, T. Gerfin, R. Humphrey-baker, L. Forro, and D. Ugarte, *Science* 268, 845 (1995).
435. M. Terrones, N. Grobert, J. Olivares, J. P. Zhang, H. Terrones, K. Kordatos, W. K. Hsu, J. P. Hare, P. D. Townsend, K. Prassides, A. K. Cheetham, H. W. Kroto, and D. R. M. Walton, *Nature* 388, 52 (1997).
436. Z. F. Ren, Z. P. Huang, J. W. Xu, J. H. Wang, P. Bush, M. P. Siegal, and P. N. Provencio, *Science* 282, 1105 (1998).
437. C. C. Zhu, W. H. Liu, and L. J. Huangfu, *J. Vac. Sci. Technol. B* 19, 1691 (2001).
438. Y. R. Cho, J. H. Lee, Y. H. Song, S. Y. Kang, M. Y. Jung, C. S. Hwang, and K. I. Cho, *J. Vac. Sci. Technol. B* 19, 1012 (2001).
439. L. Schlapbach and A. Zuttel, *Nature* 414, 353 (2001).
440. P. M. Ajayan and T. W. Ebbesen, *Rep. Progr. Phys.* 60, 1025 (1997).
441. S. M. Lee, K. S. Park, Y. C. Choi, Y. S. Park, J. M. Bok, D. J. Bae, K. S. Nahm, Y. G. Choi, S. C. Yu, N. G. Kim, T. Frauenheim, and Y. H. Lee, *Synth. Met.* 113, 209 (2000).
442. S. M. Lee and Y. H. Lee, *Appl. Phys. Lett.* 76, 2877 (2000).
443. S. Talapatra, A. Z. Zambano, S. E. Weber, and A. D. Migone, *Phys. Rev. Lett.* 85, 138 (2000).
444. Y. Ren and D. L. Price, *Appl. Phys. Lett.* 79, 3684 (2001).
445. Q. Y. Wang and J. K. Johnson, *J. Chem. Phys.* 110, 577 (1999).
446. Q. Y. Wang and J. K. Johnson, *J. Phys. Chem. B* 103, 4809 (1999).
447. Y. C. Ma, Y. Y. Xia, M. W. Zhao, R. J. Wang, and L. M. Mei, *Phys. Rev. B* 6311, 115422 (2001).
448. C. Liu, Y. Y. Fan, M. Liu, H. T. Cong, H. M. Cheng, and M. S. Dresselhaus, *Science* 286, 1127 (1999).
449. N. Nishimiya, K. Ishigaki, H. Takikawa, M. Ikeda, Y. Hibi, T. Sakakibara, A. Matsumoto, and K. Tsutsumi, *J. Alloy Compd.* 339, 275 (2002).
450. C. Liu, Q. H. Yang, Y. Tong, H. T. Cong, and H. M. Cheng, *Appl. Phys. Lett.* 80, 2389 (2002).
451. Y. Ye, C. C. Ahn, C. Witham, B. Fultz, J. Liu, A. G. Rinzler, D. Colbert, K. A. Smith, and R. E. Smalley, *Appl. Phys. Lett.* 74, 2307 (1999).
452. P. Chen, X. Wu, J. Lin, and K. L. Tan, *Science* 285, 91 (1999).
453. G. E. Froudakis, *Nano Lett.* 1, 531 (2001).
454. R. T. Yang, *Carbon* 38, 623 (2000).
455. M. S. Dresselhaus, K. A. Williams, and P. C. Eklund, *Mater. Res. Soc. Bull.* 24, 45 (1999).
456. J. E. Fischer, *Chem. Innov.* 30, 21 (2000).
457. R. G. Ding, G. Q. Lu, Z. F. Yan, and M. A. Wilson, *J. Nanosci. Nanotechnol.* 1, 7 (2001).
458. C. Zandonella, *Nature* 410, 734 (2001).
459. E. V. Barrera, *JOM-J. Miner. Met. Mater. Soc.* 52, A38 (2000).
460. D. Qian, E. C. Dickey, R. Andrews, and T. Rantell, *Appl. Phys. Lett.* 76, 2868 (2000).
461. M. J. Biercuk, M. C. Llaguno, M. Radosavljevic, J. K. Hyun, A. T. Johnson, and J. E. Fischer, *Appl. Phys. Lett.* 80, 2767 (2002).
462. C. A. Cooper, S. R. Cohen, A. H. Barber, and H. D. Wagner, *Appl. Phys. Lett.* 81, 3873 (2002).
463. C. A. Mitchell, J. L. Bahr, S. Arepalli, J. M. Tour, and R. Krishnamoorti, *Macromolecules* 35, 8825 (2002).
464. Y. Breton, S. Delpeux, R. Benoit, J. P. Salvetat, C. Sinturel, F. Beguin, S. Bonnamy, G. Desarmot, and L. Boufendi, *Mol. Cryst. Liq. Cryst.* 387, 359 (2002).
465. R. H. Xie and J. Jiang, *Appl. Phys. Lett.* 71, 1029 (1997).
466. S. R. Mishra, H. S. Rawat, S. C. Mehendale, K. C. Rustagi, A. K. Sood, R. Bandyopadhyay, A. Govindaraj, and C. N. R. Rao, *Chem. Phys. Lett.* 317, 510 (2000).
467. L. Vivien, D. Riehl, F. Hache, and E. Anglaret, *Physica B* 323, 233 (2002).
468. J. M. Planeix, N. Coustel, B. Coq, V. Brotons, P. S. Kumbhar, R. Dutartre, P. Geneste, P. Bernier, and P. M. Ajayan, *J. Am. Chem. Soc.* 116, 7935 (1994).
469. B. Coq, J. M. Planeix, and V. Brotons, *Appl. Catal. A* 173, 175 (1998).
470. Y. Zhang, H. B. Zhang, G. D. Lin, P. Chen, Y. Z. Yuan, and K. R. Tsai, *Appl. Catal. A* 187, 213 (1999).
471. M. Freemantle, *Chem. Eng. News* 74, 62 (1996).
472. D. T. Colbert and R. E. Smalley, *Trends Biotechnol.* 17, 46 (1999).
473. M. Musameh, J. Wang, A. Merkoci, and Y. H. Lin, *Electrochem. Commun.* 4, 743 (2002).
474. Y. D. Zhao, W. D. Zhang, H. Chen, and Q. M. Luo, *Anal. Sci.* 18, 939 (2002).

www.aspbs.com/enn

Sliding, Rotation, and Rolling of Nanoparticles

K. Miura

Aichi University of Education, Kariya, Japan

CONTENTS

1. Introduction
2. Sliding, Rotation, and Rolling of Single-Walled Carbon Nanotubes (SWNTs) and Multiwalled Carbon Nanotubes (MWNTs)
3. Sliding of Flakes
4. C_{60} Molecular Bearings
Glossary
References

1. INTRODUCTION

This chapter describes the sliding, rotation, and rolling of nanoparticles on the nanometer scale using scanning probe microscopy, in particular, an atomic force microscope (AFM) and a frictional force microscope (FFM). Nanoparticles described in this chapter are single-walled carbon nanotubes (SWNTs), multiwalled carbon nanotubes (MWNTs), graphite flakes, and MoS_2 flakes. Motions of nanoparticles on the nanometer scale depend strongly on the substrates where nanoparticles perform, exhibiting translational and zigzag sliding, rotation around a pivot, and rolling. The sliding, rotation, and rolling of nanoparticles exhibit novel phenomena, which are different from macroscopic events.

There exist tremendous developments of scanning probe microscopy including an AFM and a scanning tunneling microscope, in the background where one can study the movement of nanoparticles on the nanometer scale. Over the past decade, atomic and molecular manipulations have become possible through techniques derived from scanning probe microscopy, which makes a tip travel on a surface on the order of an angstrom and controls arbitrarily the motions of atoms and molecules. Thus, the technique makes nano- and microlithography possible, thereby leading to applicability for nano- and microdevices. Furthermore, simultaneous measurements of forces during manipulation give crucial knowledge of the changes of materials occurring on the nanometer scale. Hence, it becomes possible to investigate the similarity and difference of behaviors between nanoscopic and macroscopic particles.

The forces needed to manipulate atoms and molecules are consequently equivalent to the lateral force (or frictional force) of atoms and molecules on the nanometer scale when moving atoms and molecules along the surface. Hence, this study leads to that of the friction on the nanometer scale. It is interesting to note that the relationship between nanoscopic and macroscopic sliding, rotation, and rolling is similar to that between nanoscopic and macroscopic friction.

In the second section, sliding, rotation, and rolling of SWNTs and MWNTs are described, where the sliding of the SWNTs on graphite does not occur because the SWNTs form a bundle of self-assembled cables and a nanotube array in the bundle contacts with graphite. In contrast, the SWNTs on a KCl(001) surface slide easily. Natural rolling of MWNTs appears in commensurate contact with a graphite surface. This strongly means that the substrate where nanotubes perform influences their movements.

In the third section, it is also shown that graphite and MoS_2 flakes move so as to keep in commensurate contact with a substrate. The graphite flake moves on graphite such that the stacking of graphite layers is maintained. The MoS_2 flake also moves on MoS_2 such that the stacking of the sulfur basal plane of MoS_2 is maintained. In the case where the flake movement is parallel to the pulling direction, the frictional force is smallest. Furthermore, the appearance of the zigzag movement would originate in the rotation of the flake around a pivot point in addition to translation.

In the last section, C_{60} molecular bearings are described. In this section, the applicability and realization of C_{60} molecular bearings, and the detailed mechanism of their movement, are described. C_{60} molecules grow on graphite such that the hexagonal face of a C_{60} molecule pairs with the hexagonal face of the second graphite layer, so as to continue the natural stacking of the graphite. Thus, we can construct a system confining a C_{60} monolayer between graphite plates, which forms the nanogears of six-membered rings between C_{60} molecules and graphite. C_{60} molecules roll by their nanogears. Then, the novel frictional mechanism is realized in the graphite/C_{60} monolayer/graphite system, in which static frictional forces have a finite value but mean dynamical frictional forces are zero. It is emphasized that

ISBN: 1-58883-065-9/$35.00
Copyright © 2004 by American Scientific Publishers
All rights of reproduction in any form reserved.

the graphite/C_{60} monolayer/graphite system is very promising for the realization of nano- and micromachines.

2. SLIDING, ROTATION, AND ROLLING OF SINGLE-WALLED CARBON NANOTUBES (SWNTs) AND MULTIWALLED CARBON NANOTUBES (MWNTs)

Carbon nanotubes have attracted increasing interest due to their mechanical as well as electronic properties. Because of their high tensile strength, flexural rigidity, and high aspect ratio, they are promising candidates for strong fibers. SWNTs are known to be arranged mostly in the form of fibers with closely packed stacking [1], which thus form self-assembled cables on the nanometer scale. The elastic and shear moduli of SWNTs are measured to be of the order of 1 TPa and 1 GPa, respectively [2]. Consequently, we expect large elastic and shear moduli for SWNTs. Because of their shapes, which are similar to those of C_{60} and C_{70} fullerenes, the nanotubes are also promising as an ideal means of lubrication. We need to control the alignment of nanotubes to enable their use as a lubricant. Sliding and rolling behaviors of nanotubes including fullerene molecules depend strongly on their interactions with the substrate [3–8]. For example, SWNTs are bent on a KCl(001) surface but are aligned on graphite resulting in a threefold symmetry [8]. This is the reason why six carbon rings of the nanotube pair with those at the basal graphite surface, in the same manner as the stacking of graphite layers. In this case, the sliding of the SWNTs on graphite does not occur because the SWNTs form a bundle of self-assembled cables and a nanotube array in the bundle contacts with graphite. In contrast, the SWNTs on a KCl(001) surface slide easily, resulting in a shear stress of 3 MPa. Falvo et al. [3, 4] show the stick-slip lateral force behavior for rolling of MWNTs. Let us show in detail sliding, rotation, and rolling of SWNTs and MWNTs in this section.

2.1. Bundle Structure of SWNTs and their Sliding

2.1.1. Bundle Structure of SWNTs

FFM images of nanotubes on the KCl(001) surface and of those on the graphite (highly oriented pyrolytic graphite: HOPG) surface are shown in Figure 1a and b, respectively. An appreciable difference between these images appears in the configuration of the nanotubes on the substrate. Many nanotubes on the KCl(001) surface (Fig. 1a) are usually observed to be bent. In contrast, many nanotubes on the graphite surface (Fig. 1b) are aligned. The periodic lattice image of the graphite substrate is shown in the inset of the FFM image in Figure 1b. The crystallographic direction of the graphite surface is estimated from the periodic lattice image (which roughly means hollow site contrast) of the graphite [9, 10] and is illustrated on the right-hand side (top) of Figure 1b. The angle between the axial direction of the nanotubes and the $[12\bar{3}0]$ direction of the graphite substrate is estimated to be 60°. The geometrical relationship between the nanotubes and the graphite substrate is found

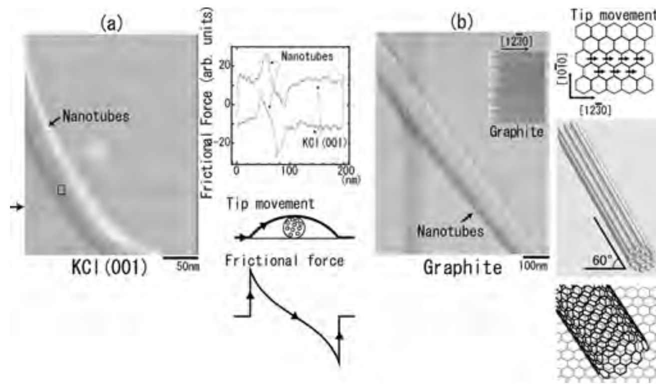


Figure 1. FFM images of nanotubes on a KCl(001) surface (a) and on a graphite surface (b). The periodic lattice image of the graphite substrate is shown in the inset of the FFM image in (b). The line profile (indicated by an arrow) of the FFM image of the nanotubes on KCl(001) is shown on the right-hand side (top) of (a). The line profile consists of components from the nanotubes and the KCl(001) substrate. The FFM contrast due to surface corrugations is described by the slope (on the right (bottom)) of tip movement (on the right (middle)).

to exhibit threefold symmetry from many samples. It is of interest to note that the angle between the axial direction of the nanotubes and the $[12\bar{3}0]$ direction of the graphite substrate becomes 60°, when the lower hexagonal faces of nanotubes of the armchair type become commensurate with the hexagonal faces of the second graphite layer of the graphite substrate, in the same manner as the stacking of graphite layers, where only a single nanotube in a bundle is illustrated on the right-hand side (bottom) of Figure 1b. This consideration is consistent with the result that the hexagonal face of a C_{60} molecule pairs with the hexagonal face of the second graphite layer [11, 12]. The line profile of the FFM image of the nanotubes on KCl(001) is shown on the right-hand side of Figure 1a. The line profile consists of components from the nanotubes and KCl(001) substrate, as illustrated in the figure. The FFM contrast consists of a component produced by surface corrugations without energy dissipation and one produced by intrinsic friction with energy dissipation [13]. The contrast due to surface corrugations is described by the slope of tip movement. Assuming the tip shape and the shape of nanotubes to be a sphere and a cylinder, respectively, the tip movement and its FFM contrast when the load is constant are illustrated in the figure.

An enlarged image of the local area (square) shown in Figure 1a is shown in Figure 2. The enlarged FFM image exhibits stripe structures representing the periodic spacing w of 2.1 nm along the scanning direction. Thus, the periodic spacing between stripes a is estimated to be 1.8 nm using $a = w\sin\theta$, where $\theta = 60.1°$ is the angle between the axial direction of the nanotubes and the scanning direction. The line profile in Figure 2 can be discussed in the same manner as Figure 1a and indicates nanotubes of the same diameter. Assuming the bundle of closely packed SWNTs to be as shown on the right-hand side of Figure 2, the height of the bundle H is given as follows using the radius of SWNT, R:

$$H = 2R + \frac{\sqrt{3}}{2}a(\mathrm{N}-1) = \left(\frac{\sqrt{3}}{2}(\mathrm{N}-1)+1\right)a - b \quad (1)$$

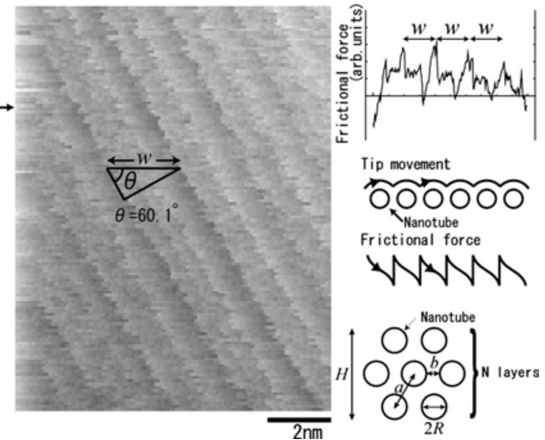


Figure 2. Enlarged image of the local area (square) shown in Fig. 1a. The enlarged FFM image exhibits stripe structures representing the periodic spacing w of 2.1 nm along the scanning direction. Thus, the periodic spacing between stripes a is estimated to be 1.8 nm using $a = w \sin\theta$, where $\theta = 60.1°$ is the angle between the axial direction of the nanotubes and the scanning direction. The line profile (indicated by an arrow) in the figure can be discussed in the same manner as in Fig. 1a and indicates nanotubes of the same diameter. A schematic view of the closely packed SWNTs is shown on the right (bottom), where $2R$, N, a, and b are the diameter of a nanotube, the layer number of nanotubes, the nearest distance between any two nanotube centers, and the nearest distance between any two nanotubes, respectively.

where N, a, and b are the layer number of nanotubes, the nearest distance between any two nanotube centers, and the nearest distance between any two nanotubes, respectively. Here, H and a were estimated to be 9.2 nm based on AFM data and 1.8 nm based on the periodic spacing of stripes, respectively. Thus, taking N = 6, the diameter of the nanotubes $2R$ and the nearest distance between any two nanotubes b are estimated to be 1.4 nm and 0.3 nm, respectively.

2.1.2. Rotation of SWNTs

Next, the sliding of the nanotubes on the KCl(001) surface and the graphite surface is described. Only on the KCl(001) surface is the sliding of the nanotubes observed. The FFM image that shows the sliding of the nanotubes on KCl(001) is shown in Figure 3. As in Figure 2, stripe features due to nanotubes are again observed. At point A, sliding is found to occur because the stripe features do not change. If rolling occurs, the striped rows of nanotubes will not coincide at sliding point A with the increase in sliding distance. The sliding distance d and the rotation angle around pivot point B, θ, are estimated to be 6.8 nm and 3.5°, respectively. Thus, the distance between sliding point A and pivot point B, l, is estimated to be 111.2 nm using $l = d/\tan\theta$. Then, the frictional force during sliding is estimated to be about 11 nN from scan line (b) where sliding occurs. Thus, the threshold value of the torque for rotation around pivot point B is estimated to be 1.2×10^{-15} Nm (11 nN × 111.2 nm). The energy loss for rotation, or the energy dissipation, is also estimated to be 0.75×10^{-16} J (11 nN × 6.8 nm), which is much smaller than that in the case of MWNTs on mica (3×10^{-16} J) [3]. This sliding occurs very often on the KCl(001) surface but not on graphite. Thus, this demonstrates that the nanotube-KCl(001) shear stress

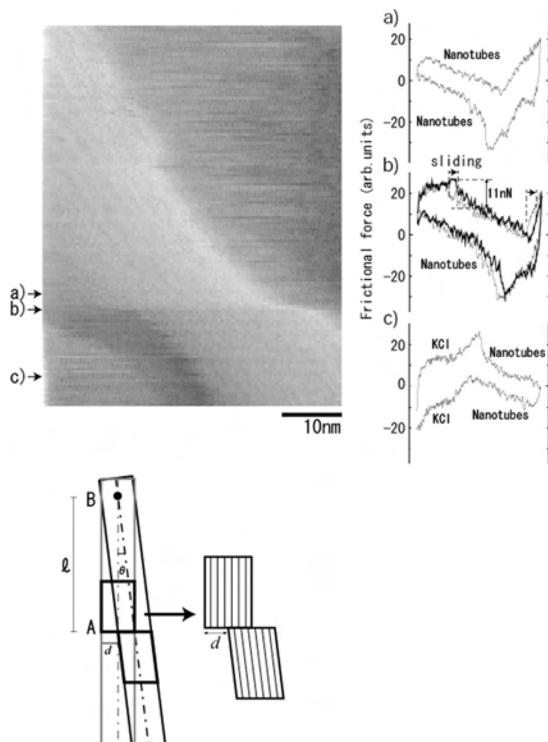


Figure 3. FFM image of the sliding of nanotubes on KCl(001). As in Fig. 2, stripe features due to nanotubes are again observed. On the right, line profiles (a) before, (b) during, and (c) after sliding are given. In line profile (b), both the scan line during sliding (thick line) and that before sliding (thin line) are depicted. At point A, sliding was found to occur. The sliding distance d and the rotation angle around pivot point B, θ, are estimated to be 6.8 nm and 3.5°, respectively. Thus, the distance between sliding point A and pivot point B, l, is estimated to be 111.2 nm using $l = d/\tan\theta$.

$\tau_{\text{nanotube-KCl(001)}}$ would be smaller than the nanotube-graphite shear stress $\tau_{\text{nanotube-graphite}}$; $\tau_{\text{nanotube-KCl(001)}} < \tau_{\text{nanotube-graphite}}$. However, the sliding and/or rolling often occurs in the case of MWNTs on graphite, as shown in the next subsection [14]. It should be noted that strictly speaking, the sliding discussed above indicates in-plane rotation around a pivot point, that the shear stress around the pivot point is not uniform, and that the essential quantity in the control of in-plane rotation movement is not the sliding force F but the moment of force Fl. In the case of a large shear stress, such as for nanotubes on graphite, it is reported that rolling is preferred over sliding due to in-plane rotation [3, 5].

2.2. Rolling of MWNTs on Graphite

2.2.1. Type of MWNTs

The FFM image of a MWNT on graphite is shown in Figure 4. Many MWNTs on graphite are aligned, resulting in a threefold symmetry, in a manner similar to SWNTs. The enlarged images of local area **a** in the graphite substrate and local area **b** in the MWNT are shown on the bottom. The enlarged image of local area **a** exhibits a one-dimensional stick-slip motion, which shows that the scanning direction of a tip is identical with the $[12\bar{3}0]$ direction of the graphite substrate [10]. Also, it is noted that the enlarged

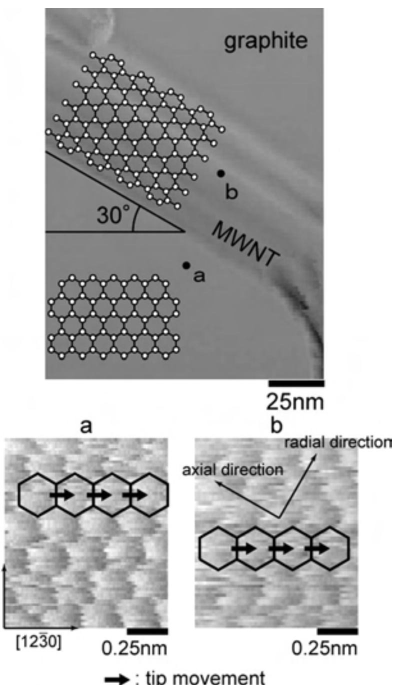


Figure 4. FFM image of the MWNT on the graphite surface. The enlarged images of local area **a** in the graphite substrate and local area **b** in the MWNT are shown on the right-hand side of the figure. The enlarged image of local area **a** exhibits a one-dimensional stick-slip motion, which shows that the scanning direction of a tip is identical with the $[12\bar{3}0]$ direction of the graphite substrate. Also, it is noted that the enlarged image local area **b** exhibits is the same as that of local area **a**.

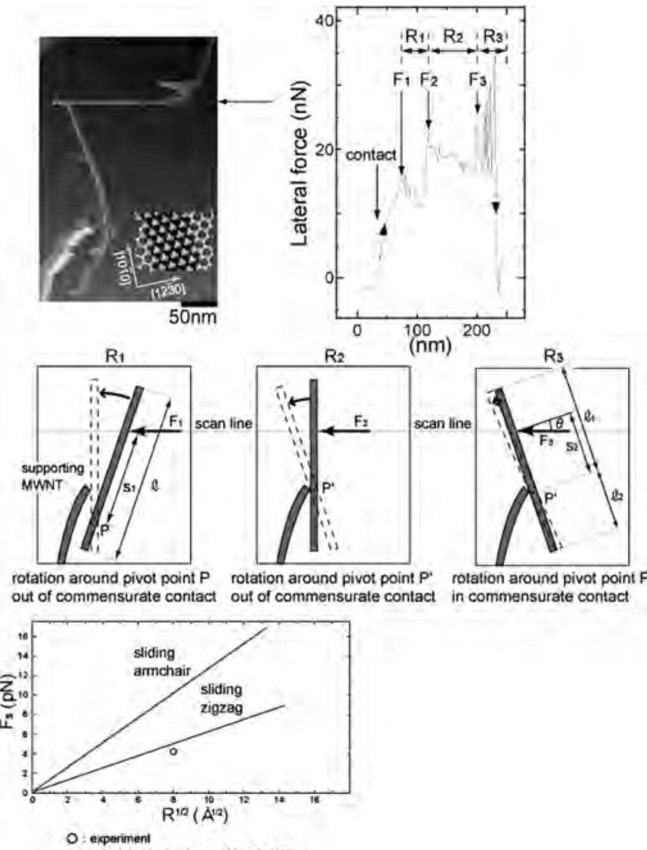


Figure 5. FFM image from the zigzag MWNT on graphite. The line profile (indicated by an arrow) is depicted on the right-hand side of the figure. First, a tip approaches the MWNT from the right and then comes into contact with it. The MWNT begins to rotate around the pivot point P out of commensurate contact (R_1). After rotating about 14°, the MWNT begins to rotate around the pivot point P′ (different from pivot point P) out of commensurate contact (R_2). Furthermore, the MWNT begins to rotate around the pivot point P′ in commensurate contact (R_3). The bottom figure shows the forces (solid lines: Buldum and Lu [5]) per angstrom needed to slide the armchair and zigzag nanotubes and one (circle) estimated from experimental data.

image of local area **b** exhibits the same image as that of local area **a**. It is of interest to note that the angle between the axial direction of the nanotube and the $[12\bar{3}0]$ direction of the graphite substrate is 30°. This reveals that the chirality of this MWNT is a zigzag type. Assuming that the interaction between the MWNT and graphite relies on the outer graphene sheet of the MWNT, it is possible that the MWNT stacks on the graphite substrate in the same manner as the stacking (AB stacking) of graphite layers. This assumption is consistent with those of other cases such as C_{60} molecules on graphite [7], SWNTs on graphite [8], and a graphite flake on graphite [15].

2.2.2. Rotation of MWNTs

Figure 5 shows the FFM image from the zigzag MWNT on graphite. First, a tip approaches the MWNT from the right and then comes into contact with it. The MWNT begins to rotate around a pivot point P out of commensurate contact (R_1). After rotating about 14°, the MWNT begins to rotate around a pivot point P′ (different from the pivot point P) out of commensurate contact (R_2). Furthermore, the MWNT begins to rotate around the pivot point P′ in commensurate contact (R_3). Vibration behavior shown in figure R_3 is probably due to fluctuations in a supporting MWNT building up the pivot point P′. Thus, this in-plane rotation of the MWNT consists of both rotations out of commensurate and in commensurate contacts. It is noted that the force needed to begin rotating in commensurate contact is larger than that out of commensurate contact. Now, assuming that the shear stress σ between the MWNT and graphite is constant around the pivot point P′, the force per unit length required to rotate the MWNT, F_l, is given as follows:

$$Fs_2 = \int_0^{l_1} \sigma l \, dl + \int_0^{l_2} \sigma l \, dl \tag{2}$$

$$F_l = \sigma d = \frac{2Fs_2}{l_1^2 + l_2^2}, \qquad l = l_1 + l_2 \tag{3}$$

where d, s_2, and l are the contact width between the MWNT and graphite, the distance between the push point and the pivot point P′, and the length of the MWNT, respectively. In commensurate contact, F_l is estimated to be approximately 4 pN/Å using $F = 25$ nN ($F_3 \times \cos 17°$), $s_2 = 189$ nm, $l = 600$ nm, and $l_1 = 433$ nm, where the radius of this MWNT is 6.5 nm. This value is in excellent agreement with that of theoretical calculation [5] for a zigzag nanotube. Using $\sigma = 0.2$ GPa [7], d is estimated to be approximately 0.2 nm, which corresponds to the lattice constant of graphite.

2.2.3. Rolling of MWNTs

Figure 6 shows the FFM image for rolling of the zigzag MWNTs on graphite. The upper two figures, A and B, show the characteristics of natural rolling. Both cases show that the original position and the final position keep in parallel after rolling. First, a first peak (F) appears which exhibits the lateral force required to push the MWNT. Then, the energies for MWNTs A and B used to push MWNTs are estimated to be 9.6×10^{-17} J and 3.4×10^{-17} J from the first peak area, respectively. After being pushed, the MWNT rolls naturally without a driving force. After rolling about two revolutions, the MWNT loses kinetic energy and stops. The tip catches up with the MWNT and scans over the MWNT, which is denoted by S in Figure 6 [14]. However, the lower figure C in Figure 6 exhibits stick-slip features. Even in this case, the MWNT translates in parallel. As reported by Falvo et al. [3, 4], the stick-slip features reveal that the tip again pushes the MWNT which, however, continues to roll. Even in this case, the rolling is about two revolutions. In the atomic force microscope (AFM) mode where a scan direction is parallel to the axis direction of a cantilever, the stick-slip features very often appear. Also, natural rolling only occurs under an argon atmosphere although the stick-slip behaviors little change under the relative humidity (0–30%). Thus, rolling behaviors seem to depend on the relative humidity and the relative coordinate between the scanning direction of a tip and the axial direction of a MWNT, which controls the chance or probability where a tip again comes into contact with a MWNT during rolling. Furthermore, the fact that natural rolling only occurs under an argon atmosphere indicates that the adsorbed water molecules on graphite and/or a MWNT influence the distance of natural rolling; for example, a MWNT easily loses kinetic energy and easily stops at a higher relative humidity.

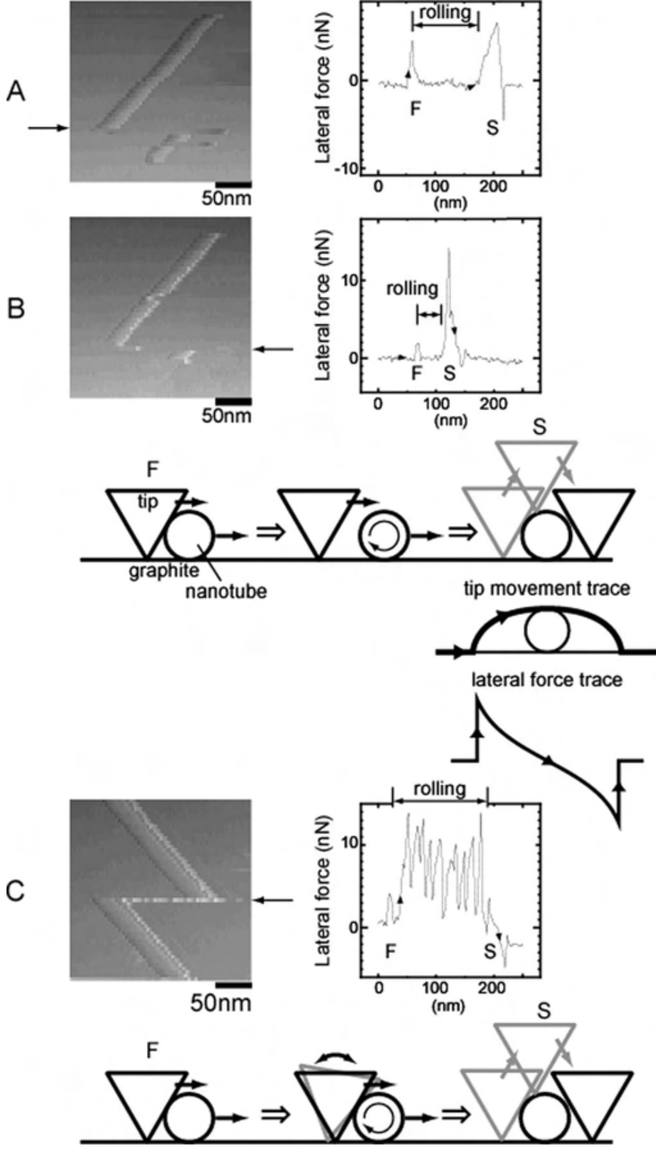


Figure 6. FFM image for rolling of the zigzag MWNTs on graphite. The line profile (indicated by an arrow) is depicted on the right-hand side of the figure. The upper two (A and B) and the bottom (C) figures show the characteristics of natural rolling, and rolling with the several contacts with the tip during rolling, respectively.

3. SLIDING OF FLAKES

In this section, we describe the sliding of a graphite flake on graphite and a MoS_2 flake on MoS_2. The graphite flake moves on graphite such that the stacking of graphite layers is maintained. In the case where the flake movement is parallel to the pulling direction, the frictional force is smallest. The MoS_2 flake also moves on MoS_2 such that the stacking of the sulfur basal plane of MoS_2 is maintained. The frictional forces between MoS_2 surfaces are clearly proportional to loading force although they depend strongly on the movement of the MoS_2 flake. The Amontons–Coulomb law is excellently satisfied at loading forces in the range of 1 to 120 nanonewtons.

3.1. Sliding of a Graphite Flake on Graphite

Figure 7 shows two x frictional force maps at different areas from a graphite flake on graphite scanned along the x direction. It should be noted that these exhibit a scaled pattern, clearly different from that of a tip on graphite from the standpoint of a periodic pattern and contrast [10, 15, 16]. Noting that this system consists of the friction between a tip and a graphite flake and that between a graphite flake and graphite, it is revealed that the friction from this system includes two different frictional mechanisms. However, it is reasonable that the frictional features from a flake on graphite exhibit only the friction between a graphite flake and graphite because they do not include the frictional force map from a tip on graphite. This indicates that the friction between a tip and a graphite flake is larger than that between a graphite flake and graphite. Thus, the graphite flake moves together with the tip when the tip is scanned. It is noted that the frictional force maps from a flake on graphite exhibit various patterns. The three types (A–C) of movement of the graphite flake are illustrated in Figure 7. These maps are transient during scanning from top to bottom. The size of the unit cell of the lower pattern is larger

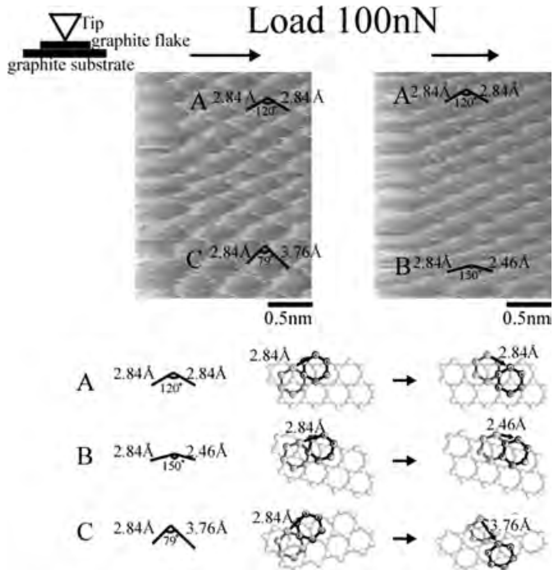


Figure 7. Two x frictional force maps from a graphite flake on graphite obtained by scanning along the x direction, where the arrows mean the scanning direction. The three types (A–C) of movement of the graphite flake are shown in the lower part of figure.

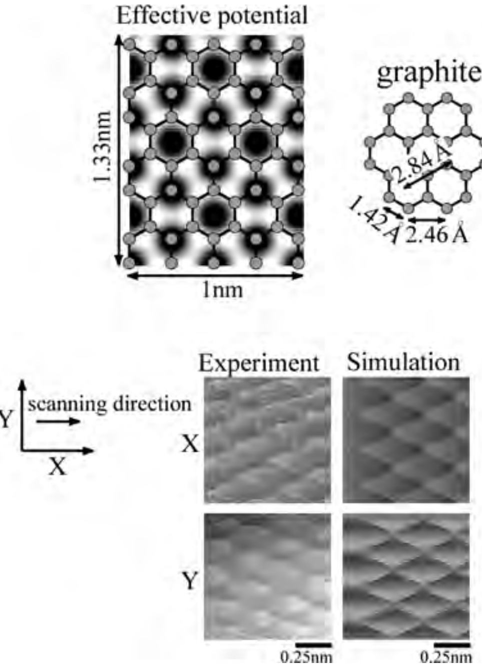


Figure 8. Experimental data and theoretical simulation for type A of Fig. 7.

than that of the upper ones. Now, we consider type A of the frictional force map, which is the standard pattern appearing in this experiment. The movement of the flake on graphite is substituted into the movement of a single particle moving with a constant velocity in the effective potential as follows:

$$V = V_0\left\{2\cos\left(\frac{2\pi}{a}x\right)\cos\left(\frac{2\pi}{a\sqrt{3}}y\right) + \cos\left(\frac{4\pi}{a\sqrt{3}}y\right)\right\} \quad (4)$$

where $a = 0.142$ nm $\times$ 2 = 0.284 nm. This effective potential is determined so that it takes the minimum values at the natural stacking relation of graphite. This simple preliminary simulation is in excellent agreement with the experimental data, as shown in Figure 8. Here, the spacing of $a = 0.284$ nm indicates a movement to the next stable position holding AB stacking of graphite. Similarly, the other patterns reveal that their movements are carried out so as to maintain the natural stacking relation (AB stacking) of graphite, as shown in Figure 7. It has been reported that the energetic barrier (the energetic difference between AA stacking and AB stacking of graphite) to the interlayer sliding is approximately 12 meV per carbon atom [17]. Assuming that the energetic barrier to the interlayer sliding in this experiment is on the order of electronvolts because the frictional force in flake friction appears on the order of nanonewtons, the number of carbon atoms contributing to the interlayer sliding is on the order of hundreds, leading to tens of unit cells of graphite. Thus, it is expected that the actual contact area between the flake and graphite becomes very small and only occupies an extremely small part of the flake. In order to observe flake movement, the relationships between the flake position and the lever support position are shown in Figure 9. Compared to the tip movement on graphite, that of the flake exhibits sharper peaks. This feature is also identical with the pattern theoretically given by Tomanek et al. and Weiss et al. [18, 19], which indicates no jump (or no energy dissipation) when the flake moves on a graphite substrate. This suggests that a driving force (a static frictional force) appears when the flake moves on a graphite substrate but a dynamical force (a dynamical frictional force) disappears. We think that this gives an explanation to the problem why the friction of the system including various flakes is generally small; for example, a friction force acting on many flakes with different movements may be close to zero.

The frictional forces in relation to loading force for a tip on graphite (D and E on the left-hand side) and for a flake on graphite (F and A on the right-hand side) are shown in Figure 10. The case F exhibits that the flake movement is parallel to the pulling direction or scanning direction, in which the frictional forces are very weakly proportional to the loading force. Thus, the proportional coefficient μ_F for case F is estimated to be approximately 0.001, which corresponds to the frictional coefficient $\mu_{12\bar{3}0}$ because the movement of case F is along the crystallographic direction [$12\bar{3}0$]. However, in case A (which is the same as case A in Fig. 7)

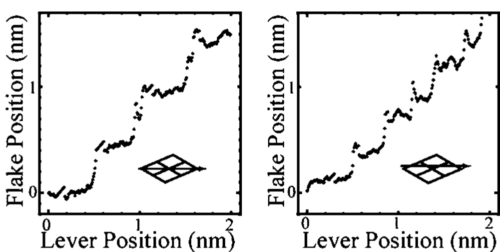


Figure 9. The relationships between the flake position and the lever support position at different lines in scaled patterns.

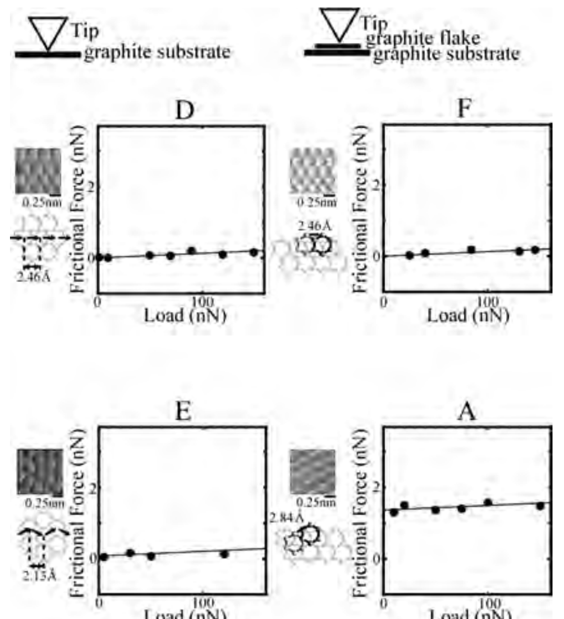


Figure 10. The mean frictional forces in relation to loading force for a tip on graphite (D and E on the left-hand side) and for a flake on graphite (F and A on the right-hand side). Images and arrows on the left-hand side represent frictional force maps and the tip movement, respectively. Image and arrows on the right-hand side represent frictional force map and the flake movement, respectively.

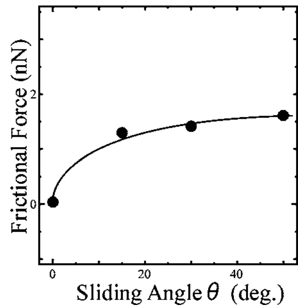


Figure 11. The mean frictional forces versus sliding angle at a loading force of 100 nN.

where the direction of flake movement is not parallel to that of the pulling force, the frictional forces satisfy the relationship that $\alpha + \mu_A F_n$, where $\alpha = 1.3$ nN, $\mu_A = 0.001$, and F_n represents the loading force. Here, it is expected that the graphite (0001) flake is easy to shear along the direction $[12\bar{3}0]$ of the graphite (0001) surface because the path to the next stable point in this case does not take the path of AA stacking and its energetic barrier is the smallest. However, cases A, B, and C always occur. The proportional coefficients μ_D and μ_E for cases D and E are also together estimated to be approximately 0.001. Now, it is interesting to note that the difference between case D(E) and case F(A) reveals the change in the frictional behavior when the frictional mechanism is transient from a single-atomic tip to a large-area flake tip. When the flake does not move parallel to the scanning direction, there exists the difference between a tip on graphite and a flake on graphite (compare E with A). However, when the flake moves parallel to the scanning direction, it is found to be identical with that from a tip on graphite (compare D with F).

Now, the anisotropy of the frictional force for different directional movement of the flake versus pulling direction is shown in Figure 11 from the results of Figure 10, which appears depending on the sliding angle θ, denoted as the angle between the direction of flake movement and the pulling direction. The frictional forces are found to increase with increase of the sliding angle, which, interestingly, appears similar to the behavior of the frictional mechanism between flat surfaces discussed by Gyalog and Thomas [20]. These results conclude the following two points: (i) in the case where the flake movement is parallel to the pulling direction, the frictional force is smallest; (ii) in the case where the direction of the flake movement is not parallel to that of the pulling force, the flake does not move toward the next stable point until the projection of the pulling force is coincident with the force to begin moving toward the next stable point. Thus, the larger the sliding angle, the larger the pulling force. Here, it should be noted that preferred slidings have been observed in other cases of MoO_3 nanocrystals on MoS_2 surfaces [21] and MoS_2 flakes on MoS_2 surfaces [22]. Thus, it is very interesting to note that these findings provide us important information of how to decrease friction between surfaces.

3.2. Sliding of a MoS_2 Flake on MoS_2

Figure 12 shows three different x frictional force maps from a MoS_2 flake on MoS_2 scanned along the x direction. It should be noted that these exhibit scaled (types A and B) and square (type C) patterns, clearly different from that of a tip on MoS_2 from the standpoint of a periodic pattern and contrast (see figures on the left-hand side of Fig. 14). Noting that this system consists of friction between a tip and a MoS_2 flake and that between a MoS_2 flake and MoS_2, it was found that the friction from this system includes two different mechanisms. However, it can be concluded that the frictional features of a flake on MoS_2 exhibit only the friction between a MoS_2 flake and MoS_2 because they do not include the frictional force map from a tip on MoS_2. This means that the friction between a tip and a MoS_2 flake is greater than that between a MoS_2 flake and MoS_2. Now, we consider type A of the frictional force map, which is the standard pattern appearing in this experiment. The movement of the flake on MoS_2 is substituted into the movement of a single particle moving with a constant velocity in the effective potential for type A, as follows:

$$V = V_0\left\{2\cos\left(\frac{2\pi}{a}x\right)\cos\left(\frac{2\pi}{a\sqrt{3}}y\right) + \cos\left(\frac{4\pi}{a\sqrt{3}}y\right)\right\}$$

where $a = 0.316$ nm. The result of this simple preliminary simulation is in excellent agreement with the experimental data, as shown in Figure 13. Here, the spacing of $a = 0.316$ nm indicates a jump to the next stable position maintaining the stacking of MoS_2. Similarly, it is also revealed that the movements for the type B and C are performed such that the natural stacking relation of MoS_2 is maintained, as shown in Figure 12.

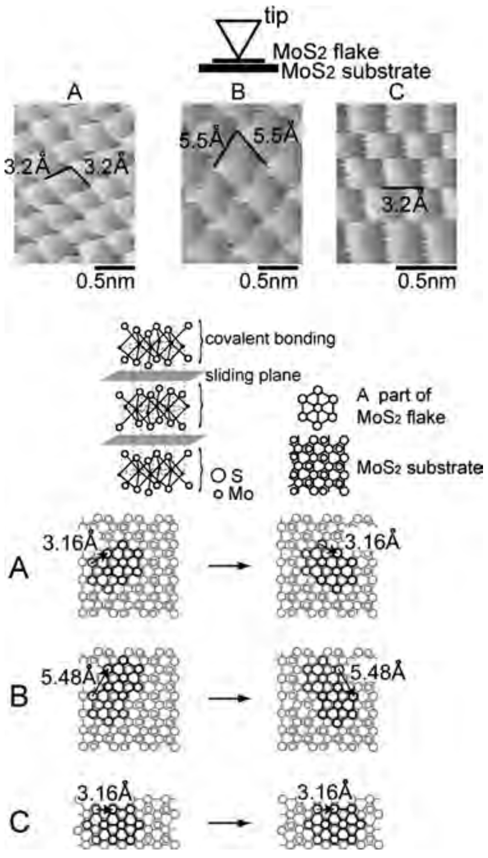


Figure 12. Three x frictional force maps at different areas from a MoS_2 flake on MoS_2 obtained by scanning along the x direction, where the arrows represent the scanning direction. The three types (A, B, and C) of movement of the MoS_2 flake are shown in the lower part.

The frictional forces for different movements in relation to loading force are shown on the right-hand side in Figure 14. These are all proportional to loading force although they depend strongly on the movement of the MoS_2 flake. Thus, the proportional coefficients defined as μ_A, μ_B, and μ_C for cases A, B, and C are estimated to be approximately 0.01, 0.02, and 0.003, respectively. The proportional coefficient μ_C for case C is much smaller than μ_A and μ_B for cases A and B because the flake movement in case C is parallel to the pulling direction or scanning direction. However, in the cases A and B where the direction of flake movement is not parallel to that of the pulling force, the frictional forces become larger than those for parallel movement. Thus, anisotropy of the frictional force for different directional movements of the flake in relation to the pulling direction prominently appears as shown in Figure 15, where the angle between the direction of flake movement

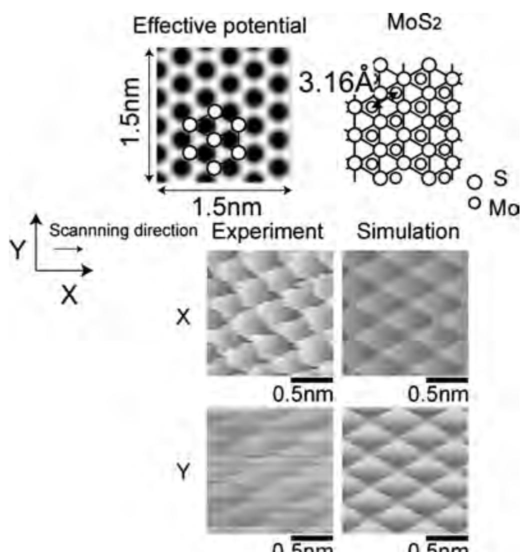


Figure 13. Experimental data and theoretical simulation for type A of Fig. 12.

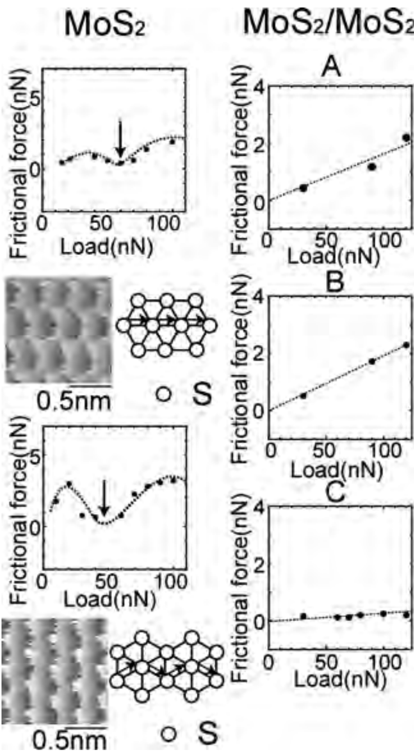


Figure 14. The frictional forces in relation to loading force for surface friction (right-hand side) and for tip friction (left-hand side). Images and arrows on the left-hand side represent frictional force maps and the tip movement, respectively.

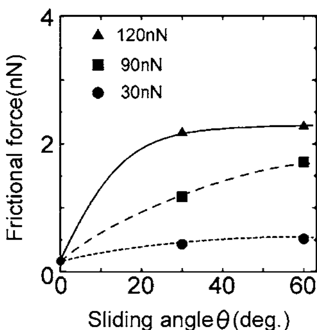


Figure 15. The frictional forces versus sliding angle deduced from the data of Fig. 14, where the sliding angles for cases A, B, and C are 30°, 60°, and 0°, respectively.

and the pulling direction is denoted as the sliding angle θ. One notes that the frictional forces versus sliding angle increase and increase strongly with higher load. Interestingly, this behavior is consistent with the frictional mechanism between flat surfaces discussed by Gyalog and Thomas [20] and Hirano et al. [23]. Since the movement of case C is along the crystallographic direction $[10\bar{1}0]$, the frictional coefficient $\mu_{10\bar{1}0}$ (the sliding angle $\theta = 0°$) where the $MoS_2(0001)$ surface shears along the crystallographic direction $[10\bar{1}0]$ of the $MoS_2(0001)$ surface is estimated to be 0.003, which is close to the experimental value (below 0.002) obtained by Martin et al. [24]. Thus, it is found that the $MoS_2(0001)$ flake is easy to shear along the direction $[10\bar{1}0]$ of the $MoS_2(0001)$ surface. Surprisingly, the frictional coefficient for case B is eight times larger than that for case C. Here, we suppose that case C is preferred over cases A and B because its energetic barrier is the smallest. However, cases A and B occur very often. This is the reason why the size of a flake is very large and the movement such as rotation of a flake would occur in addition to the parallel movement. Proportionality of the frictional forces in relation to loading force also reveals that the contact area between MoS_2 surfaces S is proportional to the loading force F_N, using the relationship $F = \mu F_N = \tau S = \tau\alpha F_N$, where α represents the proportional constant. This indicates that surface deformation is significantly smaller in MoS_2 than in graphite because the lamellar MoS_2 consists of covalent bonding, which behaves like a hard plate. Also, it should be noted that frictional forces for tip friction are quite different from those for surface friction. Along the different scanning directions of the $MoS_2(0001)$ surface, a tip performs one-dimensional stick-slip and zigzag stick-slip motions, shown on the left-hand side in Figure 14. This behavior is similar to that of ionic surfaces [25]. In the case of an ionic surface, the minimum points (shown by the arrows) appear before large distortions of the surface occur. Thus, it is inferred that large distortions of the surface occur even in MoS_2 triple layers.

This result reveals the change in the frictional behavior when the frictional mechanism is transient from a single-atomic tip to a large-area flake tip (compare data on the left-hand side with those on the right-hand). There is a prominent difference between frictional force maps of tip friction and those of surface friction. Similarly to graphite, the MoS_2 flake moves on MoS_2 such that the stacking of sulfur basal plane of MoS_2 is maintained. However, the frictional forces between MoS_2 surfaces are proportional to loading force and depend strongly on movement of a MoS_2 flake. The frictional coefficient $\mu_{10\bar{1}0}$ at the sliding angle $\theta = 0°$ is estimated to be 0.003, which indicates that a MoS_2 flake is easy to shear along the direction $[10\bar{1}0]$ of the $MoS_2(0001)$ surface. These findings also show us how to decrease friction between MoS_2 surfaces.

4. C_{60} MOLECULAR BEARINGS

The mechanical properties of fullerenes have attracted much attention in a materials science field. Especially, C_{60} and C_{70} solids have been expected to be good lubricant materials because of their nearly spherical shape and low surface strength. Okita et al. [6] have reported the nanotribological behavior on the C_{60} islands on KCl(001). Along different scanning directions of $\langle 1\bar{1}0\rangle$ and $\langle 11\bar{2}\rangle$ of a $C_{60}(111)$ surface, a tip has displayed one-dimensional stick-slip and zigzag stick-slip motions on the order of a load of nanonewtons, respectively, although at a larger loading force it destroys the $C_{60}(111)$ surface. Also, water adsorption on C_{60} films gives a lower frictional force, indicating that C_{60} molecules rotate or translate at the (111) surface. Thus, the C_{60}films are expected to exhibit various behaviors depending on loading forces, scanning direction, and relative humidity. C_{60} molecules on graphite also begin to grow with a monolayer. Then, a shear force between a single C_{60} molecule and graphite is shown.

In this section, the frictional behavior of C_{60} films on an ionic crystal and graphite is described, and furthermore the applicability and realization of C_{60} molecular bearings is described first. A rolling stick-slip model with a step rotation of a C_{60} molecule is proposed. The novel frictional mechanism is realized in the graphite/C_{60} monolayer/graphite system, in which static frictional forces have a finite value but mean dynamical frictional forces are zero. The C_{60} film thicker than the C_{60} bilayer, sandwiched by graphite plates, behaves as an elastic body. It should be emphasized that the graphite/C_{60} monolayer/graphite system is very promising for the realization of nano- and micromachines and opens a new field of molecular bearings.

4.1. C_{60} Films on Ionic Crystals and Graphite

Normal and lateral forces are measured simultaneously under humidity controlled conditions at room temperature. The C_{60} films are prepared by evaporation on a KCl(001) surface from a boron nitride (BN) crucible. The temperature of KCl(001) substrate is kept at about 200 °C. The growth mode is island growth. The C_{60}films had face-centered cubic symmetry, and the (111) facet surrounded by (001) facets is parallel to the KCl(001) surface. The C_{60} films on highly oriented pyrolytic graphite (HOPG) and KCl(001) are prepared by evaporation from a BN crucible. The temperatures of their substrates during evaporation are kept in the range of 150 °C to 200 °C.

First, we show in Figure 16a and b the normal force (NF) and lateral force (LF) images taken simultaneously for scans along the $\langle 1\bar{1}0\rangle$ and $\langle 1\bar{1}2\rangle$ directions at a load of 0.024 nN at 20% RH, respectively. The scanning area corresponds to the (111) facet surrounded by the (010) facets. It is found that the images are periodic but are quite different from each other depending on the scanning direction. In order to analyze the images quantitatively, it is necessary to calibrate the tip displacements for NF and LF images. If we take, as shown in Figure 16a, the geometrical relationship between the x, y, and z axes and cantilever, the NF profile and the LF profile are determined by the bending of the cantilever in the y or z direction, and by the torsion of the cantilever in the x direction, respectively. For calibration of the LF profile we plotted lateral force increment versus scanning distance at the first stacking point in the y direction of scanning because the bending of the cantilever in the y direction is dominant in the NF mode. Figure 16a shows the line profiles from scans 1 to 5 for a single-period area of the $\langle 1\bar{1}0\rangle$ image. Lines 1, 3, and 5 exhibit the same profiles in both

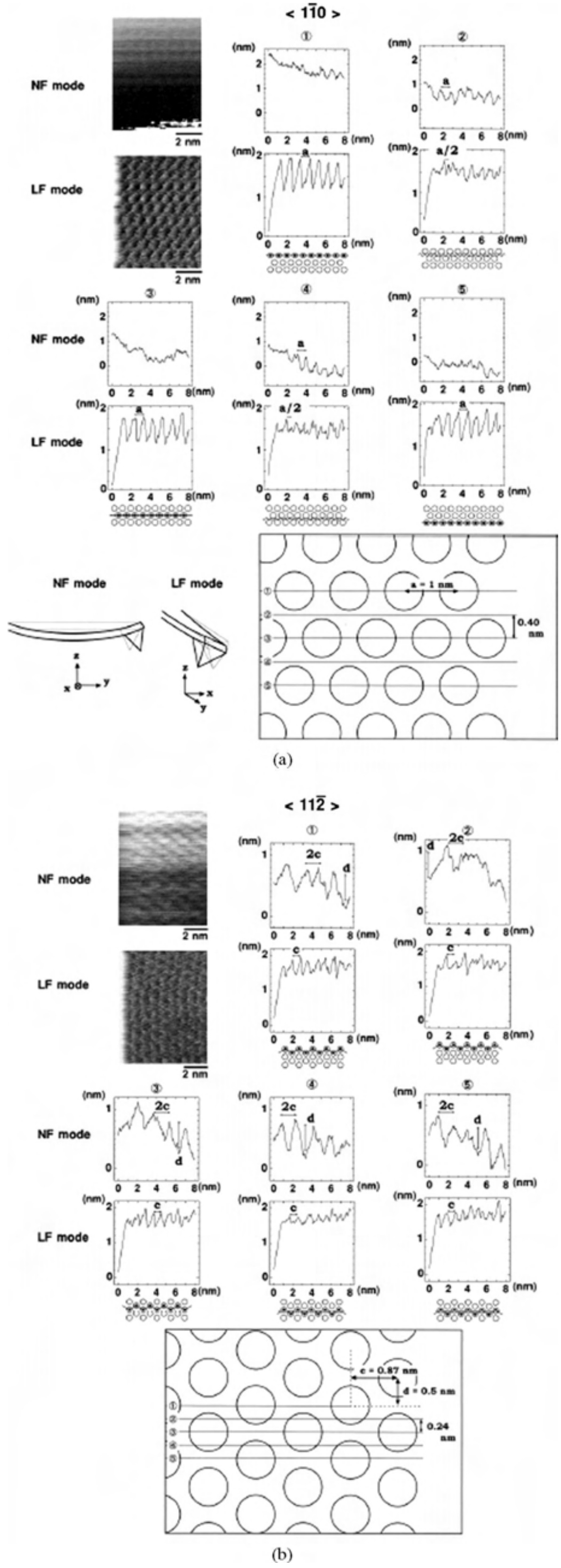


Figure 16. The normal force (NF) image and lateral force (LF) image taken simultaneously for scans in the (a) $\langle 1\bar{1}0\rangle$ and (b) $\langle 11\bar{2}\rangle$ directions at relative humidity (RH) of 20%. The line profiles from scans 1 to 5 were taken for one-period areas. The circles represent molecules.

the NF and LF profiles; the LF profile clearly has peaks with a period of 1-nm but the NF profile has no periodicity. In the LF profile of lines 2 and 4, the 1-nm period peaks become unclear and include small 0.5-nm period peaks, but in the NF profile 1-nm period peaks appear. Thus, lines 1, 3, and 5 indicate that sticking points of a tip always occur on molecules lying in a row in the $\langle 1\bar{1}0\rangle$ direction. However, lines 2 and 4 indicate that sticking points of a tip occur at zigzag positions between the neighboring molecular rows. This is consistent with small 0.5-nm period peaks in the LF profile and 1-nm period peaks in the NF profile. In the same manner as for scanning in the $\langle 1\bar{1}0\rangle$ direction, Figure 16b shows the line profiles from scans 1 to 5 for a single-period area of the $\langle 1\bar{1}2\rangle$ image. All NF profiles from lines 1 to 5 have 1.7-nm period peaks, although all LF profiles have 0.9-nm period peaks. Note that the amplitudes of 1.7-nm period peaks in the NF profiles are about 0.5-nm. Thus scans in the $\langle 11\bar{2}\rangle$ direction reveal zigzag stick-slip motions between neighboring molecular rows along the $\langle 11\bar{2}\rangle$ direction.

In Figure 17, we show frictional loops for a scan in the $\langle 1\bar{1}0\rangle$ direction at 20% RH and in dry argon, where the load was 0.024 nN for both cases. Note that the mean frictional force at 20% RH is about a quarter that in dry argon. Furthermore, the amplitude at dry argon is about three times that at 20% RH. However, both frictional loops have a typical sawtooth profile due to stick-slip motion, because they have a hysteresis during the loop, and accordingly accompany dissipation. One note is that the prominent difference between the dry argon case and the 20% RH case comes from the slope change of the sticking part, because the periodic spacing of stick-slip is the same. In general, the lateral effective spring constant, K_{eff}, which corresponds to the slope of the sticking part, is combined with the lateral contact stiffness between the tip and the sample, $K_{contact}$, the lateral stiffness of the tip, K_{tip}, and the torsional spring constant of cantilever, K_{lever}, as follows [26]:

$$K_{eff} = (1/K_{contact} + 1/K_{tip} + 1/K_{lever})^{-1} \tag{5}$$

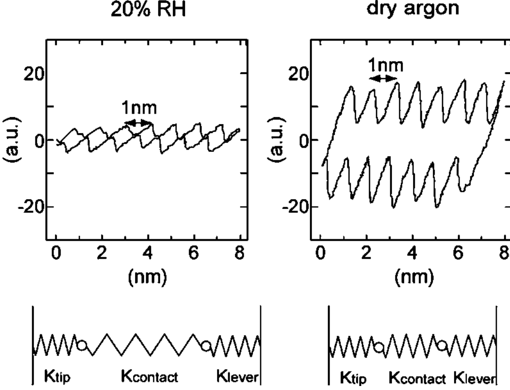


Figure 17. The frictional force loops for scanning in the $\langle 1\bar{1}0\rangle$ direction at relative humidity (RH) of 20% and in dry argon. Note that the mean frictional force at 20% RH is about one quarter than in dry argon. The change of tip-sample contact stiffness is illustrated below. The softening contact stiffness appears at 20% RH, where $K_{contact}$, K_{tip}, and K_{lever} show the lateral contact stiffness between the tip and the sample, the lateral stiffness of the tip, and the torsional spring constant of the cantilever, respectively.

If K_{tip} and K_{lever} are constant (this assumption is reasonable because the same tip is used), the decrease in lateral contact stiffness between the tip and the sample links directly to the decrease in effective spring constant. Thus, the decrease in slope of the sticking part means a softer spring for tip-sample contact. It has been shown [27] that water adsorption makes the molecules more mobile. Thus, it is now supposed that the adsorption of water molecules helps the mobility of C_{60} molecules. Here, the 'rolling' and/or 'translation' effects of a C_{60} molecule with lower resistance is considered for the mobility of C_{60} molecules, but it may be reasonable to prefer rolling rather than translation at the area of minimum energy around a C_{60} lattice point.

Figure 18 shows AFM images of C_{60} islands on a KCl(001) surface, respectively. Their islands consist of scores of molecular layers. The (111) facets appear in the C_{60} islands which form parallel to a KCl(001) substrate. Such an island growth occurs in the case where an intermolecular interaction is stronger than a molecule–substrate interaction. Thus, it is concluded that C_{60}-C_{60} interactions are stronger than C_{60}-KCl(001) interactions, respectively. Figure 19 shows AFM (topograph) images of C_{60} films on a graphite surface, respectively. Height of the C_{60} films in Figure 19 becomes about 1 nm, exhibiting a monolayer. As shown on the bottom of Figure 19, the high-resolution FFM image has a periodicity of 0.9 nm, which shows a zigzag motion along a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60} monolayer [6]. Thus, an existence of the monolayer exhibits an epitaxial growth on graphite. As shown by Gravil et al. [12], the growth of this monolayer indicates that C_{60} molecules grow on graphite such that the hexagonal face of a C_{60} molecule pairs with the hexagonal face of the second graphite layer, so as to continue the natural stacking of the graphite.

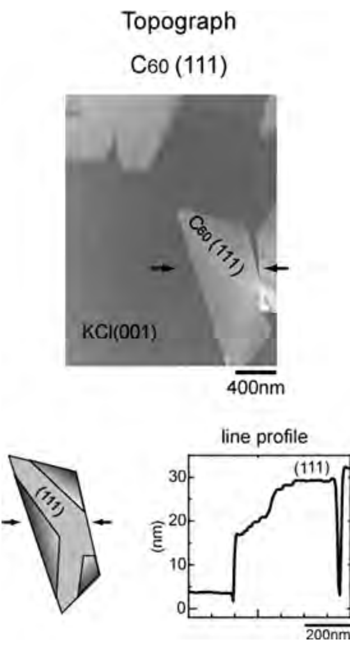


Figure 18. Atomic force microscope (AFM: topograph) images of C_{60} islands on a KCl(001) surface, respectively. The (111) facets appear in C_{60} islands which form parallel to a KCl(001) substrate. On the bottom of figure, line profiles depicted by the arrows are illustrated.

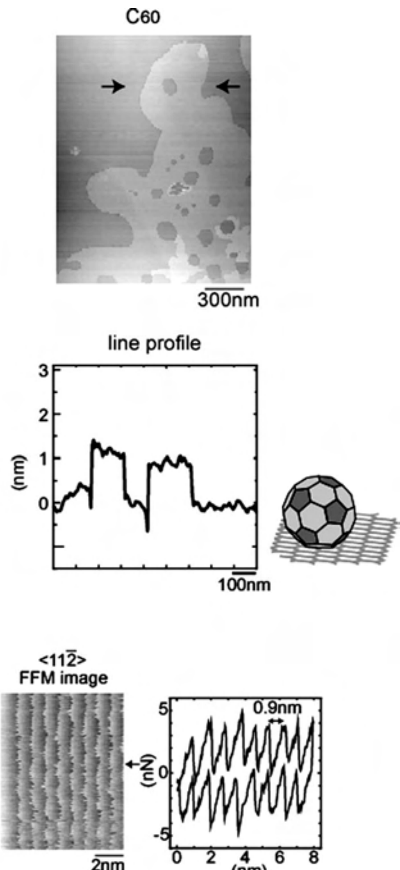


Figure 19. AFM (topograph) image of C_{60} films on a graphite surface. On the middle of figure, line profiles depicted by the arrows are illustrated. Height of C_{60} films becomes about 1 nm, exhibiting a monolayer. As shown on the bottom of figure, the high-resolution frictional force microscope (FFM) image has a periodicity of 0.9 nm, which shows a zigzag motion along a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60} monolayer.

Figure 20 shows how AFM images of Figure 19 change by increasing a loading force. One notes that the C_{60} monolayer on graphite is swept up by increasing up to a load of 0 nN. Using Herztian continuum theory [11, 28], a contact radius versus a tip load estimated using a pull-off force of −10 nN and a tip radius of 15 nm is shown on the bottom of Figure 20. At a loading force of 0 nN, a contact radius between the tip and the C_{60} monolayer surface was estimated to be about 1.3 nm, exhibiting that the contact area between the tip and the C_{60} monolayer becomes a seven molecular dimension. Then, a mean lateral force was about 3 nN. Thus, assuming that seven molecules are moved by a lateral force of 3 nN with a tip, a shear force between a single C_{60} molecule and graphite is estimated to be about 0.4 nN, indicating that a shear stress between a C_{60} monolayer and graphite is estimated to be about 0.2 GPa.

Figure 21 shows the high-resolution FFM images of (111) surfaces of C_{60} films. For $\langle 1\bar{1}0\rangle$ and $\langle 11\bar{2}\rangle$ scanning directions, both films exhibit one-dimensional stick-slip and two-dimensional zigzag motions, respectively [6]. In case of a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60}(111) films, it should be interesting to note that the image of a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60}(111) films at a loading force of −9 nN (pull-off force: −10 nN) changes from one-dimensional

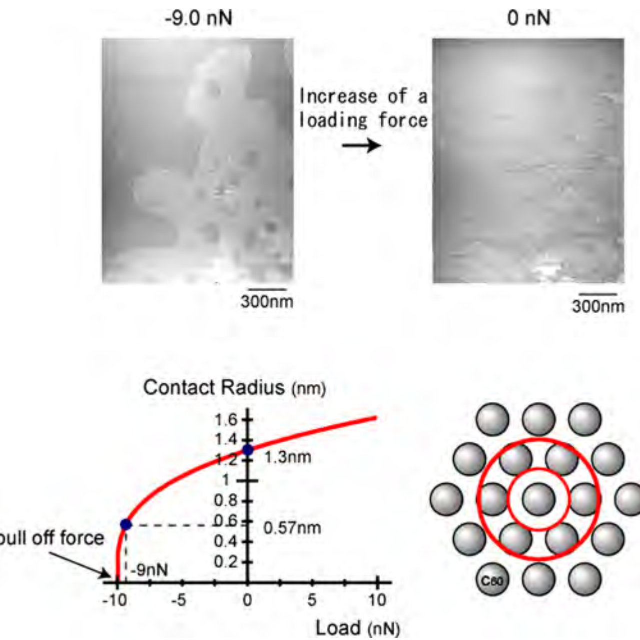


Figure 20. A change of AFM (topograph) image of Fig. 19 by increase of a loading force. One notes that a C_{60} monolayer on graphite is swept up by increasing up to a load of 0 nN. Using Herztian continuum theory, a contact radius versus a loading force was estimated and is shown on the bottom of figure, where the pull-off force is −10 nN. Contact radiuses at loading forces of 0 nN and −9 nN are 1.3 nm (large circle) and 0.57 nm (small circle), respectively.

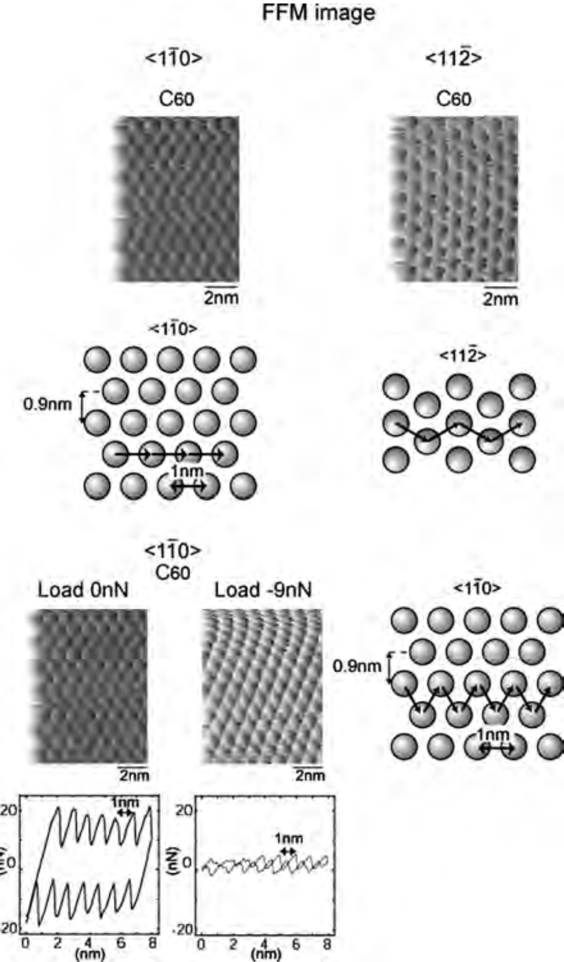


Figure 21. High-resolution FFM images of (111) surfaces of C_{60} films. For $\langle 1\bar{1}0\rangle$ and $\langle 11\bar{2}\rangle$ scanning directions, both samples exhibit one-dimensional stick-slip and two-dimensional zigzag motions, respectively. In case of a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60} (111) films, it should be interesting to note that the image of a $\langle 11\bar{2}\rangle$ scanning direction of the C_{60} films at a loading force of −9 nN (pull-off force: −10 nN) changes from one-dimensional stick-slip to clear two-dimensional motions, exhibiting a single molecule contact with a tip, which gives a mean frictional force of 1 nN.

stick-slip to clear two-dimensional motions, exhibiting a single molecule contact with a probe tip, which gives a mean frictional force of 1 nN. In other words, this is a phase transition of stick-slip motion due to a change from a multimolecular contact to a single molecular contact. Thus, a shear force between a tip and C_{60} molecule is estimated to be 1 nN. This confirms that a tip scanning at the C_{60} monolayer is unstable and it is difficult to obtain a high-resolution FFM image of the C_{60} monolayer because the shear force (1 nN) between the tip and the C_{60} molecule is larger than that (0.4 nN) between the C_{60} molecule and the graphite.

4.2. Graphite/C_{60} Monolayer/Graphite System

Nano- and micromachines have been expected to trigger the creation of a future prominent industry. As first suggested by Feynman [29], molecular bearings have been expected to be a fundamental component of many molecular mechanical devices performing a frictionless system and to be very effective for the realization of nano- and micromachines. However, previous work on micromachine bearings has produced poor results, both experimentally and theoretically. Thus far, micromachines have been considered “machines incapable of movement.” Here, we describe C_{60} molecular bearings. It has been reported [6, 7, 14] that C_{60} molecules on graphite can grow in monolayer form, which consists of close-packed C_{60} molecules. Since the six-membered ring of a C_{60} molecule stacks on that of graphite such that the AB stacking of graphite is maintained, a graphite flake placed on the C_{60} monolayer on graphite is expected to stack in the same way as the C_{60} molecules on graphite. Thus, we can construct a system confining a C_{60} monolayer between graphite plates, which forms the nanogears of six-membered rings between C_{60} molecules and graphite.

Figure 22 shows a topograph (atomic force microscope image) of graphite (A, area S_A), and C_{60} monolayers (B, area S_B) and C_{60} bilayers (C, area S_C) on graphite. When a graphite flake is placed on areas S_A, S_B, and S_C, the graphite flake on graphite (D), the graphite flake on the C_{60} monolayer/graphite (which we call graphite/C_{60}^{mono}/graphite: E), and the graphite flake on the C_{60} bilayer/graphite (graphite/C_{60}^{bilayer}/graphite: F) are obtained, respectively. As shown in Figure 22, frictional force maps for graphite (A) show that the tip exhibits one-dimensional stick-slip and two-dimensional zigzag stick-slip motions for $[12\bar{3}0]$ and $[10\bar{1}0]$ scanning directions, respectively [10], although they exhibit significant load dependence at small loads [10]. Frictional force maps for B and C show that the tip exhibits

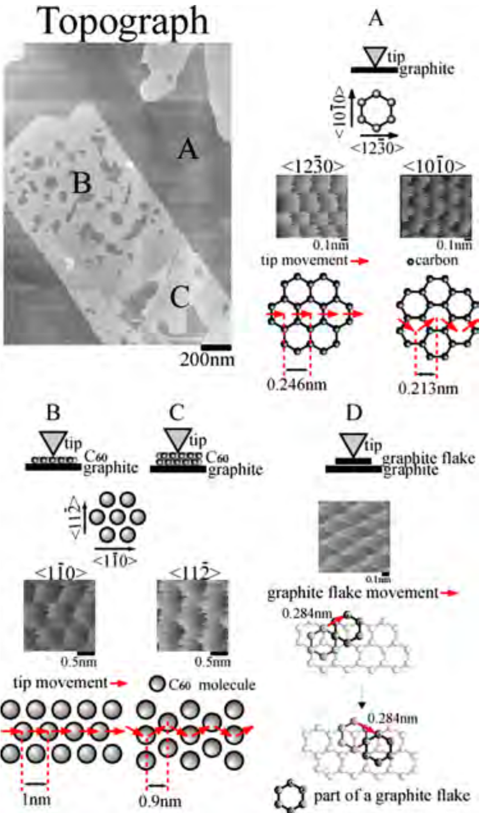


Figure 22. Topograph (atomic force microscope image) for graphite (A, area S_A), and C_{60} monolayers (B, area S_B) and C_{60} bilayers (C, area S_C) on graphite. Frictional force maps for graphite (A), C_{60} monolayers on graphite (B), C_{60} bilayers on graphite (C), and graphite flake on graphite (D) are shown in the figure.

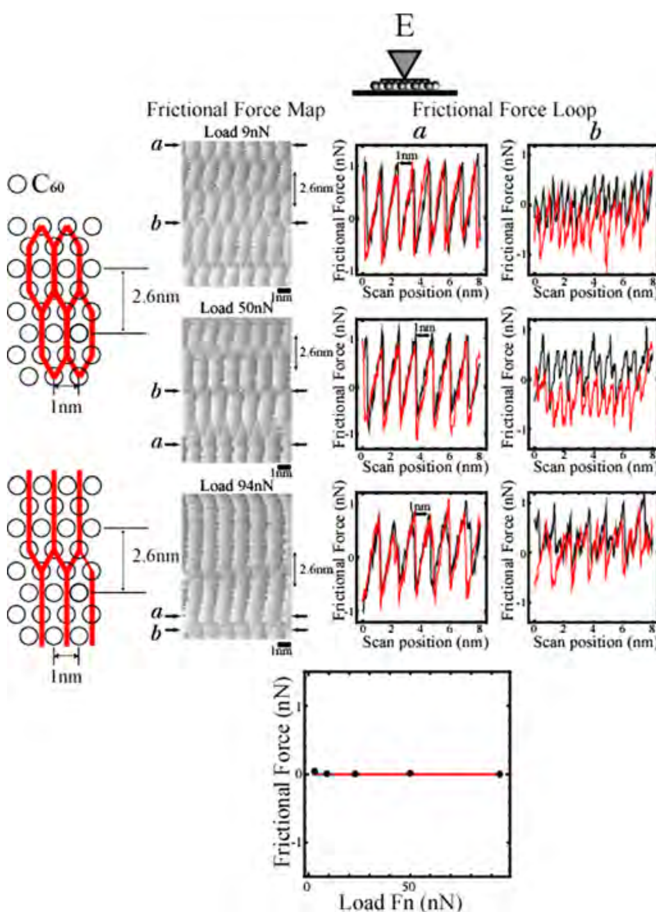


Figure 23. Frictional force maps versus loading force for graphite/C_{60}^{mono}/graphite (E). Frictional force loops obtained from line profiles shown by the arrows *a* and *b* in the frictional force maps are also represented, where the black and red lines indicate one direction and in the opposite direction, respectively. Super cell structures of the frictional force maps are illustrated on the left-hand side. Mean frictional forces versus loading force shown by the arrow *a* are shown on the bottom.

one-dimensional stick-slip and two-dimensional stick-slip motions for the $[1\bar{1}0]$ and $[11\bar{2}]$ scanning directions of a $C_{60}(111)$ surface, respectively [10]; it is more difficult to obtain clear frictional force maps for a C_{60} monolayer than for a C_{60} bilayer because the C_{60} monolayer is unstable for a tip contact. The frictional force maps for B show that the C_{60} monolayer consists of close-packed C_{60} molecules. This indicates that a six-membered ring is stacked on the graphite such that the AB stacking of graphite is maintained, and thus the six-membered ring is situated at the top of C_{60} molecules [7, 14]. The frictional force map for D shows that the graphite flake moves on the graphite such that the AB stacking of graphite is maintained [15]. Assuming that the graphite flake stacks on C_{60} molecules in the same way as C_{60} molecules stack on graphite, the projection of the first-layer six-membered net of the graphite flake is expected to be completely coincident with that of the graphite substrate through a C_{60} monolayer.

Figure 23 shows frictional force maps versus loading force for the graphite/C_{60} mono/graphite (E). It should be noted that the maps of E show transient patterns in relation to loading force, clearly different from those for graphite (A) [10], $C_{60}(111)$ surfaces (B and C) [6, 7], and the graphite flake on graphite (D) [15] from the standpoint of periodic pattern and contrast, as shown in Figure 22. Noting that this system includes friction between a tip and a graphite flake and that between a C_{60} monolayer and two graphite plates, it is considered that the friction from this system involves two different frictional mechanisms. However, it is reasonable that the frictional features of this system show only friction between the C_{60} monolayer and the two graphite plates because they do not show a frictional force map like that obtained for graphite (A) at all. This leads to the conclusion that the friction between the tip and the graphite flake is greater than that between the C_{60} monolayer and two graphite plates. Thus, the graphite flake moves together with the tip during scanning. At this point we note that the frictional force map at a load of 9 nN has a periodicity of 1 nm along the x direction and a periodicity of 2.6 nm along the y direction. This reflects the close-packed C_{60}molecular arrangement, as shown in Figure 23, but, has a superstructure (a periodicity of 2.6 nm) along the y direction. Furthermore, these maps exhibit a vertical chainlike transient pattern related to loading force. It should be noted that the frictional loops obtained from line profiles shown by the arrow *a* in the frictional force maps do not exhibit hysteresis in both directions, which indicates

that the mean frictional forces are zero and thus there is no energy dissipation, although those shown by the arrow *b* accompany some hysteresis, in which two-dimensional zigzag motions of a graphite flake exhibit. As shown on the bottom of Figure 23, the mean frictional forces at the position *a* are zero up to a high load, although the frictional forces have a finite value. Then, the maximum frictional force in this experiment is estimated to be below 1 nN, which is comparable to the shear force (0.4 nN) between a C_{60} single molecule and graphite [7].

If the nanogears of a six-membered ring between C_{60} molecules and the upper and lower graphite plates are formed in this system, it is found to be easy to apply torque to C_{60} molecules via their gears as shown in Figure 24a. Thus, it is possible that C_{60} molecules roll by means of torque, which is estimated to be approximately 1.0×10^{-19} Nm. This indicates that the rolling of C_{60} molecules occurs easily along the $[12\bar{3}0]$ direction using one side of a six-membered ring. However, there exist only four six-membered rings along a great circle of a C_{60} molecule, which does not always mean that the six-membered ring situates on the top of the C_{60} molecule after rolling. Now, the energetic barrier for rotation of a C_{60} molecule around the [0001] axis of a graphite substrate is estimated to be a few mili-electron-volts by Gravil et al. [12]. Therefore it is possible that the discrete step rotations of a C_{60} molecule

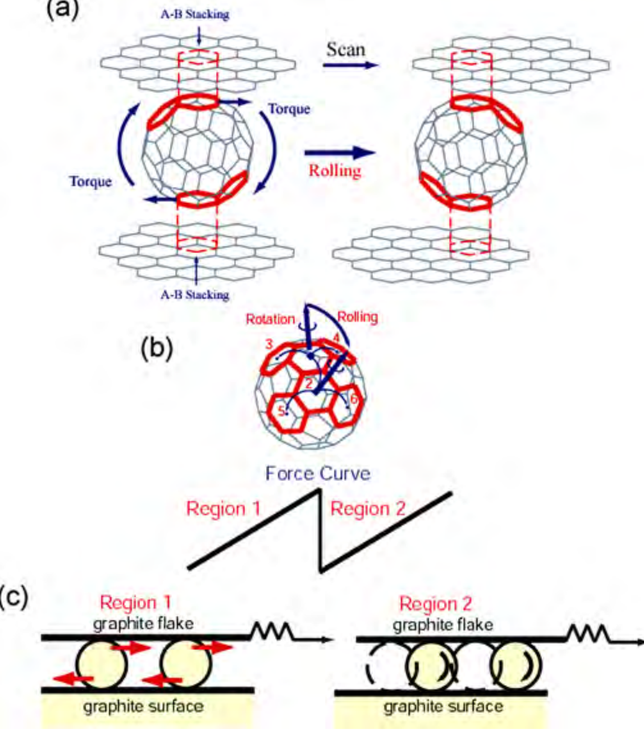


Figure 24. The stick-slip rolling model with a step rotation of a C_{60}molecule. (a) C_{60} molecules roll by means of torque, which is produced by the nanogears of six-membered rings between C_{60} molecules and the upper and lower graphite plates. (b) The discrete step rotation of a C_{60} molecule around the [0001] axis is illustrated. (c) In the region 1, C_{60} molecules stick but thermal rotations around the [0001] axis occur. In the region 2, C_{60} molecules slip in both rolling and translational motions (bottom figure).

between the positions holding AB stacking around this axis are induced by thermal excitations at room temperature. This step rotation allows C_{60} molecules to roll so that a low friction force generates. For example, when the rotation around the [0001] axis occurs, the rolling of the C_{60} molecule with the step rotation at the six-membered ring 1 occurs equivalently toward rings 2, 3, and 4. Assuming that the C_{60} molecule rolls toward the ring 2, the next rolling occurs equivalently toward rings 1, 5, and 6. Assuming that the C_{60} molecule rolls toward the ring 5, the six-membered ring of the C_{60} molecule facing to the graphite surface advances such as $1 \rightarrow 2 \rightarrow 5$ as shown in Figure 24b. Thus, a stick-slip process for rolling and translational motions is given as follows. In the region 1 in Figure 24c, C_{60} molecules stick but thermal rotations around the [0001] axis occur. In the region 2 in Figure 24c, C_{60} molecules slip in both rolling and translational motions. Then, nearly all the potential energies stored by a cantilever transform into rolling and translational motions. Thus, it should be noted that friction heat does not generate in this system. It can be expected that the van der Waals interaction between C_{60} molecules gives significant influences on the collective motion of C_{60} molecules, which can produce the super cell structure in frictional force maps.

The appearance of the vertical chainlike pattern with larger loads indicates the increase of the deformations of C_{60} molecules which leads to the vertical chainlike pattern by decreasing C_{60} nearest-neighbor distances. Here, it may be interesting to note that the deformations of the C_{60} monolayer induce a transition from a semiconductor phase to a metal phase [30].

As shown in Figure 25, the frictional force maps for graphite/C_{60}^{bilayer}/graphite (F) are also clearly different from those for graphite (A), C_{60}(111) surfaces (B and C), the graphite flake on graphite (D), and graphite/C_{60}^{mono}/graphite (E) from the standpoint of periodic pattern and contrast, as shown in Figures 22 and 23. The frictional force maps for graphite/C_{60}^{bilayer}/graphite (F) do not exhibit clear load dependence although they feature periodic patterns. As shown on the right-hand side of Figure 25, plots of mean frictional force versus loading force for graphite/C_{60}^{bilayer}/graphite (F) are shown. The curve of frictional force versus loading force from graphite/C_{60}^{bilayer}/graphite behaves in the same manner as does the Herztian contact, because the solid line is defined as $\alpha(F_n^{2/3} - \beta)$ in relation to the loading force F_n with α and β estimated to be 0.068 and 30 nN, respectively. Thus, this indicates that the C_{60} bilayer confined by graphite plates behaves as an elastic body. Furthermore, it should be noted that as C_{60} molecular layers become thick, the frictional force maps from E and F disappear but the frictional force map from graphite is reproduced. This indicates that the tip executes stick-slip motions on the graphite flake so that the C_{60} molecules and graphite substrate situated below the graphite flake do not move as well as the graphite flake because the combined shear force between the C_{60} layers and the shear force between the C_{60} molecules and graphite plate become larger than the shear force between the tip and the graphite flake.

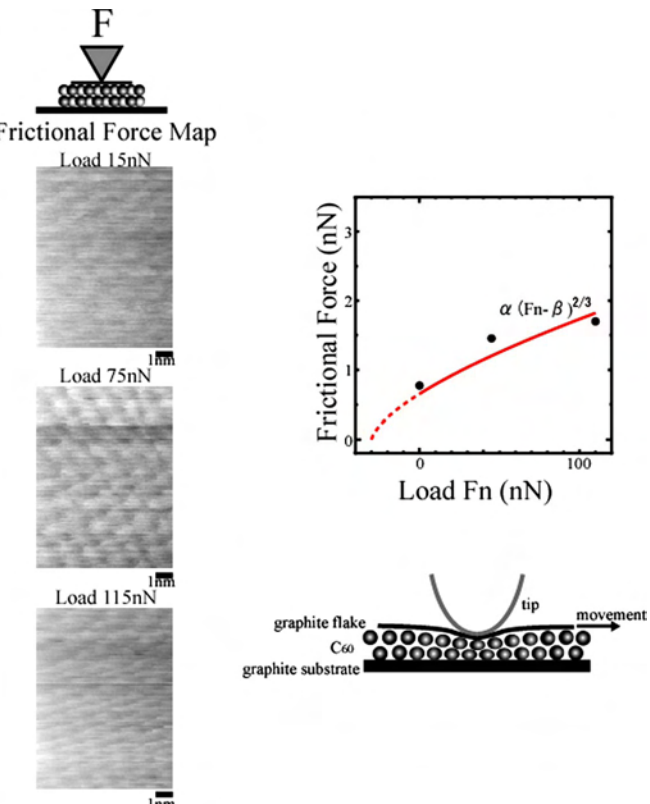


Figure 25. The frictional force maps and the curve of mean frictional force versus loading force for graphite/C_{60}^{bilayer}/graphite (F). The frictional force maps for graphite/C_{60}^{bilayer}/graphite (F) do not exhibit clear load dependence although they feature periodic patterns. The curve of mean frictional force versus loading force from graphite/C_{60}^{bilayer}/graphite behaves in the same manner as does the Herztian contact, because the solid line is defined as $\alpha\ (F_n^{2/3} - \beta)$ in relation to the loading force F_n with α and β estimated to be 0.068 and 30 nN, respectively.

GLOSSARY

AB stacking The way in which graphite layers stack up.

Amontons–Coulomb law The classical law of friction, that is, the relationship that the friction force is proportional to a loading force.

Atomic force microscope (AFM) The apparatus capable of investigating the surface topography on the atomic scale.

Frictional force microscope (FFM) The apparatus capable of investigating the lateral force on the atomic scale.

Micromachine Microscale machine.

Molecular manipulation The technique to make a molecule manipulate arbitrarily.

Multiwalled carbon nanotube (MWNT) Multiwalled nanotube consisting six-membered carbon ring network.

Nanogears Nanoscale gears.

Nanomachine Nanoscale machine.

Single-walled carbon nanotube (SWNT) Single-walled nanotube consisting of six-membered carbon ring network.

Stick-slip Periodic phenomena of sticking and slip.

REFERENCES

1. Y. Zhang, T. Ichihashi, E. Landree, F. Nihey, and S. Iijima, *Science* 285, 1719 (1999).
2. J.-P. Salvetat, G. A. D. Briggs, J.-M. Bonard, R. R. Bacsa, A. J. Kulik, T. Stöckli, N. A. Burnham, and L. Forró, *Phys. Rev. Lett.* 82, 944 (1999).
3. M. R. Falvo, R. M. Taylor II, A. Helser, V. Chi, F. P. Brooks Jr., S. Washburn, and R. Superfine, *Nature* 397, 236 (1999).
4. M. R. Falvo, J. Steele, R. M. Taylor II, and R. Superfine, *Phys. Rev. B* 62, R10665 (2000).
5. A. Buldum and J. P. Lu, *Phys. Rev. Lett.* 83, 5050 (1999).
6. S. Okita, M. Ishikawa, and K. Miura, *Surf. Sci. Lett.* 442, L9590 (1999).
7. S. Okita and K. Miura, *Nano Lett.* 1, 101 (2001).
8. K. Miura, M. Ishikawa, R. Kitanishi, M. Yoshimura, K. Ueda, Y. Tatsumi, and N. Minami, *Appl. Phys. Lett.* 78, 832 (2001).
9. H. Hölscher, U. D. Schwarz, and O. Zwörner, Wiesendanger, *Phys. Rev. B* 57, 2477 (1998).
10. N. Sasaki, K. Kobayashi, and M. Tsukada, *Phys. Rev. B* 54, 2138 (1996).
11. Z. Y. Li, *Surf. Sci.* 441, 366 (1999).
12. P. A. Gravil, M. Devel, Ph. Lambin, X. Bouju, Ch. Girard, and A. A. Lucas, *Phys. Rev. B* 53, 1622 (1996).
13. "Nanoscience: Friction and Rheology on the Nanometer Scale" (E. Meyer, R. M. Overney, K. Dransfeld, and T. Gyalog, Eds.), p. 123. World Scientific, Singapore, 1999.
14. K. Miura, T. Takagi, S. Kamiya, T. Sahashi, and M. Yamauchi, *Nano Lett.* 1, 163 (2001).
15. K. Miura, http://133.96.34.82/.
16. S. Fujisawa, K. Yokoyama, Y. Sugawara, and S. Morita, *Phys. Rev. B* 58, 4909 (1998).
17. J.-C. Chalier, X. Gonze, and J.-P. Michenaud, *Europhys. Lett.* 28, 403 (1994).
18. D. Tomanek, W. Zhong, and H. Thomas, *Europhys. Lett.* 15, 887 (1991).
19. M. Weiss and F.-J. Elmer, *Phys. Rev. B* 53, 7539 (1996).
20. T. Gyalog and H. Thomas, *Europhys. Lett.* 37, 195 (1997).
21. P. E. Sheehan and C. M. Lieber, *Science* 272, 1158 (1996).
22. K. Miura and S. Kamiya, *Europhys. Lett.* 58, 603 (2002).
23. M. Hirano, K. Shinjo, R. Kaneko, and Y. Murata, *Phys. Rev. Lett.* 67, 2642 (1991).
24. J. M. Martin, Le. C. Donnet, Th. Mogne, and Th. Epicier, *Phys. Rev. B* 48, 10583 (1993).
25. M. Ishikawa, S. Okita, N. Minami, and K. Miura, *Surf. Sci.* 445, 488 (2000).
26. W. Carpick, D. F. Ogletree, and M. Salmeron, *Appl. Phys. Lett.* 70, 1548 (1997).
27. H. Yoshizawa, Y.-L. Chen, and J. Israelachvili, *J. Phys. Chem.* 97, 4128 (1993).
28. A. Fogden and L. R. White, *J. Colloid Interface Sci.* 138, 414 (1990).
29. R. Feynman, *Eng. Sci.* 23, 22 (1960).
30. J. Nakamura, T. Nakayama, S. Watanabe, and M. Aono, *Phys. Rev. Lett.* 87, 48301 (2001).