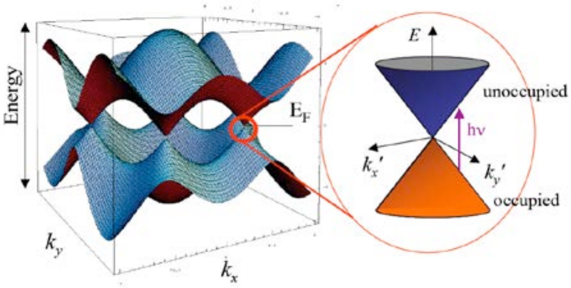
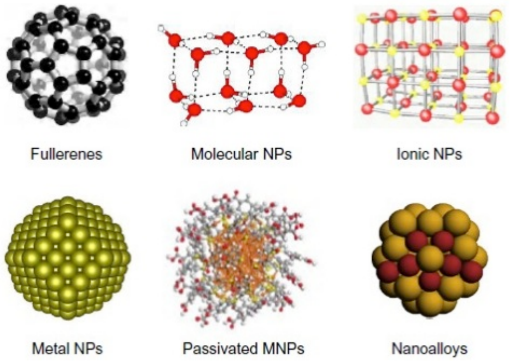
№530785-TEMPUS-1-2012-1-PL-TEMPUS-JPCR

Institute of High Technology, Taras Shevchenko National University of Kyiv

Advansed Materials of Micro and Nano Technology

Valeriy Skryshevsky, Anatoliy Evtukh, Valeriy Lozovski, Oleg Tretyak





Textbook

2016

Textbook " **Advansed Materials of Micro and Nano Technology**" developed to help higher education institutions in Ukraine to introduce new master's educational program "Designing microsystems".

Textbook " **Advansed Materials of Micro and Nano Technology**" was created with the support of the European Union within the Joint European Project "Curricula Development for New Specialization: Master of Engineering in Microsystems Design" (MastMST), identification number 530785-TEMPUS-1-2012-1-PL-TEMPUS-JPCR.

Project Coordinator prof. Zbigniew Lisik, Lodz University of Technology, Lodz, Poland.

TEMPUS teams:

- LvivPolitechnicalNationalUniversity, Lviv, Ukraine ,
  Coordinator prof. Mykhailo Lobur.
- Taras Shevchenko National University of Kyiv, Ukraine,
  Coordinator prof. ValeriySkryshevsky.
- Kharkiv National University of Radioelectronics, Ukraine ,
  Coordinator prof. Vladimir Hahanov.
- Donetsk National Technical University, Krasnoarmiysk,
  Coordinator prof. VolodymyrSviatny.
- Ilmenau University of Technology, Germany,
  Coordinator prof. Ivo Rangelow.
- Lyon Institute of Applied Sciences, France,
  Coordinator prof. Alexandra Apostoluk.
- University of Pavia, Italy,
  Coordinator prof. Paolo Di Barba.

The Textbook was approved by Editorial Committee (prof. Paolo Di Barba (University of Pavia) - Co-Chair, prof. Alexandra Apostoluk (LyonInstitute of Applied Sciences) – Co-Chair, members: prof. Zbigniew Lisik (Lodz University of Technology), Dr.Jacek Podgorski (Lodz University of Technology), Dr.Janusz Wozny (Lodz University of Technology), Dr.Valentyn lshchuk (Ilmenau University of Technology), Dr.Maria Evelina Mognaschi (University of Pavia), Dr.Roberto Galdi (University of Pavia) May 6, 2016, Pavia, Italy.

The authors express their deep gratitude to the aforementioned universities for full support of the project.

Content

Numerical Laboratory Works

Chapter 1. History of development and modern trends in materials fabrications for MEMS and NEMS.

Micro- and nano-electromechanical systems represent rather broad and diverse technological areas, such as optical systems (micromirrors, waveguides, optical sensors, integrated subsystems), life sciences and lab equipment (micropumps, membranes, lab-on-chip, membranes, microfluidics), sensors (bio-sensors, chemical sensors, gas-phase sensors, sensors integrated with electronics) and RF applications for signal transmission (variable capacitors, tunable filters and antennas, switches, resonators). From a scientific viewpoint, this is a very multi-disciplinary field, including micro- and nano-mechanics (such as stresses in structural materials), electronic effects (e.g. charge transfer), general electrostatics, materials science, surface chemistry, interface science, (nano)tribology, and optics. It is obvious that in order to overcome the problems surrounding next-generation MEMS/NEMS devices and applications it is necessary to tackle them from different angles: theoreticians need to speak with mechanical engineers, and device engineers and modelers to listen to surface physicists.

Micro/nanorobots and microactuators find their applications in various domains: microUAV in military, microrobots for in-body exploration in medicine, microactuators for microassembly and micromanipulation and for surface characterization with nanometric resolution, etc. In order to reach the severe performances required for these “micro/nano” applications—such as very high resolution, micrometric or submicrometric accuracy, and high bandwidth—convenient design of the actuators and convenient control of them are necessary. Smart materials like piezoelectric and electroactive polymers and flexible structures are among the best candidates to design the actuators, but their characteristics (nonlinearities, badly damped vibrations, etc.) require the use of efficient control techniques. In addition to these characteristics, the particularity of working at the micro/nano-scale (lack of embeddable sensors, high sensitivity to the environment, difficulty to directly sense, general uncertainties on the model used) makes their control even more challenging. Several researches have been carried out since many years with tremendous success; however, limitations still exist in terms of precision, operating speed, and reliability. Furthermore, emerging and new requirements in micro/nano-scale positioning bring additional challenges for the design and the control of the systems.

These challenges include the high axiscoupling in multi-degrees of freedom precise positioning systems, the limitations of the existing sensors to measure the signals in them, the high environmental (thermal, vibration...) sensitivity and the high noiseto- signal ratio as they are more and more small. During these last years, in order to tackle these challenges, advanced design concepts, where some are jointly coupled with control theory to include the performances at the design level, novel sensing and actuation combined technologies, noise measurement and resolution estimation techniques, and advanced control techniques with or without sensors have been developed.

Advanced probes and new fabrication techniques enable nanomaterials to pervade multiple disciplines, including physics, chemistry, engineering and biology. Nanomaterials have been extensively investigated with various kinds of morphologies (nanoparticles, nanowhiskers, nanorods, nanowires, nanoclusters, quantum dots, etc.) and compositions (semiconductor, metal, polymer, etc.). Impressive progress has been made on directed assembly and synthesis, structure, and property characterization, as well as nanoscale device concepts and performance by a diverse group of experts. However, in spite of continued advancements in various aspects of functional nanomaterials, numerous challenges must still be overcome at different stages for practical applications to be realized.

Concepts of micromechanical devices

Silicon deforms elastically and is a very robust mechanical material. Silicon has small thermal expansion, high speed of sound and very low intrinsic mechanical losses. There exist various methods for both isotropic and anisotropic etching of silicon. The well controlled anisotropic etching of silicon especially has enabled the fabrication of microscale mechanical structures and devices. When these mechanical characteristics and the micromachining are combined with the mass scale manufacturing solutions from the IC industry, we have a platform for producing micromechanical sensors and actuators in huge volumes at low cost.

First mechanical characterization and experiments in chemical wet etching of silicon were already carried out in the 1950s; and the industrial scale MEMS was born in the late 1970s. Today the various devices and even smaller objects of our physical world are becoming extensively connected to the digital information networks. Small processors and communication capabilities can be integrated into a variety of small physical objects. The microelectromechanical components and systems (MEMS) developed over the past 30 years form the technology platform that enables sensing and even actuation in these small devices.

Most of the concepts of micromechanical devices and fabrication technologies go back in time to the 1960s and the early 1970s. The mechanical characteristics of single crystalline silicon, the anisotropic etching of silicon by potassium hydroxide (KOH) and several concepts of micromechanical devices were studied. An interesting example of early device concepts is shown in Fig. 1- a resonant gate transistor based on a micromechanical resonator made of electroplated gold in 1967.

Integration of micromechanical structures, such as resonators, was a natural continuum from the pre-silicon electronics to the early concepts of silicon integrated devices. The first commercial application by Hewlett Packard used a MEMS cantilever based frequency detector in frequency synthesizers in 1980. In spite of these early concepts, the development of practical commercial applications based on MEMS resonators has taken nearly 30 years, even today very few commercial devices exist besides the scientific instruments, such as, the resonating cantilevers of the atomic force microscopes and the biosensors .

The capability to create large arrays of similar components with precise dimensions enabled new device concepts. Furthermore, the possibility to benefit from the fabrication capabilities that were scaling up in the IC industry created the starting point towards the commercial applications of MEMS.

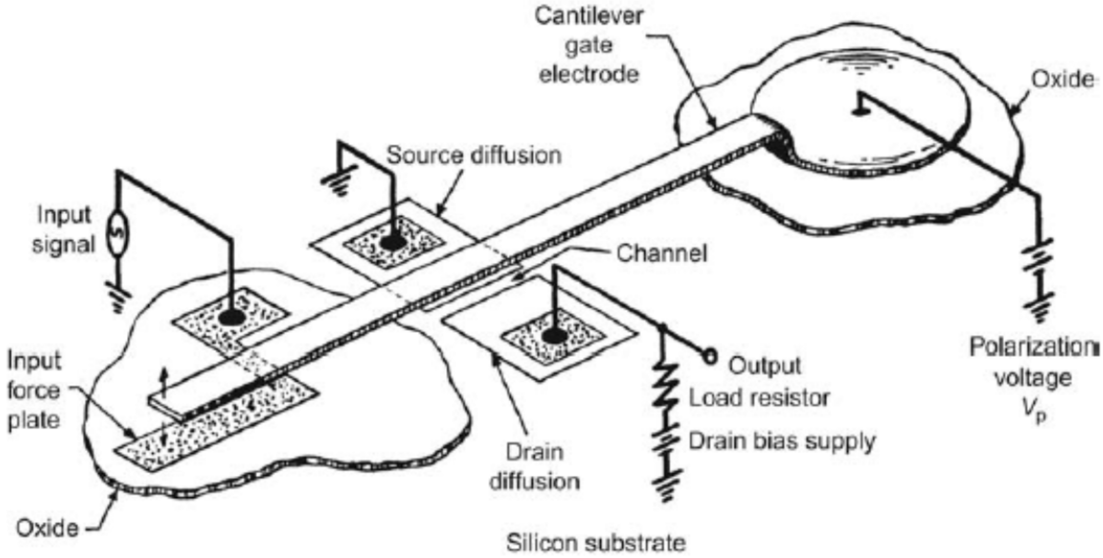


Fig 1 Resonant gate transistor published in 1968

Ink Jet Printer Nozzles Create the Industry

This capability to manufacture precise components and arrays of micromachined structures practically enabled the ink jet printers during the 1980s. IBM demonstrated the value of silicon micromachining to achieve the necessary printing precision with the integrated methods to control and manage the inks in the same micromachined device. The first mass volume application of silicon MEMS was created, and the ink jet printers became the main stream of printing in the growing information technology market.

The electrostatic control of ink jets for printing purposes was studied in the beginning of the 1970s. The capability to etch the ink jet nozzles into the silicon wafer using anisotropic KOH wet etching and to integrate the control electrodes into the same device using thin film and semiconductor processes enabled the sufficient miniaturization to improve the quality of ink jet printing to the level where the commercial solutions were possible.

Still today the ink jet printer application forms one third of the total MEMS market. Today the printing is expanding as a paradigm to electronics manufacturing. Reel-to-reel manufacturing solutions based on either inorganic or organic inks will be one of the future solutions for manufacturing low cost electronics.

Automotive Applications Drive the Reliability and the Quality

The automotive applications of pressure and motion sensors practically created the MEMS industry. The manifold air pressure (MAP) sensor introduced by Ford in the mid seventies was the first micromechanical sensor in mass volumes. The accelerometers were introduced to replace mechanical switches in airbag launchers, and later they enabled sophisticated chassis control systems. The automotive industry already in the 1980s was characterized by advanced project management and quality control that included all the module manufacturers and subcontractors. The duration of product development projects was long and required a commitment for several years. The MEMS developers soon learned to apply these strict rules of project and quality management. The high requirements for reliability created long development projects with careful testing and verification phases.

Even the simplest MEMS devices require insulating layers. The successful practical method was to use a sandwich of silicon and borosilicate glass wafers. The wafers were bonded together by a so called anodic bonding process. However, the difference in the thermal expansion of the glass and silicon wafers was a problem causing strong temperature dependence and even warping or buckling of the micromechanical structures. The solution was to use only very thin insulating glass layers between silicon wafers. It was introduced glass wafers with thermal expansion characteristics that matched the silicon.

Bulk micromachining enabled several different products: pressure sensors, accelerometers, lab-on-chip devices, etc. As the bulk micromachining provided an inherent wafer level packaging of the micromechanical structures, the very accurate control of the gas damping of accelerometers or the reference pressure of the absolute pressure sensors was feasible. The other paradigm was to grow a polysilicon thin film on top of a sacrificial silicon dioxide layer. The polysilicon film was anchored to the underlying silicon wafer and patterned to form the particular mechanical structure, e.g., the proof mass and the capacitor structures of an accelerometer. When the sacrificial silicon oxide layer was removed, the mechanical structure was released to move.

The promise of the polysilicon surface micromachining was in the integration of mechanical structures with CMOS electronics. The approach was very successful in the development of accelerometers for the airbag application that was not as demanding on the acceleration resolution. The thickness of the polysilicon layer determines the proof mass of the accelerometer, and the intrinsic acceleration resolution is inversely proportional to the square root of the proof mass. Thus the smaller mass of the polysilicon structures became a limiting factor for the application that required high acceleration resolution. After solving the challenge of wafer level encapsulation, the polysilicon surface micromachining provided smaller dimensions at lower cost, possibility for monolithic integration with CMOS devices and allowed more complex mechanical structures.

In the mid 1990s two disruptive technologies appeared. The deep reactive ion etching (DRIE) of silicon using an inductively coupled plasma source, the Bosch process, made it possible to etch deep high aspect ratio trenches into silicon. Secondly, the development of silicon on insulator (SOI) wafers enabled high quality relatively thick monocrystalline silicon layers for micromechanical structures. These technologies brought together the advantages of the bulk and surface micromachining. Through these innovative processes the MEMS technologies are becoming a true generic technology and manufacturing platform.

The CMOS electronics integration with MEMS devices has been an intensive topic of research and development. Both pre- and post- CMOS integration have been used. The key challenge has been the different pace in the development of MEMS and CMOS processes. In addition, the cost of the area on a CMOS wafer is getting increasingly more expensive; the CMOS components get smaller but the miniaturization of the MEMS devices is constrained by the sensor resolution and the necessary physical size.

New Trends

At the end of the 1990s the micromechanical pressure sensors and accelerometers existed in relatively large volumes for the automotive industry. The sensor elements, signal conditioning electronics and the packaging were developed according to the automotive system and environmental requirements. When consumer electronics and mobile phone manufacturers became interested in sensors and other MEMS devices suitable for their products, immediately they faced some challenges:

i) The interface electronics with sufficiently low power consumption and low supply voltage level did not exist.

ii) Even though a large variety of packaging standards existed, the miniaturization had not addressed the critical requirement of thinness.

iii) ☐Strict requirements for sensor performance and reliability increased the cost, and the testing and verification during both product development and production were very expensive procedures.

iv) Manufacturing volumes were not yet sufficient for consumer applications, and the lack of long-term commitment from the consumer industry did not support the investments to scale up the production capabilities.

v) Industry dynamics of consumer products and mobile phones were disruptive to the operations of the sensor manufacturers: emphasizing faster ramp-ups, shorter commitments, speed of innovation, and extreme cost consciousness.

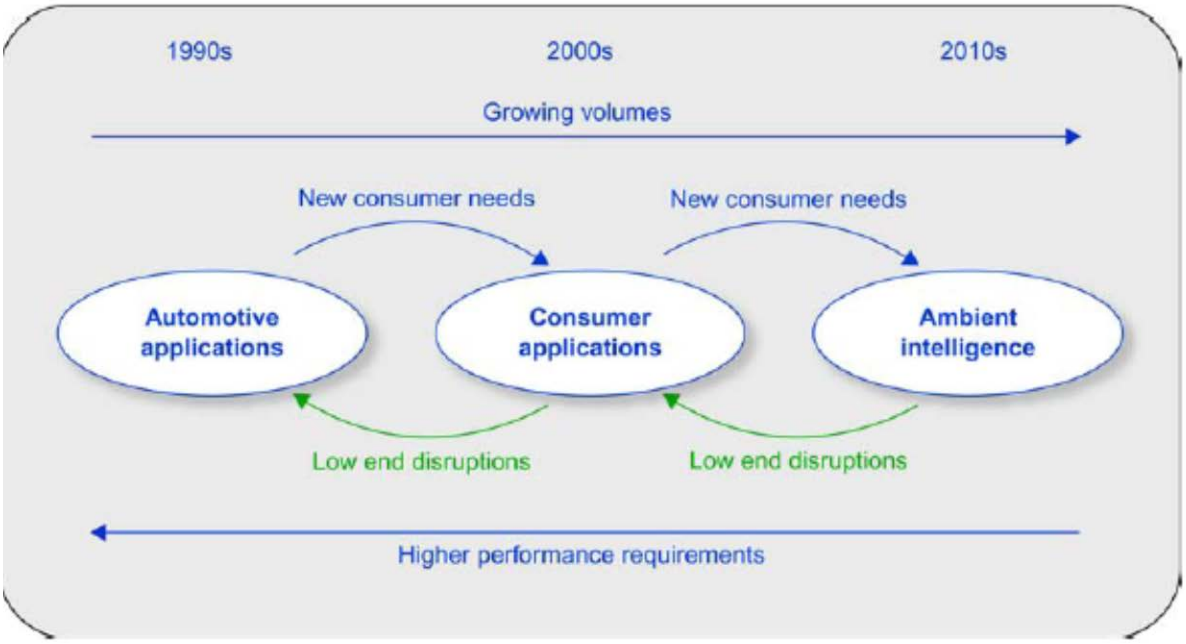


Fig 2 MEMS devices are moving towards consumer products and will enable the embedding of intelligence into human environments, i.e., the ambient intelligence

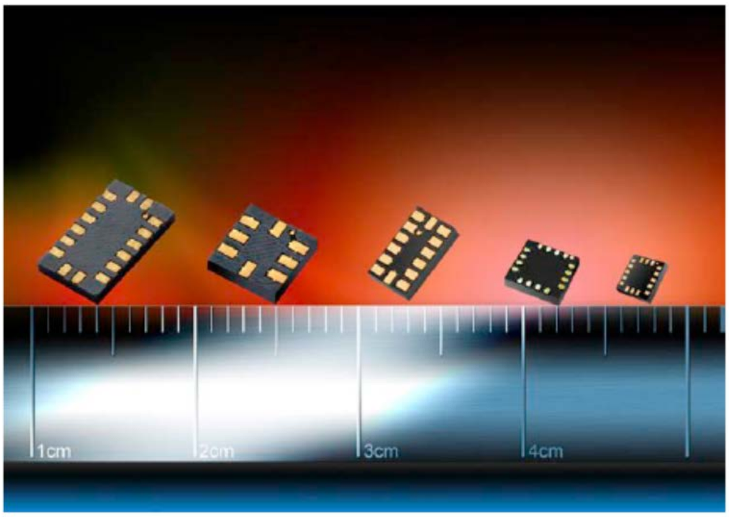


Fig 3 Miniaturization of packaged MEMS accelerometers by ST Microelectronics.

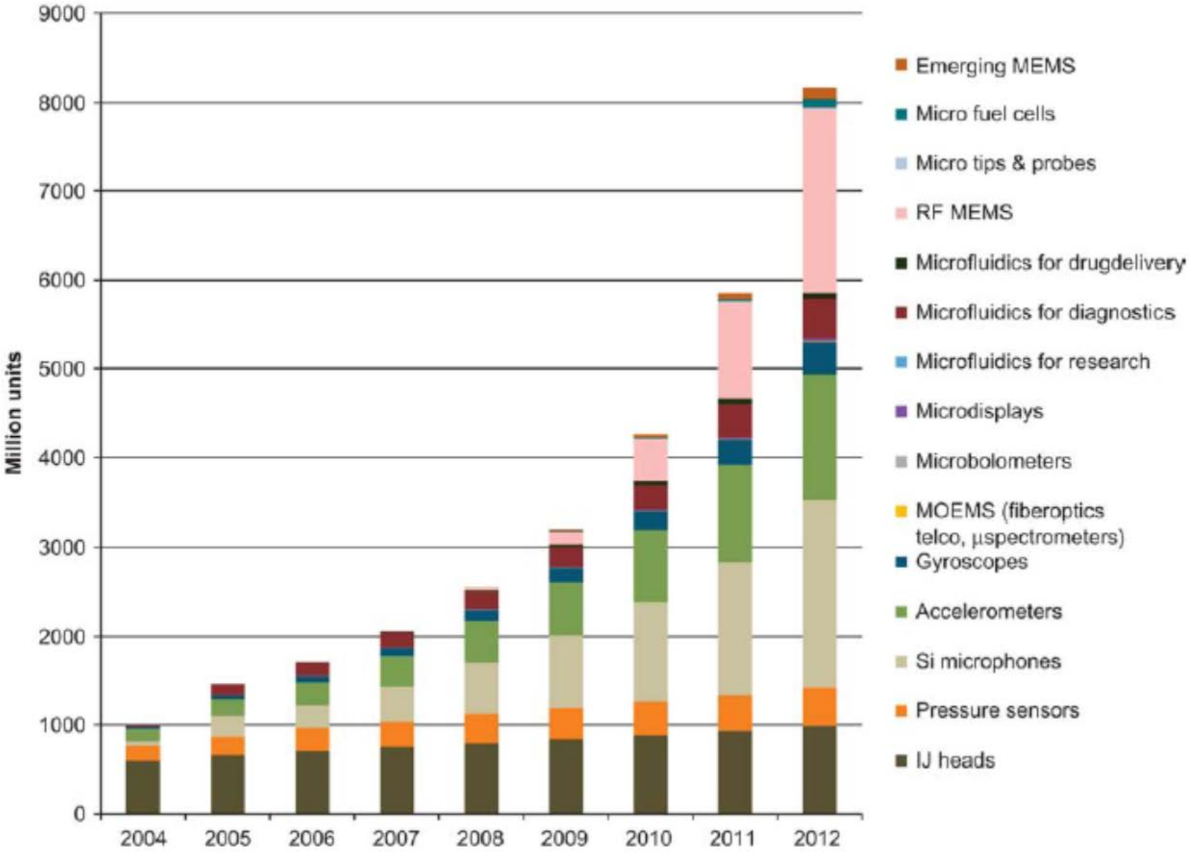


Fig 4 The estimated MEMS market trend.

Figure 2 illustrates the major shifts in the MEMS industry. The consumer products created a demand of lower cost MEMS devices with more relaxed requirements on performance. Especially, measurement electronics and packaging required rethinking. Figure 3 illustrates an example of the miniaturization of accelerometer modules for consumer products, for the game controllers and the mobile phones. In the early 2000s the ground was very fruitful for new innovation, and the global MEMS industry was very optimistic about these new application areas.

Figure 4 shows a recent estimate of the growth of the MEMS component volumes. Even the total unit volume of MEMS components is expected to nearly triple during the coming five years, the total annual growth of the revenue is clearly smaller with roughly 14%. Most of the growth is expected to come from the consumer applications of inertial sensors, silicon microphones and RF MEMS. In addition, the significant growth of microfluidics systems in medical applications, such as, the drug delivery, may happen during the coming five years. The price erosion of the MEMS components can actually be even more dramatic due to the extreme cost competition in consumer products and mobile phones. Eventually, the new cost structure of MEMS devices will affect the automotive sensor market as well.

Consumer Products

From Wristwatch to Wearable Electronics

The measurement of human scale motion by accelerometers and angular rate sensors can be used to extract and recognize relatively complex patterns of human scale motion, such as, walking, running, climbing stairs. Figure 5 shows an example of a wearable pilot device developed by Nokia in 2003. The device was capable of detecting motion patterns and recording the daily activity of a person. Even more complex motion patterns, e.g., playing golf or exercising yoga, can be recognized, recorded and used to help the person to learn, repeat and to correct these complex motion patterns. Professional training methods thus become possible for any serious amateur sportsman, and eventually, the sports and gaming will merge into various interesting combinations of physical and virtual realities.

Fig 5 Nokia Fitness Monitor pilot product that measures and recognizes various motion patterns.

The data projectors and projection displays became a new driver of MEMS development after the invention of micromirror devices at the end of the 1980s. The DLP technology by Texas Instruments was ahead of its time in complexity when it was first introduced in 1993. Based on digitally controllable aluminum micromirrors, the DLP chip enabled the miniaturization of video projectors when the personal computers became the primary tool for creating presentations. The miniaturization of data projectors has made them portable. The further miniaturization depends primarily on the power consumption and efficiency of the light source and the thermal management of the device. In addition to LED projectors also the laser sources, even RGB laser sources , are appearing. The micromirror based reflection devices will, however, meet a lot of competition from other technologies, such as, holographic displays that can be much more power efficient.

Gaming

The future of user interface and human machine interaction may already be visible in computer games and game controllers. Ideas for using sensors and haptics in the user interfaces of gaming devices has been developed by several research groups during the last ten years.

Figure 6 presents this kind of prototype device developed and published by Nokia Research Center in 2004. The detection of movements and gestures can be used to create an immersive gaming experience; a combination of a near-eye display, advanced audio functionality, gesture and body movement detection and a haptic feedback interface are ways to create virtual reality games or to augment the physical reality by virtual features.

Fig 6 3D accelerometer and 2D angular rate sensor-based game controllers created by Nokia Research Center in 2004

Medical Applications of MEMS Devices

Today the most important MEMS devices in medical applications are the microfluidics systems, pressure sensors for blood pressure monitoring and the accelerometers for cardiac pacemakers. The largest potential for new applications and business growth is related to microfluidics used in diagnostics and drug delivery.

The micromachined structures have made it possible to handle very small samples of liquids in a reliable way. Furthermore, the MEMS miniaturization creates a possibility to use simultaneous parallel measurements in the same system. These highly integrated lab-on-chip solutions change the nature of the *in vitro* diagnostic devices. Multichannel integrated biosensors, advanced sample handling and new sensitive transducers will enable point-of-care devices that are capable of, e.g., cancer diagnostics or any other analysis of the genomes.

The continuous monitoring of physiological parameters of a patient opens opportunities for more precise diagnostics. In addition to monitoring of heart rate, ECG and blood figures the possibility to monitor and register the daily activity and context of the patient during these measurements gives more information that may have a correlation to the physiological data and that can help in the diagnostics. The continuous monitoring and the possible remote connectivity will open a new segment of wearable health care products.

Mobile Phones and Mobile Multimedia Computers

The mobile phones have become digital cameras, music players, internet browsers, mobile TVs in addition to being communication devices that people carry with them always and everywhere. The mobile phones integrate a lot of functionality from consumer appliances and personal computers into a small integrated device. Figure 7 presents potential new arising trends in the mobile communication industry. Currently, the industry is shifting towards Internet-based mobile services. This is a clear consequence of new Internet-based business models and the ongoing convergence of mobile communication, information technologies and consumer electronics.

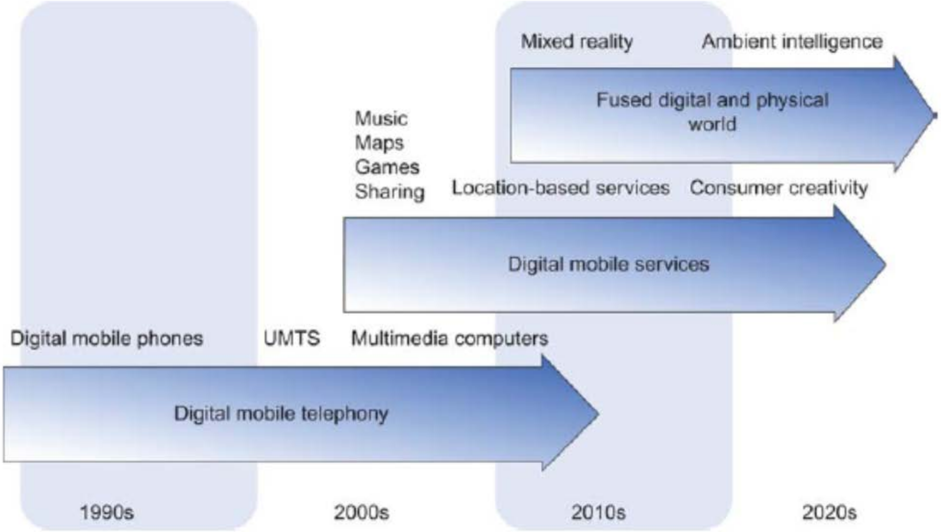


Fig 7 Mobile industry in shift towards services and ambient intelligence.

RF MEMS

The development of reliable RF MEMS devices has, however, taken much longer than expected in the beginning. There are some fundamental dimensional constraints that make the integration of RF MEMS components to mobile devices challenging:

i) The conductivity of even highly doped bulk silicon or polysilicon is not sufficient for a high enough quality factor for capacitors, inductors or switches in the frequency range of 1–5 GHz.Thus the use of metal film based micromechanical structures is required, leading to several other challenges.

ii) MEMS resonators are limited to roughly 10 MHz frequency. At higher frequencies, e.g., in 1–5 GHz range, the fabrication tolerance of narrow electrode gaps, very high control voltages and the lower Q values of resonators limit the development of practical devices. Recently, the development of piezoelectric actuators for MEMS resonators has improved the efficiency of electrical coupling.

Sensors and Actuators

The number of possible sensor applications in mobile devices is huge. Table 1 summarizes some possible mobile device applications and their requirements for the sensor system. Even if it is possible to list a large number of use cases and applications, the key challenge is how to make the functionality so generic that it enables this multitude of applications for various consumers.

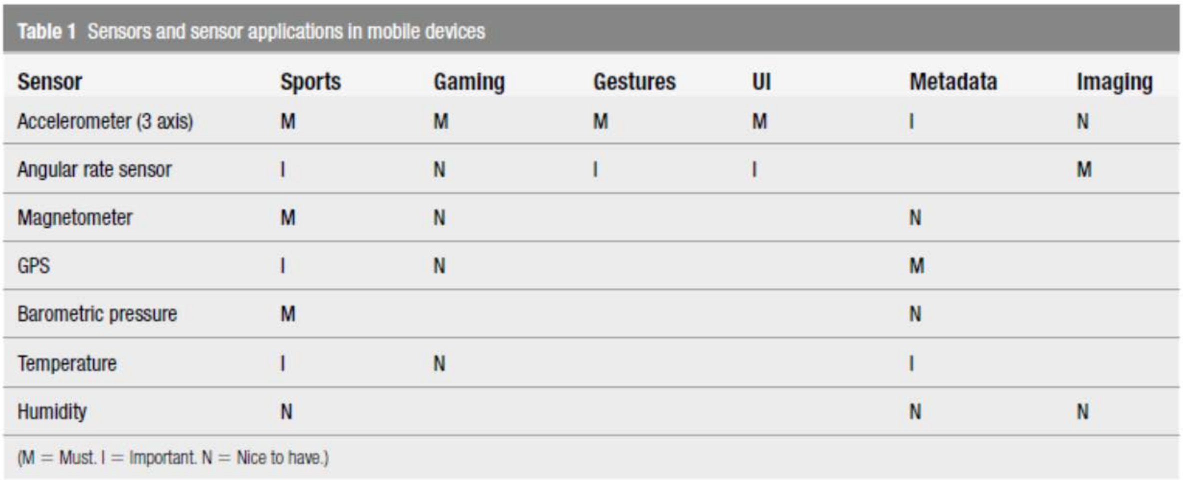


Table 1 Sensors and sensor applications in mobile devices

Sensor	Sports	Gaming	Gestures	UI	Metadata	Imaging
Accelerometer (3 axis)	M	M	M	M	I	N
Angular rate sensor	I	N	I	I		M
Magnetometer	M	N			N	
GPS	I	N			M	
Barometric pressure	M				N	
Temperature	I	N			I	
Humidity	N				N	N

(M = Must. I = Important. N = Nice to have.)

Based on the context information it is possible to build the applications for sports, gaming, gesture-based device control, user interface control and adaptation, automatic generation of image metadata, more sophisticated imaging algorithms, etc. Various sensor applications can be created using a generic software library that consists of routines to compute the various abstraction levels of context.

Towards Modular Architectures in Consumer Products

The requirements for the MEMS-based technologies in the consumer electronics can be contradictory. There is a clear tendency towards modularity; the new functionality should be easy to implement in the devices. On the other hand, the large volumes create severe cost constraints. MEMS products with different levels of integration and complexity are needed, e.g., sensors varying from simple indicators to integrated complex intelligent wireless sensors.

Figure 8 presents a possible configuration of an intelligent sensor module for a consumer product. The typical features of such an implementation are that the sensors and their primary signal conditioning circuitry are integrated into the same module. However, the system of Fig. 8 has several other interesting features. An integrated low power DSP core and a specific hardware accelerator are used for sensor signal processing. The feedback to the actuators controlled by the sensor signals is computed locally to minimise the control loop delay. The efficient power management circuitry is used to control the various sleep and active states of the device. The regulation of MEMS control voltages and the system clock are integrated into the system. Optionally, a general purpose signal processor can be added for more complex computational tasks, such as, the asynchronous communication with the host processor.

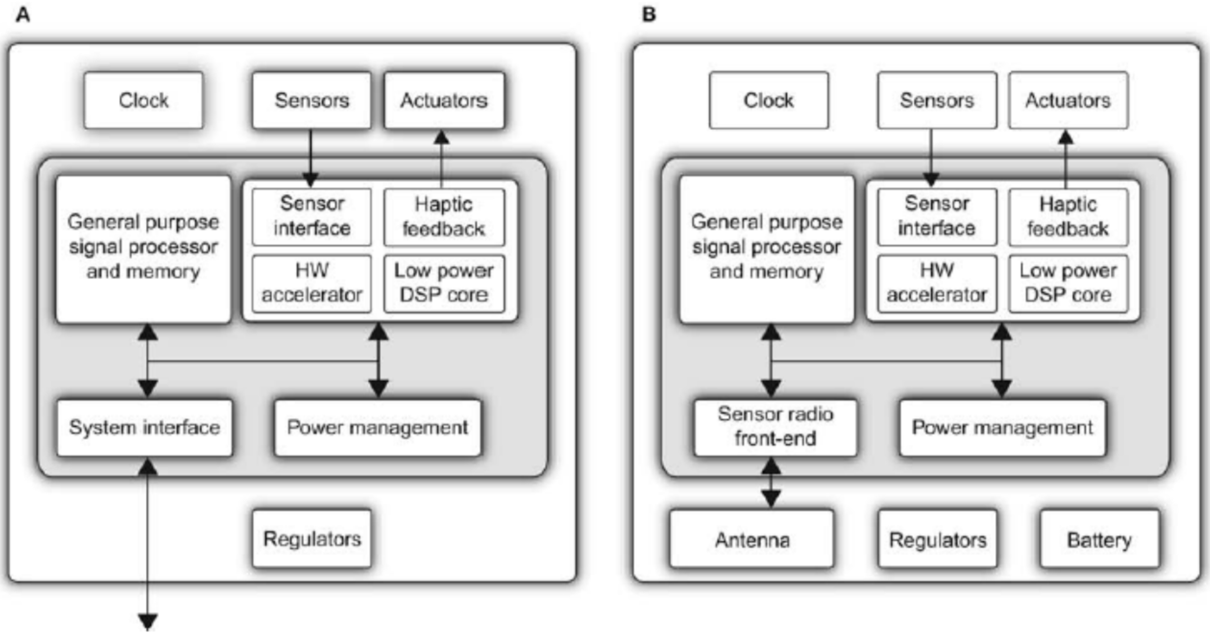


Fig 8 Architecture of an autonomous smart sensor module. (a) Sensor module that can be integrated via asynchronous digital communication interface. (b) Wireless autonomous device with its own energy sources and storage.

Ubiquitous Sensing, Computing and Communication

Merged Physical and Digital Worlds

The next possible phase of this development has many names: Ambient Intelligence, Pervasive Computing, Ubiquitous Computing (see Fig. 7). The networked devices embedded in our physical environment provide sensing, computing and communication services that can be accessed locally. When this localized information is connected to the global information networks that are capable of data aggregation, we will live in a responsive environment that knows our preference and adapts to serve our particular needs. The microsystems of sensors, processors and radios are essential enablers of these intelligent environments.

Wireless Sensing and Sensor Networks

The wireless sensor networks are enablers for the Internet of Things and the future ambient intelligence. The sensor networks can be used for optimizing the logistics and transport; the possibilities to localize and to measure the condition of transported goods can be used to optimize logistics cost and energy consumption. The safety and security can be increased by more efficient sensor networks.

Table 2 summarizes the key figures of merit of the most important initiatives of wireless sensor radios, the Bluetooth, Zigbee and the Near Field Communication (NFC) standards. These standards have not been developed and optimized for actual multi-hop sensor network applications. The basic concept of these radios is related to connecting various sensors into a host device, such as, a mobile phone, that can be a gateway into the information networks.

A wireless sensor consists typically of a radio, a microprocessor, a small memory, and an energy source in one single robust package. Figure 8 presents a possible architecture of wireless sensor that has specific requirements related to the radio communication and autonomous operation: extremely low power consumption enabling energy autonomy, capability to wake-up the system by an external sensor signal or an external radio (i.e., a wake-up radio), and an efficient control of the measurement and signal processing duty cycles. Furthermore, the wireless sensor nodes need to be easy to deploy, self-configuring, extremely robust and fault tolerant

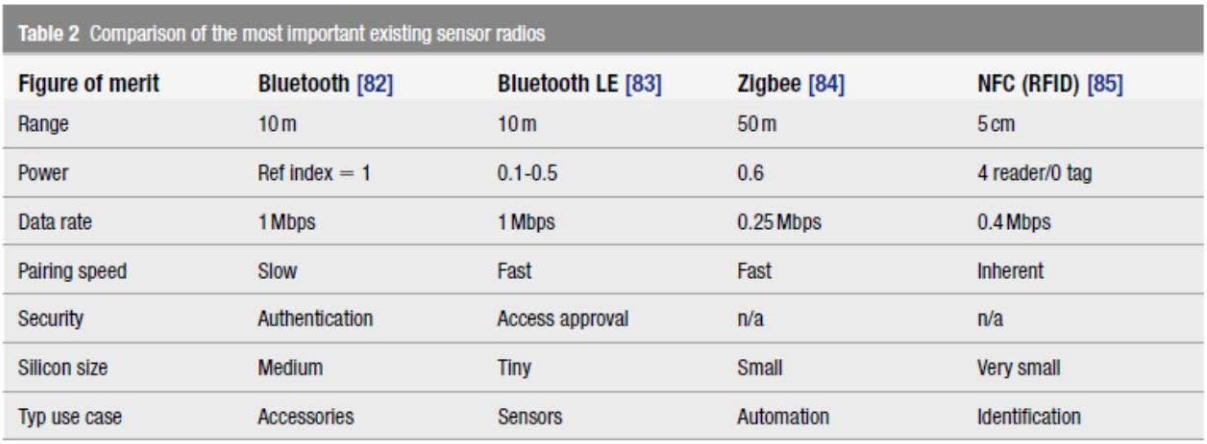


Table 2 Comparison of the most important existing sensor radios

Figure of merit	Bluetooth [82]	Bluetooth LE [83]	Zigbee [84]	NFC (RFID) [85]
Range	10 m	10 m	50 m	5 cm
Power	Ref index = 1	0.1-0.5	0.6	4 reader/0 tag
Data rate	1 Mbps	1 Mbps	0.25 Mbps	0.4 Mbps
Pairing speed	Slow	Fast	Fast	Inherent
Security	Authentication	Access approval	n/a	n/a
Silicon size	Medium	Tiny	Small	Very small
Typ use case	Accessories	Sensors	Automation	Identification

Mobile Phone as a Sensor

As discussed before, soon 90% of the global human population will have mobile coverage. In some years these devices have become also sensors and computers that are inherently connected to the global information network. The mobile devices can have several roles in the context of ambient intelligence and sensor networks.

The mobile device is already today a user interface towards the global Internet based services and digital content, and the user interfaces of the mobile devices will have improved capabilities for this multimodal information. Secondly, based on the efficient radio solutions (see Table 2) the mobile devices are also becoming a user interface to access the sensors and other devices in the local environments. Figure 9 presents the paradigm of a mobile device as a trusted personal user interface towards the local and global information.

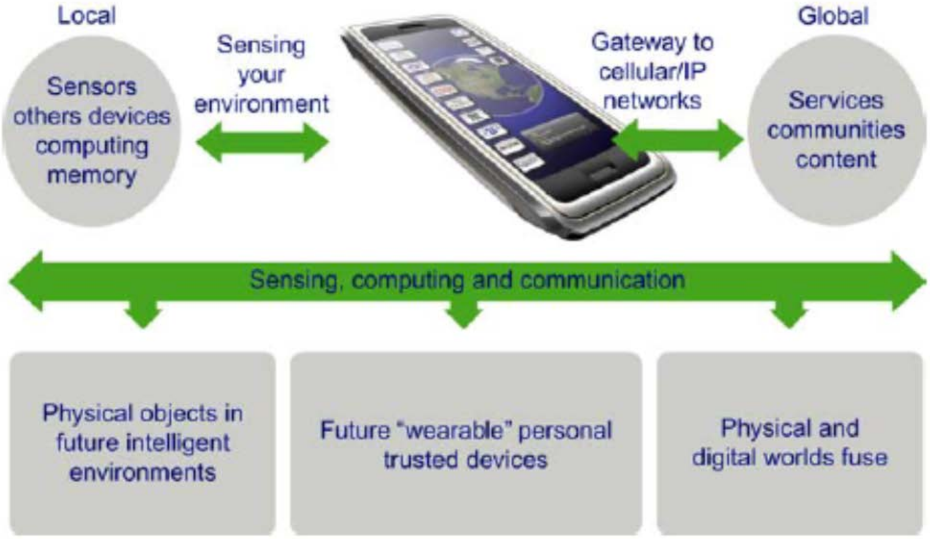


Fig 9 Mobile phone as a gateway between local intelligent environment and global digital services and content.

Future of MEMS Technologies

Silicon is a nearly perfect mechanical material and enables the micromachining of very precise structures. There are also some limitations related to the cost and the ultimate material properties. In microfluidics and optical applications, e.g., in diffractive optics, the use of polymer substrates and structures is much more inexpensive in large volumes. The sophisticated technologies for replicating the microstructures into polymer surfaces have been developed. The basic idea of the process is to fabricate the primary mechanical structure into the surface of the silicon wafer, deposit a metal film on top of the silicon structure and electroplate a thick metal layer (for example, a nickel layer). When the thick metal layer is removed from the surface of the silicon wafer, we have a mold to replicate the structure on a polymer surface, for example, using injection molding or hot embossing.

In RF applications the conductivity of even the very highly doped silicon or polysilicon is not sufficient for a high enough quality factor for tunable RF capacitors, switches or inductors. The use of metal films has been a way to increase performance of the micromechanical components. The use of free standing metal thin films for tunable or switchable structures is shown to be possible. The behavior of metal thin films differs from bulk metals; for example, inelastic deformations are smaller.

The novel nanomaterials will create a possible impact in the future MEMS. The use of carbon nanotubes as piezoelectric transducers has been demonstrated with good results. The use of bottom up grown silicon nanowires for resonators has been studied. Nanotechnologies will create new material solutions for MEMS, and possibly some new nanoscale mechanical structures, so called NEMS structures, will become commercially feasible some day.

Figure 10 illustrates a concept of a future consumer device based on polymer electronics, stretchable electronics and new functional surface materials . The key attributes of such a future device are transformability, compliancy in terms of controllable flexibility and stretchability, extreme thinness and transparency. The integration of functionality will be based on reel-to-real assembly and printing of interconnects together with many active and passive printed components . This kind of device is still very far in the future but the first steps towards the novel integration solutions will be taken within the coming years based on printed electronics.

Fig 10 A transformable and wearable device, based on stretchable electronics. Industrial design by Nokia

Platform for Nanoscience and Nanotechnologies

The MEMS has also created capabilities for the probing and manipulation of matter in the nanoscale. The atomic force microscopy based on micromechanical cantilevers and tips is one of the key methods for scanning, imaging and measuring nanoscale objects. The micromechanical probes can be used for accessing nanoscale systems and their dynamical properties. The micromechanical systems will play a significant role in developing the nanoscale electronics, functional nanomaterials and surface structures.

The explosion of digital content in terms of images, music and videos creates an increasing need for high density, low power and intelligent mass storage systems. The magnetic hard disks will not scale to the memory densities that would fulfill these requirements.

The micromechanical cantilever array has already been developed for nanoscale mass storage concept. Several companies, such as, IBM, HP, ST Microelectronics and Seagate, have been working for a concept called probe storage. The concept is based on reading, writing and erasing information on the surface of a polymer film, metal film or a magnetic film using a microelectromechanical cantilever and a micromachined tip.

The IBM Millipede has a capability to read, write and erase dots on the surface of the polymer film with the pit resolution of 20 nm. The micromachined system is relatively complex with the array of cantilevers and the actuators to move the plate hosting the polymer film. However, the manufacturing of the MEMS system is feasible today; the challenge of the probe memory technologies is the stability of the nanostructures used for the memory

Chapter.2 CLASSICAL EFFECTS IN LOW DIMENSIONAL SYSTEMS

There are two kinds of effects connected with low dimension of the systems. First of all there are the effects of space quantitisations. Second set of effects are the classical effects caused by interfaces. From this point of view the next systems should considered as low dimension: small particles (mesoparticles); thin films and interfaces.

Indeed the interfaces should be the causes of the changes of properties of solids. For example, the specific effects of the energy structure of semiconductors such as вигин зон can be observed. The surface polaritons can be excited at the metal surface *etc*. These phenomena we shall consider in detail below and we shall see that decreasing of the dimensions of the systems leads us to the new world with the other wonderful properties characterized by bright effects.

2.1. The conductivity of thin metal films

As it is well known the band electron in the perfect crystal moves under external electric field without scattering which lead to infinity conductivity. Indeed the relaxation time in this case is infinitely large $\tau \to \infty$, the electric conductivity becomes infinit $\sigma = \left(ne^2 / m\right)\tau \;\to\; \infty$ (see Fig.2.1). Because under non zero temperature the perfect crystal does not exist, electrons at least will be scattered by the crystal lattice oscillations. The numerous mechanisms of electron scattering are in

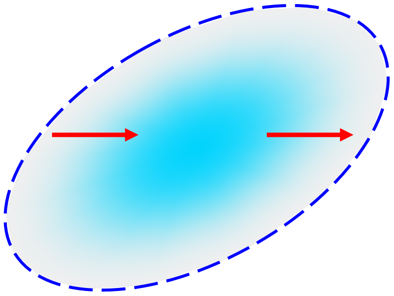


Fig.2.1. The electron moves through the perfect crystal without scattering.

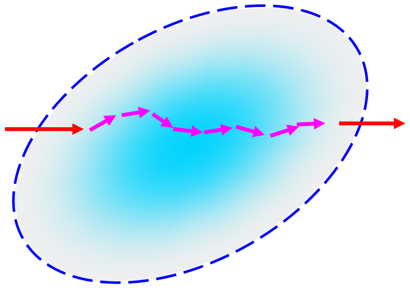


Fig.2.2. The electron moves trough the real crystal and scatters at the different defects

the real crystals. In general the electron scattering processes caused by lattice defects. The processes lead to any relaxation time. This is the cause of electric resistance. Then, the electric conductivity in the crystal metal arises due to scattering processes. Image that any bulk sample of a metal or a semiconductor becomes thinner and thinner. When one obtains the thin film which thickness are about the free path of electrons the processes of scattering at the interfaces become play an essentional role. These scattering processes can influence on the electric conductivity formation. When the characteristic relaxation times of the surface scattering are rather small the processes can determine the conductivity of thin film. Obviously, the processes should lead to increase of thin film conductivity. Then, one can hope to obtain the decreasing of electric conductivity of thin film. The effect of decreasing electric conductivity of thin film can be considered as nonquantum effect in low-dimension systems. The effect can arise when the additional scattering on the interfaces processes are characterized by rather small relaxation times. These scatterings obviously occur simultaneously with the electron scatterings on the lattice defects. The effect of increasing of electric conductivity in thin films arises obviously when the thickness of the film is about or smaller that the free path of electrons and relaxation time of scattering on the interfaces are smaller or about the smallest relaxation time of scatterings on the lattice defects.

The general idea of change of conductivity of thin film can be simply formulated as the next: The conductivity of metal or semiconductor is determined by equation $\sigma = ne^2\tau / m^*$. The relaxation time one evaluates as $\tau \approx l / v$, with l the free path, and v averaging velocity of electron. When the thin film interfaces put a geometric restrictions for the electron moving, the free path defends by the film thickness d. Then, relaxation time should be evaluated as $\tau \approx d / v$. These facts lead to changing of electric conductivity compare with the conductivity of bulk sample.

Taking into account these circumstances K.Fuchs proposed the simple model for evaluation of conductivity of thin films. Analyze, following Fuchs, the influence of electron scattering by interfaces on the electric conductivity in thin films forming. As it was mentioned above, the spatial quantization processes will be neglected. Namely, one shall suppose that the film thickness is longer for de Broigle wavelength of electron.

To evaluate the electron scattering by interfaces influence on the electric conductivity one writes Boltzmann kinetic equation

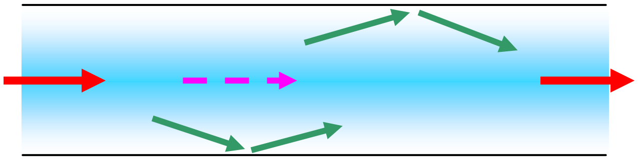


Fig. 2.3. Electron scattering by defects and surfaces in the thin film.

$$-\frac{e}{m}\mathbf{E}\cdot\mathrm{grad}_v f+\mathbf{v}\cdot\mathrm{grad}_r f=\left(\frac{\partial f}{\partial t}\right)_{coll}, \tag{2.1}$$

where f is the distribution function of electrons, $\mathbf{E}$ is an electric field. If f_0 is a distribution function of unexcited state, then $\partial f_0/\partial t=0$. When the shift from equilibrium state is small, one can use the relaxation time approximation

$$\left(\frac{\partial f}{\partial t}\right)_{coll}=-\frac{f-f_0}{\tau}. \tag{2.2}$$

Because the film is inhomogeneous along its thickness (normally to the film plane – OZ axis of Cartesian coordinate system), one can write for the electrons distribution function

$$f=f_0+f_1(v,z), \tag{2.3}$$

with де f_1 nonequilibrium part of distribution function. Then, Boltzmann equation has a form

$$\frac{\partial f_1}{\partial z}+\frac{f_1}{\tau v_z}=\frac{eE}{mv_x}\frac{\partial f_0}{\partial t}. \tag{2.4}$$

Let one shall find the solution of Eq.(2.4) in the form

$$f_1(v,z)=\frac{eE\tau}{m}\frac{\partial f_0}{\partial v_x}\left[1+F(v)\exp(-z/\tau v_z)\right], \tag{2.5}$$

with $F(v)$ any function, which should be defined from the boundary conditions. To obtain the result which can якісно describe the effects of contributions of scattering by interfaces one can use simple boundary conditions diffusion scatterings of electrons by the surface. These boundary conditions mean that each of electron trajectories after electron collision with a lattice defect is finished by collision with surface. Electron distribution function in this case does not depend on direction of electron moving. Moreover relaxation processes of electron scattering by the surface occur similarly to the processes in the volume. Mathematically these circumstances are written as the next:

$$\begin{aligned} f_1(v,0) &= 0\,, \qquad v_z > 0\,, \\ f_1(v,d) &= 0\,, \qquad v_z < 0\,. \end{aligned} \tag{2.6}$$

Taking into account the boundary conditions [Eq.(2.6)], one obtains the solution in the form

$$f_1(v,z) = \frac{eE\tau}{m}\cdot\frac{\partial f_0}{\partial v_x}\times\begin{cases} 1-\exp\left(-\dfrac{z}{v_z\tau}\right), & v_z > 0, \\ 1-\exp\left(\dfrac{d-z}{v_z\tau}\right), & v_z < 0. \end{cases} \tag{2.7}$$

The current density defines by non-equilibrium part of electron distribution function

$$j(z) = -2e\left(\frac{m}{h}\right)^3 \iiint v_x\, f_1(v,z)\,dv_x dv_y dv_z\ . \tag{2.8}$$

Then, total current through the film is obtained by integration along the film thickness

$$J = \int_0^d dz\; j(z)\,, \tag{2.9}$$

and conductivity is defined by standard equation

$$\sigma = J\,/\,E\,. \tag{2.10}$$

When $d \to \infty$ one can define from Eq.(2.9) the conductivity of bulk sample. As a result, one obtains

$$\frac{\sigma_{bulk}}{\sigma_{film}} = \frac{\varphi(\gamma)}{\gamma}, \tag{2.11}$$

with

$$\varphi^{-1}(\gamma) = \frac{1}{\gamma} - \frac{3}{8\gamma^2} + \frac{3}{2\gamma^2}\int_1^\infty \left(1/x^3 - 1/x^5\right)e^{-\gamma x}dx, \quad \gamma = d/l. \tag{2.12}$$

Let us analyze obtained result. In the case of thick film (when $\gamma \sim 1$), the relative change of the film conductivity is

$$\frac{\sigma_{bulk}}{\sigma_{film}} \approx 1 + \frac{3}{8\gamma}. \tag{2.13}$$

Then, the difference between conductivities of a bulk sample and thick film is $(3/8)(l/d)$. Namely, when film thickness is more the free path of electrons for 10 times, the conductivity correction is not more than 0,04. Then the conductivity of thick film is defined by conductivity of material from which the film is fabricated with the accuracy 4 %. In the case of thin film (when $\gamma \ll 1$) one obtains from Eq.(2.11)

$$\frac{\sigma_{bulk}}{\sigma_{film}} \approx \frac{4}{3\gamma \ln(1/\gamma)}. \tag{3.14}$$

The dependence of relative conductivity of thin film on the parameter γ in the case of diffuse scattering is shown in Fig.2.4 (curve 1). As we can see at the film thickness about 0.2*l* film conductivity decreases relatively the bulk conductivity for three times. Because the electron free path for the metals under room temperature is about 10-50 nm, the помітний difference between the bulk and thin film conductivities can be observed only for very thin films. The electron free path increases up to few hundred nanometers under low temperatures. Then the effects of difference of electric conductivity of thin film from bulk conductivity can be observed at the film thickness 10-50 nm.

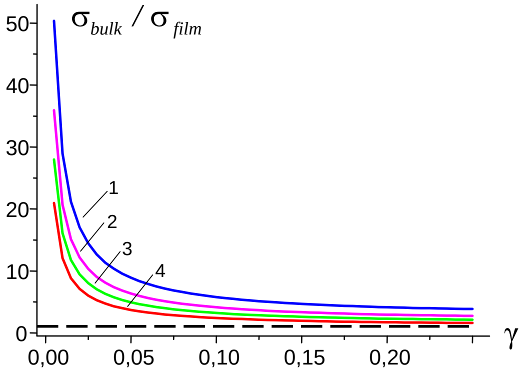


Fig. 2.4. The dependence of relative conductivity of thin film on ratio film thickness to electron free path. Curve 1 corresponds to diffuse scattering by interfaces; curves 2-4 demonstrate the contribution of mirror electron reflecting by interface: 2 - $p = 0,2$;
3 – $p = 0,4$; 4 – $p = 0,7$

The method demonstrated above can be generalized for the case when the part of electrons are elastically scattered by the film interfaces (so-called mirror reflection). Let the part of electrons (*p*) are scattered by the upper and bottom surfaces of the film with the change of velocity direction (parallel to OZ axis) by opposite. The other (1- *p*) part of electrons are scattered diffusely with full loosing of drift velocity at the film surfaces. Approximately the value *p* can be supposed as constant. Then, the boundary condition can be written in generalized form

$$f_0 + f_1^+(v, z=0) = p\left[f_0 + f_1^-(-v_z, 0)\right] + (1-p)\, f_0 \quad , \qquad z = 0\,, \tag{2.15}$$

$$f_0 + f_1^-(v, z=d) = p\left[f_0 + f_1^+(-v_z, 0)\right] + (1-p)\, f_0 \quad , \qquad z = d\;. \tag{2.16}$$

Tking into account these boundary conditions, one can find from Eqs.(2.4-2.5)

$$f_1^+(v,z) = \frac{eE\tau}{m}\frac{\partial f_0}{\partial v_x}\left[1 - \frac{1-p}{1-p\exp(-d/v_z\tau)}\exp\left(\frac{-z}{v_z\tau}\right)\right], \qquad v_z > 0\,, \tag{2.17}$$

$$f_1^-(v,z) = \frac{eE\tau}{m}\frac{\partial f_0}{\partial v_x}\left[1-\frac{1-p}{1-p\exp(d/v_z\tau)}\exp\left(\frac{d-z}{v_z\tau}\right)\right], \qquad v_z<0. \tag{2.18}$$

Then, solution of Eqs.(2.4-2.5) can be written in the form of Eq.(2.11) with

$$\varphi_p^{-1}(\gamma) = \frac{1}{\gamma} - \frac{3}{2\gamma^2}(1-p)\int_1^\infty dx\left(1/x^3 - 1/x^5\right)\left(1-e^{-\gamma x}\right)\left(1-pe^{-\gamma x}\right)^{-1}. \tag{2.19}$$

For the case of thick films (when $\gamma > 1$) one obtains

$$\frac{\sigma_{bulk}}{\sigma_{film}} \approx 1+\frac{3}{8\gamma}(1-p), \qquad \gamma>1. \tag{2.20}$$

When thin film, the difference of film conductivity from the bulk conductivity one can find

$$\frac{\sigma_{bulk}}{\sigma_{film}} \approx \frac{4}{3}\frac{1}{\gamma(1+2p)}\frac{1}{\ln(1/\gamma)}, \qquad \gamma \ll 1;\ p<1. \tag{2.21}$$

The elastic scattering processes influence on the thin film conductivity is shown in Fig,2,4 (curves 2 – 4). One can see that elastic scatterings strongly decrease the difference between the film conductivity and bulk conductivity. Note, the more relative quantity of elastically scattered electrons (p), the less the difference. But the conductivity of very thin (γ = 0,05) film is in many time less than bulk conductivity.

Considered method can be generalized for the case the elastic scattered processes as by interface as by upper film surface (Fig,2,5). This case, for example, corresponds to the film situated at the substrate. Let p is the part of electrons – which elastically scateered by interface but q is the part of electrons which elastically scattered by upper surface of the film. In this case instead Eqs.(2.20) and (2.21) one writes

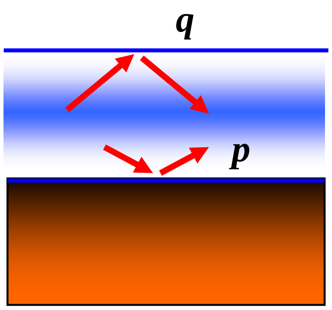


Fig. 2.5. Elastic electron scattering by upper film surface and interface.

$$\frac{\sigma_{film}}{\sigma_{bulk}} \approx 1-\frac{3}{8\gamma}\left(1-\frac{p+q}{2}\right), \qquad \gamma \gg 1, \tag{2.22}$$

$$\frac{\sigma_{film}}{\sigma_{bulk}} \approx \frac{3}{4}\frac{(1+p)(1+q)}{1-pq}\gamma \ln(1/\gamma), \qquad \gamma \ll 1. \tag{2.23}$$

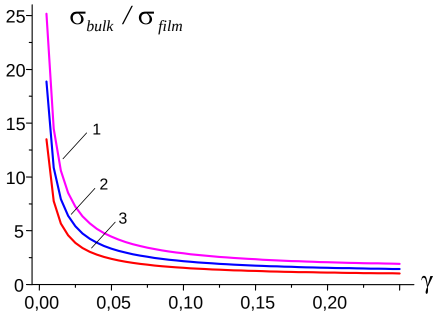


Fig.2.6. The dependence of relative conductivity of thin film on γ: curve 1 corresponds to p = 0,2, q = 0,2; curve 2 corresponds to p = 0,5, q = 0,2; curve 3 corresponds to p = 0,7 , q =0,3

As one can see from the dependence of conductivity of thin film at the substrate, when the part p of electrons are elastically reflected by upper film surface and the part q of electrons are elastically reflected by interface "film-substrate", the difference of film cnductivity from bulk conductivity is less than when only diffuse scattering by the film surfaces or/and elastic scattering only by film surface (Fig.2.6). As we can see in (Fig. 2.6), conductivity of thin film less differs from bulk conductivity when parts p and q of electrons mirrored by upper and bottom interfaces, respectively than diffuse reflection and when only one interface reflects part of electrons mirrorly. Fig.2.7 demonstrates that simple Fuchs model adequate describes the behavior of conductivity of thin film which thickness is about (60 нм < d < 160 нм). Then in this chapter using simple Fuchs model, we demonstrated the classical dimension effect of conductivity dependence on the film thickness. The dependence origins of additional electron scattering by the film interfaces when the film thickness is about the electron free path. Note, the simple model considered here does not take into account many other aspects of conductivity formation. For example the model does not take into account the processes of electron scattering by

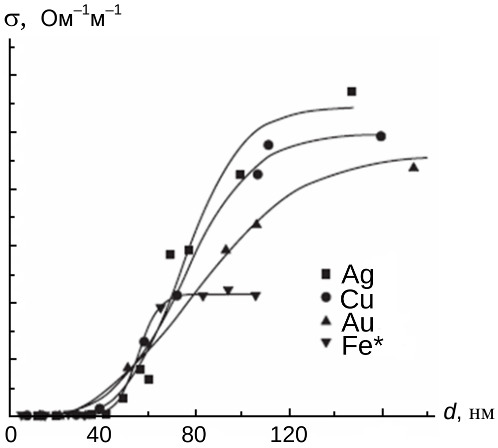


Fig. 2.7. The dependence of conductivity of thin metallic films on its thickness.

interfaces of grain boundaries in the polycrystalline thin films. Nevertheless the model could be essentially improved by introducing the phenomenological parameter describing the electron scattering by the interfaces.

2.2. The optical properties of nanostrutures

Remember, we define the low-dimension object as such which dimension at least along one axis is much less than characterisctic length parameters of external factors. To low-dimension effects existence the linear dimension of the system should be about the characteristic length which play an essential role in formation of the effects. For example, when the linear dimensions of the system are less/about de Broglie wavelength the space quantization effects can be observed. To observe the 'classic' dimension effect the wavelength of external radiation or thickness of the transitional layer at the surface of the object, or electron free path in the effect of enhancement of thin film conductivity considered above.

The nano-technologies which nowadays demonstrate impetuous development has led to the necessity of consideration of electrodynamics of the systems, characterized by small linear dimension [1-3]. In particular this problem became actual for consideration of elctrodynamical properties of quantum dots [4,5] and quantum dot arrays [6,7], development of scanning near-field optical microscopy [8-10], electrodynamics of nano-composite systems [11-13], etc. In any aspect of these one can point to some achievements in solving these problems. Namely, some ideas of the local field were used as base for numerous works in which the nonlocal electrodynamics of quantum well [14-16] and quantum dot structures [17-19] were studied. The main idea of the most of these works is finding the local field at arbitrary point of the system under consideration. The distillation of local-field method was introduction of local-field factor which connects the short-range local field with long-range external field (see, for example, the reviews [20-23]). The method is widely used and many results were obtained in the frame of the local-field method. The next step of nano-particle electrodynamics is reformulation of constitutive equations from connection between the local current and local field to connection between the local current and external (long-range) field. This connection lies in the base of effective susceptibility concept discussed in the present work. This approach has a clear physical argument – only external long-range field is controlled in experiment. A local-field having the short-range nature can not be measured directly in experiment. One should note that term “effective” here has any different mean from term “effective” in the effective medium theory [24-27]. The effective medium theory is based on the idea of replacing the inhomogeneous (composite) medium by an equivalent homogeneous medium such that the fluctuations induced by restoring the heterogeneity average to zero. Then, roughly speaking, the idea of effective medium theory is replacing one (having complicated structure) medium by another one – homogeneous simple medium characterized by some effective permeability [25]. Contrary to this, the idea of effective susceptibility concept consists in consideration an object (the nano-particle, thin film, etc.) which responds to external (probing) signal as a single whole. Then the linear response to the external field (effective susceptibility) is obviously the characteristic of the system. This characteristic is not trivial. Indeed, on the one hand the idea is to consider the system as a single whole but on the other hand the effective susceptibility connects the local characteristic of the system – the local current, which can strongly change from point to point inside the object – with the external long-range field acting on the system as a single whole. The advantage of effective susceptibility introduction is obviously the possibility to calculate the most of electrodynamical characteristics of the system because the local current distribution which is the source of the field, can be calculated in the frame of discussed approach.

2.2.1. Linear responses to the local and an external field

Contemporary electrodynamics dealt with small (nano) objects needs to distinguish local and external fields [20]. Let the external long-range field acts on the bulk sample made from homogeneous material which electrodynamical properties are described by dielectric constant $\varepsilon_{ij}(\omega)$. It is well known, in any bulk homogeneous material the local field $E_i(\mathbf{R},\omega)$ differs from the external ones $E_i^{(0)}(\mathbf{R},\omega)$ by dielectric constant $\varepsilon(\omega)$. Then, when the medium does not characterised by spatial dispersion, the local field can be defined via external field as

$$E_i(\mathbf{R},\omega) = \varepsilon(\omega) E_i^{(0)}(\mathbf{R},\omega). \qquad (2.24)$$

In this case, when the external field, measured somewhere at infinity, depended on the coordinate as $E_i^{(0)}(\mathbf{R},\omega)=E_i^{(0)}(\omega)e^{i\mathbf{kR}}$, the local field inside the bulk sample depends on coordinate of observation point only via phase factor

$$E_i(\mathbf{R},\omega)=E_i(\omega)e^{i\mathbf{k}'\mathbf{R}}\,, \qquad (2.25)$$

with $\mathbf{k}'=\sqrt{\varepsilon(\omega)}\mathbf{k}$ wave vector of the electromagnetic wave propagating inside a bulk. This local field is usually named the electric induction [28-29]. The dielectric constant shows the field enhancement (weakening) by the local currents inside the bulk induced by external field. Differently other situation arises when one considers the acting long-range external field on the small particle. In this case, because near-surface transitional layer fills rather essential part of the volume of the particle, the local field will be strongly inhomogeneous. Then, the connection between the local and external fields can not be presented in a simple form of Eq.(2.24). To find this connection one should remember, that induced current and local field are connected by linear response to the local filed (or electrical susceptibility)[1]

$$j_i(\mathbf{R},\omega)=-i\omega\chi'_{ij}(\omega)E_j(\mathbf{R},\omega)\,. \qquad (2.26)$$

This relation is usually named as constitutive equation [28, 29]. The constitutive equation is not convenient for study of electrodynamics of small systems because the local field inside the particle can not be measured and controlled experimentally. Only external field can be controlled in experiment. Then, analogously to constitutive equation one can introduce the relation connecting the local current with external field.

$$j_i(\mathbf{R},\omega)=-i\omega X'_{ij}(\mathbf{R},\omega)E_j^{(0)}(\mathbf{R},\omega)\,, \qquad (2.27)$$

---

[1] Of course, *j* in this equation is current density. But for simplicity here and below we shall suppose that the volume of the systems under consideration is equal unit and, then, say *current* instead *current density*.

One should note that the connection between the local current and external field [Eq.(2.27)] has an universal form. Indeed, in general case, because of nonlocal nature of the electrodynamic interactions this connection should be written in the form

$$j_i(\mathbf{R},\omega) = -i\omega \int_V d\mathbf{R}' \tilde{\chi}_{ij}(\mathbf{R}-\mathbf{R}',\omega) E_j^{(0)}(\mathbf{R}',\omega) , \qquad (2.28)$$

with $\tilde{\chi}_{ij}(\mathbf{R}-\mathbf{R}',\omega)$ nonlocal tensor of linear response to external field. Taking into account that integration in Eq.(2.28) is over the particle volume and the external field in the most cases is the long-range and very weakly changes in the distance about linear dimension of the particle, one can remove the external field from integrand and obtain Eq.(2.28), where the effective susceptibility has a form

$$\mathrm{X}'_{ij}(\mathbf{R},\omega) = \int_V d\mathbf{R}' \tilde{\chi}_{ij}(\mathbf{R}-\mathbf{R}',\omega) \ . \qquad (2.29)$$

Then, the linear response to a local field is the characteristic of the material and linear response to an external field is the characteristic of the object made of this material. As a result, the knowledge of induced currents inside the system under consideration allows us to find the local field at any point in the system, which means the solution of the electrodynamic problem. Then one can to insist that effective susceptibility characterizes the most electrodynamical properties of the nano-system.

2.2.2. Low-Dimension Systems. Electrodynamical point of view

What is the reason of the difference in local field inside the bulk and inside the small particle? Obviously the differences are connected with contribution of interfaces. In the bulk materials the interfaces and transitional layers at interfaces fill only very small part of the volume of the object. On the contrary, the transitional layer fills the significant part of the volume of small particle (Fig.2.6). Then, one can think that when the object is acted by long-range external field, the local field inside a small particle strongly changes from point to point, but the local field inside a bulk strongly changes only in small region near the interface. Then, the thickness of transition layer is the parameter of linear dimension with the help of which one should compare the linear dimension of the system under consideration. Then, when the characteristic linear dimension of the object at any direction is about of length of transition layer, the object can be considered as low-dimensional. From this point of view the thin film with thickness *h* about of transition layer length *l* should be supposed as a quasi-two-dimension object. Its electrodynamical characterization should be considered with taking into account the local field inhomogeneity. For example, the electrodynamics of such ultra-thin films could be developed in the frame of effective susceptibility concept. What we should mean when say – transition layer? Of course, transition layers are caused, firstly, by changing lattice constant near the surface. Second, transition layer is caused by forth surface states and dangling bonds. The next reason

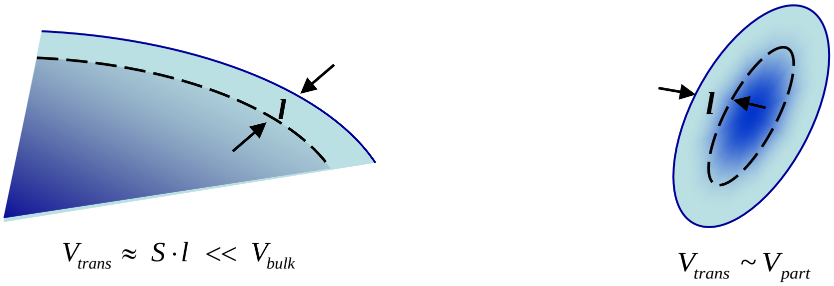


Fig.2.6 Share of transition layer in the bulk (left) and small particle (right)

of transition layer existence is surface impurities and defects, causing surface charges. From other hand, transition layer, or more precisely, the region near perfect interface may occur only due to existence of field gradients caused by interface. One can point out some additional reasons of transition layers existence. Anyway, the interfaces are the reason of the transition layers existence. Then, one can point out, at least, two kinds of electrodynamically low-dimension systems. There are thin films (analogue of two-dimension objects), and small (nano-) particles which are the analogue of zero-dimensional objects. All of these systems are under rapt attention in this work.

Because the knowledge of effective susceptibility means the knowledge of solution of self-consistent problem of local field characterization, calculating the effective susceptibility can be considered as the solution of main problem of the electrodynamics of nano-systems. To calculate the effective susceptibility of the systems under consideration one needs to calculate the self-consistent local field in the system. Then one should to solve the equation which the local field is obeyed.

2.2.3. Self-consistent equation

Let the system consisting of the object embedded in any medium is acted by external field $E_i^{(0)}(\mathbf{R},\omega)$. It is well known [20-23], the local field at any point in the system under consideration can be found with equation

$$E_i(\mathbf{R},\omega) = E_i^{(0)}(\mathbf{R},\omega) - i\omega\mu_0 \int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega) J_l(\mathbf{R}',\omega), \qquad (2.30)$$

where $G_{ij}(\mathbf{R},\mathbf{R}',\omega)$ is a photon propagator (Green function of Maxwell equations) [30-32] describing the light propagation in the medium in which the object of volume V is situated. Because the local current and local field are connected by constitutive equation [Eq.(2.26)], one can rewrite Eq.(2.30) in the form

$$E_i(\mathbf{R},\omega) = E_i^{(0)}(\mathbf{R},\omega) - (\omega/c)^2 \int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega) \chi_{jl}(\omega) E_l(\mathbf{R}',\omega), \qquad (2.31)$$

with dimensionless electrical susceptibility (linear response to local field) $\chi_{jl}(\omega) = (1/\varepsilon_0)\chi'_{jl}(\omega)$. The equation for self-consistent field is usually named as Lippmann-Schwinger equation [20-23, 33]. One should note, that the electrodynamical properties of the material of which the object is made, are described by linear response to local field $\chi_{ij}(\omega)$ and should be found either from microscopical calculations in the frame of quantum theory (see, for example, the problem N159 in textbook [34], and refs. [37-39]) or from experimental measurements. Introducing the effective susceptibility $\mathrm{X}_{ij}(\mathbf{R},\omega) = (1/\varepsilon_0)\mathrm{X}'_{ij}(\mathbf{R},\omega)$, which connects the current inside the object and external long-range field (see, Eq.(2.28)), one can write the solution of Eq.(2.31) in the form

$$E_i(\mathbf{R},\omega) = E_i^{(0)}(\mathbf{R},\omega) - (\omega/c)^2 \int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega) \mathrm{X}_{jk}(\mathbf{R}',\omega) E_k^{(0)}(\mathbf{R}',\omega). \qquad (2.32)$$

From this point of view the effective susceptibility describes the electrodynamical properties of the system. Then, knowledge of effective susceptibility of the system allows us to calculate the local self-consistent field at any point of the system under consideration.

2.2.4. Effective susceptibility

The effective susceptibility concept can be useful for developing of electrodynamics of low dimensional systems, because allows to calculate the scattered (reradiated) field by the system via equation

$$E_i^{(scatt)}(\mathbf{R},\omega) = -k_0^2 \int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega) \mathrm{X}_{jl}(\mathbf{R}',\omega) E_l^{(0)}(\mathbf{R}',\omega)\,, \tag{2.33}$$

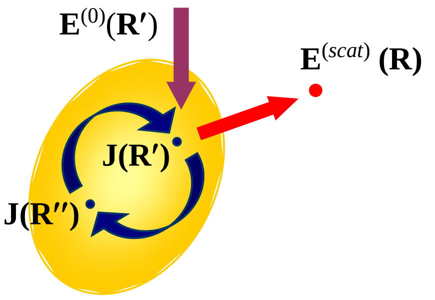


Fig.2.8. Sketch demonstrates the scattered by small particle field formation

where $k_0 = \omega / c$. Integration in Eq.(2.33) is over volume in which the currents induced by external field are located. The formation of scattered field is schematically demonstrated in Fig.2.8. Namely, the external field acting on the particle generates the self-consistent current inside the particle. The current, for one's turn, is the origin which induces the scattered field. In general case this volume can be multiply connected. Calculation of effective susceptibility in this case is rather difficult problem and can be provided by direct solving the electrodynamical problem or approximately in the frame of so-called “pseudo-vacuum” Green function method [20, 37]. When the wavelength of external field is more longer then the characteristic dimension of the system, Eq.(2.33) can be written in the form

$$E_i^{(scatt)}(\mathbf{R},\omega) = \Re_{il}(\mathbf{R},\omega) E_l^{(0)}(\mathbf{R},\omega)\,, \tag{2.34}$$

with

$$\Re_{il}(\mathbf{R},\omega) = -k_0^2 \int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega) \mathrm{X}_{jl}(\mathbf{R}',\omega) \tag{2.35}$$

generalized reflecting coefficient.

When one studies the light reflection by thin film, one can make Fourier transformation in the film plane, or transmit to so-called **k**-z representation. As a result, the reflected field at cross-section z can be calculated as

$$E_i^{(scatt)}(\mathbf{k},z,\omega)=\Re_{il}^{(film)}(\mathbf{k},z,\omega)E_l^{(0)}(\mathbf{k},z,\omega)\,, \tag{2.26}$$

$$\Re_{il}(\mathbf{k},z,\omega)=-k_0^2\int_0^h dz' G_{ij}(\mathbf{k},z,z',\omega)\mathrm{X}_{jl}(\mathbf{k},z',\omega)\,. \tag{2.37}$$

2.2.5. Effective susceptibility of small particle[2]

Let us consider the interaction of small particle with external electromagnetic field radiating the particle. The self-consistent electromagnetic field at any point inside the system “medium-particle” obeys to Eq.(2.31). To find the effective susceptibility one should to use Eqs.(2.26) and (2.28) and write connection between local and external fields, supposing that reciprocal matrix $\left(\mathrm{X}_{ij}(\mathbf{R},\omega)\right)^{-1}$ exists.

$$E_j^{(0)}(\mathbf{R},\omega)=\left(\mathrm{X}_{ij}(\mathbf{R},\omega)\right)^{-1}\chi_{ik}(\omega)E_k(\mathbf{R},\omega)\,. \tag{2.38}$$

Then, one can substitute this connection in Eq.(2.31) and obtain

$$\begin{aligned}E_j(\mathbf{R},\omega)=\left(\mathrm{X}_{kj}(\mathbf{R},\omega)\right)^{-1}\chi_{ki}(\omega)E_i\,(\mathbf{R},\omega)\\ -k_0^2\int_V d\mathbf{R}' G_{ij}(\mathbf{R},\mathbf{R}',\omega)\chi_{jk}(\omega)E_k\,(\mathbf{R}',\omega)\quad.\end{aligned} \tag{2.39}$$

This equation is true for all points in the system, including all points inside the particle. Then, one can integrate both parts of equation over the particle volume

---

[2] Chapters 2.2.5-2.2.8 the students which do not interest the mathematical details could omit

$$\int_V d\mathbf{R}E_j(\mathbf{R},\omega) = \int_V d\mathbf{R}\left(\mathrm{X}_{kj}(\mathbf{R},\omega)\right)^{-1}\chi_{ki}(\omega)E_i\,(\mathbf{R},\omega)$$
$$- k_0^2\int_V d\mathbf{R}'\int_V d\mathbf{R}G_{ij}(\mathbf{R}',\mathbf{R},\omega)\chi_{jk}(\omega)E_k(\mathbf{R},\omega)\ , \qquad (2.40)$$

where replacing the integration variables in last term of right part was performed. Supposing that local filed can be represented in the form

$$E_i\,(\mathbf{R},\omega) = \sum_{\mathbf{k}} E_i\,(\mathbf{k},\omega)e^{i\mathbf{kR}}\ , \qquad (2.41)$$

one obtains from Eq.(2.40)

$$\sum_{\mathbf{k}}\ \int_V d\mathbf{R}\left\{-\left(\mathrm{X}_{kj}(\mathbf{R},\omega)\right)^{-1}\chi_{ki}(\omega) + \right.$$
$$\left. +\omega^2\int_V d\mathbf{R}'G_{ik}(\mathbf{R}',\mathbf{R},\omega)\chi_{kj}(\omega) + \delta_{ij}\right\}E_j\,(\mathbf{k},\omega)e^{i\mathbf{kR}} = 0\ . \qquad (2.42)$$

Because exponents are the complete set of orthonormal functions, to satisfy this equation one needs to put

$$-\left(\mathrm{X}_{kj}(\mathbf{R},\omega)\right)^{-1}\chi_{ki}(\omega) + k_0^2\int_V d\mathbf{R}'G_{ik}(\mathbf{R}',\mathbf{R},\omega)\chi_{kj}(\omega) + \delta_{ij} = 0\,. \qquad (2.43)$$

Solution of this equation is

$$\mathrm{X}_{ij}(\mathbf{R},\omega) = \chi_{il}(\omega)\left[\delta_{jl} + k_0^2\int_V d\mathbf{R}'G_{jk}(\mathbf{R}',\mathbf{R},\omega)\chi_{kl}(\omega)\right]^{-1}. \qquad (2.44)$$

One should to note that obtained expression for effective susceptibility [Eq.(2.44)] is true under condition of nonzero of imagine part of determinant of the matrix $\left[\delta_{jl}+k_0^2\int_V d\mathbf{R}' G_{jk}(\mathbf{R}',\mathbf{R},\omega)\chi_{kl}(\omega)\right]$, and equation $\mathrm{Re}\det\left[\delta_{jl}+k_0^2\int_V d\mathbf{R}' G_{jk}(\mathbf{R}',\mathbf{R},\omega)\chi_{kl}(\omega)\right]=0$ defines the eigenmodes in the system.

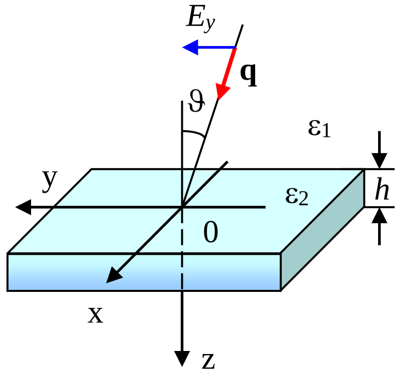


Fig.2.9 Sketch of thin film

2.2.6. Effective susceptibility of a film

Analogously to the case of nano-particle, one can calculate the effective susceptibility of a film. Discussed conception allows us to take into account the inhomogeneous along the thickness film, namely when dielectric constant is depended on z-coordinate (Fig.2.9). Then, one can write that linear response to a local field is depended on *z*

$$\chi_{ij}(z,\omega)=\varepsilon_{ij}(z,\omega)-\delta_{ij}\ . \qquad (2.45)$$

Because the film is homogeneous in its plane (XOY plane in Cartesian coordinate system), one can perform the partial Fourier transformation in the film plane, then one can transform to the **k**-z representation. In this representation Eq.(2.31) has a form

$$E_i(\mathbf{k},z,\omega)=E_i^{(0)}(\mathbf{k},z,\omega)-k_0^2\int_h dz' G_{ij}(\mathbf{k},z,z',\omega)\chi_{jl}(z',\omega)E_l(\mathbf{k},z',\omega)\ . \qquad (2.46)$$

To calculate the effective susceptibility of the thin film one needs to differ at least two different cases. First, the film can be very thin – the case of ultra-thin film. It means that one can not say that light propagates across the film thickness. From this point of view, the film and light acting the film should be considered as a single whole. Namely one should consider the single excited system consisting of the film and electromagnetic field localized at the film. Then, when one would like to calculate the effective susceptibility of the ultra-thin film, one should consider the film as secondary source of the "scattered" field, and make the calculation similar to demonstrated in Chapt.2.2.5 of this textbook. As a result, one can obtain for effective susceptibility of the ultra-thin film

$$X^{(uf)}_{ij}(\mathbf{k},z,\omega)=\chi_{il}(z,\omega)\left[\delta_{jl}+k_0^2\int_0^h dz'G_{jk}(k,z',z,\omega)\chi_{kl}(z,\omega)\right]^{-1}\ ,\quad 0\le z\le h, \qquad (2.47)$$

where the integration is over the film thickness.

Definitely other situation arises when the thickness of the film is rather large, when the phase shift can plays the role. In this case, one should take into account a possibility of light propagation across film thickness. Other words, one should to solve Eq.(2.46) explicitly with taken into account processes of light propagation along the thickness of the film. The correct procedure of solution this problem was proposed in Ref.[42] when homogeneous isotropic material of film ($\varepsilon_{ij}(\omega)=\varepsilon_2(\omega)\delta_{ij}$). Unfortunately we can not point out to the method of finding effective susceptibility of the inhomogeneous film where the phase shifts can play the role. From other hand it is clear that only when thickness of the film is more (about) than λ/2 (λ is the wavelength of probing light) the phase shifts can contribute to optical properties of the film. Then, one has dealings with the thick film when its thickness $h\ge\lambda/2$. It means that one can use standard approach to calculate the field inside the film, taking into account the border conditions and reflection and transmission coefficients [28, 29]. To obtain the effective susceptibility of the homogeneous film with taken into account the wave propagation along the thickness of the film, one should use two equations of self-consistent fields [42]

$$E_i(\mathbf{k},z,\omega)=E_i^{(0)}(\mathbf{k},z,\omega)-(\varepsilon_2(\omega)-\varepsilon_1)k_0^2\int_h dz'G_{ij}^{(1)}(\mathbf{k},z,z',\omega)E_j(\mathbf{k},z',\omega)\ , \qquad (2.48)$$

$$E_i(\mathbf{k},z,\omega)=(\varepsilon_1-\varepsilon_2(\omega))k_0^2\int dz'G_{ij}^{(2)}(\mathbf{k},z,z',\omega)E_j(\mathbf{k},z',\omega)\ , \qquad (2.49)$$

where $k_0=\omega/c$ and $\int dz'...=\int_{-\infty}^{0}dz'...+\int_h^{\infty}dz'...$, because the integral is over all space outside the film. Performing solution of Eqs.(2.48) and (2.49), one obtains the connection between the current and external field. In the case of s-polarized incident field $E_y^{(0)}(k,z,\omega)=E_y^{(0)}(k,\omega)e^{i\eta_1 z}$ this connection has a form

$$J_y(k,z,\omega)=\left(\varepsilon_2(\omega)-\varepsilon_1\right)\left\{A^+e^{i(\eta_2-\eta_1)z}+A^-e^{-i(\eta_2+\eta_1)z}\right\}E_y^0(k,\omega)e^{i\eta_1 z}\ , \qquad (2.50)$$

where $\eta_{1,2}=k_0\sqrt{\varepsilon_{1,2}-\sin^2\vartheta}$, $A^\pm$ are a functions of incident angle ϑ and frequency ω, depending on film thickness h and dielectric constants of environment ε_1 and film material ε_2. These functions can be easily found, but because of its very cumbersome form they are not shown here. As it easily can be seen the yy-component of effective susceptibility of the film has a form

$$\mathrm{X}_{yy}(k,z,\omega)=(i/\omega)\left(\varepsilon_2(\omega)-\varepsilon_1\right)\left\{A^+e^{i(\eta_2-\eta_1)z}+A^-e^{-i(\eta_2+\eta_1)z}\right\}. \qquad (2.51)$$

Analogously one can find the other components of effective susceptibility.

2.2.7. Effective susceptibility of sub-monolayer of the particles covering a surface

The electrodynamic properties of sub-monolayer covers of surface of a solids should be studied when the different practice problem are considered. It, for example, can be the problems of SPR-sensors of bio-molecules [43,44] , other problem is connected with problem of plasmons propagations along the surface covered by metal nano-particles [45-47]. The all these problems can be reduced to consideration the surface covering by the particles having any determined form, say, ellipsoidal-like (Fig.2.10). The equation of self-consistent field has the next form

$$E_i(\vec{R},\omega)=E_i^{(0)}(\vec{R},\omega)-k_0^2\sum_{\alpha=1}^{N}\int_{V_\alpha}d\vec{R}'G_{ij}(\vec{R},\vec{R}',\omega)\chi_{jl}(\omega)E_l(\vec{R}',\omega)\,, \qquad (2.52)$$

where summation is over position each of the particles. Because the system is macroscopically homogeneous in the plane of a substrate, and linear dimension of the particles is much less than wavelength of testing light, and averaged distance between neighbouring particles are more than particle dimension, one can consider the system as quasi-point-like dipoles at the surface each of them polarizes as, for example, ellipsoidal particle. Then, one can write

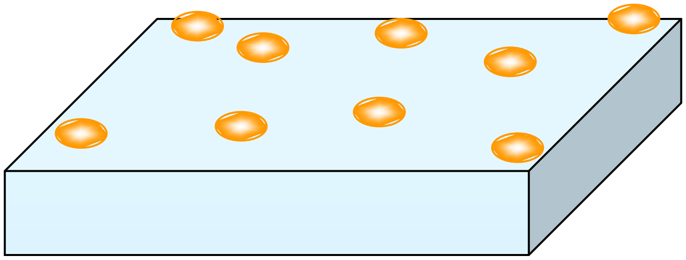


Fig.2.10. The surface covered by ellipsoidal particles

$$\sum_{\alpha}\int_{V_\alpha} d\vec{R}' G_{ij}(\vec{R},\vec{R}',\omega)\chi_{jl}(\omega)E_l(\vec{R}') \approx \sum_{\alpha} G_{ij}(\vec{r}-\vec{r}_\alpha, \quad , \qquad (2.53)$$

with $\tilde{\chi}_{jl}(\omega) = V_M \chi_{jl}(\omega)$, and $\chi_{jl}(\omega)$ the tensor of linear response of the ellipsoidal particle. Averaging over positions of the particles performed according to

$$\overline{\sum_{\alpha=1}^{N} G_{ij}(\mathbf{r}-\mathbf{r}_\alpha, z, z_\alpha, \omega)\tilde{\chi}_{jl}(\omega)E_l(\mathbf{r}_\alpha, z_\alpha, \omega)} =$$

$$\frac{1}{S^{N-1}}\int d\mathbf{r}_1 d\mathbf{r}_2 ... d\mathbf{r}_N \sum_{\alpha=1}^{N}\int \frac{d\mathbf{k}}{(2\pi)^2} e^{-i\mathbf{k}(\mathbf{r}-\mathbf{r}_\alpha)} G_{ij}(\mathbf{k}, z, z_\alpha, \omega) \times \qquad (2.54)$$

$$\times \tilde{\chi}_{jl}(\omega)\int \frac{\mathrm{d}\mathbf{k}'}{(2\pi)^2} e^{-i\mathbf{k}'\mathbf{r}_\alpha} E_l(\mathbf{k}', z_\alpha, \omega)$$

gives

$$\overline{\sum_{\alpha=1}^{N} G_{ij}(\mathbf{r}-\mathbf{r}_\alpha, z, z_\alpha, \omega)\tilde{\chi}_{jl}(\omega)E_l(\mathbf{r}_\alpha, z_\alpha, \omega)} =$$

$$= n\int \frac{d\mathbf{k}}{(2\pi)^2} e^{-i\mathbf{k}\mathbf{r}} G_{ij}(\mathbf{k}, z, z_\alpha, \omega)\tilde{\chi}_{jl}(\omega)E_l(\mathbf{k}, z_\alpha, \omega) \ , \qquad (2.55)$$

where n is concentration of the particles at the surface of a substrate, **r** is the coordinate in the layer plane, $\mathbf{r}_\alpha$ is the position of α-th particle in the layer plane. Then, the equation of self-consistent field written in the **k**, z - representation has a form

$$E_i(\mathbf{k}, z_\alpha, \omega) = E_i^{(0)}(\mathbf{k}, z_\alpha, \omega) - nk_0^2 G_{ij}(\mathbf{k}, z, z_\alpha, \omega)\tilde{\chi}_{jl}(\omega) E_l(\mathbf{k}, z_\alpha, \omega). \tag{2.56}$$

Taken into account that $\tilde{\chi}_{jl}(\omega)$ is a response to a local field, one can write

$$E_l(\mathbf{k}, z_\alpha, \omega) = \left(\tilde{\chi}_{jl}(\omega)\right)^{-1} P_j(\mathbf{k}, z_\alpha, \omega), \tag{2.57}$$

where $P_j(\mathbf{k}, z_\alpha, \omega)$ is the polarization of the sub-monolayer film in **k**-z representation. Substitution of Eq.(2.57) into Eq.(2.56) gives us the equation connecting the layer polarization and an external field

$$\left(\tilde{\chi}_{ij}(\omega)\right)^{-1} P_j(\mathbf{k}, z_\alpha, \omega) = E_i^{(0)}(\mathbf{k}, z_\alpha, \omega) - nk_0^2 G_{ij}(\mathbf{k}, z, z_\alpha, \omega) P_j(\mathbf{k}, z_\alpha, \omega). \tag{2.58}$$

It should be noted that one supposes in Eqs.(2.56)-(2.58) that all particles are situated at the same distance z_α from the surface. The solution of Eq.(68) has a form

$$P_j(\mathbf{k}, z_a, \omega) = \mathrm{X}_{ji}(\mathbf{k}, z_\alpha, \omega) E_i^{(0)}(\mathbf{k}, z_\alpha, \omega) \quad , \tag{2.59}$$

where the effective susceptibility

$$\mathrm{X}_{ji}(\mathbf{k}, z_a, \omega) = \left[\left(\tilde{\chi}_{ji}(\omega)\right)^{-1} + nk_0^2 G_{ij}(\mathbf{k}, z_\alpha, z_\alpha, \omega)\right]^{-1}. \tag{2.60}$$

One should to note that obtained expression for effective susceptibility of the submonolayer cover is remarkable because of as concentration as shape (and dimension) of the particles define the susceptibility. This, for example allows develop the methods of modeling of submonolayer covers optical properties which can be useful for determination of concentration, shape, and dimensions of the particles by simple optical measurements [44].

The optical response of the quasi-pointness dipole two-dimensional lattice was considered in [48]. The effective susceptibility of the dipole layer was calculated too.

2.2.8. Calculation of self-action field inside of nano-object

The self-action field (local-field reaction) is the important characteristic of the nano-particle. It defines, in part, the Lamb shifts of the energy levels of quantum dot, and, generally, the linear response to the external field. Let us suppose that one needs to calculate the field, which is caused by the currents induced in the particle, acted at any point inside the particle. Then one needs to calculate the field scattered by the particle inside the particle. Other words, one should to calculate the field which in the frame of effective susceptibility concept has a form

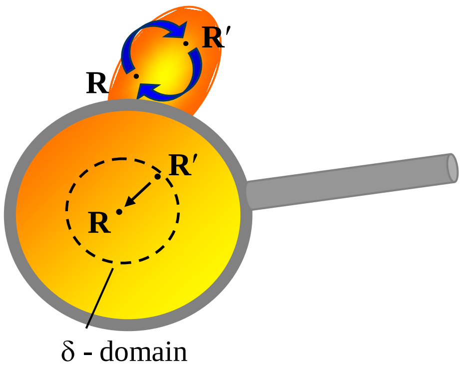


Fig. 2.11. Sketch illustrating the creation of δ-domain to calculate the self-action field

$$E_i^{(sa)}(\mathbf{R},\omega) = -k_0^2 \int_V d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega) \mathrm{X}_{jl}(\mathbf{R}',\omega) E_l^{(0)}(\mathbf{R}',\omega), \quad \mathbf{R} \in V \ . \qquad (2.61)$$

Because Green function has an singularity at $\mathbf{R} = \mathbf{R}'$, the problem of calculation of self-action field, when $\mathbf{R} \in V$ exists. The method of calculation of integrals similar to

$$\int_V d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega) \chi_{jl}(\omega) E_l(\mathbf{R}',\omega), \qquad (2.62)$$

when $\mathbf{R} \in V$ was discussed in Refs. [49, 50]. The method proposed in these refs. named as inclusion volume method is widely used, especially when numerical calculations in near-field [32, 51,52]. Calculation of self-action field in the frame of effective susceptibility concept has some specific features, which we demonstrate here. To calculate the self-action field let one choose the small region at point **R** (see, Fig. 2.11), and rewrite right part of (2.61) in the form

$$E_i^{(sa)}(\mathbf{R},\omega) = -k_0^2 \int\limits_{V-\delta} d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega) \mathrm{X}_{jl}(\mathbf{R}',\omega) E_l^{(0)}(\mathbf{R}',\omega) \\ - k_0^2 \int\limits_{\delta} d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega) \mathrm{X}_{jl}(\mathbf{R}',\omega) E_l^{(0)}(\mathbf{R}',\omega) \tag{2.63}$$

In the near-field limit one can suppose that

$$\lim_{\mathbf{R}'\to\mathbf{R}} G_{ij}(\mathbf{R}-\mathbf{R}',\omega) = \frac{1}{k_0^2} \frac{g_{ij}}{(\mathbf{R}-\mathbf{R}')^3} \ . \tag{2.64}$$

Taking into account this circumstance, one can write for effective susceptibility

$$\lim_{\mathbf{R}'\to\mathbf{R}} \mathrm{X}_{jl}(\mathbf{R}') = \left[\left(\chi_{lj}(\omega) \right)^{-1} + \int\limits_{V} dR'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3} \right]^{-1} . \tag{2.65}$$

Because

$$\left(\chi_{lj}(\omega) \right)^{-1} + \int\limits_{V} d\mathbf{R}'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3} = \\ \left(\chi_{lj}(\omega) \right)^{-1} + \int\limits_{V-\delta} d\mathbf{R}'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3} + \int\limits_{\delta} d\mathbf{R}'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3} \ , \tag{2.66}$$

One can see, that two first terms in the right part of (2.66) is not singular but the third term is singular. Then, one can evaluate

$$\left(\chi_{lj}(\omega)\right)^{-1} + \int_V d\mathbf{R}'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3} \approx \int_\delta d\mathbf{R}'' \frac{g_{jl}}{(\mathbf{R}''-\mathbf{R})^3}. \tag{2.67}$$

From this equation one immediately obtains

$$\lim_{\mathbf{R}'\to\mathbf{R}} \mathrm{X}_{jl}(\mathbf{R}') = \left[\int_\delta dR'' \frac{1}{(\mathbf{R}''-\mathbf{R})^3}\right]^{-1} g_{lj}{}^{-1}. \tag{2.68}$$

As a result last term in (2.63) can be rewritten in the form

$$\begin{aligned} &\int_\delta d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega)\mathrm{X}_{jl}(\mathbf{R}',\omega)E_l^{(0)}(\mathbf{R}',\omega) \\ &\approx \int_\delta d\mathbf{R}' \frac{g_{ij}g_{jl}{}^{-1}}{(\mathbf{R}-\mathbf{R}')^3}\left[\int_\delta d\mathbf{R}'' \frac{1}{(\mathbf{R}''-\mathbf{R})^3}\right]^{-1} E_l^{(0)}(\mathbf{R}) = E_i^{(0)}(\mathbf{R}) \ . \end{aligned} \tag{2.69}$$

It means that self-action field in the frame of effective susceptibility concept will be reduced to calculation of nonsingular terms

$$E_i^{(sa)}(\mathbf{R},\omega) = -k_0^2 \int_{V-\delta} d\mathbf{R}' G_{ij}(\mathbf{R}-\mathbf{R}',\omega)\mathrm{X}_{jl}(\mathbf{R}',\omega)E_l^{(0)}(\mathbf{R}',\omega) - E_i^{(0)}(\mathbf{R},\omega) \ , \quad \mathbf{R}\in V \tag{2.70}$$

Because of effective susceptibility can be calculated in many instances analytically, Eq.(2.70) can be useful for numerical calculations.

2.2.9. Effects of low dimension in nanooptics

To demonstrate the discussed concept we will calculate the absorption profiles when inhomogeneous thin nano-composite films with different distributions of inclusions along the film absorbs the incident light beam, dispersion of electromagnetic waves localized at the submonolayer film of metallic particles situated on the surface. First, we consider the *light absorption by thin nano-*

composite film with inhomogeneous distribution of inclusions. To calculate absorption profiles one should to find the energy of external field which is absorbed by the film per unit of a time. The energy of monochromatic radiation of wave vector **k** absorbed by the film can be presented as Joule heat and calculated according to

$$Q(\mathbf{k},\omega)=\frac{1}{4}\left\langle\overline{\left(J_i(\mathbf{k},z,\omega)+J_i^*(\mathbf{k},z,\omega)\right)\left(E_i(\mathbf{k},z,\omega)+E_i^*(\mathbf{k},z,\omega)\right)}\right\rangle , \tag{2.71}$$

with local current $J_i(\mathbf{k},z,\omega)$ and local field $E_i(\mathbf{k},z,\omega)$, which are induced inside the sample by external radiation of the field $E_i^{(0)}(\mathbf{k},z,\omega)$. The next designation

$$\overline{(...)}=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}dt(...) \tag{2.72}$$

which means the averaging over long-time period, is used in Eq.(2.71) , and

$$\langle...\rangle=\frac{1}{h}\int_{-h/2}^{h/2}dz(...) \tag{2.73}$$

means the averaging over film thickness.

Then, making these averagings and taking into account that Eq.(2.47) can be rewritten in the form

$$X_{ij}^{(uf)}(\mathbf{k},z,\omega)=\chi_{il}(z,\omega)\Omega_{lj}(\mathbf{k},z,\omega) \ , \tag{2.74}$$

with

$$\Omega_{lj}(\mathbf{k},z,\omega)=\left[\delta_{jl}+k_0^2\int_0^h dz'G_{jk}(\mathbf{k},z',z,\omega)\chi_{kl}(z,\omega)\right]^{-1} \tag{2.75}$$

local-field correction factor, one obtains

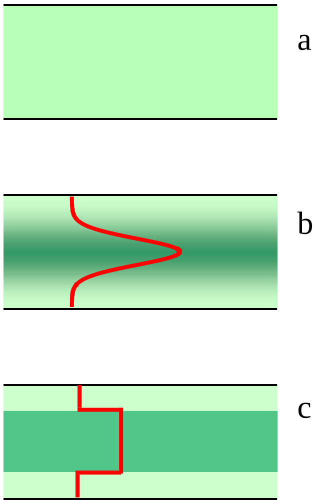


Fig.2.12. Sketch of the films with different distribution of inclusions

$$Q(\omega)=-\frac{1}{4}i\omega\Big\langle\chi_{ip}(z,\omega)\Omega_{pl}(\mathbf{k},z,\omega)\left(\Omega_{il'}(\mathbf{k},z,\omega)\right)^{*}$$

$$-\left(\chi_{ip}(z,\omega)\Omega_{pl'}(\mathbf{k},z,\omega)\right)^{*}\Omega_{il}(\mathbf{k},z,\omega)\Big\rangle E_{l}^{(0)}(\mathbf{k},z,\omega)\left(E_{l'}^{(0)}(\mathbf{k},z,\omega)\right.$$

(2.76)

One should take into consideration that due to in-plane isotropy of the film, the linear response to a local field tensor has the structure

$$\chi_{ij}(z,\omega)=\begin{pmatrix}\chi_{\parallel}(z,\omega) & 0 & 0\\ 0 & \chi_{\parallel}(z,\omega) & 0\\ 0 & 0 & \chi_{\perp}(z,\omega)\end{pmatrix}.$$

(2.77)

When absorption of normally incident light (XOZ is the incidence plane), the absorption profile can be written in a form

$$I_S(\omega)=\frac{1}{2h}\omega\int_0^h dz\,\mathrm{Im}\,\chi_{\parallel}(z,\omega)\left|\Omega_{yy}(k,z,\omega)\right|^2,\quad k=0,$$

(2.78)

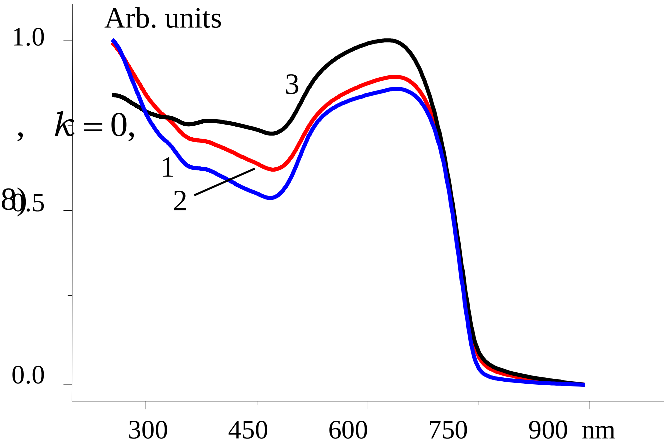


Fig.2.13. Absorption profiles of the Teflon films with homogeneous (1), Gaussian-like (2), and key pattern (3) distributions of gold inclusions.

where $I(\omega)=Q(\omega)/I_0(\omega)$, and $I_0(\omega)=\left|E_y^{(0)}(\omega)\right|^2$ is the intensity of incident beam. As it can be easily seen from Eq.(2.78), the absorption of external field energy is determined by imagine part of linear response to the local field tensor. The local-field effects, for one's turn, are described by local-field correction factor $\Omega_{yy}|k,z,\omega|^2$. Now we shall calculate the absorption profiles for the thin film with different distributions of inclusions along the film thickness. For example, we consider homogeneous (Fig.2.12a), Gaussian-like (Fig.2.12b), and key pattern (Fig.2.12c) distributions of inclusions inside the film. It was supposed that films of thickness 98 nm consists of Teflon matrix in which the gold spherical inclusions are embedded. The volume part of inclusion was supposed 0.1. The half-width of Gaussian-like distribution and width of the step of key pattern were supposed 0.4 h. The results of calculations of absorption profiles are presented in Fig.2.13. All curves are normalized on the particle number. We can see that different distributions of inclusions lead to different spectral properties of the nano-composite film. As a results, one can assert that effective susceptibility concept is adequate approach to describe the optical properties of nano-composite thin films with different distribution of the inclusions along the thickness. Because of proposed approach is based on the using of initial susceptibility which can be calculated in the frame of effective medium theory, one easily can calculate the absorption spectra of thin nano-composite films with shelling particles [57] or three-component nano-composite thin films [58].

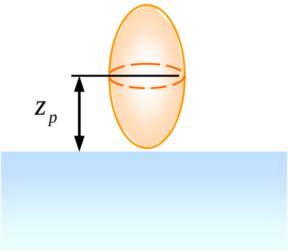


Fig.2.14. Elipsoid-like metal particle at the surface

Very important characteristic of optical properties of thin films is *the dispersion of electromagnetic waves* localized at the submonolayer film of metallic particles at the solid surface. The approach based on a concept of effective susceptibility can be easily

used for description of resonant properties of the nano-systems. Indeed, as it can be seen from Eq.(2.60) , the equation

$$\det\left[\left(\alpha_{ji}(\omega)\right)^{-1}+nk_0^2G_{ij}(\mathbf{k},z_\alpha,z_\alpha,\omega)\right]=0\ , \tag{2.79}$$

describes the condition of existence of the eigenmode in the system under consideration. Here $\alpha_{ji}(\omega)$ is the electric susceptibility of a single particle at the surface. This eigenmode is the electromagnetic wave localized at the submonolayer film situated at the surface. Let us suppose that the surface is covered by the particles of ellipsoidal shape. The particles are oriented of its main axis normally to the surface (Fig.2.14). In this case the initial susceptibility tensor (linear response of a single particle at the surface) has a form

$$\alpha_{ij} = \begin{pmatrix} \alpha_{\parallel} & 0 & 0 \\ 0 & \alpha_{\parallel} & 0 \\ 0 & 0 & \alpha_{\perp} \end{pmatrix} . \qquad (2.80)$$

The normal ($\alpha_{\perp}$) and lateral ($\alpha_{\parallel}$) components of the tensor have the next form

$$\alpha_{\parallel,\perp} = \varepsilon_r V_p \frac{(\varepsilon_p - \varepsilon_r)}{\varepsilon_r + (\varepsilon_p - \varepsilon_r) m_{\parallel,\perp}} L_{\parallel,\perp} \ , \qquad (2.81)$$

with

$$L_{\parallel,\perp} = \left[1 + \frac{(\varepsilon_r - \varepsilon_m)(\varepsilon_p - \varepsilon_r)}{3(\varepsilon_r + \varepsilon_m)(\varepsilon_r + (\varepsilon_p - \varepsilon_r) m_{\parallel,\perp})} U_{\parallel,\perp} \right]^{-1} , \quad U_{\parallel,\perp} = \vartheta, 2\vartheta \ , \quad \vartheta = h_{\perp} h_{\parallel}^2 / (2 z_p)^3 , \qquad (2.82)$$

and $m_{\parallel,\perp}$ is the depolarization factor, which describes the influence of the particle shape on its electrodynamical properties [54]. In the case of the particle shape as a prolate ellipsoid ($h_{\perp} > h_{\parallel}$) the components of depolarization factor have a form [54]

$$m_{\parallel} = \frac{1}{2}(1 - m_{\perp}) , \quad m_{\perp} = \frac{h_{\parallel}^2 / h_{\perp}^2}{(1 - h_{\parallel}^2 / h_{\perp}^2)^{3/2}} \left(\frac{1}{2} \ln \frac{1 + \sqrt{1 - h_{\parallel}^2 / h_{\perp}^2}}{1 - \sqrt{1 - h_{\parallel}^2 / h_{\perp}^2}} - \sqrt{1 - h_{\parallel}^2 / h_{\perp}^2} \right) , \qquad (2.83)$$

Using Cartesian coordinate system in which the XOY plane coincide with the surface of a substrate and wave vector of “surface” wave is directed along OX axes, Eq.(2.79) reduces to two simple dispersion relations. One of them describes the s-polarized wave

$$\left[\alpha_{\parallel}(\omega) \right]^{-1} + n k_0^2 G_{yy}(k, \omega) = 0 . \qquad (2.84)$$

Another equation describes a dispersion relations for p-polarized wave

$$\left\{\left[\alpha_{\parallel}(\omega)\right]^{-1}+nk_0^2G_{xx}(k,\omega)\right\}\left\{\left[\alpha_{\perp}(\omega)\right]^{-1}+nk_0^2G_{zz}(k,\omega)\right\} - n^2k_0^4G_{xz}(k,\omega)G_{zx}(k,\omega)=0 \quad . \tag{2.85}$$

This electromagnetic waves are an evanescent waves, localized at the covered by the particles surface. The cover of nano-particles strongly influences on the dispersion properties of surface plasmon polariton. In particular, the additional branches of p-polarized waves arise. There are four branches of dispersion curves when ellipsoid particles cover the surface [55]. At that the shift of the dispersion curve of surface plasmon polariton can be observed.   In Fig.2.15 the influence of the surface cover by the particles having shape as prolate ellipsoids ($h_{\perp}=2nm, h_{\parallel}=1nm$) on dispersion curve of surface plasmon polariton is shown. Concentration of the particles were chosen n =10^{12} part/cm^2 . Fig.2.15b demonstrates that small shift of the dispersion curve due to covering layer can lead to rather

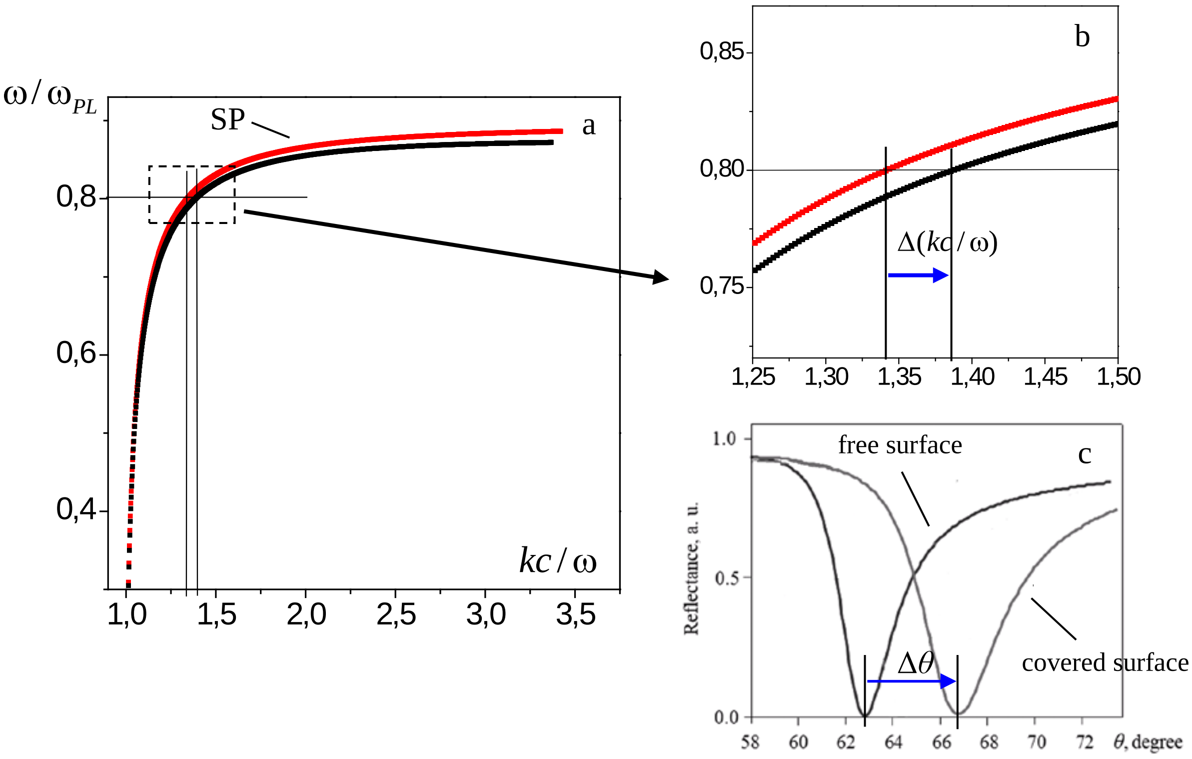


Fig.2.15. The influence of nano-particle cover on dispersion of surface plasmon (a). Shift of the dispersion curve caused the cover of prolate particle (b). The shift of the SPPR curve (c)

appreciable shift of the curves of surface plasmon polariton resonance (SPPR), which can be observed experimentally [44]. Because the shift of SPPR curve is strongly depended on the shape of the particle, there is a possibility to define the concentration and shape of the particles with measurement of dispersion curves of surface plasmon polariton and SPPR. In particularly, one can think that the experiments of this kind can give some information about albumen state at the surface. This idea can be used for sensorics. For example, if golden surface is modified with special kind of molecules which interact only with certain type molecules (say, red one), the surface plasmon polariton resonance will be observed at different angles with and without adsorbed to the surface molecules (Fig.2.16).

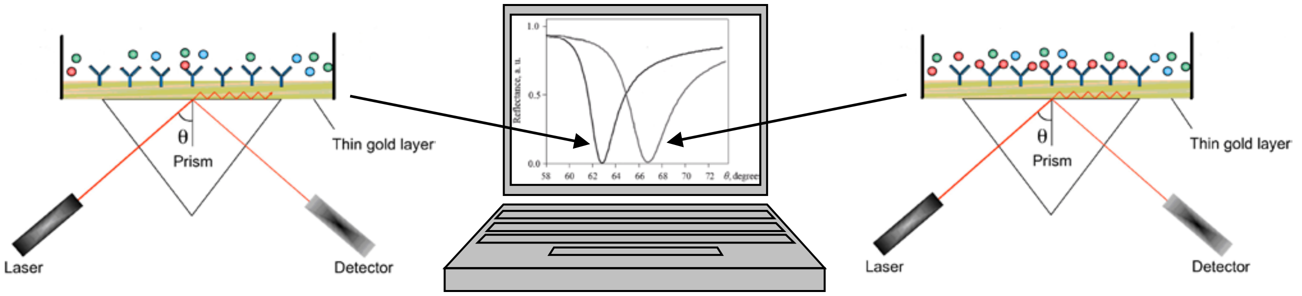


Fig.2.16. The SPPR sensor of certain type molecules. Solution with very low concentration of red molecules characterizes of SPPR at small resonance angle (about 63^0). When concentration of red molecules increases The resonant curve shifts to greater angles (about 67^0).

Chapter. 3 Quantum-size effects in semiconductor heterostructures

In the case when linear dimensions of the system become about of de Brogile wavelength of electron, the electron properties of the system can drastically change. Namely, the electron properties become strongly depend on boundary conditions, which demand the the continuity of the wave function and its derivatives at the interfacses. These demands are the result of wave nature of quantum particles and reflect the formation of standing wave at the domain of electron confinement (the region of electron localization). Then, the electron confinement arizing is depended on as the boundary conditions as linear dimensions of the system. It is clear that the systems in which the spatial quantization effects can be obsevrated should be, at first, the linear dimension about de Brogile wavelength, and, second, the quantum coherence conditions have to be satisfied. Do the space quantization effects observe in the real semiconductor systems? Or, could the real semiconducting systems with linear dimensions about of electron de Brogile wavelength be fabricated. Consider the Fig.3.1 in which the dependence of electron de Brogile wavelength on the effective mass is presented. We can see that the characteristic values of de Brogile wavelength lie in the range of 10 – 80 nm. Then, the effects of spatial quantization can be observed in the semiconductor systems linear dimensions of which are about 10 – 80 nm. Then, to observe the spatial quantization effects one needs to establish the

3.1. Electron fundamental lengths

As it is well known, electron inside the semiconductor characterizes by effective mass m*, which is usually smaller than free electron mass. Then electron de Brogile wavelength in semiconductor is more thnan the free electron de Brogile wavelength

$$\lambda_B = \frac{h}{p} = \frac{h}{\sqrt{2m^* E}} = \lambda_B^0 \sqrt{\frac{m}{m^*}} . \tag{3.1}$$

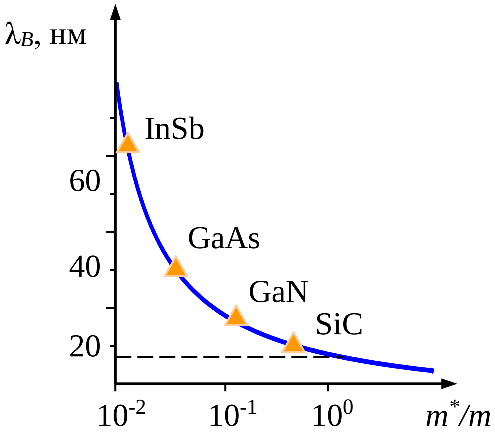


Fig. 3.1. The dependence of de Brogile wavelength of electrons in semiconductors at T=300 K on effective mass

The ratio of the electron effective masses for InSb m^*/m = =0,014, for GaAs m^*/m = 0,067, for GaN

m^*/m = 0,172 , and for SiC m^*/m = 0,41. Then we can calculate using Eq.(3.1) the absolute values of electron de Brogile wavelength for these semiconductors. The results are shown in Fig.3.1.

Consider at which dimensions of the systems one can hope the spatial quantization effects. Let the geometrical dimensions of semiconductor sample are $L_x \times L_y \times L_z$. Suppose for definitely that $L_x < L_y < L_z$. When the system under consideration is homogeneous and scattering processes are weak, then one can think that the electron moving is quasiballistic. Then only linear dimensions of the system one should to compare with electron de Brogile wavelength. Because integer half-wave de Brogile can be along the thickness of any confines system only discret numbers of electron states can be in the system. Then the electron moving in the restricted sample can be quantized. The conditions of quantizations can be defined with the next hierarchy of lengthes:

– three dimension (3D) electron moving can be in the bulk sample when

$$\lambda_B << L_x, L_y, L_z \quad . \tag{3.2}$$

In this case electron behaves like a free particle with effective mass m^*;

– two dimension (2D) system, or quantum well is the system in which quantization occurs along one of the axis (say Ox). Along the zOy plane the electron moves as free particle with effective mass m^*

$$\lambda_B \cong L_x << L_y, L_z \, ; \tag{3.3}$$

– one dimension (1D) system, or quantum needle (quantum wire) is the system in which the quantization occurs along two dimensions, as well electron moves as a free particles along quantum needle direction:

$$L_x \cong L_y \cong \lambda_B << L_z \, ; \tag{3.4}$$

– zero-dimension system, or quantum dot forms when spatial quantization occurs along all three axes. In this case:

$$\lambda_B \cong L_x \cong L_y \cong L_z \, . \tag{3.5}$$

Take into account now, that in the real systems electrons are under action of scattering processes. The scattering leads to stochastization of elecrom movies. Due to this reason, electrons leave off be quantum objects. Scattering processes could be devided on two kinds – ellestic and nonellastic.

When the electrons are scattered ellastically only pulse can be changed. Opposite to ellestic processes unelastic scattering result change as pulse as energy.

Consider the influence of different scattering processes on the electron quantization. Let the electron ellestic scattering by the lattice imperfection, say by impurity, (see, Fig.3.2). If initial electron state was characterized by wave function $\Psi_i(\mathbf{r},t)$ and wave vector $\mathbf{k}_i = \mathbf{p}_i / \hbar$. After scattering its wave vector becomes and will be equal to $\mathbf{k}_{sc} = \mathbf{p} / \hbar$. Due to ellestic scattering electron energy does not change. Namely, $E(p) = E(p_i) = \hbar\Omega$. Then, electron wave function which before scattering has a form

$$\Psi_i(\mathbf{r},t) = e^{-i\Omega t} e^{i\cdot\frac{\mathbf{p}_i\mathbf{r}}{\hbar}} \tag{3.6}$$

transforms to the next

$$\Psi_{sc}(\mathbf{r},t) = e^{-i\Omega t} \sum_{\mathbf{p},|\mathbf{p}|=p_i} A_{\mathbf{p}} e^{i\cdot\frac{\mathbf{p}\cdot\mathbf{r}}{\hbar}} = e^{-i\Omega t}\psi(\mathbf{r})\,. \tag{3.7}$$

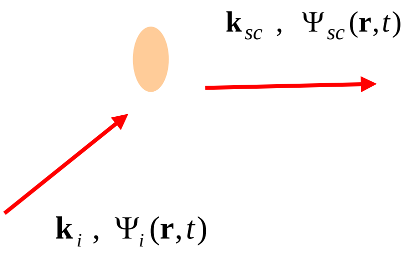


Fig. 3.2. The changing of wave vector and wave function at elastic electron scattering

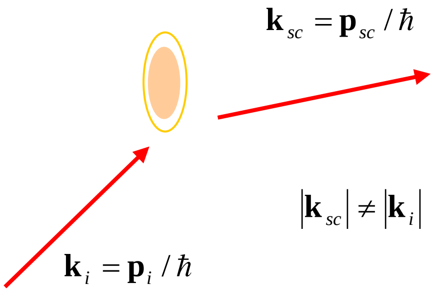


Fig. 3.3. The changing of wave vector and wave function at unelastic electron scattering

As it is well known, $|A_{\mathbf{p}}|^2$ defines the probability to find the particle with the pulse $\mathbf{p}$ after scattering. Incident and scattering waves form the complex wave picture. At elastic electron scattering the essential feature of the picture is the wave function phase conservation. As a result the square of modulus of wave function (3.7) is undependent on the time:

$$|\Psi_{sc}(\mathbf{r},t)|^2 = |\psi(\mathbf{r})|^2 . \tag{3.8}$$

Then, the ellestic scattering does not disturb of phase coherency of electron moving. If τ_e is an average time between two acts of elastic scattering, the average mean free path of electron is

$$l_e = v\tau_e , \tag{3.9}$$

with v average electron velocity. Then, there is an electron coherency at the distances larger than l_e . When the electrons are scattered unellastically (Fig.3.3), enstead Eq.(3.7) one obtains

$$\Psi_{sc}(\mathbf{r},t) = \sum_{\mathbf{p},|\mathbf{p}|=p_i} A_{\mathbf{p}} e^{i\cdot\frac{\mathbf{p}\cdot\mathbf{r}}{\hbar}} e^{-i\Omega(\mathbf{p})t} . \tag{3.10}$$

In this case the square modulus of wave function depends on as coordinate as time. Then, in this case the coherency is broken.

Let τ_E is an average time between of two inelastic collision and average electron free path at inelastic scattering is L_E . Then, electron conserves the quantum coherence at the distances less than L_E . The distance L_E usually is more large than the distance l_e. It means that electron scatters many time before its energy noticeably decreases (see, Fig.3.4). As aresult the electron moving becomes chaotic (the stochatization effect occurs) and its average displacement for the time τ_E is defined by diffusion coefficient D

$$L_E = \sqrt{D\tau_E} \quad , \qquad (\tau_\mathrm{E} >> \tau_e) . \tag{3.11}$$

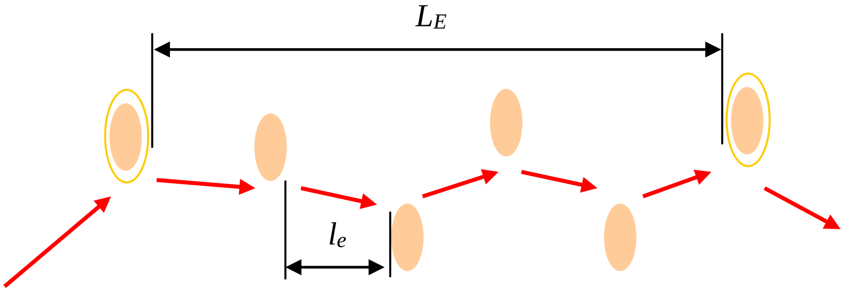


Fig. 3.4. The gain of stochastic electron moving due to elastic and inelastic scatterings.

Diffusion coefficient can be written as

$$D = \frac{v^2\tau_e}{\alpha} , \tag{3.12}$$

with parameter α = 3 for 3D electron gas, α = 2 for 2D and α =1 for 1D electron gas. When the temperature incteases, the number of electrons, which can take part in the electron transport, increases and the more numbers of electron with different energy can take part in the cynetic processes. The electron coherence in this case is broken. Indeed, if $k_B T << E_F$, each electronic state has similar amplitude but slightly different phases. Up to the difference is rather weak, the averaging over the electronic states inside the distribution function of electrons (which are slightly broadening by temperature) does not disturb its phase. As a result the coherence is conserved. When the temperature increases when electrons with different energies take part in the kinetic processes, the coherence of the system become broken due to the averaging over electron energy states occurs with rather wide broadening. Let L_T is the characteristic length at which the coherence of the system becomes broken due to nonzero temperature. The temperature broadening of distribution function is proportion to $k_B T$ and results to phase broadening which changes with time as $t \cdot (k_B T / \hbar)$. Then, to evaluate the broadening time one can use the equation

$$\tau_T = \frac{\hbar}{k_B T} . \qquad (3.13)$$

When scattering processes are elastic, electrons diffuse to the distances about $\sqrt{Dt}$ for the time t . The time are usually more than averaged time of free path τ_e . It means that during the time τ_T electron diffuces by temperature moving for the distance

$$L_T = \sqrt{D\tau_T} = \sqrt{D\hbar / k_B T} . \qquad (3.14)$$

Then the electron coherence is briken at the distances mure than L_T .
Note here we will concider the electron systems the electron states of which are coherent.

3.2. Heterostructures. Quantum wells, quantum needles, quantum dots

The heterostructure is formed when more that two different materials connected to one structure with interfaces between the materials. The conditions under which the heterostructures are foermed are considered in other chapters of the textbook.

The simplest amd most prevalent heterostructures are quantum wells. The modern technologies allow to fabricate the interfaces between the different materials of almost prefect. Then, transition layer firming the interface fabricated with molecular beam epytaxy technology can be monoatomic. It means that in the most cases to describe the heterostructure electronic properties one can suppose the perfect interface. The simplest example of heterostructure are two semiconductors with one interface. The interface often is named as heterocontact. Well known heterostructures is SiO_2-Si heterostructure with very small dencity of the defects at the interface. There are wide using the pairs of 3-5 semiconductors GaAs-AlGaAs; GaInAs-InP; GaInAs-AlInAs; GaSb-AlSb; GaN-AlN *etc*. There other heterostructures formed with 2-6 semiconductor compounds, namely CdZnSe-ZnSe; ZnSTeSe-ZnSSe etc.

Note, the main purpose of heterstructures fabrication is the controlled modification of energy bands in the structures. Other words, the heterostructures could be considered as an artifical materials with predetermined properties. The different kinds of heterostructures energy diagramms are shown in Fig.3.5. The two types of heterostructures with a single interface are shown in Figs.3.5 (a) and (b); the different types of heterostructures with two interfaces are shown in Figs.3.5 (c) and (d); The multiinterface heterostructures are shown in Figs.3.5 (i) and (f);

The idealized pictures of three types of heterostructures – quantum wells, quantum needles and quantum dots will be considered below.

3.3. Quantum wells

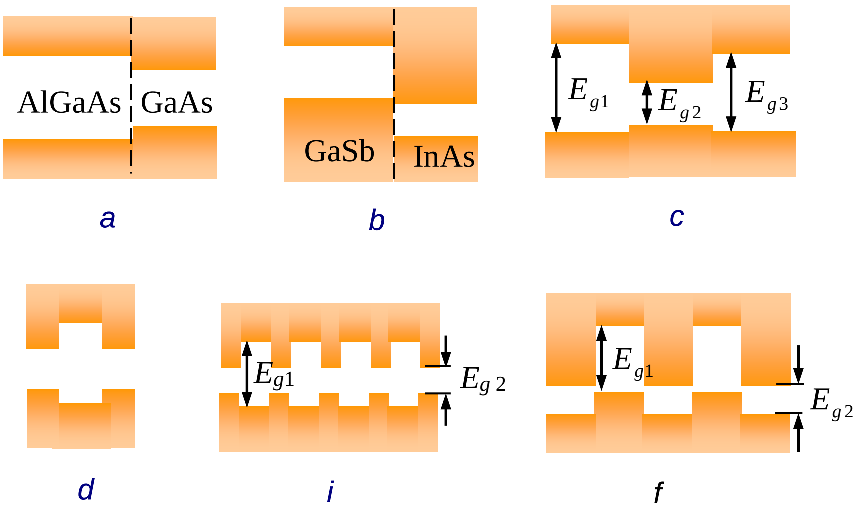


Fig. 3.5. The energy diagrams of different heterostructures

Quantum well can be fabricated when a single thin layer of the semiconductor material is grown between of two other (usually wide-gap semiconductors) materials (See, Fig.3.6). The energy gap dissolubility leads to electron confinement inside a well. Then the electron movement will occur in the restricting potential. One can without loss of generality consider the rectangular potential as the model for electron properties of quantum well (QW) studies. Then we consider the idealized potential shown in fig.3.6:

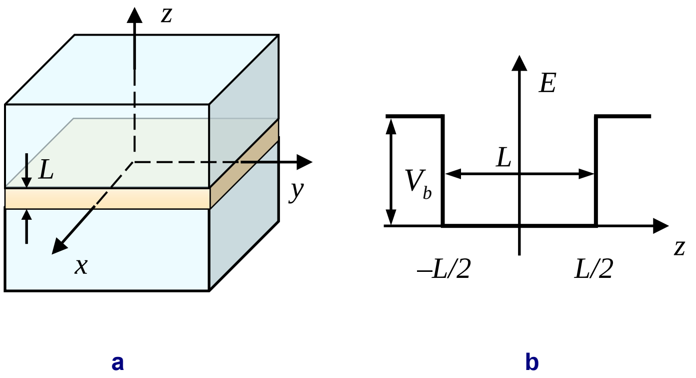


Fig. 3.6. Quantum well as thin semiconductor film between thick layers of wide-band materials (a) and idealized energy diarram of the structure (b)

$$V(z) = \begin{cases} 0\ , & \text{when} \quad |z| \le L/2\ , \\ \mathrm{V_b}\ , & \text{when} \quad |z| \ge L/2\ , \end{cases} \qquad (3.15)$$

with V_b and L the depth and wide the quantum well, respectively. Because of the electrom moving inside the QW plane is a free, one canwrite the wave function in the form

$$\Psi(x, y, z) = e^{ik_x x + ik_y y}\chi(z)\,. \qquad (3.16)$$

Then, Schrödinger equation can be written as equation for unknown function $\chi(x)$

$$\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + V(z)\right]\chi(z) = \left(E - \frac{\hbar^2 k_{\|}^2}{2m^*}\right)\chi(z)\,. \qquad (3.17)$$

Here $k_{\|} = (k_x, k_y)$, and $\left(\hbar^2 k_{\|}^2\right)/(2m^*)$ is a kinetic energy of electron moving with effective mass m^* along the quantum well plane (xOy). Denote

$$\in = E - \frac{\hbar^2 k_{\|}^2}{2m^*}\,. \qquad (3.18)$$

Because of E is a full energy of the particle, then $\in$ is an energy of the electron moving in the direction normal to xOy plane, or transverse moving. Then one should to solve the one-dimension problem

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2} + V(z)\right]\chi(z) = \ \in\chi(z)\,. \qquad (3.19)$$

3.3.1. *The bound states of electrons*

According to general properties of Schrödinger equation one can expect the two types of the equation solutions. Namely, at $\in < V_b$ electron moving will be restricted by energy well barriers. But at $\in > V_b$ the electron moving will be infinite which corresponds to unbounded electron states. The exsistance of nonzero solutions of Eq.(3.19) when $\in < V_b$ is possible only electron energy will be descrete.

The solution of Eq.(3.19) outside the well has a form

$$\chi(z) = \begin{cases} Ae^{-k_b(z-L/2)} & , \quad \text{when} \quad z \ \geq \ L/2, \\ Be^{k_b(z+L/2)} & , \quad \text{when} \quad z \leq -L/2, \end{cases} \tag{3.20}$$

where $k_b = \sqrt{-2m^*(\in - V_b)/\hbar^2}$. The solution inside the quantum well has a form

$$\chi(z) = C \cdot \cos k_w z + D \cdot \sin k_w z \quad , \text{when} \quad |z| \leq L/2\,, \tag{3.21}$$

where $k_w = \sqrt{-2m^* \in /\hbar^2}$, A, B, C, D are the constants which are determined from the standard boundary conditions for wave functions. As a result, one obtains the algebraic system for unknown values A and C for even solution

$$C \cos k_w L/2 = A, \tag{3.22}$$

$$C k_w \sin k_w L/2 = A k_b. \tag{3.23}$$

To exsistance the nonzero solutions this system its determinant should be equal to zero. This condition leads to

$$\mathrm{tg} k_w L/2 = k_b / k_w. \tag{3.24}$$

Analogously one obtains for odd solution

$$\mathrm{ctg} k_w L/2 = -k_b / k_w. \tag{3.25}$$

The analysis of Eqs.(3.24) and (3.25) gives us that in the case when

$$\frac{2\pi(l-1)}{L} < k_w < \frac{2\pi(l-1/2)}{L} \tag{3.26}$$

should be satisfied

$$\cos k_w L/2 = \pm k_w / k_0 , \tag{3.27}$$

and when

$$\frac{2\pi(l-1/2)}{L} < k_w < \frac{2\pi l}{L} , \tag{3.28}$$

the equation

$$\sin k_w L/2 = \pm k_w / k_0 \tag{3.29}$$

should be satisfied. Note, in these equation $k_0 = \sqrt{2m^* V_b / \hbar^2}$.

The left (L) and right parts of Eqs.(3.27) and (3.29) one can represent in the graphic form shown in Fig.3.7. First, one can see that there sre discrete values of energy because there are the discrete solution of Eqs.(3.27) and (3.29). Second, there is one value of energy level (corresponds to cross-section line 1 with cos(…) (Eq.3.27).

$$\in_1 = \frac{\hbar^2 k_{w,1}^2}{2m^*} . \tag{3.30}$$

The second energy level can be evaluated with the equation

$$\in_2 \approx V_b . \tag{3.31}$$

When parameter k_0 (say the depth of energy well) increases, the first and second energy levels become deeper and the third level arises. Each new energy level arises when parameter

$$q = \sqrt{\frac{2m^* V_b L^2}{\pi^2 \hbar^2}} \tag{3.32}$$

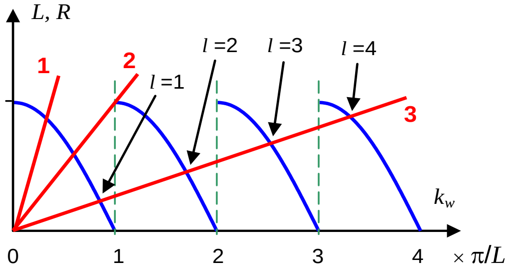


Fig. 3.7. Graphic solution of Eqs.(3.27) and (3.29). Line 2 corresponds to critical value of k_0, when the second energy level arises. Line 3 correspons to four energy levels

becomes integer value. Then the number of energy levels with the energy $\in < V_b$ can be calculated by the next equation

$$N = 1 + \text{int}\left[\sqrt{\frac{2m^* V_b L^2}{\pi^2 \hbar^2}}\right]. \tag{3.33}$$

Then, varying the parameters of quantum well, namely, shoosing components of the structure with specified wide of band gap and varying the geometric wide of quantum well (thickness of semiconductor thin film) one can obtain the system characterized preassigned structure of energy levels. Then one can obtain the new artificial material. The numbers of bounded electron states in the quantum wells as function of well wide for electrons (V_b = 224 meV, m^*=0,067m) and holes (V_b = 150 meV, m^*=0,4m) of heterosystem AlGaAs-GaAs are shown in Fig.3.8. One can see, that at the same wide of wuantum well the munbers of bpounded states is larger for larger effective mass of band electrons. This fact is very important from technological point of view. Indeed, one can see that, the weak variation of wide quantum well during fabrication of structures with one-dimension electron confinement does not lead to changing number of energy levels when effective mass is small. The energy levels of bounded states one can rigorously obtain in the two limits. Fist, when the quantum well is rather shallow, there is only one energy level

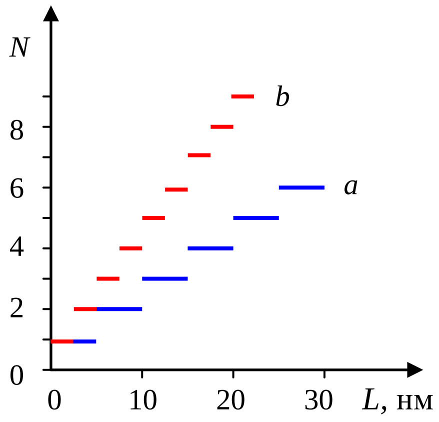


Fig. 3.8. The number of the bound states of a square well plotted as a function of well thickness: *a* – V_b = 224 meV, m^*=0,067m; *b* – V_b = 150 meV, m^*=0,4m

$$\in_1 \approx \frac{m^* L^2 V_b^2}{2\hbar^2}. \tag{3.34}$$

Second, when the quantum well is very deep

$$\in_n = \frac{\hbar^2 \pi^2 n^2}{2m^* L^2}, \qquad n = 1, 2, 3, \ldots \tag{3.35}$$

In this case the distance between the energy levels increases when quantum number n increases. For example, for quantum well that is the thin GaAS film of thickness L = 12,5 нм, where $m^* = 0{,}067m$ one obtain $\in_1 \approx 35$ меВ, $\in_2 \approx 140$ меВ, …

Then, in contrast to the electrons in the bulk semiconductor, the energy structure of the system with one-dimension electron confinemet (along OZ axis), the electron energy will be quantized and due to free motion along the quantum well plane the subbands will be formed. As a result the energy dispersion will be described by the equation

$$E_{n.k_{\parallel}} = \in_n + \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2)\,, \tag{3.36}$$

with $\in_n$ quantized energy associated with the transverse (perpendicular to the heterostructure) electron confinement. Then, the electron states for quantum dots are described by two quantum numbers – the discrete number of quantum subband n andcontinuous value of wave number $k_{\parallel}$, which describes the free motion of electron along the quantum well plane. The energy dispersion law has a view the set of paraboloids one inside other as it shown in Fig.3.9. Note, the minimal energy value corresponds to lowest energy level of bound state at $\mathbf{k}_{\parallel} = 0$. It means that at $n = 1$ the electron energy is high than energy of bulk electron in the center of Brillouin zone – one can obcerve so-clled blue shift.

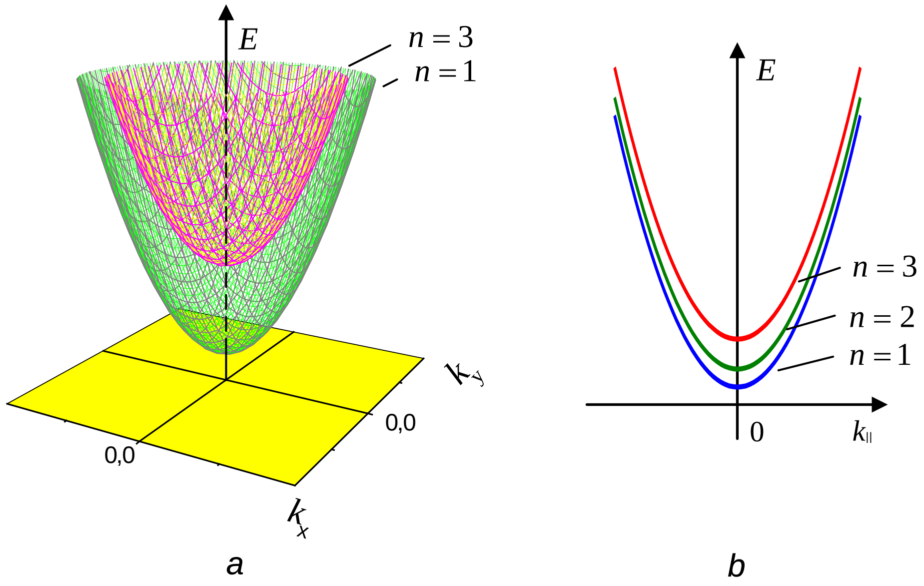


Fig. 3.9. Schematic of the energy spectrum of the two-dimensional electron in the first three subbands: *a* – 3D view of the paraboloids putted one into another; *b* – cross-sections of dispersion surfaces for fixed direction of quasipulse

Note that the quasi-twodimension electron systems can be fabricated by forming the heterointerface with modulation doping method. The idea behind modulation doping consists in the next: Two materials with almost identical lattice constants but different bandgaps are grown on top of each other to form a heterojunction. One example of a well-behaved heterojunction is Ga As/AlGaAs. The lattice

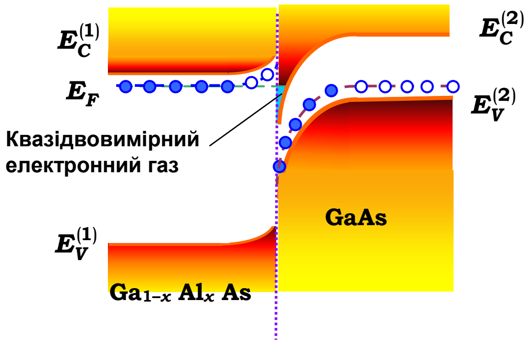


Fig. 3.10. Formation of channel of quasi-twodimension electrons at the interface of modula-tion doping heterostructure. The impurities of donor type are marked as blue circles. The acceptors are marked with white circles. Fremi level coinsites with the Fremi level of donor impurities in AlGaAs

constants of these two semiconductors differ by less than 1 %. The bandgap of AlGaAs with less than 40% of Al is direct and larger than that of GaAs. The difference between their bandgaps is divided in an approximately 60/40 split between the conduction and valence bands. The results are very abrupt discontinuities, known as band offsets, in their energy bands at the interface, as shown schematically in Fig. 3.10. If the material with the larger bandgap (AlGaAs) is then doped with shallow donors, the Fermi level is shifted from the middle of the bandgap of AlGaAs to the donor level. In order to maintain a constant chemical potential throughout the two materials, electrons will flow from AlGaAs to GaAs. This causes the band edges to bend at the interface as shown in Fig. 3.10, a phenomenon known as band bending. The band bending also occurs near the surface of semiconductors as a result of the existence of surface states. Due to band bending the electrons in GaAs are now confined by an approximately triangular potential near the interface and form a *two-dimensional (2D) electron gas.* These 2D electrons are physically separated from the ionized impurities in AlGaAs, hence they are only weakly scattered by the charged impurities. Then the electrons in so-called electron pocket will be characterized by high mobility. Then, these electrons can be considered as quasi-two-dimension systems with properties very close to the properties of the system (quantum well with quadrate potential) considered abpove. It is clear that one should to consider the electron moving in the triangular potential.

3.3.2. Quantum effects in a continuum-electron spectrum. Transversal transport

So far, in studying the energy states of the quantum wells, one has concentrated on the discrete spectra generated by the potential profiles of heterrostructures. Here one considers another important case of continuous-electron spectra.

Consider the quantum well with the rectangular potential profile (Fig.3.6) given by Eq.3.15. Consider the electron states in the quantum well, as well as the reflection of wavelike solution from the heterointerfaces. The quantum-well potential discontinuities affect not only the elrctron solution in the well but also those above the well, although to a lesser extent. Analyzing the Schrödinger equation (3.19) for energy $\in > V_b$, one can conclude that for regions far away from the quantum well, there are two plane-wave solutions propagating in opposite directions. The first wave corresponds to the propagation of an incident electron from left to right; then at $z \to -\infty$ there are incident and reflected plane waves, and at $z \to \infty$ there are only transmitted plane waves. The other situation is related to the opposite direction of the electron propagation. Both cases can occur at the same energy. Since both directions are identical, one needs to study obly one of them, namely the case in which an elecgtron propagates from left to right.

Then, introduce a wave vector in the region of the barrier as

$$k_b = \sqrt{\frac{2m}{\hbar^2}(\in - V_b)}\,. \tag{3.37}$$

Note that for the energies more that barrier high $(\in > V_b)$, wave vector is a real quantity. Then the wave function is

$$\chi(z) = \begin{cases} e^{ik_b(z+L/2)} + re^{-ik_b(z+L/2)}\ , & z \le -L/2; \\ ae^{ik_w z} + be^{-ik_w z}\ , & -L/2 \le z \le L/2; \\ te^{ik_b(z-L/2)}\ , & z \ge L/2. \end{cases} \tag{3.38}$$

The constants a, b, r, and t can be found by matching these solutions and their derivatives at the well, $z = -L/2$ and $z = L/2$. As a result, one find that a continuum-energy spectrum all constants a, b, r, and t can be expressed in terms of the coefficient of the incident wave, which we choose to equal 1. Then, one obtains the system of algebraic equations for the constants r, and t

$$\begin{cases} \dfrac{1}{t}+\dfrac{r}{t}\dfrac{k_w-k_b}{k_w+k_b}=\cos k_wL-i\sin k_wL\ , \\ \\ \dfrac{1}{t}+\dfrac{r}{t}\dfrac{k_w+k_b}{k_w-k_b}=\cos k_wL+i\sin k_wL\ . \end{cases} \tag{3.39}$$

These equations give us the connection between coefficients r and t:

$$r(\in)=\frac{i}{2}\left(\frac{k_w}{k_b}-\frac{k_b}{k_w}\right)t(\in)\sin k_wL\ . \tag{3.40}$$

The coefficient determinating transmitted electron wave is

$$t(\in)=\left[\cos k_wL-(i/2)(k_w/k_b+k_b/k_w)\sin k_wL\right]^{-1}. \tag{3.41}$$

Then, one can calculate the incident, reflected and transmitted electron currents according to wellknown equation

$$\mathbf{j}=-\frac{i\hbar}{2m}(\psi^*\nabla\psi-\psi\cdot\nabla\psi^*)\,. \tag{3.42}$$

Namely, current dencity of the incident electron flow is

$$j_{in}=\frac{\hbar k_b}{m}\ , \tag{3.43}$$

current dencity of the reflected electron flow is

$$j_r=\frac{\hbar}{m}k_b\,|\,r\,|^2\ , \tag{3.44}$$

and current dencity of the transmitted electron flow is

$$j_{tr} = \frac{\hbar}{m} k_b \, |t|^2 \quad . \tag{3.45}$$

The transmittion and reflectrion coefficients are

$$T(\in) = \frac{j_{tr}}{j_{in}} \quad , \qquad\qquad R(\in) = \frac{j_r}{j_{in}} , \tag{3.46}$$

or

$$T(\in) = |t|^2 = \left[1 + \frac{1}{4} \left(\frac{k_w}{k_b} - \frac{k_b}{k_w} \right)^2 \sin^2 k_w L \right]^{-1} , \tag{3.47}$$

$$R(\in) = |r|^2 = 1 - T(\in) . \tag{3.48}$$

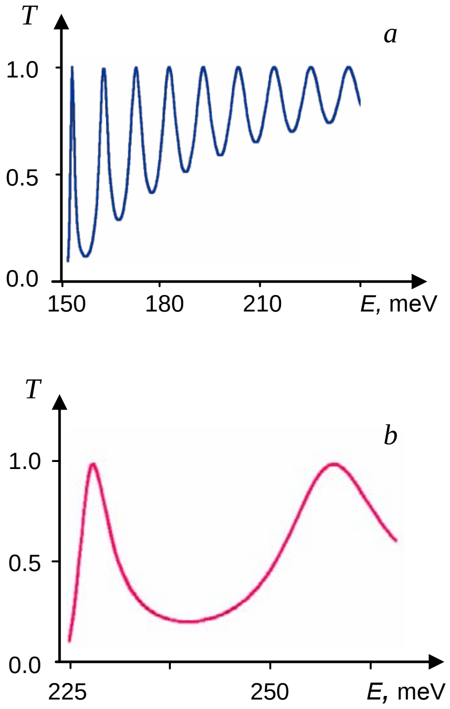


Fig.3.12. Energy dependence of the trans-mission coefficient in a square quantum well of thickness L = 25 nm in hetero-structure AlGaAs/GaAs/AlGaAs:

panel *a* – holes, V_b=150 меВ, *m**=0,48*m*;
panel *b* – electrons, V_b=224 меВ, *m**=0,067*m*

Note Eq.(3.48) is evident because the total probability of transmission and reflectrion is equal to 1. The equations that one has obtained are interesting because they illustratepurely quantum-mechanical features. First, one can see that for electron energy $\in > V_b$ the reflection coefficient *R* is nonozero. This fact reflects the purely quantum efefctb of reflections from the boundaries of the potential well, which vanishes in the classical limit. Second, the denominator of Eq.(3.47) oscillates as a function of energy. At energies satisfaying the condition

$$k_w L = \sqrt{\frac{2m\in}{\hbar^2}} L = n\pi \,, \tag{3.49}$$

The transmission coefficient equals to 1. Consequently, the reflection coefficient R equals zero. For other energies the transmission coefficient is less than 1. At small differencies $\Delta = \in - V_b$ ona can $T(\in) \sim \Delta$, and transmission coefficient vanishes at $\Delta \to 0$. Other interesting quantum-mechanical effects are related to the behaviour of the electron wave function. The probability of finding a propagating particle in the region of a quantum well oscillates as a function of the energy. This behaviour affects the technologically important processes of electron capture of and escape from the quantum wells. These processes are important for numerous applications of quantum structures including the quantum-well semiconductor laser. In Fig.3.12, numerical results are presented for typical depth of potential wells in AlGaAs/GaAs structures for both holes (*a*) and electrons (*b*).

3.3.3. Density of states of a two-dimension electron gas in quantum wells

The important characteristic of a quantum system is the density of states ρ(*E*) that gives the number of quantum states in a small interval of energy (*E, E*+δ*E*):

$$dN(E) = \rho(E)dE \,. \tag{3.50}$$

If the set of quantum numbers corresponding to a certain quantum state is α, the general expression for the density of states is defined by

$$\rho(E)=\sum_{\alpha}\delta(E-E_{\alpha})\,. \tag{3.52}$$

In the case under consideration the set of quantum numbers includes a spin quantum number *s*, a quantum number *n* characterizing ther transverse quantization of the electron states, and a continuous two-dimensional vector $\mathbf{k}=(k_x,k_y)$. Hence set of the quantum numbers are: $\alpha\equiv\left\{s,n,k_x,k_y\right\}$. There is a twofold spin degeneracy of each state $(s=\pm 1/2)$ so that

$$\rho(E)=2\sum_{k_xk_y}\sum_{n}\delta\left[E-\in_n-\frac{\hbar^2(k_x^2+k_y^2)}{2m^*}\right]\quad. \tag{3.53}$$

To calculate the sum over k_x and k_y one can define the area of the surface of the quantum well $L_x\times L_y$, where L_x and L_y are the sizes of the quantum well in the OX and OY directions, respectively. If cyclic boundary conditions are assumed in the OX and OY directions, the possible values of wave vectors k_x and k_y

$$k_x=m\frac{2\pi}{L_x}\,,\qquad k_y=n\frac{2\pi}{L_y}\,. \tag{3.54}$$

Other hand, the large planar dimensions of the quntum well structure allow to use the rule

$$\sum_{k_xk_y}f(k_x,k_y)\;\to\;\frac{L_xL_y}{4\pi^2}\int dk_xdk_y\,f(k_x,k_y)\,. \tag{3.55}$$

By using Eq.(3.55) one transforms Eq.(353) into

$$\rho(E)=\frac{S}{2\pi^2}\sum_{n}\int dk_xdk_y\delta\left[E-\in_n-\frac{\hbar^2(k_x^2+k_y^2)}{2m*}\right]\,. \tag{3.56}$$

Here one shows in which manner the integrals in Eq.(3.56) can be calculated[1]. Consider an integral

$$j = \int dxdy\delta(\varepsilon - ax^2 - ay^2)\,. \tag{3.57}$$

Note at first, that at $\varepsilon < 0$ and $a > 0$ the integral [Eq.(3.57)] equals to zero. When $\varepsilon > 0$, one can use the wellknown folmula for δ-function

$$\delta(A^2 - X^2) = \frac{1}{2A}\left[\delta(X - A) + \delta(X + A)\right] \tag{3.58}$$

and

$$\delta(ax) = \frac{1}{a}\delta(x)\,. \tag{3.59}$$

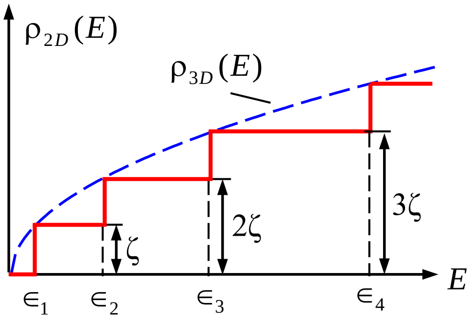


Fig. 3.13. Density of states for two-dimensional electrons in an infinitely deep potential well.

Than, one rewrites Eq.(3.57) as

$$j = \int dydx\frac{1}{a}\delta(\varepsilon / a - x^2 - y^2) =$$
$$= \frac{1}{a}\int dy\frac{1}{2\sqrt{\varepsilon / a - y^2}}\int dx\left[\delta(x - \sqrt{\varepsilon / a - y^2}) + \delta(x + \sqrt{\varepsilon / a - y^2})\right]\,. \tag{3.60}$$

Because argument of integrand varies in the range $[-\infty, +\infty]$, the both terms with delta-function contributes to integral and give 1. As a result, one obtains

$$j = \frac{1}{a}\int dy\frac{1}{\sqrt{\varepsilon / a - y^2}}\,. \tag{3.61}$$

To calculate this integral one needs to determine the integration range. Argument of delta-function in Eq.(3.60) can be equal to zero only when

$$\varepsilon / a - y^2 \geq 0\,, \tag{3.62}$$

or $-\sqrt{\varepsilon / a} \leq y \leq \sqrt{\varepsilon / a}$. Then, one obtains

---

[1] Further mathematical procedures can be ommited by students do not interest of calculation detals

$$j = \frac{1}{a} \int_{-\sqrt{\varepsilon/a}}^{\sqrt{\varepsilon/a}} dy \frac{1}{\sqrt{\varepsilon/a - y^2}} = \frac{1}{a} \arcsin\left(\frac{y}{\sqrt{\varepsilon/a}} \right) \Bigg|_{-\sqrt{\varepsilon/a}}^{\sqrt{\varepsilon/a}} = \frac{\pi}{a} . \tag{3.63}$$

Recalling the $a = \hbar^2 / 2m*$, one can write

$$\rho(E) = \frac{m*S}{\hbar^2 \pi} \sum_n \vartheta(E - \epsilon_n) \ , \tag{3.64}$$

where $\vartheta(x)$ is the Heaviside step function

$$\vartheta(x) = \begin{cases} 1 \ , & x > 0, \\ 0 \ , & x < 0. \end{cases} \tag{3.65}$$

Each term in the sum of Eq.(3.64) corresponds to the contribution from one subband. The contribution of all subbands are equal and independent of energy. Often the density of states per unit area is used to eliminate the size of a sample. As a result, the density of states of two-dimensional electrons exhibits a staircase-shapped energy dependence, with each step being associated with one of the energy states ϵ_n. Figure 3.13 depicts the two-dimensional density of states. It is instructive to compare obtained results with the density of states of electrons in bulk crystals

$$\rho^{3D} = (m^* / \hbar^2)^{3/2} (V / \pi^2) \sqrt{2E} \tag{3.66}$$

Comparing the dencity states for 3D – $\rho^{3D}(E)$ and 2D – $\rho^{2D}(E)$ electron gases, one sees that the diffrency in the behaviours of $\rho^{2D}(E)$ and $\rho^{3D}(E)$ displays maximally for low-energy subbands. Increasing number of subband leads to decresing the difference between $\rho^{2D}(E)$ and $\rho^{3D}(E)$. The difference vanishes at big numbers of subbands n.

Note that real systems are characterized by a dissipative processes, i.e.the scattering of electrons on the defects of crystal lattice. The dissipative procceses lead to smoothing of each of step the function given by Eq.(3.64). The effect rather good describes by empirical roole when Heaviside function should be replaced by

$$\vartheta(E-E_n) \;\to\; \frac{1}{\pi}\left\{\frac{\pi}{2}+\mathrm{Arctg}\left[\frac{(E-\epsilon_n)}{\Gamma_n}\right]\right\} , \qquad (3.67)$$

with Γ_n damping parameter. Note once more, than according to Eq.(3.64) $\rho(E)=0$ at $E<\ \epsilon_1$ and when for the bulk material optical with band gap E_g light absorption can be observed only for photons with energy $\hbar\omega$ for which $\hbar\omega \geq E_g$, then because $\epsilon_1(L) \geq E_g$ for quantum well the absorption edge will be shifted to blue side. This shift depends on the well wide and varies from zero up to 100 meV.

3.4. Transverse transport through multiheterostructure

The electron ststes and its muving in the structures consisting of quantum wells were considered above in this chapter. The considered structures can be named as transverce ones because it was supposed only electron moving directed mainly normally to interfaces planes. The multiheterostructures which form the barriers for the transverse electron moving are the simplest planar systems with spatial quantization. These structures will be considered in this point of the chapter.

3.4.1. Transverse electron moving through double heterostructure

Consider the electron moving trough heterostructure GaAs-AlGaAs-GaAs, in which two semiconductor layers GaAs are separated by thin film of compound AlGaAs with wide band gap. As a result the electron moves under action of square potential (Fig.3.14). Let electron energy $E > V_b$. Because potential left and right of barrier equals to zero, in the both these domains wavevectors are

$$k=\sqrt{2m^{*}E/\hbar^{2}} \, . \qquad (3.68)$$

Inside the barrier domain the electron moving is described by wavevector

$$k_b = \sqrt{2m^*(E - V_b)/\hbar^2}\ . \qquad (3.69)$$

Using well-known T-matrix method one finds the transmission coefficient

$$T = \left\{1 + \frac{V_b^2}{4E(E - V_b)}\sin^2 k_b a\right\}^{-1}, \quad E > V_b\,. \qquad (3.70)$$

For the case $E < V_b$ one should make in Eq.(3.70) the changing $k_b \to \ i\sqrt{2m^*(V_b - E)/\hbar^2} = i\tilde{k}_b$. Then

$$T = \left\{1 + \frac{V_b^2}{4E(V_b - E)}\mathrm{sh}^2\, \tilde{k}_b a\right\}^{-1}, \quad E < V_b\,. \qquad (3.71)$$

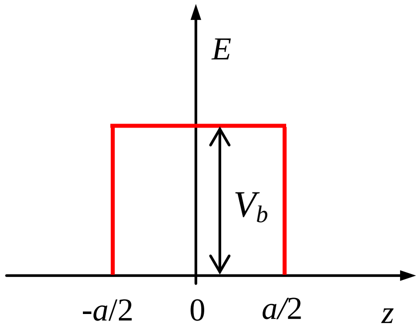


Fig. 3.14. Square potential barrier

In the case when $k_b a$ is large, or when the barrier energy is large, or the barrier wide is large, the exponent of hyperbolic cosine will play a main role in Eq.(3.71), then the transmission coefficient can be represent approximately as

$$T \cong \frac{16E}{V_b} e^{-2k_b a}\,. \qquad (3.72)$$

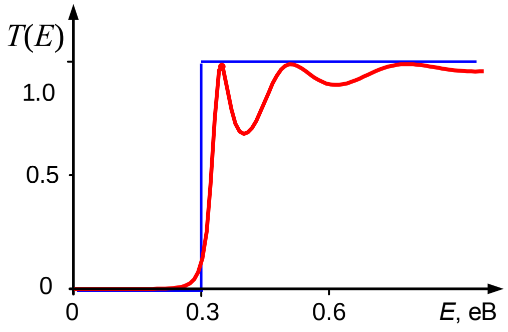


Fig. 3.15. Transmission coefficient of quantum particle (red curve) of barrier with heigh of 0,3 eV and thickness 10 nm in double

heterostructure GaAs-AlGaAs-GaAs. The transmission coefficient for classical particle is shown as blue.

This is dominated by the exponent term, and we obtain the simple estimate for the probability of tunneling through any barrier as $\sim e^{-2k_b a}$. The transmission coefficient $T(E)$ as a function of energy for a square potential barrier of height of 0,3 eV and thickness 10 nm in GaAs. As we can see the transmission coefficient qualitatively differs from the same in the classic case when $T(E) = 1$ for the energies $E > V_b$ and $T(E) =$ 0 for the energies $E < V_b$.

Pay our attention to oscillations of transmission coefficient which clearly manifested at the energies about V_b. Then it should be expected the interesting effects in the tranverse transport. The effects are connected with the quantum nature of carriers moving
in the inhomogeneous at the distances about deBroil wavelength.

3.4.2. Resonant Tunneling

Consider the quantum well between of two barriers, which, for example be formed in the structure consisting of thin semiconductor (GaAs) layer between two thick layers of AlGaAs playing a role of barriers. When the barriers are very thick the singl quantum well formed. The discret energy levels excist in the quantum well. Let the energy level $\in_1$ formed inside the quantum well as it shown in Fig.3. Ther wave function is strongly localized at the well and the quantum state is well defined. When the barrier domains become desrease, the other situation arrises. There are no longer true bound states because an electron can tunnel through one of the barriers and escape from the well. However, the electron may remain in the well for a long time if the barriers are thick enough, and a remnant of the bound state persists as a resonant or quasi-bound state. The energy of this state cannot be precisely defined but is spread into a range $\hbar / \tau$, where τ is the lifetime of an electron in the well before it tunnels away.

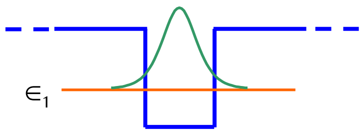


Fig. 3.16. A finite square potential well with a true bound state.

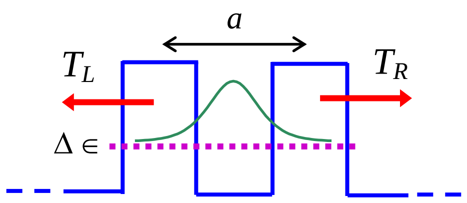


Fig. 3.17. Broadening of localized level into narrow band in the two-barrier structure

Resonant states have a clear signature in the transmission coefficient. In general, the transmission probability T of two barriers is roughly the product of the values for the two individual barriers. Near resonance, however, T rises dramatically above the product and reaches its maximum value of unity if the structure is symmetric. There is perfect transmission through the double barrier, however opaque the individual barriers. This is a resonant tunneling. It is a widely observed effect, seen also in microwaves and in light where it is used in Fabry-Perot etalon. We shall calculate the form of the transmission coefficient for a one-dimensional system and then consider the resonant-tunneling diode.

Let the right-hand barrier has transmission and reflection amplitudes t_R and r_R which are, depended on wave vector k, so its transfer matrix is

$$\mathbf{T_R} = \begin{pmatrix} 1/t_R^* & -r_R^*/t_R^* \\ -r_R/t_R & 1/t_R \end{pmatrix}. \tag{3.73}$$

The results are clearer if the equations are written to appear symmetric from the point of wiev electron trapped between the two barriers. Such an electron is traveling in the usual direction if it tries to escape to the right, but it is going in the opposite direction if it tries to escape to the left. It is therefore better to use the outward going transmission and reflection coefficients for the left-hand barrier. The usual T-matrix applies to an electron impinding from the left can be written in terms of the left-going amplitudes

$$\mathbf{T}_L^{(21)} = \begin{pmatrix} 1/t_L^* & r_L/t_L \\ r_L^*/t_L^* & 1/t_L \end{pmatrix}. \tag{3.74}$$

Taking into account that the left barrier is shifted to – (1/2)a and the right barrier is shifted to (1/2)a, one obtains the T-matrix for the resonant-tunneling structure

$$\mathbf{T}^{(31)}=\begin{pmatrix} e^{-ika/2} & 0 \\ 0 & e^{ika/2} \end{pmatrix}\begin{pmatrix} 1/t_R^* & -r_R^*/t_R^* \\ -r_R/t_R & 1/t_R \end{pmatrix}\begin{pmatrix} e^{ika/2} & 0 \\ 0 & e^{-ika/2} \end{pmatrix}\times$$
$$\times\begin{pmatrix} e^{ika/2} & 0 \\ 0 & e^{-ika/2} \end{pmatrix}\begin{pmatrix} 1/t_L^* & r_L/t_L \\ r_L^*/t_L^* & 1/t_L \end{pmatrix}\begin{pmatrix} e^{-ika/2} & 0 \\ 0 & e^{ika/2} \end{pmatrix}. \tag{3.75}$$

The making product of matrixes gives

$$\mathbf{T}^{(31)}=\begin{pmatrix} \left(1-r_L^*r_R^*e^{-2ika}\right)/t_l^*t_R^* & \left(r_L e^{ika}-r_R^*e^{-ika}\right)/t_R^*t_L \\ \left(r_L^*e^{-ika}-r_Re^{ika}\right)/t_L^*t_R & \left(1-r_L r_Re^{2ika}\right)/t_l\,t_R \end{pmatrix}. \tag{3.76}$$

Amplitude of the wave transmitting the barrier region equals to $1/T_{22}$, then

$$t=\frac{t_Lt_R}{1-r_Lr_Re^{2ika}}\,. \tag{3.77}$$

To analize obtained result one should to write complex amplitude reflection coefficients in the polar form with the phases ρ_R and ρ_L:

$$t=\frac{t_Lt_R}{1-|r_L||r_R|e^{2ika+i\rho_L+i\rho_R}}\,. \tag{3.78}$$

Note, the squaring the modulus of amplitude of reflection and transmission coefficients determine the flux reflection and transmittion coefficients, one obtains from Eq.(3.78)

$$T=\left|\frac{t_Lt_R}{1-|r_L||r_R|e^{2ika+i\rho_L+i\rho_R}}\right|^2=\frac{T_LT_R}{1+R_RR_L-|r_R||r_L|\left(e^{i\varphi}+e^{-i\varphi}\right)}\,, \tag{3.79}$$

with $\varphi=2ka+\rho_L+\rho_R$ the phase. Using that $\cos\varphi=1-2\sin^2\varphi/2$, one can rewrite obtained equation in the form

$$T = \frac{T_L T_R}{\left(1 - \sqrt{R_R}\sqrt{R_L}\right)^2 + 4\sqrt{R_R}\sqrt{R_L}\sin^2 \varphi / 2}. \tag{3.80}$$

All terms in this equation are depended on energy. Then the analisys of the tranmsmittion coefficient is rather hard. But when one takes into account that usually the most rapid variation near a resonance is due to the change in phase of the wave between the barriers, one can assume that the other terms vary slowly with respect to this. It means, in particular, T has a peak when the sine in the dominator vanishes. Then, the condition for resonant states requires $\varphi = 2n\pi$. In this case

$$T = \frac{T_L T_R}{\left(1 - \sqrt{R_L R_R}\right)^2}. \tag{3.81}$$

Whereas the transmission coefficients are small in the real two-barrier structures, one can write

$$\left(1 - \sqrt{R_L R_R}\right)^2 = \left(1 - \sqrt{(1 - T_L)(1 - T_R)}\right)^2 \approx \left(1 - \left(1 - \frac{1}{2}(T_L + T_R)\right)\right)^2 = \frac{1}{4}(T_L + T_R)^2 . \tag{3.82}$$

Then Eq.(3.81) is rewritten to

$$T = T_{ph} \approx \frac{4T_L T_R}{(T_L + T_R)^2}. \tag{3.83}$$

We can see from Eq.(3.83) that in the case of two identical barriers when $T_L = T_R$ transmission coefficient through two barrier structure becomes equal to 1. This is an excellent result of resonance. Indeed each of barrier transmits the electron wave with small probability, but the two-barrier structure can transmit the wave perfectly when $T = 1$.

Note, the condition of resonance

$$2ka + \rho_L + \rho_R = 2n\pi \tag{3.84}$$

Is the requirement for constructive interface within the well. Consider an electron bouncing back and forth between the barriers. It pick up a phase of ka in each direction, with additional phases of ρ_L and

ρ_R when it reflects from the barriers. Using the fact that the individual transmission coefficients for the left and right barriers are small, one obtains from Eq.(3.80)

$$T \approx \frac{T_L T_R}{\left(T_L + T_R\right)^2 / 4 + 4\sin^2 \varphi / 2} = T_{ph} \left[1 + \frac{16}{(T_L + T_R)^2} \sin^2 \frac{\varphi}{2}\right]^{-1} . \tag{3.85}$$

Analising obtained expression, note that because prefactor near the sine, second term in the square brackets can be dominanting if sin is not so small. Suppose that the typical averaged value of sine is ½. Then, one obains form Eq.(3.85) $T \approx (T_L T_R)/2$. Thus the overall transmission coefficient is tipically the product of those for the two barriers, as might be expected. In the case of resonance situation eccentially changes. Strong deviationsn from this occur when the sine vanishes. Put $\varphi = 2n\pi + \delta\varphi$. Expand the sine in Eq.(3.85) to first order over parameter $\delta\varphi$:

$$T \approx T_{ph} \left[1 + \frac{4(\delta\varphi)^2}{(T_L + T_R)^2}\right]^{-1} = \frac{T_{ph}}{1 + \left(\dfrac{\delta\varphi}{\varphi_0 / 2}\right)^2} , \tag{3.86}$$

with $\varphi_0 = T_L + T_R$. One can see that, transmission coefficient decreases in twice compare with its maximum value T_{ph} , when $\delta\varphi = \pm\varphi_0 / 2$. It shows that φ_0 is the full width at half-maximum. The resonant peak has a Lorenzian shape. Rewritting the Lorenzian as function of energy, one obtains

$$\delta\varphi = \left(d\varphi / dk\right)\left(dk / dE\right)\delta E = \frac{\delta E}{\left(dE / dk\right)\left(dk / d\varphi\right)} = \frac{E - E_{ph}}{\left(dE / dk\right)\left(dk / d\varphi\right)} . \tag{3.87}$$

Remember that $k = \sqrt{2m^* E / \hbar^2}$, $\varphi = 2ka + \rho_L + \rho_R$ and calculate the derivations $dk / d\varphi = 1/\left[d\varphi / dk\right] = 1/2a$ та $dE / dk = k\hbar^2 / m^* = v\hbar$. Then

$$\delta\varphi = \frac{E - E_{ph}}{\left(\hbar v / 2a\right)} , \tag{3.88}$$

where v is the velocity of the electron between the barriers. Using Eq.(3.88), one can rewrite Eq.(3.86) in the form

$$T = \frac{T_{ph}}{1 + \dfrac{(E - E_{ph})^2}{(\hbar v / 2a)^2 \left[(T_L + T_R)/2\right]^2}}. \tag{3.89}$$

Introduce the notation

$$\Gamma = \frac{\hbar v}{2a}(T_L + T_R). \tag{3.90}$$

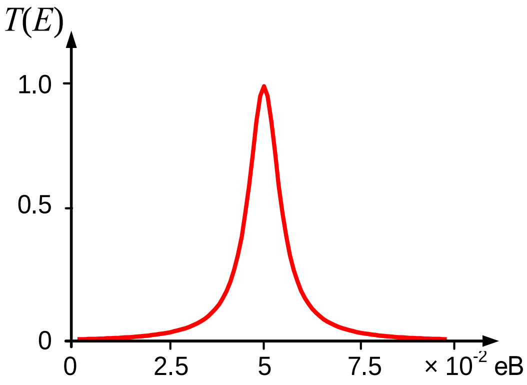


Fig. 3.18. The dependence of transmission coefficient for two-barrier structure on the electron energy. Resonant tunneling

Then the transmission coefficient as a function of energy is

$$T(E) \approx T_{ph}\left[1 + \left(\frac{E - E_{ph}}{\Gamma/2}\right)^2\right]^{-1}, \tag{3.91}$$

with E_{ph} the enrgy of resonance. This Lorenzian shape with a full width at half-maximum of Γ is typical resonant phenomenon. The shape changes if the flux transmission coefficients cannot be taken as constant over the width of the resonance. This might occur in a resonant state near the top of the barrier, where T changes rapidly. The behaviour of transmission coefficient of the electron in two-barrier structure AlGaAs-GaAs-AlGaAs when resonant tunneling, is shown in Fig.3.18. The structure consists of two similar barriers where E_{ph} = 0,05 eB, Γ = 0,01 eB.

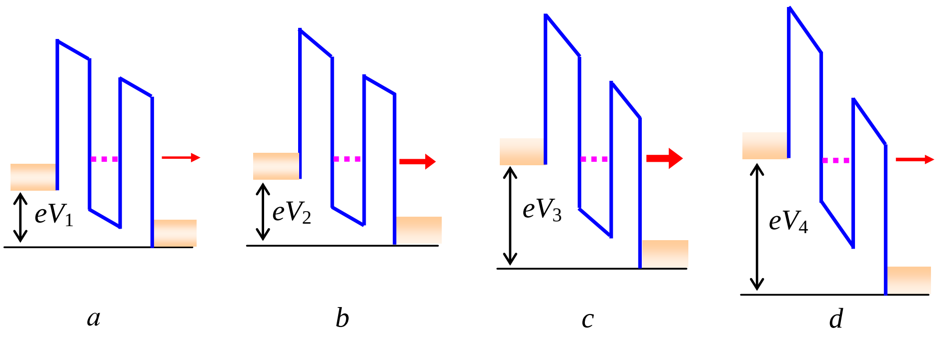


Fig.3.19. The energy profiles trough a resonant-tunneling diode. The bias increases from a to d, giving rise the I(V). The shaded areas on the left and right are the Fermi seas of electrons

One can see, once more, that the behaviour of transmission coefficient has brightly resonant character. This resonant dependence of transmission coefficient allows to expect of obtaining the system with nonlinear current-voltage characteristic. The effect of bias on a resonant-tunneling structure is shown in Fig.3.19. One supposes that only one localized energy level inside the wuantum well E_l. The bias across the diode is small (Fig.3.19 a). In this case the energy level inside the quantum well between the barriers E_l is above the sea of incoming electrons, and little currant flows through diode structure. When the bias has brought down the energy of the resonant level (Fig.3.19 b), so the sea of electrons on the left can pass through it and a larger current flows. In three dimension the current increases linearly as E_l approaches the bottom of the sea of electrons on the left (Fig.3.19 c). A further increase of the bias , as in (d), pulls the energy of the resonant level down so far that it is no longer available to electrons and the current decreases abruptly. The result is a current-voltage characteristic shows negative differential conductivity as it is shown in Fig.3.20, which can be utilized in an amplifier or oscillator. Figure 3.19 shows that the bias has at least three major effects on the electronic structure: it changes the Fermi levels, shifts the energy of the resonant state, and alters profiles and transmission properties of the barriers. The first two are essential, but for simplicity we shall neglect the changes T_L and T_R although this approximation can rarely be justified in practice. When to central layer of tunneling resonant structure one puts the contact with which one can control the energy of resonant level, one obtains new device – tunnel resonant transistor. The resonant-tunneling diode was afirst sample of real device fabricated on the base of quantum well structure with spatial quantization effect. The idea of such diode was proposed in 1963 by Johansen. Esaki and Chang have created operating unit.

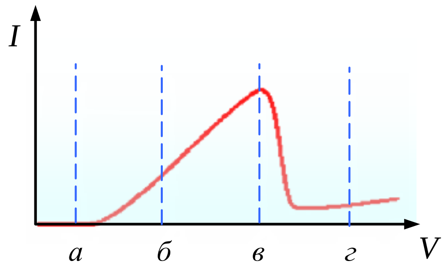


Fig.3.20. Current-voltage characteristic of resonant tunneling diode

3.4.3. Coupling between quantum wells. Superlattices

Consider the electron properties of the structure consisting of few quantum wells, formed with thin semiconductor (say, GaAs) layers separated by thing dielectric (or semiconductor with wide band gap, say AlGaAs) layer shown in Fig.3.21. These structures are characterized by unusual electron properties and can be in interest not only from fundamental point of view but as the systems on which the electronic and optic units are based. Superlattice miniband structures depend on the heterostructure type, either type I, type II or type III. For type I the bottom of the conduction band and the top of the valence subband are formed in the same semiconductor layer. The example of such heterostructure ($Al_xGa_{1-x}As$-GaAs) is shown in Fig.3.22a. In type II the conduction and valence subbands are staggered in both real and reciprocal space, so that electrons and holes are confined in different layers. The example of type II heterostructure is $In_xGa_{1-x}AS$-$GaSb_yAs_{1-y}$ shown in Fig.3.22b. Type III superlattices involve semimetal material, such as HgTe/CdTe. Although the bottom of the conduction subband and the top of the valence subband are formed in the same semiconductor layer in Type III superlattice, which is similar with Type I superlattice, the band gap of Type III superlattices can be continuously adjusted from semiconductor to zero band gap material and to semimetal with negative band gap.

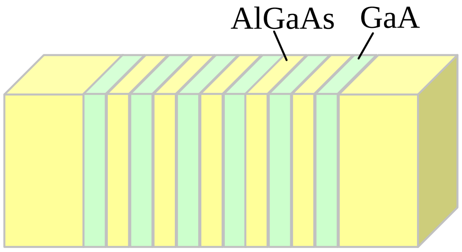


Fig. 3.21. Heterostructure consisting of quantum wells (GaAs) separated by AlGaAs layers form the superlattice

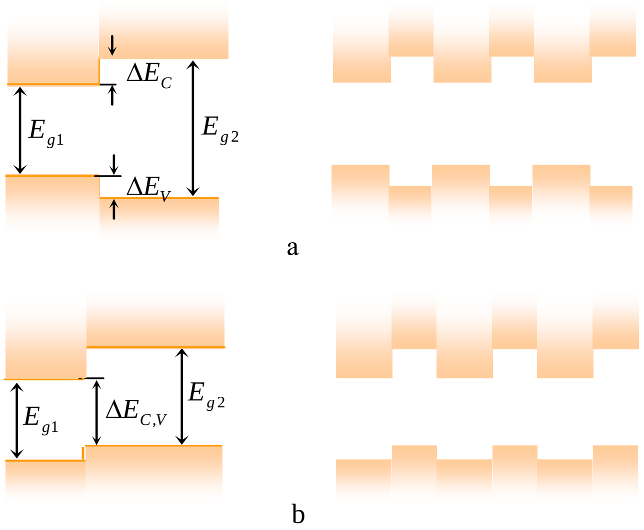


Fig. 3.22. The energy bands in the heterojuntions (left) and superlattices (right) I type (a) and II type (b)

Whether to which type of heterostructure is the superlattice this is characterised by some general features, the main of which is the periodical potential. Emphasize, the characteristics of the potential can be controlled during the structure fabrication. Due to superlattice potential eccentially changes the band structure of semiconductors of which the superlattice is fabricated, the superlattice, per se, is a new type of semiconductor without analogues characteristics in the nature. Then, superlattices are characterized by special properties. There are:

- the electron spectra different from the spectra of initial semiconductors;
- numerous energy bands;
- very strong anisotropy of electron properties;
- suppression processes of electron-hole recombination;
- carriers concentration can controlled vary in the wide range and does not definite by doping;
- wide possibilities for electronic band reconstructions.

Before more detail studies of heterogeneous structures, consider at first the effect of electron interaction in two quantum wells.

3.4.4. Coupling between two quantum wells

When we considered a properties of a single quantum well, we found that the electron wave function is not zero in the barrier region. Accordngly, the electron has some finite probability of penetrating into the barrier (dielectric layer) of a heterostructure. This may be illustrated by calculation of the total probability to find the electron in the barrier region

$$P_b = \int_{|z|>L/2} dz \left|\chi(z)\right|^2, \tag{3.92}$$

where integration is over the barrier layer. The results of calculations are shown in Fig.3.23. This tunneling behaviour is one of the important manifestations of quantum mechanics and it leads to range of physical phenomena that are exploited in various electronic and optoelectronic devices. Study the role of tunneling as a means of providing the coupling between quantum wells. Let us consider the double quantum well structure consisting the two quantum wells, separated by barrier. Let, for simplicity, the quantum wells have identical rectangular shapes and are characterized by the same depth V_b and width L. Suppose the distance between wells is d , and the wells are situated symmetrically with respect to the plane z = 0 (Fig.3.24). Let us consider the one dimensional propagation along Oz axis of the electron across the barrier.

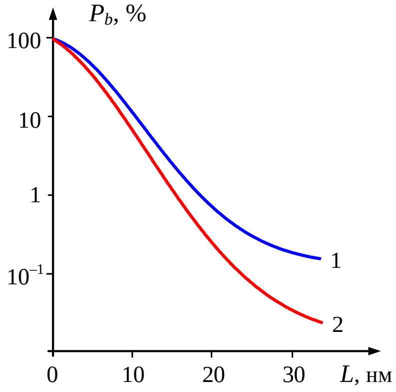


Fig.3.23. Probability of finding the electrons in the barrier region as function of the well thickness L. The electron state is assumed to be the ground state of the well. Curve 1: 1:$V_b = 224$ meV , $m^* = 0,067m_e$; curve 2 correspons to $V_b = 150$ meV , $m^* = 0,4m_e$

For such a double-well system the Shrödinger equation of the transverse motion has a simple form

$$\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2}+V\left(z+\frac{d}{2}\right)+V\left(z-\frac{d}{2}\right)\right]=\in\chi(z)\,. \tag{3.93}$$

The term $V(z+d/2)$ is the potential of the left well centered at z = – d/2, wile the term $V(z-d/2)$ correspondes to the right well centered at z = d/2. The potential is equal to

$$V(z)=\begin{cases} V_b>0, & \text{inside barrier,} \\ 0\,, & \text{inside the wells.} \end{cases} \tag{3.94}$$

One can respect the wave function of the double-well system through superposition of the solutions of the single quantum well (being centered at z = – d/2 and at z = d/2, respectively) problem $\chi_\nu(z)$

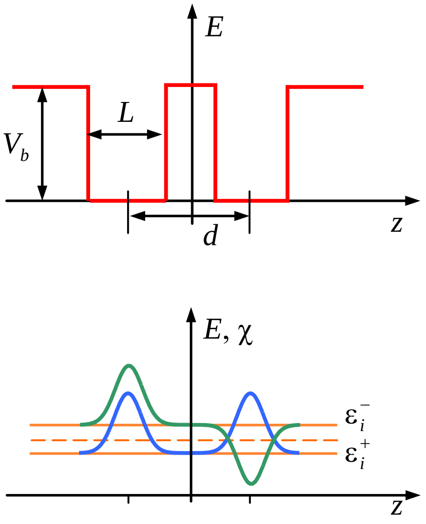


Fig. 3.28. Energy levels and wave functions for a double-well heterostructure (above): the dashed line corresponds to the position of the energy level ε_i in separated quantum wells. For coupled wells the splitting of this level is marked by $\varepsilon_i^{\pm}$.

$$\chi(z) = \sum_{\nu} \left\{ A_\nu \chi_\nu (z - d/2) + B_\nu \chi_\nu (z + d/2) \right\}, \tag{3.95}$$

where ν runs over all possible bound and continuum states of the single quantum well problem. However, if the separation between wells is large enough so that the expected tunneling from one well to another has a small probability, one can restrict himselves to only one level, say "*i*-th", and takes from Eq.(3.95) the combination

$$\chi(z) = A_i \chi_i (z - d/2) + B_i \chi_i (z + d/2) \ . \tag{3.96}$$

This is the wave function for the electron states with energy near one of the bound states ε_i of the separated wells. One should now to find the coefficients A_i and B_i . Because the distance *d* between the wells is supposed to be large, in the first approximation one can conclude that electron state ε_i is twofold degenerate. Then, the electron in state *i* may occupy either the left or the right well with equal probability. To correct the energies and find the coefficients A_i and B_i , we should use the following procedure of the perturbation theory for degenerate states. Let us use that ε_i is an energy of electron localized inside the single quantum well. Then Schrödinger equation for electrons in two potential wells we can write in the form

$$\left[-\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + V(z - d/2) \right] \chi_i (z - d/2) = \varepsilon_i \chi_i (z - d/2) \, , \tag{3.97}$$

and

$$\left[-\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + V(z + d/2) \right] \chi_i (z + d/2) = \varepsilon_i \chi_i (z + d/2) \ . \tag{3.98}$$

Substitute the wave function (3.96) into equation (3.93) and use with equations (3.97) and (3.98). We obtain the next equation

$$\left[\varepsilon_i + V(z + d/2) - \in \right] A_i \chi_i (z - d/2) + \left[\varepsilon_i + V(z - d/2) - \in \right] B_i \chi_i (z + d/2) = 0 \, . \tag{3.99}$$

Then we multiply this equation first by $\chi_i^*(z-d/2)$, and then by $\chi_i^*(z+d/2)$. Obtainined two equations integrate over coordinate z. We obtain

$$\left[\varepsilon_i + \int\limits_{-\infty}^{\infty} dz\chi_i^*(z-d/2)V(z+d/2)\chi_i(z-d/2) - \in\right]A_i +$$
$$+\left[(\varepsilon_i - \in)\int\limits_{-\infty}^{\infty} dz\chi_i^*(z-d/2)\chi_i(z+d/2) + \int\limits_{-\infty}^{\infty} dz\chi_i^*(z-d/2)V(z-d/2)\chi_i(z+d/2)\right]B_i = 0$$

and

$$\left[(\varepsilon_i - \in)\int\limits_{-\infty}^{\infty} dz\chi_i^*(z+d/2)\chi_i(z-d/2) + \int\limits_{-\infty}^{\infty} dz\chi_i^*(z+d/2)V(z+d/2)\chi_i(z-d/2)\right]A_i +$$
$$+\left[(\varepsilon_i - \in + \int\limits_{-\infty}^{\infty} dz\chi_i^*(z+d/2)V(z-d/2)\chi_i(z+d/2)\right]B_i = 0.$$

Let us introduce the next designations

$$\int dz\chi_i^*(z-d/2)\chi_i(z+d/2) = \int dz\chi_i^*(z+d/2)\chi_i(z-d/2) = r_i\,, \tag{3.100}$$

$$\int dz\chi_i^*(z\pm d/2)V(z\mp d/2)\chi_i(z\pm d/2) = s_i\,, \tag{3.101}$$

$$\begin{aligned}&\int dz\chi_i^*(z-d/2)V(z-d/2)\chi_i(z+d/2) = \\ &= \int dz\chi_i^*(z+d/2)V(z+d/2)\chi_i(z-d/2) = t_i\end{aligned} \tag{3.102}$$

and obtain two homogeneous linear equations for A_i and B_i

$$\left[\varepsilon_i + s_i - \in\right]A_i + \left[(\varepsilon_i - \in)r_i + t_i\right]B_i = 0\,, \tag{3.103}$$

$$\left[(\varepsilon_i - \in)r_i + t_i\right]A_i + \left[(\varepsilon_i - \in + s_i\right]B_i = 0\,. \tag{3.104}$$

To existing nontrivial solutions of this system the determinant of the system have to equals to zero, or

$$\left(\varepsilon_i + s_i - \in\right)^2 - \left[(\varepsilon_i - \in) r_i + t_i\right]^2 = 0 . \tag{3.105}$$

Finally, we obtain

$$\left(\varepsilon_i + s_i - \in\right) = \pm\left[(\varepsilon_i - \in) r_i + t_i\right] . \tag{3.106}$$

Then, the electron energy of two binding quantum wells obtains as modified energy level of a single well by shifted and spllitted the initial energy level.

$$\in_{(i)}^{\pm} = \varepsilon_i + \frac{s_i}{1 \pm r_i} \pm \frac{t_i}{1 \pm r_i} . \tag{3.107}$$

The energy levels of two binding quantum wells are defined by the set of integrals (3.100)–(3.102). The first integral (r_i) is called an overlap integral, and describes the measure of wave functions from different wells overlap. The second (s_i) is named a shift integral. It characterizes the shift of degenerate energy level due to tunneling interaction between the states in the different wells. The third integral (t_i) is named a transfer integral. It reflects the measure of splitting the initial energy level into two. When the distance between the wells is rather big, the overlapping is small. It means that weak penetration of the electron wave function into barrier region and the overlap integral is small compared with 1, so that we can neglect the corresponding terms in Eq.(3.107). Then we obtain the simple equations for the energies

$$\in_{(i)}^{\pm} \cong \varepsilon_i + s_i \pm t_i . \tag{3.108}$$

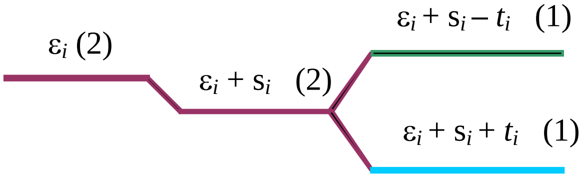


Fig. 3.29. Diagram of an energy level modification in a double well heterostructure when shift and transfer integrals are taken into account. The degeneracy of levels is indicated in parentheses.

Sinse integrals s_i and t_i are always less than zero, the first is responsible for shifting down the levels and leads to the energy decrease down from the initial energy position of the single quantum well. Likewise, the second determines the splitting of the initial twofold degenerate level (see, Fig.3.29). For the lowest level, which is designated as $\in^{+}_{(i)}$. The level $\in^{+}_{(i)}$ corresponds to symmetric combination in Eq.(3.107), i.e. $A_i = B_i$. The correspondinhg wave function schematically shown in Fig.3.25 with blue color. The upper level $\in^{-}_{(i)}$ is antisymmetric with $A_i = -B_i$ which schematically shown in Fig.3.25 by green curve. From Eqs.3.100-102 we can see that both shifting and splitting are caused by tunneling phenomenon. The same consideration is valid for any electron states i localized in coupled quantum wells. Eqs.(3.107-108) describe the energies of the transverse electron motion in a double-well structure. The total energies include the in-plane electron motion kinetic energy are

$$E_i^{(\pm)}(\vec{k}_{\|}) = \in^{\pm}_{(i)} + \frac{\hbar^2}{2m*} k_{\|}^2 . \tag{3.109}$$

The main result of the above consideration is the existence of a special type of interaction between coupled quantum wells, that is there is a coupling associated with tunneling between adjacent quantum wells. Note that in the considered structure the specific character of electron moving remains unchanged, namely, in-plane the electrons behave like free particles with an effective mass m^*. The electron motion in the transverse direction is quantized and spatially confined. One can see from Fig.3.28 the both symmetric and antisymmetric states are localized at the both quantum wells. The wave functions of the both these states have the extremums. The extremums correspond to maximum of the square modulus of wave function $|\chi|^2$.

Chapter 4. Monocrystalline, multicrystalline, polycrystalline and amorphous silicon

4.1. Introduction

Silicon is the second most abundant element on earth. It always occurs in the form of a compound, most often as the oxide (silica) or as a silicate. These compounds form a wide variety of rocks and minerals. The common mineral quartz is a form of silica.

During 20th century, the properties now associated with semiconductors were first observed. These properties include the negative temperature dependence of resistivity due to the exponential variation of the free carrier density with temperature; photoconductivity, a decrease in resistance on exposure to light; various thermoelectric effects; the Hall and other magnetoelectric effects; and rectification between the semiconductor and a metal contact. This last property was widely used for detectors in crystal radio receivers in the first half of the 20th century.

Silicon is a chemical element with symbol Si and atomic number 14 (Fig. 4.1). It is a tetravalent metalloid, more reactive than germanium, the metalloid directly below it in the table. Controversy about silicon's character dates to its discovery; it was first prepared and characterized in pure form in 1823. In 1808, it was given the name silicium (from Latin: silex, hard stone or flint), with an-ium word-ending to suggest a metal, a name which the element retains in several non-English languages. However, its final English name, first suggested in 1817, reflects the more physically similar elements carbon and boron.

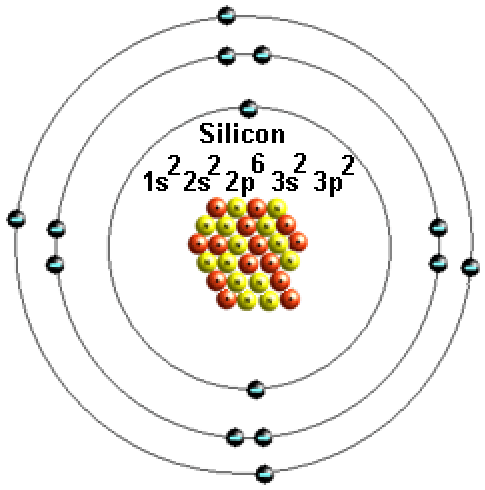


Fig. 4.1. Silicon atom structure.

Elemental silicon also has a large impact on the modern world economy. Although most free silicon is used in the steel refining, aluminium-casting, and fine chemical industries (often to make fumed silica), the relatively small portion of very highly purified silicon that is used in semiconductor electronics (< 10%) is perhaps even more critical. Because of wide use of silicon in integrated circuits, the basis of most computers, a great deal of modern technology depends on it.

Silicon is an essential element in biology, although only tiny traces of it appear to be required by animals. However, various sea sponges as well as microorganisms like diatoms and radiolaria secrete skeletal structures made of silica. Silica is often deposited in plant tissues,

Silicon is a solid at room temperature, with relatively high melting and boiling points of 1414 and 3265 °C, respectively. Like water, it has a greater density in a liquid state than in a solid state, and so, like water but unlike most substances, it does not contract when it freezes, but expands. With a relatively high thermal conductivity of 149 $W \cdot m^{-1} \cdot K^{-1}$, silicon conducts heat well.

In its crystalline form, pure silicon has a gray color and a metallic luster. Like germanium, silicon is rather strong, very brittle, and prone to chipping. Silicon, like carbon and germanium, crystallizes in a diamond cubic crystal structure, with a lattice spacing of 0.5430710 nm.

The outer electron orbital of silicon, like that of carbon, has four valence electrons. The 1s, 2s, 2p and 3s subshells are completely filled while the 3p subshell contains two electrons out of a possible six.

Silicon is a semiconductor. It has a negative temperature coefficient of resistance, since the number of free charge carriers increases with temperature. The electrical resistance of single crystal silicon significantly changes under the application of mechanical stress due to the piezoresistive effect.

Silicon is a metalloid, readily either donating or sharing its four outer electrons, allowing for many forms of chemical bonding. Like carbon, it typically forms four bonds. Unlike carbon, it can accept additional electrons and form five or six bonds in a sometimes more labile silicate form. Tetra-valent silicon is relatively inert, but still reacts with halogens and dilute alkalis, but most acids (except for some hyper-reactive combinations of nitric acid and hydrofluoric acid) have no known effect on it. However, having four bonding electrons gives it, like carbon, many opportunities to combine with other elements or compounds in the right circumstances.

Measured by mass, silicon makes up 27.7% of the Earth's crust and is the second most abundant element in the crust, with only oxygen having a greater abundance. Silicon is usually found in the form of complex silicate minerals, and less often as silicon dioxide (silica, a major component of common sand). Pure silicon crystals are very rarely found in nature.

4.2. Types of Crystalline and Amorphous Silicon

The manufacture of the hyperpure silicon for electronics and photovoltaics occurs in two stages. The oxygen is removed to produce metallurgical grade silicon. It is further refined to produce semiconductor grade silicon. An intermediate grade with impurity levels between metallurgical silicon and semiconductor grade silicon is often termed solar grade silicon.

4.2.1. Metallurgical Grade Silicon

Silica is the dioxide form of silicon (SiO_2) and occurs naturally in the form of quartz. While beach sand is also largely quartz, the most common raw material for electronic grade is high purity quartz rock. Ideally the silica has low concentrations of iron, aluminum and other metals. The silica is reduced (oxygen removed) through a reaction with carbon in the form of coal, charcoal and heating to 1500-2000 °C in an electrode arc furnace.

$$SiO_2 + C \rightarrow Si + CO_2 \qquad (4.1)$$

The resulting silicon is metallugical grade silicon (MG-Si). It is 98% pure and is used extensively in the metallurgical industry.

An even greater production of silicon is in the form of ferrosilicon that is manufactured using a similar process to that described above but is done in the presence of iron. Ferrosilicon is used extensively in metals manufacture. In 2013 the total production of silicon was 7.6 million tones and 80 % of that was in the form of ferrosilicon (Fig. 4.2) [1].

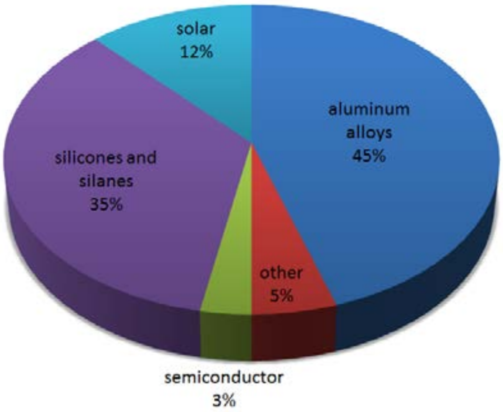


Fig. 4.2. From the 1.8 million tones of metallurgical silicon produced in 2010, 12 % was for the production of silicon solar cells [2].

Metallurgical (MG) silicon is produced at the rate of millions of tons/year at a low economic cost of few $/kg and an energy cost of 14–16 kWh/kg. As such, it is 98–99% pure, with a major contamination of carbon, alkali-earth and transition metals, and hundreds of ppmw of B and P (Fig. 4.3) [3].

The transition metals in the silicon result in deep levels in the bandgap and the high recombination activity make metallurgical grade silicon unsuitable for use in electronics. In addition, boron and phosphorous dopant impurities are much too high in concentration (>50–100 ppmw) to allow suitable compensation procedures [4].

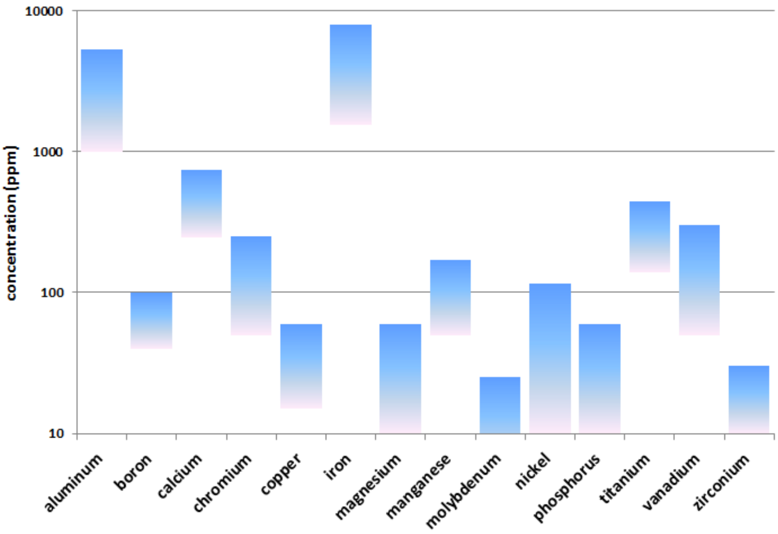


Fig. 4.3. The impurity levels in metallurgical grade silicon vary widely due to process variation and the source raw materials of silica and carbon [5].

4.2.2. Electronic Grade Silicon

A small amount of the metallurgical grade silicon is further refined for the semiconductor industry. Powdered MG-Si is reacted with anhydrous HCl at 300 °C in a fluidized bed reactor to form $SiHCl_3$

$$Si + 3HCl \rightarrow SiHCl_3 + H_2 \qquad (4.2)$$

During this reaction impurities such as Fe, Al, and B react to form their halides (e.g. $FeCl_3$, $AlCl_3$, and BCl_3). The $SiHCl_3$ has a low boiling point of 31.8 °C and distillation is used to purify the $SiHCl_3$ from the impurity halides. The resulting $SiHCl_3$ now has electrically active impurities(such as Al, P, B, Fe, Cu or Au) of less than 1 ppba.

Finally, the pure $SiHCl_3$ is reacted with hydrogen at 1100°C for ~200 – 300 hours to produce a very pure form of silicon.

$$SiHCl_3 + H_2 \rightarrow Si + 3HCl \qquad (4.3)$$

The reaction takes place inside large vacuum chambers and the silicon is deposited onto thin polysilicon rods (small grain size silicon) to produce high-purity polysilicon rods of diameter 150-200 mm. The process was first developed by Siemens in the 60's and is often referred to as the Siemens process.

The resulting rods of semiconductor grade silicon are broken up to form the feedstock for the crystallisation process. The production of semiconductor grade silicon requires a lot of energy. Solar cells can tolerate higher levels of impurity than integrated circuite fabrication and there are proposals for alternative processes to creat a "solar-grade" silicon.

4.2.3. Terms for Silicon

The allotropic forms of silicon range from a single crystalline structure to a completely unordered amorphous structure with several intermediate steps in between. Silicon or other semiconductor materials used for electronics and solar cells can be single crystalline, multicrystalline, polycrystalline, microcrystalline, nanocrystalline or amorphous (Table 4.1, Fig. 4.4) [6]. The key difference between these materials is degree to which the semiconductor has a regular, perfectly ordered crystal structure, and therefore semiconductor material may be classified according to the size of the crystals making up the material.

Table 4.1. Terms for silicon.

Descriptor	Symbol	Grain Size	Common Growth Techniques
Single crystal	sc-Si	>10cm	Czochralski (CZ) float zone (FZ)
Multicrystalline	mc-Si	1mm-10cm	Cast, sheet, ribbon
Polycrystalline	pc-Si	1µm-1mm	Chemical-vapour deposition
Microcrystalline	µc-Si	<1µm	Plasma deposition
Nanocrystalline	nc-Si	<100 nm	Plasma deposition
Amorphous	a-Si	0	Plasma deposition

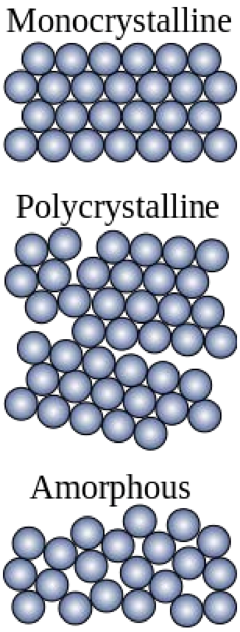


Fig. 4.4. Schematic of allotropicforms of silicon.

Monocrystalline silicon (mono c-Si) is a form in which the crystal structure is homogenous throughout the material; the orientation, lattice parameter, and electronic properties are constant throughout the material [7]. Dopant atoms such as phosphorus and boron are often incorporated into the film to make the silicon n-type or p-type respectively. Monocrystalline silicon is fabricated in the form of silicon wafers, usually by the Czochralski Growth method, and can be quite expensive depending on the radial size of the desired single crystal wafer. This monocrystalline material, while useful, is one of the chief expenses associated with producing photovoltaics where approximately 40% of the final price of the product is attributable to the cost of the starting silicon wafer used in cell fabrication [8].

Multicrystalline silicon (multi c-Si) is composed of many smaller silicon grains of varied crystallographic orientation, typically >1 mm in size. This material can be synthesized easily by allowing liquid silicon to cool using a seed crystal of the desired crystal structure. Additionally, other methods for forming smaller-grained polycrystalline silicon (poly-Si) exist such as high temperature chemical vapor deposition (CVD).

Nanocrystalline silicon nc-Si), sometimes also known as microcrystalline silicon (μc-Si), is an allotropic form of silicon with paracrystalline structure - is similar to amorphous silicon (a-Si), in that it has an amorphous phase. Where they differ, however, is that nc-Si has small grains of crystalline silicon within the amorphous phase. This is in contrast to polycrystalline silicon (poly-Si) which consists solely of crystalline silicon grains, separated by grain boundaries. The difference comes solely from the grain size of the crystalline grains. Most materials with grains in the micrometre range are actually fine-grained polysilicon, so nanocrystalline silicon is a better term. The term Nanocrystalline silicon refers to a range of materials around the transition region from amorphous to microcrystalline phase in the silicon thin film.

Amorphous silicon (a-Si) has no long-range periodic order. The application of amorphous silicon to photovoltaics as a standalone material is somewhat limited by its inferior electronic properties [9]. When paired with microcrystalline silicon in tandem and triple-junction solar cells, however, higher efficiency can be attained than with single-junction solar cells [10]. This tandem assembly of solar cells allows one to obtain a thin-film material with a bandgap of around 1.12 eV (the same as single-crystal silicon) compared to the bandgap of amorphous silicon of 1.7-1.8 eV. Tandem solar cells are then attractive since they can be fabricated with a bandgap similar to single-crystal silicon but with the ease of amorphous silicon.

4.3. Single Crystalline Silicon

The atoms in crystalline materials are arranged in a periodic pattern which is characteristic of the material (Fig. 4.5). This pattern can be visualized as a number of parallelepipeds with atoms at the corners. These parallelepipeds, or unit cells, repeat themselves throughout the structure. There are seven different crystal systems; some of these have variants with extra atoms between the corner atoms so that 14 different atomic structures are possible [11]. Cubic crystals are characterized by unit cells with three equal lengths along three mutually perpendicular axes. The unit cell of a cubic crystal is described by the lattice constant, *a*, which is the length of a side of the cube.

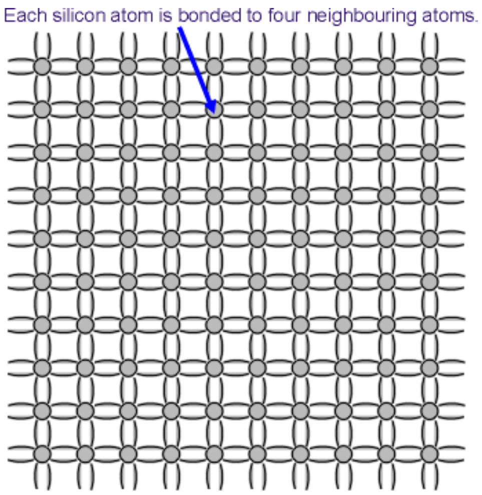


Fig. 4.5. The regular arrangement of silicon atoms in single-crystalline silicon produces a well-defined band structure. Each silicon atom has four electrons in the outer shell. Pairs of electrons from neighbouring atoms are shared so each atom shares four bonds with the neighbouring atoms.

Single-crystalline wafers typically have better material parameters but are also more expensive (Fig. 4.6). Crystalline silicon has an ordered crystal structure, with each atom ideally lying in a pre-determined position. Crystalline silicon exhibits predictable and uniform behaviour but because of the careful and slow manufacturing processes required, it is also the most expensive type of silicon.

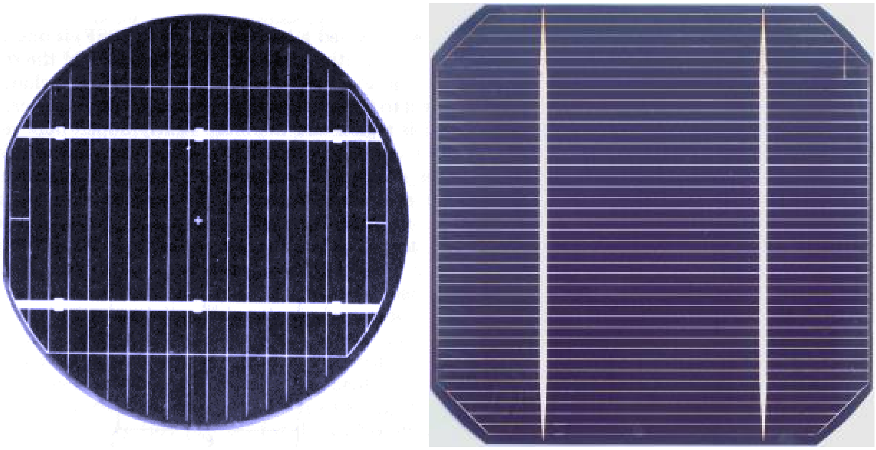


Fig. 4.6. Single crystalline silicon is usually grown as a large cylindrical ingot producing circular or semi-square solar cells. The semi-square cell started out circular but has had the edges cut off so that a number of cells can be more efficiently packed into a rectangular module.

In single crystalline silicon material the crystal orientation is defined by Miller indices. A particular crystal plane is noted using parenthesis such as (100). Silicon has a cubic symmetrical cubic structure and so (100), (010) etc are equivalent planes and collectively referred to using braces {100}. Similarly, the crystal directions are defined using square brackets, e.g. [100] and referred collectively using triangular brackets, <100>.

In solar cells the preferred orientation is <100> as this can be easily textured to produce pyramids that reduce the surface reflectivity. However, some crystal growth processes such as dendritic web <111> produce material with other orientations.

To denote the crystal directions, single crystal wafers often have flats to denote the orientation of the wafer and the doping. The most common standard is the SEMI standard:

1. If the minor flat is 180° from the major flat the wafer is n-type <100>
2. If the minor flat is 90° to the left or right the wafer is p-type <100>.
3. If the minor flat is 45° up on the left or right the wafer is n-type <111>
4. If there are no minor flats the wafer is p-type <111>

Single crystalline substrates are typically differentiated by the process by which they are made. Czochralski (Cz) [12] wafers are the most commonly used type of silicon wafer, and are used by both the solar and integrated circuit industry (Fig. 4.7). The process of making a large single crystalline silicon ingot by the Czochralski process is shown in Fig. 4.8. The use of quartz crucibles in the manufacture of Cz substrates causes the incorporation of ppm (10^{18} cm^{-3}) oxygen into the silicon ingot. The oxygen itself is relatively benign but creates complexes with boron doping that degrades the carrier lifetime [13, 14]. N-type ingots fabricated with phosphorous dopants have similar oxygen concentrations but do not show the degradation effect nor do wafers with lower resistivity or gallium dopants [15].

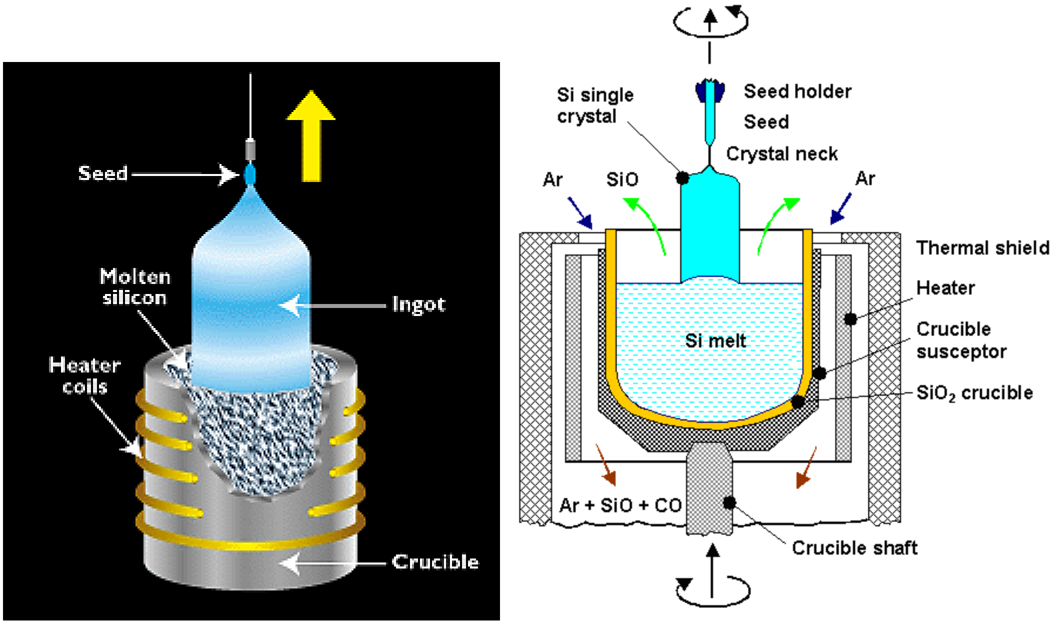


Fig. 4.7. Schematic of Czochralski wafer growth.

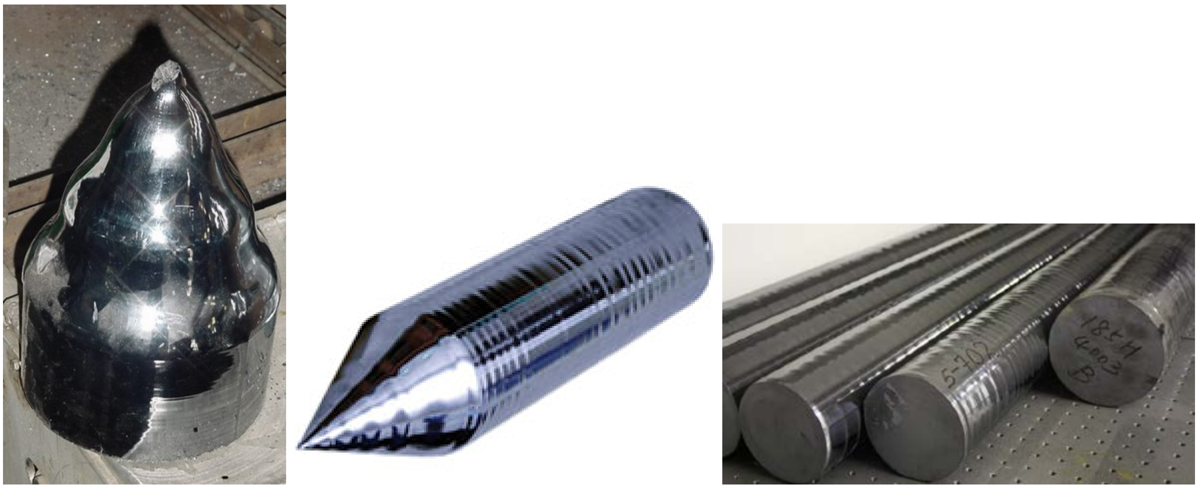


Fig. 4.8. Czochralski grown silicon. The bottom cylindrical section has been cut off to make wafers. Such "tops and tails" left over from growing the semiconductor industry are a large source of silicon supply for the photovoltaics industry.

Even though the CZ process is commonly used for commercial substrates, it has several disadvantages for high efficiency laboratory or niche market solar cells. CZ wafers contain a large amount of oxygen in the silicon wafer. Oxygen impurities reduce the minority carrier lifetime in the solar cell, thus reducing the voltage, current and efficiency. In addition, the oxygen and complexes of the oxygen with other elements may become active at higher temperatures, making the wafers sensitive to high temperature processing. To overcome these problems, Float Zone (FZ) wafers may be used (Fig. 4.9) [16]. In this process, a molten region is slowly passed along a rod or bar of silicon. Impurities in the molten region tend stay in the molten region rather than be incorporated into the solidified region, thus allowing a very pure single crystal region to be left after the molten region has passed. Due to the difficulty in growing large diameter ingots and the often higher cost, FZ wafers are typically only used for laboratory cells and are less common in commercial production [17].

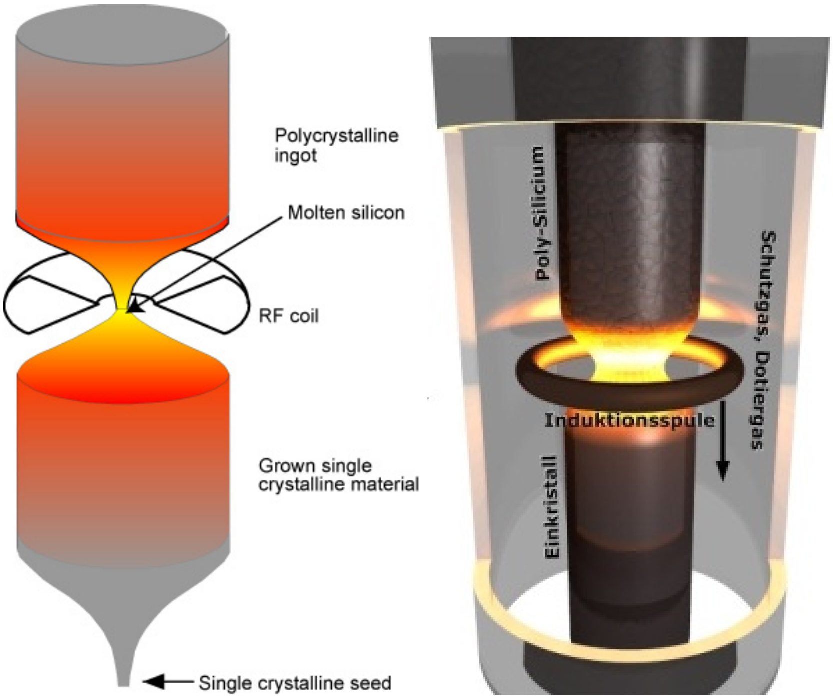


Fig. 4.9. Schematic of Float Zone wafer growth.

4.3.1. Silicon Crystal Structure

Silicon, like other group IV insulators and semiconductors including diamond, germanium, and gray tin, crystallizes in the diamond cubic lattice structure [18]. Each silicon atom has four bonds, one to each of its four nearest neighbors, as shown in Fig. 4.10. This lattice structure can be constructed from two interpenetrating face-centered cubic lattices displaced from each other along the body diagonal by a distance equal to one-fourth of its length. Group III-V compounds and some II-VI compounds crystallize in a similar arrangement, called the zinc-blende lattice, where one of the constituents occupies sites on one of the face-centered cubic lattices and the other constituent occupies sites on the other.

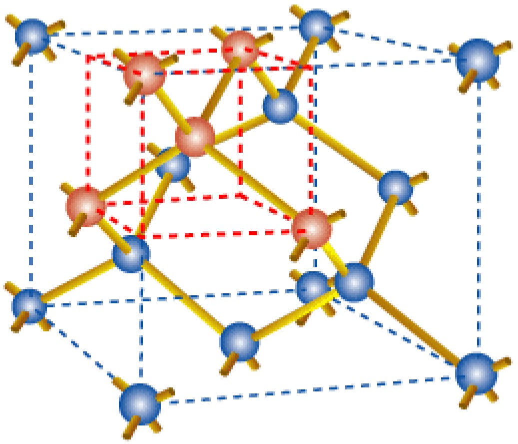


Fig. 4.10. Unit cell of the diamond cubic lattice structure showing the lattice constant, a, and the locations of the silicon-silicon bonds [19].

The lattice constant, *a*, is 0.5430710 nm in pure, floatzoned silicon [20]. The diamond cubic lattice is much more open than most crystal configurations. Five large interatomic voids in which interstitial atoms are easily accommodated are located along the body diagonals as shown in Fig. 4.11**.** However, inclusion of atoms in these voids may change a; for example, if about **10l8** interstitial oxygen atoms/cm^3 are present, *a* = 0.5430747 nm [20]**.** Though it is quite small, this difference becomes very important in the utilization of silicon samples for accurate determination of the Avogadro constant by the **X**-ray crystal density method [21]. Although oxygenfree, semiconductor-grade silicon represents the most perfect material available for this purpose, refinements in the determination are still being made as material with fewer and fewer crystal defects become available.

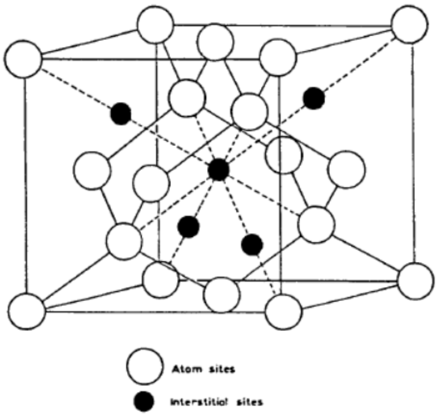


Fig. 4.11. Unit cell of the diamond cubic lattice structure showing the five large interatomic voids along the body diagonals [22].

It is convenient to describe the various planes which pass through the atomic sites in a crystal by their Miller indices. These indices, h, k, and 1, are the reciprocals of the intercepts of the plane with the x, y, and z axes, respectively. A plane is denoted by the symbol (hkl); negative intercepts are denoted by a bar over the index. A family of planes, e.g., (l00), (010),(001), (001^{-}), (01^{-}0), and (1^{-}00), is denoted by a bracketed symbol: (100). In general, the indices are expressed as the smallest integers having the same ratio. Selected low index planes are shown in Fig. 4.12.

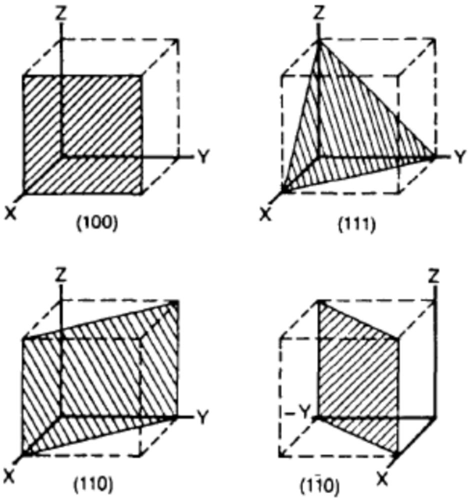


Fig. 4.12. Examples of low-index planes.

The Miller indices may also be used to indicate a direction. In a cubic crystal, the x, y, and **z** axes form an orthogonal set; therefore the direction perpendicular to a particular plane has the same Miller indices as the plane. A direction is denoted by a square bracketed symbol: [hkl]; a family of directions is denoted by angular brackets: <hkl>. For example, the x axis is the [100] direction and the three coordinate axes are the <l00> directions. The body diagonals are the <111> directions.

In a simple cubic lattice, the spacing between planes, d_{hkl} is given by:

$$d_{hkl} = \frac{a}{[(h^2 + k^2 + l^2)]^{1/2}} \qquad (4.4)$$

The distance between the (100) planes (which pass through the corner atoms) is *a*. However, because there are atoms between the corner atoms in the diamond cubic structure, planes which pass through the atoms in the crystal are actually closer together than this if the Miller indices are expressed in the usual form as the smallest integers. There are eight atoms in the unit cell of the diamond lattice. These are located at the following points: (0,0,0), (0,1/2,1/2), (1/2,0,1/2), (1/2,1/2,0), (1/4,1/4,1/4), (1/4,1/4,3/4), (3/4,1/4,3/4), and (3/4,3/4,1/4). Therefore the (100) planes which include all the atoms are the (400) planes, and so the actual distance between the atomic planes along any of the cubic axes is *a*/4 or 0.1358 nm.

4.3.2. Crystal defects

The atoms in a silicon crystal are stacked in regular order. Along the [lll] direction, adjacent pairs of atomic planes are displaced horizontally from each other. There are three different locations for the atoms; the cycle repeats every third group. Atoms in planes identified by the same symbol lie directly under each other along the [lll] direction. Any disturbance in this regular order of atoms is a crystal defect.

A wide variety of defects can form in a silicon crystal during and after growth. Their properties are summarized in the following paragraphs. In addition, it should be noted that defects in silicon have been the object of intensive research efforts in the past decade or so.

Point Defects. Vacancies, silicon self interstitials, and impurity atoms in substitutional or interstitial sites, are defects which occur at a lattice point and do not distort the crystal over an extended distance.

Vacancies and self interstitials. The combination of a self interstitial and an associated vacancy is known as a Frenkel defect while an independent vacancy is known as a Schottky defect. These point defects exist in thermodynamic equilibrium in the crystal in relatively large concentrations at temperatures near the melting point of silicon. Their concentrations at room temperature are governed both by the thermal history of the crystal and by the presence of sources and sinks for the defects. Vacancies may be quenched in during cooling or they may be generated by the plastic deformation which may occur as a result of thermal stresses during the cool down process. Vacancies which diffuse to the surface of the crystal disappear. The surface is the principal source and sink for vacancies, but extended defects within the crystal can also serve as sinks. A pair of vacancies can also combine to form a divacancy which appears to be a stable defect in silicon. Originally, vacancies were thought to be the dominant form of point defect in silicon crystals. However, at the present time, interstitials appear to be assuming a greater significance. Silicon self interstitials are generated during the oxidation of silicon, not only at the surface but also during the formation of oxide precipitates in the bulk. Both vacancies and self interstitials interact extensively with other defects and play a dominant role in impurity diffusion in silicon. Studies of the effects of surface oxidation on impurity diffusion have led to a considerable increase in our understanding of the characteristics of vacancies and self interstitials in silicon [23, 24].

Atomic impurities. Impurity atoms may exist in the silicon crystal either substitutionally on sites usually occupied by silicon atoms or in interstitial sites. Solubilities of various atomic impurities in silicon [25] are summarized in Fig. 4.13 together with additional data on silver [26], nickel [27], and sodium and potassium [28]. More recent solubility data [29] for some of the transition metals are given in Fig. 4.14.

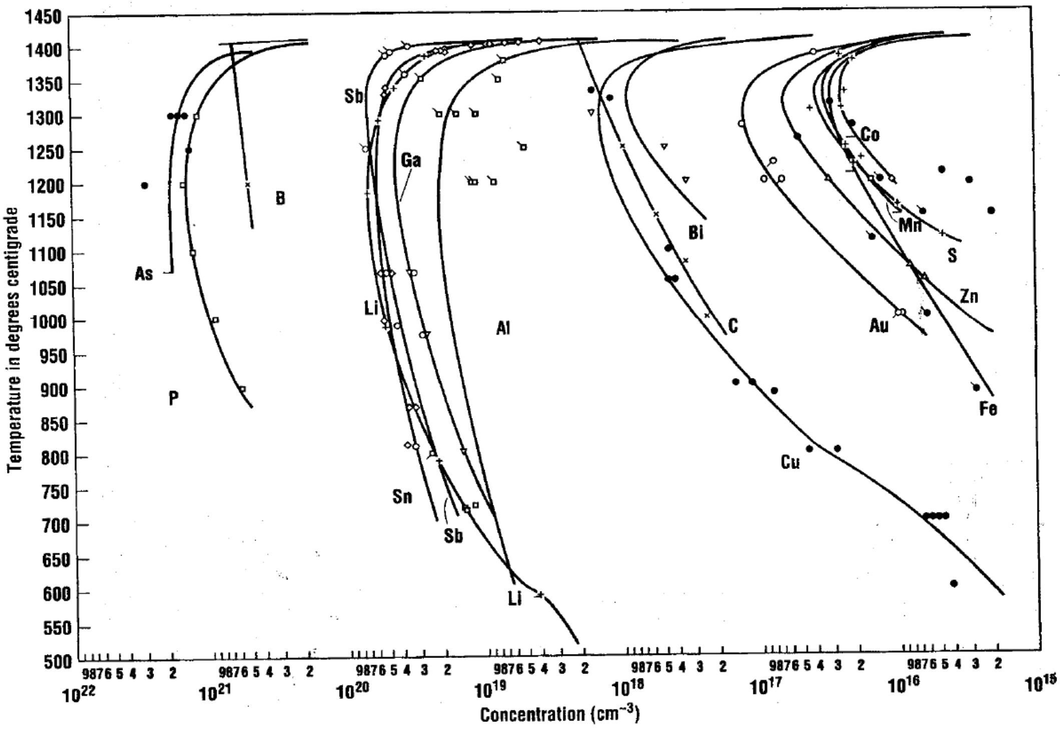


Fig. 4.13. Solid solubilities of impurities in silicon as a function of temperature [30].

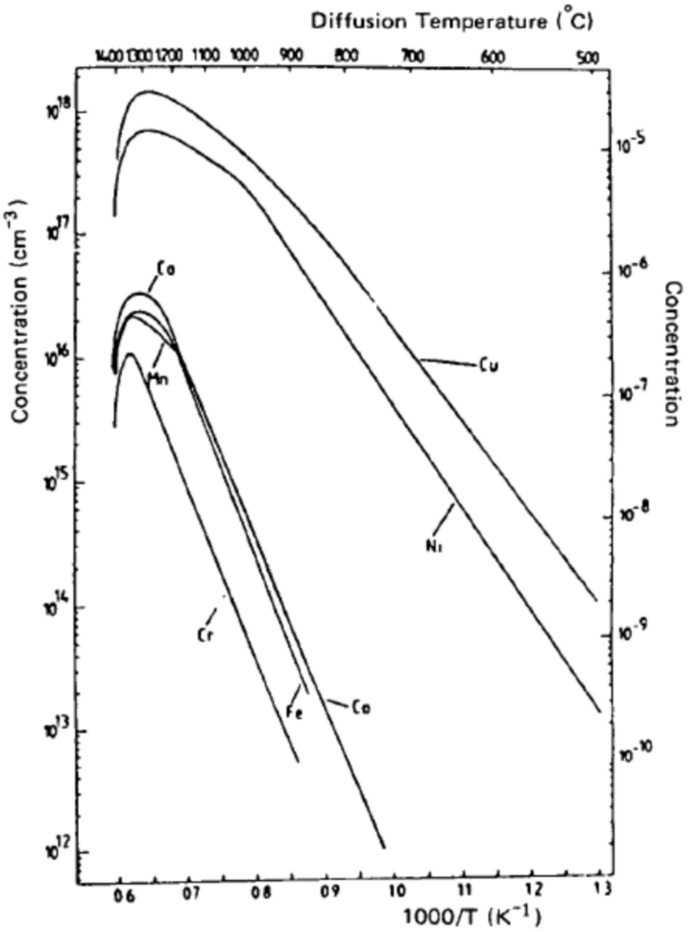


Fig. 4.14. Solubilities of transition metals in silicon [29].

Dopant impurities. When introduced into a silicon crystal, impurities from the third column of the periodic table (such as boron, aluminum, or gallium) or impurities from the fifth column (such as phosphorus, antimony, or arsenic) occupy substitutional sites and control the electrical characteristics of silicon. These impurities are intentionally introduced in concentrations ranging from less than 2 parts per billion atomic (ppba) ($<10^{14}$ atoms/cm^3) to more than 200 parts per million atomic (ppma) ($> 10^{19}$ atoms/cm^3), depending on the requirements.

Oxygen. In silicon pulled from melts contained in quartz crucibles by the modified Czochralski technique, oxygen occurs in concentrations ranging from 15 to 30 ppma, as measured by charged particle activation analysis [31]. The oxygen concentration in silicon grown by the float-zone technique is very much lower, of the order of 20 ppba or less; in crystals pulled from the melt in a magnetic field (Magnetic Czochralski, or MCz, growth) the oxygen concentration may be low, <5 ppma, or very high, >35 ppma, depending on the growing conditions. Oxygen atoms dissolved in silicon predominantly occupy interstitial sites. Two infrared absorption bands, one at 1107 cm^{-1} and the other at 514 cm^{-1}, are associated with the interstitial oxygen impurities in silicon. The first of these is widely used for quantitative analysis of oxygen in silicon. Oxygen plays three important roles in silicon: thermal donor formation; precipitate formation and intrinsic gettering; and mechanical strength of the crystal.

Carbon. Like silicon, carbon has four valence electrons, but it is considerably smaller. It may occur in small concentrations (<0.5 ppma) in all forms of silicon and sometimes occurs in higher concentrations. In substitutional sites, it is electrically neutral, but it may complex with other atoms. An infrared absorption band at about 605 cm^{-1} on the edge of the very large two-phonon silicon lattice absorption band is associated with carbon. There is general agreement that carbon is involved in the precipitation of oxygen in silicon, but there is considerable controversy concerning the responsible mechanism(s).

It was demonstrated that carbon in excess of 0.6 ppma plays a major role in oxygen precipitation at temperatures below 750°C and point out that such precipitates may dissolve at temperatures above 1150°C. Carbon is also shown to be involved with “new donor” formation. Carbon interacts strongly with oxygen, which is introduced into float-zone silicon in significant quantities during the longtime, high-temperature diffusion cycles associated with power device fabrication, forming defects in the material which may adversely affect device properties. On the other hand, it was conclude that the presence of relatively large amounts of carbon is not harmful to integrated circuits and other shallow devices which are fabricated for shorter times and at lower temperatures. Even when such devices are fabricated from Czochralski material, there is usually an oxygen-free region (denuded zone) near the surface where the devices exist. However, it was recognized that carbon may affect the bulk precipitate formation which in turn may affect device properties. Carbon has also been identified as the defect which pairs with group III acceptor centers to form complexes called x levels [32], and interactions with metallic impurities have been reported.

Hydrogen. For some time, hydrogen has been known to affect the properties of the interface between silicon and silicon dioxide [33]. More recently, hydrogen has been shown to passivate donor [34] and acceptor [35] dopant species, oxygen-related thermal donors [36], and chalcogen double donors [37]. Hydrogen is a very rapid diffuser and can penetrate deeply into a silicon wafer even at very low temperatures [38]. Dry etching processes can frequently result in introduction of hydrogen [39] which can permeate the surface regions of the wafer [40]. Since hydrogen in silicon is the subject of intensive investigation at the present time, greater understanding of its properties in silicon can be expected to be forthcoming.

Nitrogen. Unlike the other group V donor impurities, nitrogen appears to have almost no electrical activity. Its solubility is between 20 and 200 ppba. Several infrared absorption bands are associated with nitrogen in various states [41, 42]. Even in very low concentrations, nitrogen serves as an effective means for strengthening the crystal, apparently through dislocation locking of nitrogen-oxygen complexes both in Czochralski crystals grown from silicon nitride crucibles [43] and in nitrogen-doped float-zone crystals [42]. Typical concentrations of nitrogen in doped crystals range from 40 to 150 ppba.

Transition metals. Another important class of atomic impurities is the transition metal group which includes titanium, vanadium, chromium, manganese, iron, cobalt, nickel, and copper. The properties of these impurities tend to vary systematically with increasing atomic number. They tend to form complexes with dopants, especially boron [29, 44]. Some also precipitate out as silicides [45]. All these impurities are fast interstitial diffusers but some are faster than others. In equilibrium at high temperatures, they are in interstitial sites, but whether they can be quenched into interstitial sites at lower temperatures or whether they precipitate out depends on the relative values of the solubilities and diffusion coefficients. If these impurities are present in large enough concentration, they can affect the electrical conductivity, but they are more important as recombination centers. In general, these impurities are considered to be undesirable in silicon; direct impact of the presence of iron on integrated circuit yields has been demonstrated [46]. An elegant technique for detecting iron in silicon was developed in connection with this work [47]. Recently, flat band shifts in MOS capacitors with about 800 A thick oxides were correlated with the presence of iron contamination in the 5% HF solution used for preoxidation clean [48]. Subsequent work has confirmed that iron surface contamination degrades the properties of MOS device structures [49], but this work found that while iron ions in 1% HF solutions do not adhere to the wafer surface, the iron concentration on the surface is directly proportional to the iron concentration in HNO_3 solutions. In both studies, the iron was introduced as ferric nitrate; the only apparent difference is the dilution of the HF solution. More recently, new methods based on total reflection X-ray fluorescence (TXRF) have been reported for analysis of metallic impurities on silicon surfaces [50]. The sensitivity of these methods varies with atomic number of the impurity but is of the order of 10^9 to 10^{11} cm^{-2} for a variety of transition metals. The analysis depth is of the order of 3 nm.

Other atomic impurities. Other impurities sometimes appear in silicon in significant quantities. Gold is widely used to control carrier lifetimes in power devices and high speed integrated circuits. Silver and platinum have also been used for this purpose. Lithium is used to form very high resistivity silicon for nuclear particle detection. The group VI series, sulfur, selenium, and tellurium have been investigated extensively because of its scientific interest [51]. Germanium is completely miscible in silicon; germanium-silicon films have been used to introduce misfit dislocations to provide an extrinsic gettering mechanism in epitaxial silicon wafers [52]. Such alloy films open new doors to defect engineering to control the properties of active multi-layer structures fabricated on silicon substrates [53].

Extended Defects. Dislocations are line defects which extend over many atomic spacings within the crystal. Stacking faults, localized regions in which the stacking order of the atoms in the crystal is disturbed, are bounded by partial dislocations. Precipitates are inclusions of other materials within the crystal structure.

Dislocatioirs aird slip. Dislocations arise when a portion of the crystal is plastically deformed and slips across another portion. In silicon, as in all diamond cubic crystals, the slip plane between the two portions of the crystal is a (111) plane. The stresses in the crystal cause atomic planes to move from one equilibrium position to another. In some cases, there is an extra plane of atoms on one side of the slip plane; the line of atoms at the end of this extra half plane is called an edge dislocation, as shown in Fig. 4.15. The dislocation may move through the crystal along the direction of the shearing forces. Under some conditions, the crystal may twist. In this case, there is no extra half plane of atoms. The dislocation, known as a screw dislocation, occurs along the line where the planes above and below the slip plane return to their original relationship, as shown in Fig. 4.16. Screw dislocations move through the crystal in the direction perpendicular to the shearing forces.

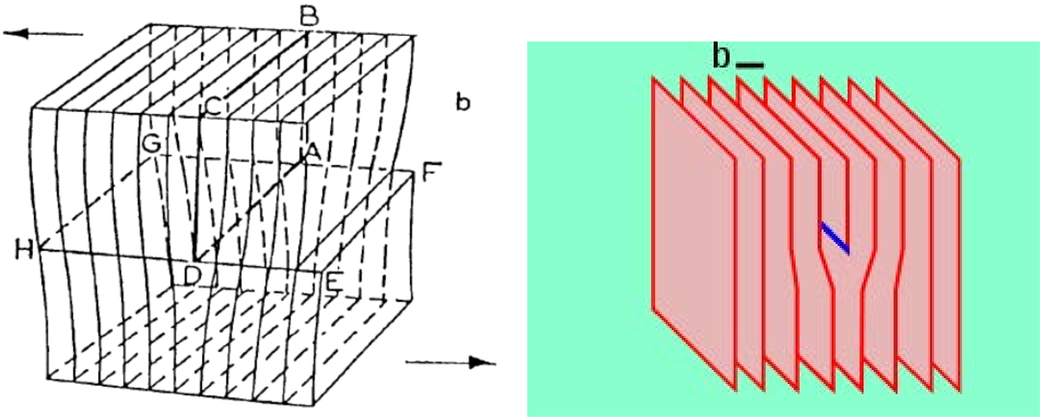


Fig. 4.15. Edge dislocation denoted by AD in the slip plane EFGH [54].

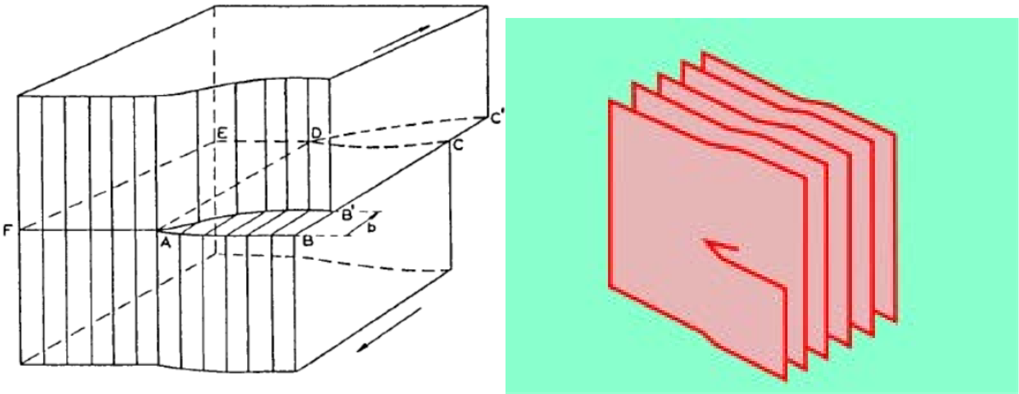


Fig. 4.16. Screw dislocation denoted by AD in the slip plane FBCE [54].

Dislocations in real crystals may be pure edge, pure screw, or, more commonly, a combination of both. Dislocations are characterized by a vector known as the Burgers vector which is the closure failure between a circuit around the dislocation and a similar circuit in a perfect crystal [55]. This vector, denoted by the symbol ***b***, is perpendicular to an edge dislocation and parallel with a screw dislocation. A dislocation cannot end within a single crystal except at another dislocation; thus dislocations form closed loops or interconnecting networks [54]. The correspondence between the intersection of a dislocation line with the surface and the etch pit formed by using a preferential etch has been demonstrated by the use of copper decoration and direct imaging of the dislocations [56]. A standard method is available for developing and counting dislocation etch pits [57]. In addition a variety of X-ray techniques can also be used in studying dislocation characteristics [54].

A procedure for growing crystals free from dislocations has been developed [58]. This is accomplished by necking down a small seed so that the few dislocations which might remain grow out to a surface and then shouldering the crystal out to the desired diameter very carefully to avoid introducing new dislocations. All commercially available silicon crystals are today grown using this method. It should be noted that in dislocation-free crystals, the vacancy supersaturation may lead to undesirable aggregation in the form of voids or stacking faults [59]. Even in dislocation-free crystals, slip may be introduced during crystal growth, generally at the bottom of the crystal, due to excessive thermal stresses. In addition, slip may occur during thermal processing of wafers, particularly at points where the wafer is confined such as boat resting points. Slip may also be introduced during epitaxial layer growth. ASTM Standard Test Method F 47 [57] covers procedures for identifying slipped regions of crystals and wafers. Slip is characterized by lines of etch pits along <110> directions.

Stacking faults. Stacking faults may occur during epitaxial layer growth, although modern methods have greatly reduced their incidence. Stacking faults are also induced by oxidation, either at the surface during device processing or in the bulk during oxide precipitate growth. Since silicon dioxide occupies more volume than the silicon consumed in its formation, an excess of silicon interstitials forms in the neighborhood, and if suitable nuclei and stress fields are present, stacking faults may form. A twin is a portion of a crystal which contains atoms stacked in one order (e.g., ABCAB) on one side of a boundary (the twin plane) and in the mirror image order (e.g., BACBA) on the other. Twins form in dendritic silicon crystals but are rarely seen in commercial crystals pulled from the melt.

Precipitates. If impurities are present in concentrations larger than their solubility at a given temperature, they may precipitate, usually as compounds of silicon, provided that suitable nuclei are present. These nuclei may arise from the presence of other types of defects (heterogeneous nucleation) or from tiny aggregates of the same atoms which form the precipitate (homogeneous nucleation) [60]. Precipitates may form during the cool down cycle after growth or during subsequent processing. Other defects, such as dislocation loops and stacking faults, sometimes form as a result of precipitation. ASTM Standard Test Method F 416 [61] covers detection of precipitates and other oxidation induced crystal defects by means of preferential etching. Identification of the chemical nature of precipitates and of precipitate morphology requires the use of transmission electron microscopy.

4.3.3. Bands and Bonds in Pure Silicon Crystals.

A silicon atom has 14 electrons. These are distributed two each in 1s and 2s states, six in 2p states, and three each in 3s and 3p states. When silicon atoms are brought together to form a crystal, the discrete levels are broadened into bands so that all the electrons originally in a given energy state have energies slightly different from each other. Because there are so many atoms per unit volume, these bands may be regarded as continuous. In the crystal, the 3s and 3p states intermingle to form two bands separated by a large energy gap without energy states. There are four quantum states per atom in the lower band, called the valence band, and four quantum states per atom in the upper band, called the conduction band. The gap between these bands is called the forbidden energy gap, or simply the forbidden gap. As noted previously, each atomic site in silicon is surrounded by four nearest neighbor sites which occur at the points of a tetrahedron. The chemical bonds in silicon are formed between nearest neighbors, each atom of the pair contributing one electron. As a result, each atom is surrounded by eight electrons, four of its own and one from each of its nearest neighbors. Such a configuration, which is similar to that of a noble gas, is chemically very stable. All of the available electrons have places in the valence bonds and none are free to move in the lattice unless some energy is applied to break the bond. Another way of looking at the situation is that all the electrons occupy all of the energy states in the valence band. For an electron to move, it must gain enough energy to be elevated across the gap to an energy state in the conduction band. The energy required may be supplied from a photon of energy equal to or greater than the forbidden energy gap (photoconductivity) as shown schematically in Fig. 4.17 or from thermal energy if the temperature is sufficiently high. Electrons which leave the broken bonds enter states in the conduction band and are free to move throughout the crystal lattice. The broken bonds left behind are free to accept electrons from neighboring bonds, and as a result the empty states in the broken bonds, known as "holes", also seem to move throughout the lattice. Note that equal numbers of holes and electrons are produced when electrons are excited across the forbidden gap.

Dopant Impurities. So far, we have considered pure silicon without impurities or defects, and we have assumed that each lattice site is occupied by one and only one silicon atom. Suppose now that a phosphorus atom is substituted for a silicon atom as shown schematically in Fig. 4.18. Phosphorus is in the fifth column of the periodic table and has five valence electrons. Four of these electrons contribute to the valence bonds in the same way as the four valence electrons of the replaced silicon atom.

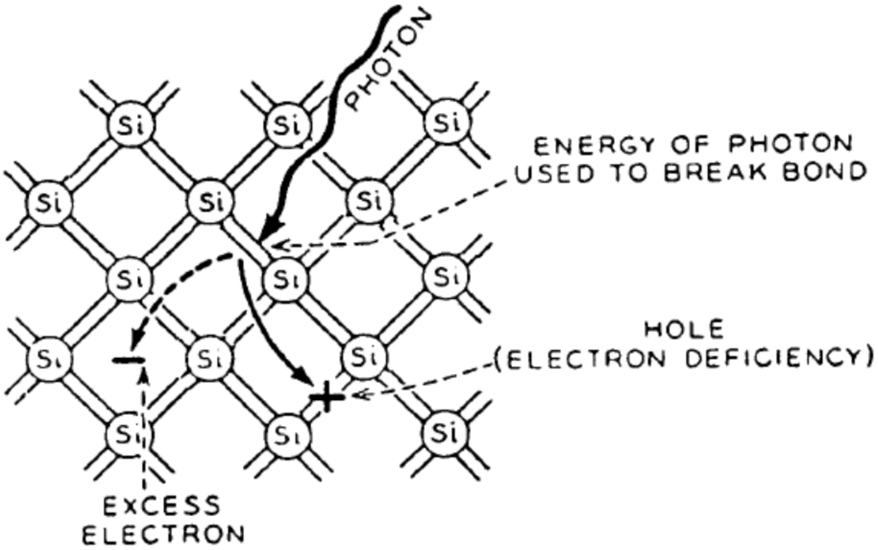


Fig. 4.17. Two dimensional representation of the covalent bonds between adjacent silicon atoms in an intrinsic (pure) silicon crystal showing the generation of a hole-electron pair by a photon of energy greater than the forbidden energy gap [9].

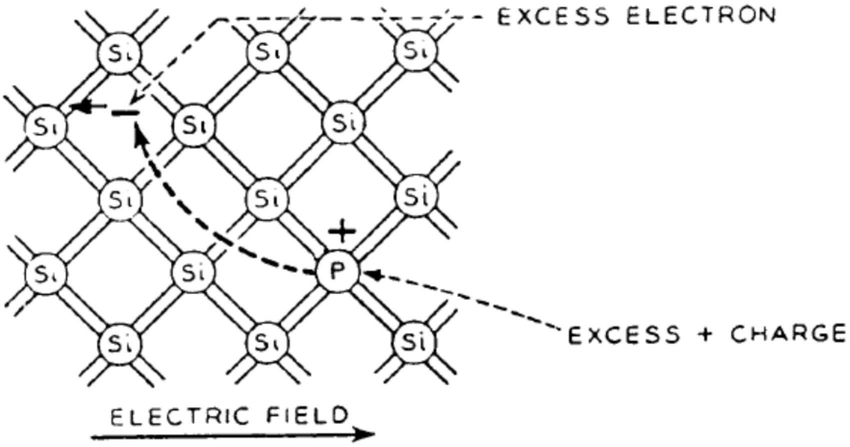


Fig. 4.18. Two dimensional representation of the covalent bonds between adjacent silicon atoms in a silicon crystal doped with phosphorus, a donor impurity, showing the generation of a free electron by thermal ionization of the donor impurity [9].

The fifth electron is not a part of the valence bond structure. It is attached to the phosphorus atom only by the electrostatic attraction of the positive charge of the nucleus. The energy required to free this electron is much smaller than that required to break a valence bond. Because one of its electrons can easily be freed to move about in the crystal, phosphorus is known as a donor impurity. In an energy band diagram, the phosphorus atom is represented by an energy level in the forbidden gap very close to the bottom of the conduction band. The energy difference between this level and the bottom of the conduction band is the energy required to free the electron from the phosphorus atom; this energy is known as the donor activation energy. Because mobile electrons have a negative charge, silicon containing phosphorus or another group V impurity is called n-type. If on the other hand a silicon atom is replaced by a boron atom, or another atom from column III of the periodic table, the impurity atom contributes only three valence electrons. These three electrons form valence bonds with three of the neighboring silicon atoms. In the fourth bond there is a missing electron, or hole. An electron from a neighboring bond may move into this hole with very little expenditure of energy, leaving a hole behind, as shown schematically in Fig. 4.19. The energies required for this are similar to those required to free an electron from a donor atom. Because they accept electrons from the valence band, impurities with three valence electrons are called acceptor impurities. An acceptor impurity is represented by an energy level in the forbidden gap near the top of the valence band; the energy difference is known as the acceptor activation energy. Since mobile holes behave similarly to "electrons" with a positive charge, silicon which contains acceptor impurities is called p-type.

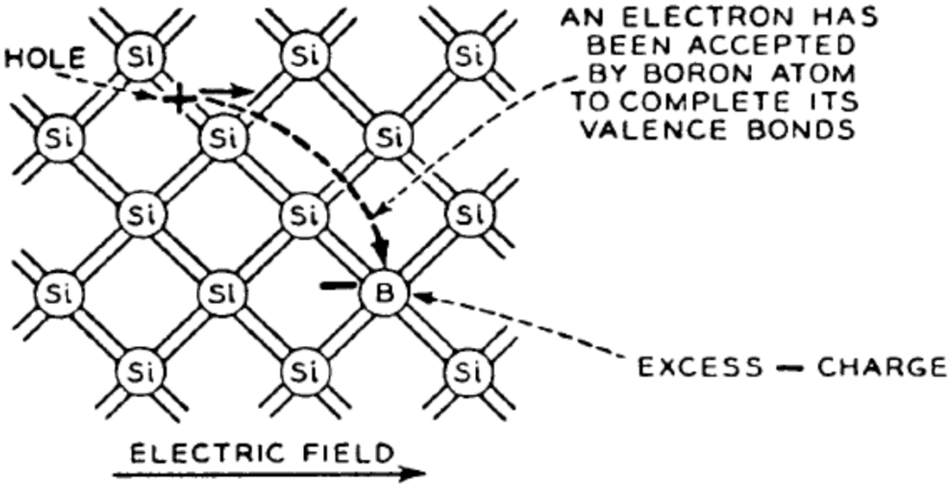


Fig. 4.19. Two dimensional representation of the covalent bonds between adjacent silicon atoms in a silicon crystal doped with boron, an acceptor impurity, showing the generation of a free hole by thermal ionization of the acceptor impurity [9].

4.4. Polycrystalline Silicon

Thin films of polycrystalline silicon have been used for a decade and a half as gate electrodes of metal-oxide-semiconductor (MOS) integrated circuits. The ability to self-align the gate electrodes of MOS field-effect transistors reduces capacitance and improves circuit speed. In addition to its use in conventional MOS integrated circuits, polysilicon is being used in new applications which demand very stringent control over its material properties; such advanced applications include high-value resistors in static RAMS, floating gates in electrically alterable ROMs, and emitters of bipolar transistors. One or even two layers of polysilicon are now used in virtually all advanced bipolar integrated circuits also, improving circuit performance. In addition to improving circuit speed, the compatibility of polycrystalline silicon with subsequent high-temperature processing allows its efficient integration into advanced integrated-circuit processing and permits fabrication of new device structures.

This excellent technological compatibility of polysilicon with integrated-circuit processing allows straightforward fabrication of convenient interconnections in VLSI circuits. Although a resistivity of less than $l0^3$ Ω cm, eight orders of magnitude less than that used for static RAM load resistors, is routinely achieved, the lower bound on the resistivity of polysilicon can limit the performance of integrated circuits which use polysilicon interconnections to conduct signals long distances across a chip. Therefore, alternate materials are being incorporated into integrated-circuit interconnection structures to replace or augment polysilicon. In the future, interconnections will limit integrated-circuit performance even more severely.

Continued advances in the design of integrated circuits and the use of polysilicon in novel structures have brought into focus the need for a greater understanding of the detailed characteristics of polycrystalline-silicon films and their application in integrated circuits.

4.4.1. Deposition

Over the many years that polysilicon has been used in commercial integrated circuits, the equipment used for its deposition has changed greatly, evolving from low-capacity, silicon epitaxial reactors operating at atmospheric pressure to high-volume, low-pressure CVD systems. In all commercial applications, however, chemical vapor deposition has been used to deposit polysilicon. In early prototype investigations, physical vapor deposition by evaporation was studied, but the step coverage provided by this technique was inadequate to cover the irregular surface topology of the integrated circuit; by contrast, CVD techniques provided material with excellent conformal step coverage, leading to the rapid acceptance of this more complex formation technology.

At the time that chemical vapor deposition was first used for polysilicon deposition, the most common CVD system was the horizontal, atmospheric-pressure epitaxial reactor widely used in the late 1960s and early 1970s. This type of system (illustrated in Fig. 4.20), allows operation over a wide temperature range, but its capacity is severely limited by the size of the susceptor, on which the wafers are placed in a single layer. The low-pressure CVD (LPCVD) reactor developed to overcome the limited capacity of the horizontal system can form layers on 100-200 wafers simultaneously, although the range of conditions over which it can operate satisfactorily is severely limited. The basic elements of the LPCVD reactor are illustrated in Fig. 4.21.

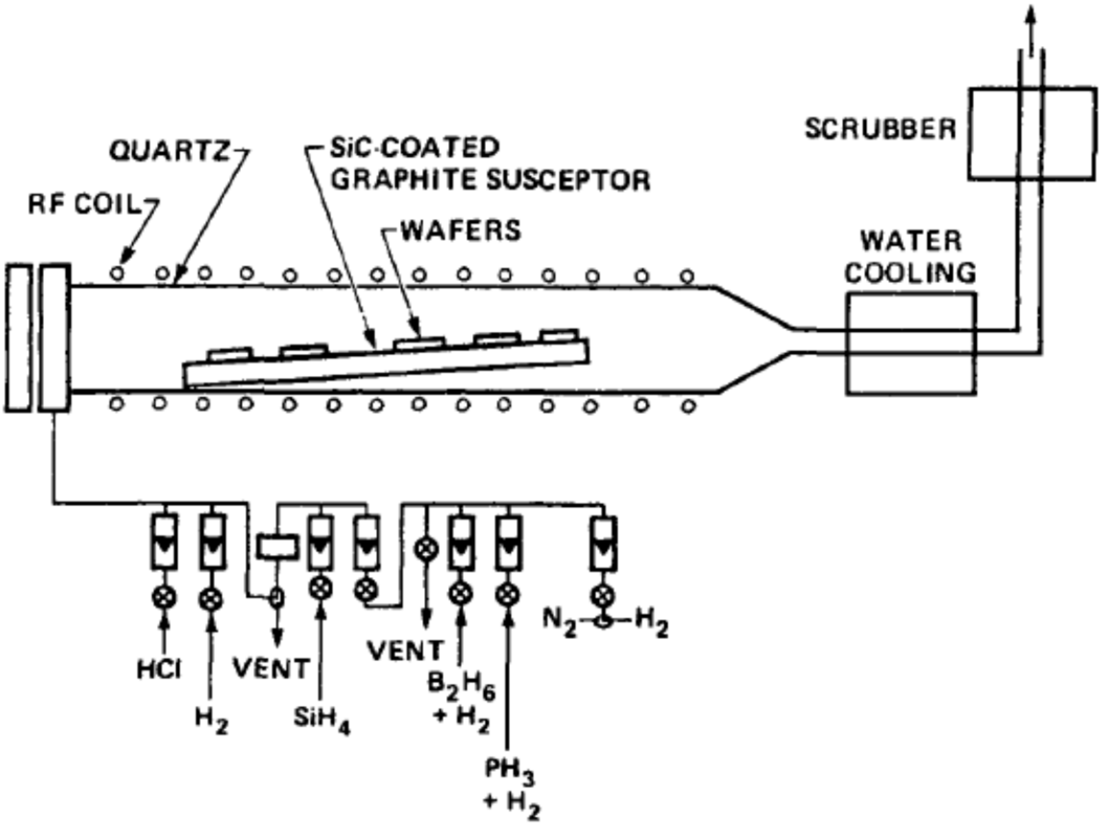


Fig. 4.20. The versatile, horizontal, cold-wall, atmosphericpressure reactor has the flexibility needed to develop many new CVD processes, but its limited wafer capacity makes it less desirable for routine manufacturing.

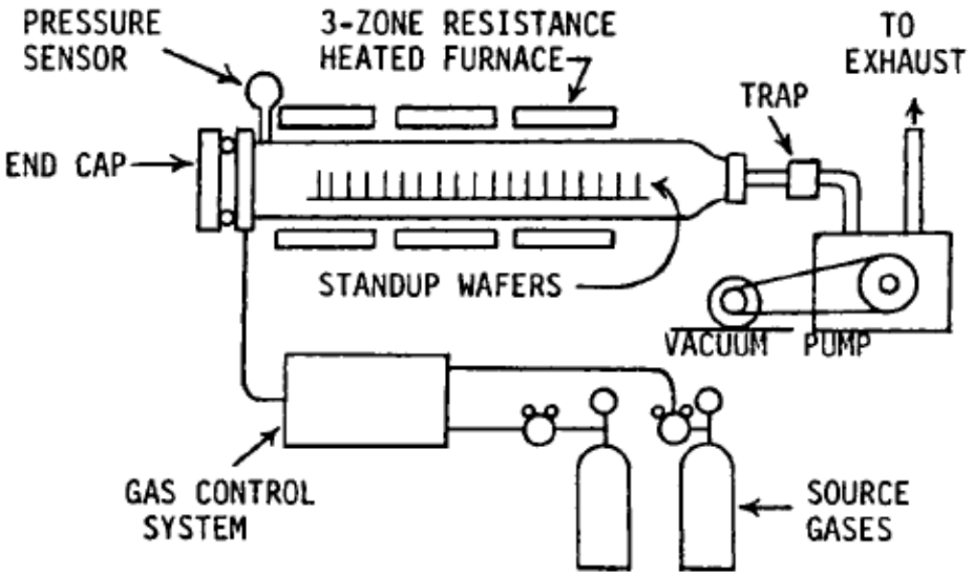


Fig. 4.21. The hot-wall, low-pressure reactor is used for routine deposition of polysilicon beof its high wafer capacity and simplicity.

Horizontal, Atmospheric-Pressure Reactor. In the horizontal, atmospheric-pressure, reactor, a small quantity of the silicon-containing gas is mixed with a large amount of carrier gas, and this gas mixture is forced through the reactor. Only the wafers and their supporting susceptor are heated; the walls of the chamber remain relatively cool and little deposit forms on them. The forced convection region is also relatively cool so that gases flowing in this region can travel long distances along the susceptor without reacting significantly. The majority of the temperature gradient occurs across the boundary layer. In this type of reactor, energy is often coupled into the susceptor by radio-frequency induction or by heating the opaque support plate with lamps without significantly heating the walls of the chamber. (The term "susceptor" is often used to describe the supporting plate on which the wafers sit even when heating methods other than rf-induction are used.) In either case, the temperature of the wafer can only be controlled within about 5 or 10°C, which would lead to unacceptable deposition rate and thickness variations if the reactor were operated in the reaction-limited regime. On the other hand, the gas flow, and especially diffusion through the well-defined boundary layer, is relatively well controlled, and operation of this reactor in the mass-transport-limited regime is preferred.

Although the horizontal reactor is very flexible and can operate over a wide temperature range because the forcedconvection flow region remains cool, its wafer capacity is severely limited by the size of quartz envelopes available and the large area needed when wafers are placed in a single layer on the susceptor surface.

Low-Pressure Reactor. To overcome the capacity limitations of the horizontal reactor, the high-capacity, low-pressure reactor was developed. In this reactor the wafers are placed in a resistance-heated furnace similar to an oxidation or diffusion furnace and are closely spaced, generally perpendicular to the axis of the tube. Figure 4.22 shows that, in this reactor, the gases flow first through the annular space between the chamber walls and the wafers. They then move between the closely spaced wafers to the wafer surfaces, where the reaction occurs. If mass transport limited the overall reaction, the deposition rate would be markedly higher near the readily accessible edges of the wafers. Therefore, in contrast to the horizontal reactor, this reactor must be operated in the reaction-limited regime. Fortunately, the temperature of the resistance-heated furnace used to heat this reactor can readily be controlled to a fraction of a degree so that the rapid variation of deposition rate with temperature in this operating regime does not degrade control of the deposited film thickness.

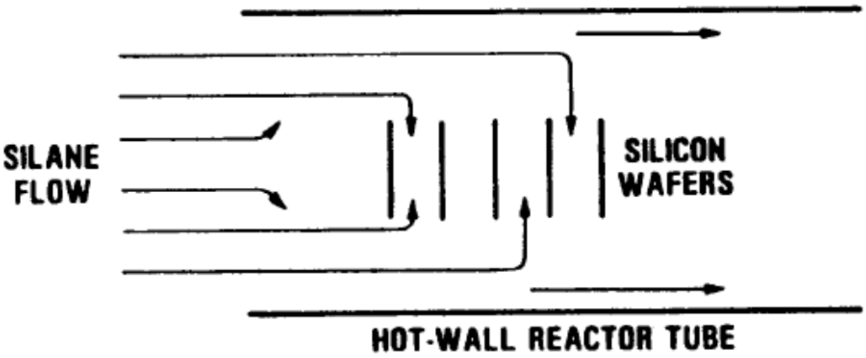


Fig. 4.22. The narrow space between wafers in the lowpressure reactor makes gas diffusion to the centers of the wafers difficult.

Because of the long, narrow space through which the reactants must diffuse, obtaining reaction-limited operation requires considerable effort. Lowering the temperature alone is not adequate; the ease of mass transport must also be enhanced. As we have already seen, mass transport can be characterized by the ratio D/S, where D is the gas-phase diffusivity and S is the boundary-layer thickness or another characteristics dimension of the deposition system. The rapid variation of the diffusivity with pressure is the key to achieving surface-reaction-limited operation in this type of reactor. The diffusivity is inversely proportional to the total pressure; reducing the pressure by a factor gf several thousand to a fraction of a torr (tens of pascals) increases the ease of gas-phase diffusion by a similar amount. Making the mass transport easier by operating at reduced pressures thus moves the deposition process into the reaction-rate-limited regime, as desired. To maximize the deposition rate, the partial pressure of the silicon-containing gas must be comparable to the total pressure, and little or no carrier gas is used in this reactor. Typical deposition conditions are shown in Table 4.2.

Table 4.2. Typical plysilicon dposition conditions.

	Atmosheric Pressure Reactors	Low Pressure Reactors	
Temperature (°C)	950	620	640
Silane partial pressure (torr)	0.3	0.2	0.2
Total pressure (torr)	760	0.2	1.0
Deposition Rate (nm/min)	120	10	15
Wafer capacity	20	100	100
Carrier Gas	H_2	None	N_2 or H_2
Throughput (wafers/hour)	40	100-150	100-150

* 1 torr = 133 pascals (Pa); 1 atm = 1.013×10^5 Pa.

One significant difference between the hot-wall LPCVD system and the cold wall reactor is the temperature of the gases in the forced-convection region. Because the gases are heated significantly in the forced-convection region of the hot-wall reactor, they may decompose or react in this region, leading to particles on the wafer surface or a porous film, as well as changing the deposition rate. Gas-phase decomposition is also promoted by the limited amount of carrier gas available to dilute the reactants and decrease the reaction probability. For example, a hydrogen carrier gas can suppress the thermal decomposition of silane because it is one of the products of the reaction

$$SiH_4(g) \rightarrow Si(s) + 2H_2(g) \qquad (4.5)$$

The presence of hot gases in the forced-convection region is even more deleterious when silicon compounds are to be deposited in the hot-wall reactor. One gas may tend to decompose before the desired reaction between different gases occurs, limiting the variety of gases that can be effectively used. For example silane decomposes more readily when it reacts with ammonia in the reaction and the less easily decomposed, silicon-containing gas dichlorosilane (SiH_2C1_2) is generally used in place of silane, leading to other problems.

4.4.2. Structure

The polycrystalline-silicon films are now almost exclusively deposited in LPCVD reactors at a temperature near 625°C, although in the past they were often deposited at higher temperatures in horizontal, atmosphericpressure systems. The operating regime is chosen to provide the greatest control over the deposition process in the particular reactor being used.

Nucleation and Surface Processes. In our previous discussion we focused our attention on the processes occurring in the gas phase because these processes depend strongly on the reactor used for the deposition; we also considered the overall chemical reactions. Now, we want to look at the processes occurring on the surface during and after nucleation because these processes influence the structure of the material formed.

As the silicon-containing species approaches the wafer surface, it may react in the gas phase, especially when the temperature in the boundary layer is high and the silicon containing gas reacts at a low temperature. The reactions may form intermediate species, such as SiH_2 or a series of reactions may proceed to form elemental silicon. In the latter case, silicon atoms may agglomerate into small silicon grains in the gas phase. These grains then migrate to the wafer surface where they form an irregular, porous deposit, as well as loosely adherent particles. This gas-phase or homogeneous reaction does not form the dense, uniform films needed for integrated-circuit applications and is usually suppressed in favor of heterogeireous reactions, which occur on the surface. In the case of heterogeneous reaction, some initial portions of the overall reaction may occur in the gas phase, especially at atmospheric pressures and higher temperatures [62-64], but the final stage of the reaction forming elemental silicon occurs on the wafer surface. The adsorbed silicon species or its precursors diffuse on the surface, and the structure of the final film depends strongly on the amount of surface migration possible before the adsorbed silicon atoms are covered and immobilized by subsequently arriving silicon atoms.

The structure is, of course, also affected by the nature of the underlying substrate. If it is crystalline and atomically clean, the diffusing silicon atoms see an array of relatively low-energy sites at which it is favorable for them to locate. If they have enough thermal energy to migrate on the surface, they preferentially locate over these low-energy sites. If enough energy is available, the ordering is complete, and a single-crystal or epitaxial layer forms. At lower temperatures, the diffusing adsorbed surface atoms do not have enough energy to arrange themselves over the low energy sites, and the epitaxial relationship is lost. Random nuclei form on the wafer surface, and these nuclei join to form a polycrystalline deposit. Surface contamination can also hinder surface diffusion so that the diffusing atoms cannot reach low-energy sites, again preventing formation of a single-crystal layer. If the contamination is irregular, some very-low-energy sites may be available to attract diffusing atoms from the surrounding regions, and thick deposits may form at these locations, leading to defects in the films. Because the excess material near the defect is supplied by diffusion from the surrounding regions, the film may be thinner in the surrounding region than in the regions distant from the defect.

In many applications, however, the surface on which the polysilicon film is to be deposited is amorphous, rather than crystalline, and initial nucleation of the silicon film is quite different. Silicon atoms adsorbed on a perfectly clean, amorphous surface diffuse randomly and have a greater opportunity to re-evaporate (*desorb*) because they are less firmly bound to the surface. As the adsorbed atoms diffuse on the surface, they may encounter other diffusing atoms and form a pair. This atom pair is more stable than is an isolated atom and is less likely to desorb. As it diffuses on the surface, it may join other atoms, forming a larger, more stable cluster, until the cluster has a low probability of desorbing, and a critical cluster or stable nucleus is formed. Because the probability of diffusing atoms encountering each other depends strongly on the number of adsorbed atoms on the surface, it is a strong function of their arrival rate (through the partial pressure of the silicon-containing gas) and their desorption rate (through the temperature of the substrate and the binding energy of the diffusing atoms to the exposed surface). Because of the dependence on the binding energy, stable deposits may form on one surface while not forming on another surface under the same deposition conditions.

After stable nuclei form, additional adsorbed atoms diffusing on the surface can either initiate additional nuclei or join existing nuclei. When the existing nuclei are close enough together, additional atoms are more likely to join an existing nucleus, and the number of nuclei saturates and remains constant as the size of each nucleus grows (Fig. 4.23) [65]. The saturation number of nuclei depends on the substrate, the arrival rate of atoms, and the temperature. At higher temperatures, adsorbed atoms, can diffuse greater distances to join an already formed nucleus rather than initiating a new nucleus, and fewer, but larger, nuclei form (Fig. 4.24) [66]. Thus, more silicon atoms must be deposited before a continuous film forms as the nuclei impinge on each other. Because the individual nuclei are randomly oriented on the amorphous surface, discontinuities form when two differently oriented or differently aligned nuclei join. These discontinuities (graiit boundaries) strongly influence the properties of the polycrystalline-silicon film.

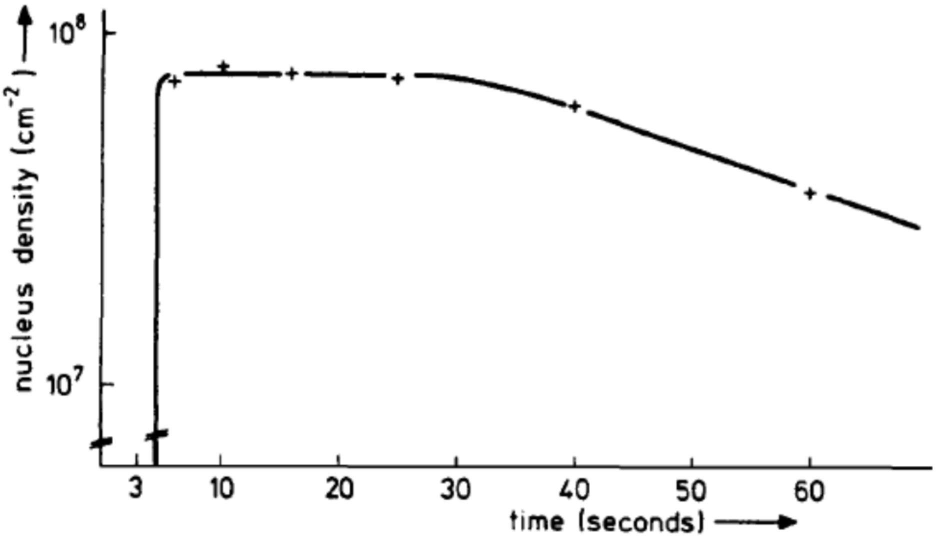


Fig. 4.23. The density of stable nuclei as a function of exposure time to SiH_4, showing the incubation period, the rapid increase to a steady state value, and coalescence [65].

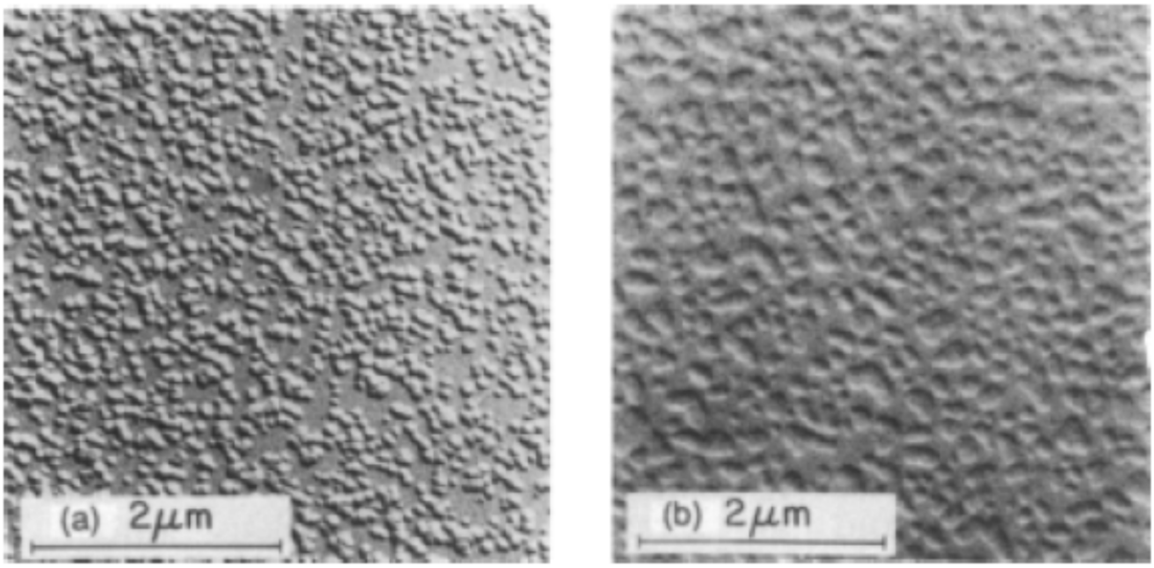


Fig. 4.24. Replica micrographs of films 14 nm thick deposited at (a) 850°C and (b) 1025°C in an atmospheric pressure reactor. At higher temperatures more surface diffusion can occur and fewer, but larger, nuclei are formed.

Even after a continuous film forms, the structure is strongly influenced by the thermal energy available for surface migration. At very low temperatures, the adsorbed silicon atoms have little thermal energy and cannot diffuse significantly on the substrate surface before they are covered by subsequently arriving silicon atoms. Once they are covered, their random arrangement is locked into place, and an amorphous structure with no long-range order forms. (The amorphous silicon films of interest to us in this chapter are formed by CVD at a temperature slightly below the temperatures at which polysilicon is deposited; they are dense, contain little hydrogen, and differ markedly from the porous, amorphous films deposited at much lower temperatures by sputtering or glow-discharge techniques. We will discuss these higher-temperature CVD amorphous films further below.) At higher temperatures, adequate surface diffusion is possible to allow a crystalline structure to form. On an amorphous substrate, or even on a crystalline substrate at intermediate temperatures, the deposited film is polycrystalline. At high temperatures on a single-crystal substrate, the individual nuclei may each align themselves with the substrate crystal so that no grain boundaries form when the individual nuclei grow together, and an epitaxial layer forms.

Although temperature is the most important variable controlling the structure, pressure and deposition rate also affect the structure through their effect on surface diffusion, as discussed above. Just as surface migration decreases when less thermal energy is available at lower substrate surface temperatures, it also decreases when more of the available adsorption sites are filled by other (non-silicon) adsorbed atoms. As the arrival rate of silicon atoms increases, the amount of time during which the adsorbed atoms can diffuse on the surface before they are immobilized by subsequently arriving atoms decreases, and the amount of migration is again limited.

Each of these variables can modify the ordering of the growing grains, and they are consequently interrelated. For example, the temperature at which a polycrystalline structure, rather than an amorphous structure, forms during deposition is lower at lower total pressures because less carrier gas is adsorbed on the surface to impede rearrangement of the atoms diffusing on the surface. At atmospheric pressure, the transition between the deposition of amorphous and polycrystalline silicon occurs at about 680-700°C; in an LPCVD reactor at a typical pressure of about 0.2 torr, the transition temperature is about 580-600°C. Similarly, a given structure is formed at lower temperatures as the deposition rate decreases so that the adsorbed atoms can diffuse for a longer time on the surface before they are immobilized. (This effect is limited, however, by impurities. At very low silicon arrival rates, impurities can be the dominant species adsorbed on the surface, where they impede the surface migration of silicon atoms. Impurities are especially harmful in systems which are not sufficiently controlled for purity and vacuum integrity.) Within the temperature range where polycrystalline films are deposited, the structure of the polycrystalline film is also dominated by the amount of surface migration possible during deposition.

4.4.3. Application

The development of polysilicon technology was driven by the use of polysilicon as a gate electrode for integrated circuits. However, once it was developed, polysilicon technology found use in an increasing number of applications, not only in MOS integrated circuits, but in bipolar technology as well.

Silicon-Gate MOS Technology. In the mid-1960s most gate electrodes were made from aluminum, which was deposited after the source and drain regions were doped. The aluminum had to overlap onto these regions to insure that a continuous channel from source to drain was formed when the gate was biased to turn on the transistor. The required alignment tolerance caused a significant Miller feedback capacitance between the gate and the drain, which slowed circuit speed.

By using a material which allows the source and drain regions to be self-aligned to the gate electrode, as shown in Fig. 4.25, this capacitance can be markedly reduced (Fig. 4.25). Providing self-aligned interconnections not only reduces the Miller capacitance, but it also makes this capacitance more uniform from one device to another; this device matching is often critical for improved circuit performance. The material used for such a self-aligned structure needs to be compatible with the high temperatures required for the source and drain diffusions and other high-temperature processing steps which follow gate-electrode formation.

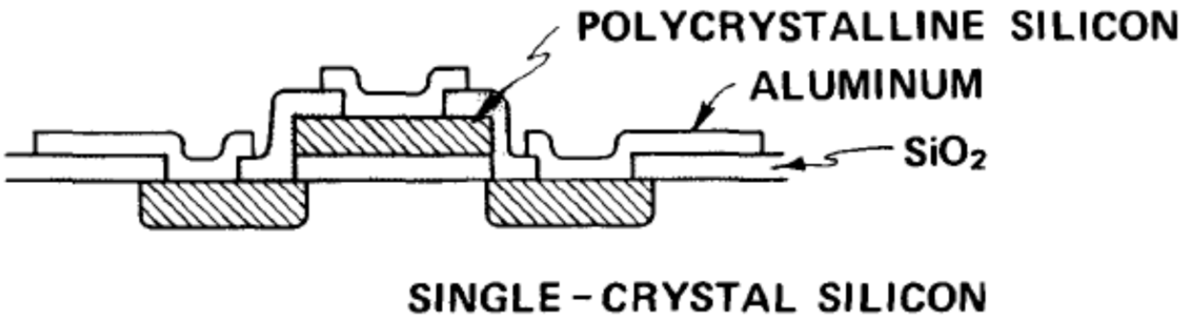


Fig. 4.25. In a silicon-gate transistor, the gate electrode can be used to define the location of the source and drain dopant atoms, thereby eliminating the need for a large overlap of the gate electrode. This self-aligned structure reduces the overlap capacitance.

Although other high-temperature materials, such as tungsten and molybdenum were investigated, the compatibility of polycrystalline silicon with integrated-circuit fabrication led to its rapid adoption [67]. In the initial attempts to use thin films of silicon for gate electrodes, the silicon layers were deposited by evaporation. However, these evaporated films could not adequately cover the steps on the integrated circuit surface between the gate oxide and the field oxide. At the time when polysilicon technology was being developed, formation of epitaxial silicon by chemical vapor deposition was widely used in bipolar integrated circuits, and it was recognized that polysilicon could be deposited by similar techniques if the substrate was an oxide-coated surface, rather than single-crystal silicon. Because of the readily available technology and the conformal step coverage of films produced by chemical vapor deposition, this method of forming polysilicon layers was quickly adopted.

One of the terms entering into the threshold-voltage expression of an MOS transistor is the difference between the work functions of the gate and the silicon (i.e., the "metal"-semiconductor work-function difference Φ_{MS}). When polysilicon is used as the gate electrode, the metal-semiconductor work-function difference depends on the dopant concentration in the polysilicon, as well as that in the silicon substrate. For maximum conductivity, the polysilicon is very heavily doped, and the Fermi level is generally assumed to be very near the edge of the conduction or valence band for n-type or p-type polysilicon, respectively. Thus, the threshold voltage changes by the band-gap of silicon (1.1 eV) when the dopant type in the polysilicon gate is changed. For the highest conductivity, the polysilicon is usually n-type; however, in CMOS circuits, it may be more convenient to dope the gate of the p-channel transistor p-type, especially if the dopant is added during the source-drain implantation. (In this case, a silicide or other low-resistance conducting material is usually placed over the polysilicon gate to lower the interconnection resistance and short the p^+n^+ diode that would otherwise form between the gates of p-channel and n-channel transistors.) Choosing the correct dopant type for the gate electrode can also produce the desired threshold voltage without using complex dopant profiles in the single crystal silicon, which can degrade transistor operation (e.g., by reducing the carrier mobility).

When the gate electrode is doped during the source drain implantation only, the dopant concentration may be lower than when a separate doping cycle is used, and the Fermi level may not be at the band edge. The flat-band voltage is usually comparable to that in devices with the polysilicon doped from the vapor phase [68], but more detailed investigation does reveal some differences [69, 70].

Not only using polysilicon gates allow self-aligning the source and drain regions to the gate electrode, but the polysilicon can also be used as an additional partial level of interconnections. In addition, placing the polysilicon interconnections over thick field oxide reduces the capacitance below that of diffused interconnections. The additional level of interconnections provided by polysilicon cannot be used as effectively as other levels, however, because the processing sequence of a typical silicon-gate integrated circuit prevents forming a heavily doped region of single-crystal silicon under the polysilicon line. However, even with this limitation, polysilicon interconnections provide additional flexibility in integrated-circuit layout. As integrated-circuits complexity increased, interconnecting devices became more critical, and the additional flexibility provided by polysilicon interconnections was of increasing benefit.

The ability of polysilicon to withstand subsequent high-temperature processing also allowed development of more reliable integrated-circuit fabrication processes. Coverage of surface steps by evaporated aluminum interconnections has always limited yield and reliability. When polysilicon gates with their ability to withstand high temperatures are used, a phosphorus-containing, silicon-dioxide layer can be deposited over the polysilicon and heated until it becomes viscous and "flows," reducing the abruptness of the steps which the aluminum must cover. This phosphosilicate glass (PSG) can be flowed before the contact windows are opened through the oxide to the underlying heavily doped source and drain regions. It can also be flowed ("reflowed") after the windows are opened to provide a more easily covered step at the edges of contact regions. In addition to its applications in normal silicon-gate structures, the high-temperature capability of polysilicon allowed development of more complex structures, which provided denser integrated circuits even with the same minimum feature size. For example, the size of the one transistor, dynamic random-access-memory cell can be reduced significantly by using two different layers of polysilicon-one as the gate of the access transistor and one as the counter electrode of the storage capacitor. Figure 4.26a [71] shows that the single-layer-polysilicon, dynamic RAM cell requires space between the gate and capacitor counterelectrodes. The space between these two polysilicon features is filled with a diffused region to prevent a barrier between the inversion layers under the electrodes and to allow charge transfer to the storage region when the access transistor is turned on. When two levels of polysilicon are used (Fig. 4.26b), the access transistor and the storage region are separated only by the thickness of the oxide grown on the side of the first level of polysilicon, rather than by the minimum horizontal feature size. Thus, the cell size is reduced appreciably, but at the expense of more complex processing. Although the electrodes are separated by the finite thickness of the oxide grown on the side of the first layer of polysilicon, fringing fields reduce the barrier between the two electrodes when both are biased to induce inversion layers in the substrate. Charge-coupled devices (CCDs) also use closely spaced, independent polysilicon electrodes above single-crystal silicon to allow charge transfer between adjacent inversion layers in a compact device structure.

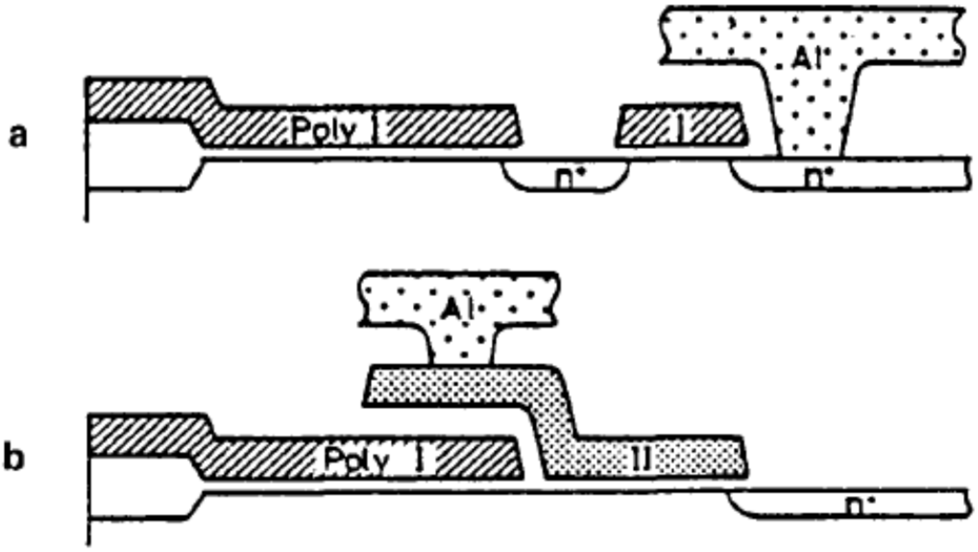


Fig. 4.26. Dynamic RAM cells achieved with (a) a single layer of polysilicon and (b) two layers of polysilicon. Using two layers eliminates the need for the lateral separation between the gate and storage electrodes and allows a more compact cell [71].

Fabrication of device structures with two levels of polysilicon often takes advantage of the difference in the oxide thickness grown on lightly doped single-crystal silicon and on heavily doped polysilicon. The oxide under the second level of polysilicon serves as the capacitor dielectric for the storage cell and should be as thin as possible for maximum charge storage. On the other hand, the oxide between the two levels of polysilicon should be as thick as possible to reduce unwanted parasitic capacitance between these two electrodes and increase the breakdown voltage. The more rapid oxidation rate of heavily doped polysilicon compared to lightly doped single-crystal silicon provides the desired difference in oxide thickness. Oxidation conditions can be chosen to maximum this difference, consistent with other limitations on processing conditions.

Gettering. In some processes, the high-temperature capability of polysilicon also allows efficient gettering of heavy metals away from critical device regions. The single crystal silicon on the back of the wafer can be exposed by removing oxide layers from the back before the polysilicon on the front of the wafer is heavily doped with phosphorus from the gas-phase so that phosphorus is also added to the back of the single-crystal wafer. The phosphorus-containing region on the back of the wafer can attract heavy metals during all subsequent high-temperature device processing. This gettering can be important in providing highly reliable integrated circuits in high-volume production. For example, removing the metals from the region under the storage capacitor of a dynamic RAM reduces the generation-recombination centers there, and increases the time allowed between refresh cycles. However, backside gettering is becoming less important today as integrated-circuit fabrication temperatures are reduced and thicker wafers are used. Alternate gettering techniques are being developed which do not rely on diffusion of heavy metals through the entire thickness of the silicon wafer.

EEPROMs. Using two layers of polysilicon can also be valuable in other devices. For example, a substantial amount of electrically erasable, programmable read-only memory (EEPROM) is frequently included on a microprocessor or other computer chip to allow storage of code specific to a particular application and other information that must be in frequently modified. This type of nonvolatile memory element requires one level of polysilicon above another in the gate region of the device. The first level is not electrically connected and "floats". The charge stored in this floating gate determines the condition of the underlying single-crystal silicon between the source and drain regions. If it contains excess electrons, it attracts holes and turns off an n-channel device. The opposite charge creates an inversion layer in the silicon and allows current to flow between source and drain. Application of a voltage of the appropriate potential allows charge to be transferred through the gate oxide to charge the floating region and "write" a cell of this programmable read-only memory. The charge may be generated by avalanching the drain-substrate junction to create free carriers, which are then attracted to or repelled from the floating gate by the potential on the overlying control gate. The charge can be removed from the floating gate by tunneling through the oxide separating the two layers of polysilicon. If this method of erasing the cell is used, the surface texture of the lower layer of polysilicon is critical. We have already seen that the conduction through oxide grown on polysilicon depends strongly on the asperities on the polysilicon surface. Control of the deposition conditions and the resulting structure of the polysilicon is more critical for this application than probably for any other.

High-Value Resistors. After polysilicon technology was developed for silicon-gate structures, it quickly found use in other applications, such as static RAMS. A typical static RAM circuit uses a cross-coupled structure that allows current to flow in one or other other of two parallel current paths. This static RAM cell can be formed by using two transistors in each leg of the circuit and two access transistors, one connected to the intermediate note of each leg of the circuit, forming a 6-transistor cell. To conserve area, the two load transistors can be replaced by high-value resistors to obtain a 4-transistor, 2-resistor cell [72]. The requirements on the resistance are not severe. It must be low enough to provide adequate current to retain charge on the intermediate node of each leg of the circuit and prevent noise from causing the cross-coupled cell to change state; it must be high enough to limit the current flowing through the driver transistor that is turned on so that the power dissipation remains acceptable. Resistors in the range of 10^9 Ω are generally satisfactory.

Using lightly doped polysilicon for this application is attractive; it is compatible with standard integrated-circuit processing and can even be placed above an active element on the circuit to reduce the cell size. It only needs to be shielded from dopant introduction during doping of the polysilicon gate or the source and drain regions. Undoped polysilicon has a resistivity in the mid-10^5 Ω-cm range. With a typical thickness of a few hundred nanometers, a sheet resistance of about 10^{10} Ω/sq. is easily obtained. In some cases, this value is satisfactory; more frequently, however, a somewhat lower resistance is needed to insure that adequate current is available to avoid stray charge from switching the state of the cell. A small amount of dopant can readily be added by ion implantation. Although control of the resistivity in the lightly doped region is difficult, the wide range of resistance allowed in this circuit makes lightly doped polysilicon acceptable.

Because hydrogen can change the resistance of high-value resistors by saturating dangling bonds at grain boundaries, the effect of hydrogen during device fabrication and operation is of concern. Moderately doped films are most sensitive to the barrier heights at grain boundaries, and are affected most strongly by the addition of hydrogen. Because the high-valued polysilicon resistors used in static RAMS are often doped in this range, the value of the resistors can be inadvertently changed by hydrogen introduced during the later stages of the fabrication cycle or during operation. Hydrogen from a plasma nitride passivation layer is of special concern; as we have seen, hydrogen from this source effectively passivates dangling bonds in polysilicon, and it can move during device operation, changing the value of the resistor. If circuit margins are not adequate, these changes pose a severe, long-term reliability problem.

Although polysilicon load resistors are adequate for many static RAM cells, improved circuit performance is obtained if the element connected to the transistor being turned off can conduct enough current to charge the intermediate node rapidly, and a voltage-variable load element is advantageous. If the perpendicular electric field provided by a gate electrode can modulate the conductivity of a polysilicon resistor to even a moderate degree, the circuit performance of the static RAM can be improved. To obtain this load element and simultaneously increase circuit density, the polysilicon layer can be placed above a thin oxide on the gate electrode of the bulk MOS transistor so that one gate electrode can control both the bulk transistor and the voltage-variable polysilicon resistor. As an extension of this concept, the possibility of constructing an MOS transistor with its channel in a layer of polysilicon has been explored for this and other applications.

MOS Transistors in Polysilicon. Operation of conventional integrated circuits requires compromises to avoid interactions between nearby elements. CMOS devices suffer from latch-up between neighboring n-channel and p-channel transistors; capacitive coupling to the high-relative-permittivity silicon can degrade circuit performance. Charge generation in the underlying silicon can reduce the radiation hardness of circuits in bulk silicon; charge injected at one node may travel to another node through the substrate, changing the charge state of dynamic memory cells. More conceptually, restricting integrated circuits to a single plane of devices at the surface of a bulk silicon wafer limits the type of devices and circuits possible. The ability to use the third dimension would allow fabricating novel device structures and should also reduce the interconnection length between devices.

The concept of placing the active layer of an MOS transistor within a thin layer of silicon has been explored for a number of years. MOS transistors with their conducting channels in polysilicon were fabricated early in the development of MOS integrated circuits [73, 74]. Figure 4.27 shows a p-channel transistor with its channel in a layer of polysilicon and the corresponding drain characteristics. Although the shape of the characteristics is similar to that of bulk MOS transistors, a large gate voltage must be applied before significant drain current flows, and the transistor has a very low transconductance. This inefficiency of the gate voltage in inducing a conducting channel can be attributed to the high concentration of allowed trapping states within the forbidden gap of polysilicon films, especially at the grain boundaries. As a gate voltage is applied, the energy bands must be bent to induce a conducting inversion layer near the silicon surface. Before a conducting channel can be induced in a material containing a high concentration of traps, the energy levels corresponding to many of the traps must be moved through the Fermi level [74]; that is, the charge state of many of the traps must be changed. Much of the applied gate voltage is used to charge or discharge trapping levels, rather than inducing conducting free carriers in an inversion layer, and excessively high gate voltages must be applied before conduction can occur between source and drain. Thus, the characteristics of transistors built in moderately doped polysilicon films are dominated by the behavior of the traps, as was the resistivity of polysilicon.

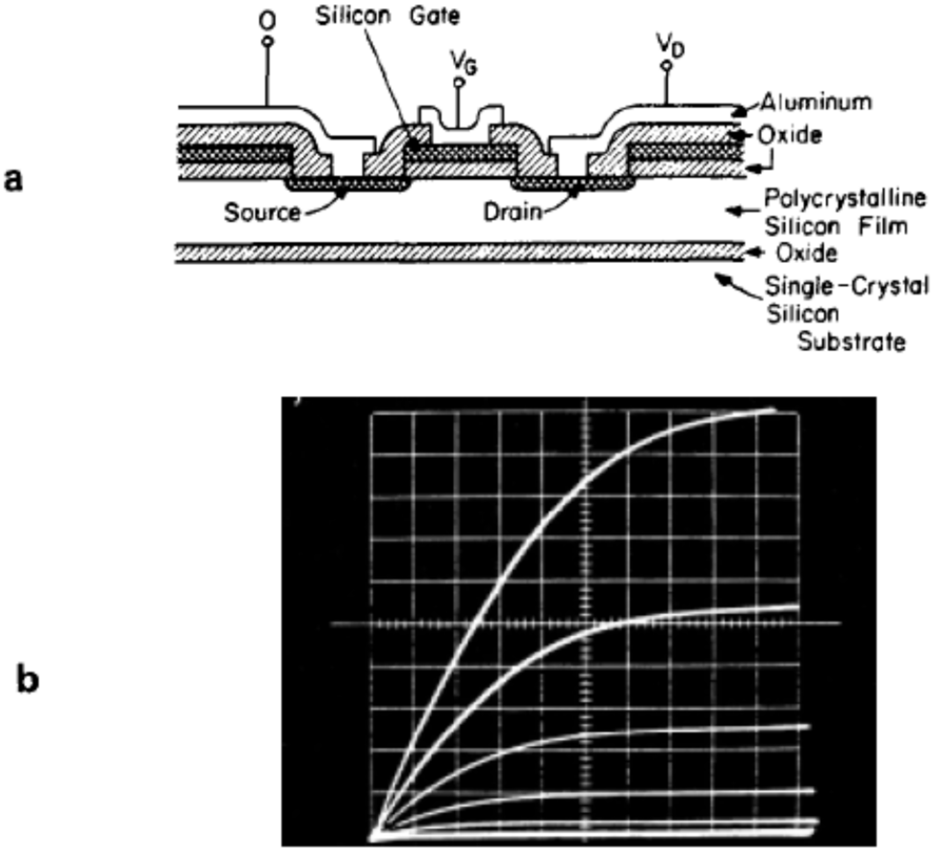


Fig. 4.27. (a) Cross section of an MOS transistor with its channel within a polysilicon film and second polysilicon layer forming the gate electrode. (b) Corresponding drain characteristics of a p-channel transistor: I_d = -50 pA/div, V_d = -2V/div; V_g = -2.5 V/step, 8 steps.

Because most of the traps are located near the grain boundaries, the transistor can be visualized as several transistors in series. Comparatively high-quality grains are separated from each other by highly defective grain boundaries. The regions above the central portions of the grains have a threshold voltage determined by the normal transistor equations for single-crystal silicon; the regions above the grain boundaries have a much higher threshold voltage. A moderate gate voltage bends the bands and induces conducting channels over the central regions of the grains, but barriers to conduction still exist at the surface near the grain boundaries, and no continuous conducting channel is formed. Conduction is not obtained until a considerably higher gate voltage is applied to change the charge state of the traps at the grain boundaries near the surface of the polysilicon and induce conducting channels where the grain boundaries intersect the surface. Consequently, the observed threshold voltage is much higher than expected from the average dopant concentration.

Even when a continuous conducting channel has been induced, the transistor characteristics are still influenced by the number of trapping states near the surface of the silicon.

If the number of states that must be moved through the Fermi level to change the surface potential slightly and increase the drain current is small, the current can increase rapidly because the effective channel length of the currentlimiting region is very short (only above the grainboundaries; the remaining regions are heavily inverted already). On the other hand, if the trap density is high, the charge state of many traps must still be changed as the gate voltage is increased, and the effective mobility remains low. The latter case applies in the fine-grain polysilicon normally deposited in integrated-circuit fabrication processes.

More useful transistors can be obtained by employing accumulation-mode transistors in place of the more common inversion-mode transistors normally used in silicon integrated circuits. In bulk silicon transistors, useful devices can only be obtained by modulating an inversion layer near the surface of the silicon; the conduction of the thick wafer under the gate cannot readily be modulated by changing the gate voltage. In addition to confining the conducting layer, the use of an inversion-mode transistor electrically separates the conducting channel from the common substrate by a depletion region and provides the needed isolation between transistors.

In a thin-film transistor, however, an oxide layer lies immediately beneath the silicon film, and parallel conduction in the bulk of the film can be small if the film is lightly doped. An accumulation layer can be induced near the surface of the film to provide the transistor action, while little current flows in the absence of this accumulation layer (especially if the bulk of the film is depleted of free carriers at zero gate voltage). Generating an accumulation layer requires less bending of the energy bands than does forming an inversion layer, so the applied gate voltage does not need to change the charge state of as many trapping levels within the polysilicon film as in an inversion-mode transistor. Conduction modulation is easier, and the magnitude of the transistor threshold voltage is lower. The characteristics of an accumulation-mode transistor depend on the thickness of the polysilicon film; a thin film can be completely depleted so that little drain current flows at zero gate voltage, while the neutral region in a thicker film can lead to significant "leakage" current.

Even in accumulation-mode transistors in fine-grain polysilicon, however, the threshold voltage remains too high to be readily compatible with the voltages commonly used in integrated circuits. Although these transistors may be useful in specialized applications, the characteristics of the polysilicon must be modified before transistors in polysilicon can be used in more standard integrated circuits.

Two different approaches can be taken to reduce the effective trap density in polysilicon to levels which allow more useful transistors to be fabricated. Because these trapping states are predominately located at grain boundaries, we must reduce the effectiveness of the trapping states at the grain boundaries or modify the grain boundaries themselves. In the first approach, the defects are passivafed (eg, by adding active hydrogen), so that they do not provide active levels within the forbidden gap to act as traps. A smaller fraction of the applied gate voltage is then used to change the charge state of the traps, and it is more effective in inducing and modulating a conducting channel; the threshold voltage decreases, and the mobility increases [75, 76]. Although their properties do not approach those of transistors in bulk wafers, hydrogen-passivated transistors may be quite satisfactory for selected applications, such as the load elements of static RAMS or the peripheral circuitry for active-matrix-addressed displays.

However, even when the grain boundaries are passivated with hydrogen to the maximum extent practical, the number of active grain-boundary states is still large, and the transistor characteristics are markedly inferior to those of transistors in single-crystal silicon. To achieve transistors of much higher quality, the grain boundaries themselves must be removed or at least reduced in number. This can be accomplished by melting the entire thickness of the finegrain polysilicon film with a scanned heat source, such as a cw laser, an electron beam, a graphite rod, or a high-intensity incoherent light source. When the silicon solidifies, only a relatively small number of grains nucleate, and these grains then grow into the rapidly cooling molten silicon to produce large-grain material containing few grain boundaries, and consequently, few grain-boundary traps. The entire thickness of the fine-grain silicon must be melted so that the solidifying silicon does not regrow on the remaining small grains to again produce a fine-grain structure. A moving heat source is generally used so that the lateral temperature gradient promotes growth of the few nucleating grains; alternatively, the desired lateral thermal gradient can be obtained by locally varying the amount of power absorbed to melt the silicon (ie, by locally varying the reflectivity and using a narrow-band optical source) or by locally varying the thermal conductance.

Transistors subsequently fabricated in this recrystallized silicon have properties approaching those of transistors in single-crystal silicon, and are much superior to transistors fabricated in fine-grain polysilicon. Field-effect mobilities are typically about 1/2 to 2/3 those of transistors in bulk silicon wafers [77]. These transistors are quite useful in a number of applications, but the few remaining grain boundaries impede their use in VLSI applications. If a grain boundary is parallel to the current-flow direction and intersects the source and drain regions, the source-drain dopant can readily diffuse along the grain boundary during device fabrication [78] to reduce the effective channel length or even cause a short circuit in a small-geometry transistor. If the grain boundary is perpendicular to the current-flow direction, it presents added impedance to current flow. Large-geometry transistors contain many grain boundaries, and the transistor characteristics of one device do not differ markedly from those of another. In small-geometry transistors, however, some transistors will contain grain boundaries and others do not. The resulting variation in transistor characteristics cannot be tolerated in high-performance circuits, and the few remaining grain boundaries must be removed.

Two techniques can be used to produce material free of grain boundaries in the active transistor regions. In one approach, windows (seeding regions) are periodically opened in the oxide above the single-crystal substrate before the fine-grain polysilicon is deposited. As the molten silicon solidifies during recrystallization, the crystal structure of the substrate propagates into the solidifying material, first vertically, then laterally over the oxide. In this manner, grain-boundary-free, single-crystal silicon is obtained over the insulating oxide. However, the distance that the grain-boundary-free material can propagate laterally is limited, making the technique difficult to control. Alternatively, regions of transparent material of varying thicknesses can be placed above the silicon before recrystallization so that the amount of power absorbed from a narrow-band optical source varies from one region to another. The regions which absorb the least power are coolest and solidify first so that any grain boundaries which subsequently form are located away from these regions. In this manner, the location of defects can be controlled photolithographically, and they can be placed outside of the active device regions (Le., in the field regions). This technique can be coupled with the "seeding technique" discussed above to control the orientation of the grain-boundary-free material. Transistors subsequently fabricated in these single-crystal films have properties very similar to those of transistors in bulk silicon wafers, although the floating neutral region beneath the channel leads to the so-called *kink-effect* [79] and other, more subtle differences [80, 81].

Bipolar-Circuit Applications. Although polycrystalline silicon is most frequently used in MOS integrated circuits, it is becoming crucial in highperformance, bipolar integrated circuits also. Its use decreases parasitic resistance and capacitance and also reduces reverse carrier injection from the base into the emitter.

In modern bipolar transistors, the minority-carrier transit time through the thin base region is very short, and the operating speed of the transistor is limited by parasitic elements. The resistance of the extrinsic base region (ie, the region between the iiitrirzsic base under the emitter and the external base contact) can degrade device performance significantly, as can the capacitance associated with the depletion region between the extrinsic base and the collector. In typical bipolar transistors, the extrinsic base region is heavily doped to reduce its resistance (while increasing the capacitance). However, in conventional structures the extrinsic base region must be separated from the emitter region by an alignment tolerance, and the decrease in base resistance provided by the heavily doped, extrinsic base region is limited; it is also variable because of varying mask alignment. To reduce the need for a trade-off between low base resistance, low capacitance, and ease of mask alignment, a polysilicon layer can be used to provide most of the extrinsic base. The size of the extrinsic-base region in the single-crystal silicon is reduced, and most of the extrinsic- base region is placed in a layer of polysilicon above an oxide, as shown in Fig. 4.28a. The extrinsic-base-collector capacitance is reduced in this structure because a thick layer of low-permittivity oxide separates the two regions over most of the area. In addition, the polysilicon of the extrinsic base can be used to self-align the emitter region to the extrinsic base so that the two are closely spaced and independent of alignment variations. Thus, using a polysilicon extrinsic-base region greatly improves the performance of bipolar transistors.

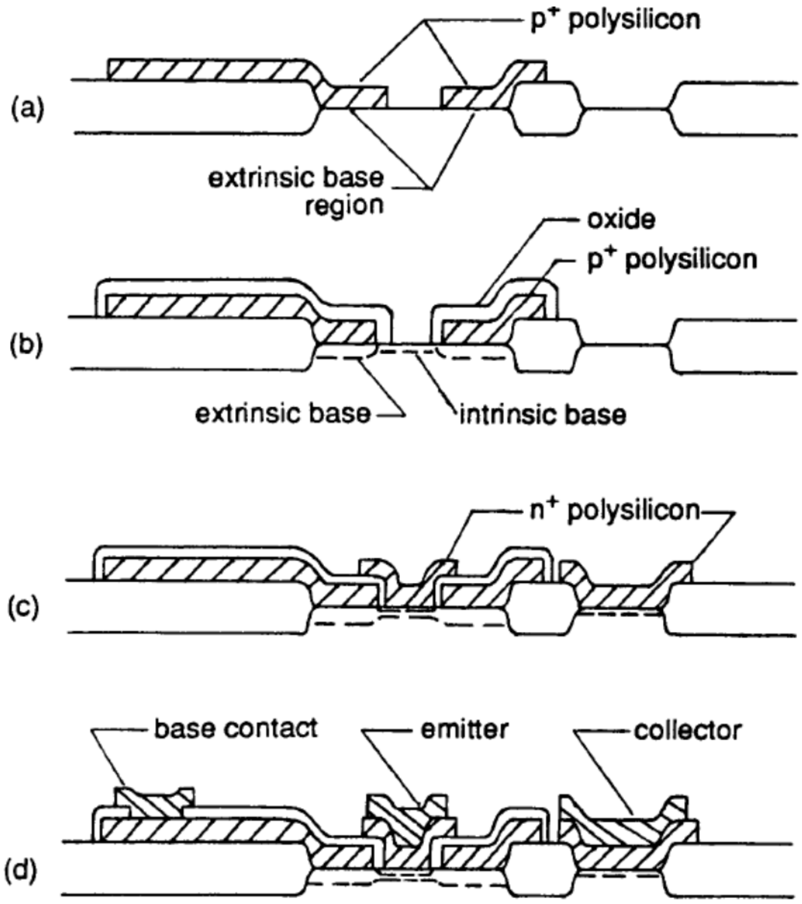


Fig. 4.28. Two layers of polysilicon can be used in the polysilicon-emitter transistor, with an oxide on the sidewall of the first layer separating the two. (a) The extrinsic-base region is formed by the first polysilicon layer which extends over the field oxide. (b) Boron is diffused from the polysilicon into the single-crystal silicon to make contact to the intrinsic base, which can be added by implantation. The polysilicon is oxidized, and the oxide is removed from the single-crystal region which is to become the emitter. (c) A second layer of polysilicon is deposited and defined either before or after adding the emitter doping to the single-crystal silicon. (d) Contact

to the base region is formed over the oxide, reducing the area of the single-crystal extrinsic -base region.

Polysilicon can also be advantageously applied to the emitter region of bipolar transistors (Fig. 4.28c). In the most straightforward application, the polysilicon serves as a source for the dopant atoms diffused into the single- crystal silicon to form the emitter. When high concentrations of dopant are added directly to single-crystal silicon by either gas-phase diffusion or by ion implantation, defects can form in the silicon. Because the heat cycles which can be used after emitter doping must be very limited to avoid excessive dopant diffusion, removing this damage adequately is often difficult. If the dopant is added to a polysilicon layer above the single-crystal silicon and then diffused from the polysilicon into the single-crystal silicon, the single-crystal region is not damaged, and a high quality emitter can be formed. Because the dopant diffusivity is much lower in single-crystal silicon than in polysilicon, an emitter formed in this manner can be very shallow.

Using a polysilicon region above the emitter offers other advantages in addition to providing a convenient source of dopant atoms. Such polysiiicort-entitter bipolar transistors can have much higher gains than do similar transistors in which metal makes direct contact to the single-crystal emitter. One of the factors degrading transistor performance is the injection of holes from the p-type base region into the n-type emitter. This reverse iitjectiorz adds to the base current a component which reduces the emitter ittjectioir efficieizcy and, therefore, the gain of the transistor. Placing a layer of polysilicon above the emitter reduces the reverse injection and increases the overall transistor gain. The detailed mechanism for this reduction in reverse hole injection is not fully understood, and several explanations have been proposed.

First, the role of the interfacial oxide between the polysilicon and the single-crystal silicon appears to be critical [82]. A thin, residual oxide layer between the polysilicon and the single-crystal silicon may impede hole transfer from the base into the emitter. Alternatively, holes may be injected into the polysilicon to diffuse toward the emitter contact, where they would recombine. Because the polysilicon layer is typically thicker than the single-crystal emitter region, the hole gradient is reduced, and less injection occurs [83]. The lower lifetime of minority carriers in polysilicon, however, limits the benefit of which can be gained by this mechanism, suggesting that the role of the polysilicon-single-crystal interface is crucial in obtaining the highest transistor gain. More detailed observations show that the gain enhancement decreases when the thin oxide between the single-crystal silicon and the polysilicon becomes discontinuous. Although the maximum gain enhancement can be obtained with a thin, continuous interfacial oxide, practical use of such a structure is questionable because control of this interfacial oxide is difficult: if it is too thin, the gain is reduced; if it is too thick, the emitter resistance is excessive. However, even without the gain related to the interfacial oxide, using polysilicon in the emitter region offers substantial advantages, and most high-performance bipolar processes use polysilicon-emitter transistors.

An alternate method of reducing minority carrier injection from the base into the emitter is to use a material with a wider bandgap which presents a large barrier to reverse minority-carrier injection. Oxygen-doped polycrystalline silicon (Semi-jrisulating mlycrystallitie silicoii or SIPOS) has been used for this purpose [84]. Because of the alignment between the bandgaps of single-crystal silicon and SIPOS, electron injection across the heterojunction is not significantly impeded, but the unwanted reverse injection of holes into the emitter is greatly reduced, improving the injection efficiency.

Polysilicon Diodes. In addition to transistors, other devices can be fabricated in polysilicon films. The simplest device is a lateral p-n junction diode [85, 86]. A structure in which both the p-type and the n-type regions are heavily doped can readily be obtained in a CMOS process which employs p^+ and n^+ polysilicon gates for the p-channel and n-channel transistors, respectively. (As mentioned above, these diodes can be undesired parasitic elements between the two types of transistors unless a silicide is formed over the polysilicon or the two gate regions are connected together by other means. However, in other cases, the diode may be used advantageously.) The diode quality factor is typically about 1.8-2 [85], indicating that recombination in the space-charge regions dominates; the breakdown voltage is low because of the heavy doping on both sides of the junction. The breakdown voltage can be increased by reducing the dopant concentration on one side of the junction. However, because of the rapid increase in resistivity of polysilicon with decreasing dopant concentration, the series resistance in the polysilicon diode can rapidly limit the device performance. The low lifetime in polysilicon does, however, reduce minority-carrier storage, and polysilicon diodes are expected to switch rapidly.

Polysilicon Pressure Sensors. Polycrystalline silicon is also used in pressure sensors. In this application, a thin layer of polysilicon is placed over an oxide, and resistors are formed in the polysilicon. A hole is etched through the substrate from the back, leaving a thin membrane of oxide and polysilicon. When pressure is applied to the membrane, the polysilicon distorts, and the values of the resistors change. In a typical circuit, unstrained resistors are located nearby over the thick silicon wafer for comparison so that small changes in resistance can readily be measured. Because of the grain boundaries, the sensitivity of these polysilicon pressure transducers is only about half that of single-crystal silicon pressure sensors, but their ease of fabrication and isolation from the substrate make polysilicon pressure transducers useful devices.

Device Isolation. In addition to forming part of the active device structure, polysilicon is also used for isolation between adjacent devices. The oldest application of polysilicon that is still in current production is its use as the supporting layer for the single-crystal islands of dielectrically isolated integrated circuits [87]. To fabricate a circuit using this technology, a single-crystal substrate of the appropriate conductivity type (e.g., n-type to serve as the collector of a bipolar transistor) is first oxidized, and the oxide is patterned. Isolation moats are etched into the substrate to a depth greater than the thickness of the final silicon pockets. If (100)-oriented silicon is used, the etching terminates when the etched grooves reach the apex of a V formed by the intersection of (111) planes; the depth of the grooves is determined by the width of the oxide removed from these regions. After a heavily doped region, which serves as the buried layer or subcollector, is added and the surface is oxidized, a thick layer of polysilicon is deposited.

Because this layer of polysilicon serves as a mechanical support, it must be approximately as thick as a starting silicon wafer, and high deposition rates are needed. Therefore, the deposition is usually carried out at a high temperature using a chlorinated silicon compound, such as dichlorosilane or silicon tetrachloride. Using a high temperature also reduces wafer deformation (warpage) during deposition by increasing the desorption of gaseous impurities and increasing surface migration. After polysilicon deposition, most of the original single-crystal wafer is removed by mechanically lapping and polishing the wafer until the previously formed grooves appear. The structure is then inverted, and transistors are fabricated in the exposed single-crystal silicon pockets. Because of wafer distortion during the process, the pockets are often displaced substantially from their original location. Registration of devices to the pockets is difficult, and this technique can only be used for low-density circuitry. However, it does provide the excellent isolation required for very-high-voltage or radiation-tolerant integrated circuits. Some of the limitations can be removed by using self-terminating etching techniques [88-90] so that grooves do not have to be formed before polysilicon deposition. The registration limitation is then removed, but controlling the polysilicon-deposition and silicon-removal processes remains difficult, restricting the technique to specialized applications.

In another isolation technique that also uses polysilicon formed at high temperatures, polysilicon and epitaxial silicon are simultaneously deposited. The substrate is first oxidized, and the oxide is defined to leave it in the regions where polysilicon is to be formed. During subsequent silicon deposition, the material grown on the exposed silicon substrate is epitaxial and that on the oxide is polycrystalline. Using this mixed structure takes advantage of the rapid diffusion of dopant atoms in polysilicon. Dopant can readily diffuse through the entire thickness of the polysilicon regions while only penetrating a small distance into single-crystal silicon. This phenomenon can be used to form p-it junction isolation regions in a bipolar integrated circuit without excessively long diffusion times. The lateral diffusion into the single-crystal regions is also significantly reduced. To further lower the diffusion times required, doped oxides can be used to nucleate the polysilicon, so that diffusion occurs into the polysilicon regions both from the top and from the bottom. A heavily doped collector sirik can also be formed by this mixed epitaxial-polycrystalline deposition to lower the parasitic collector resistance. Because rapid diffusion is essential for this technique to be useful, the structure of the polysilicon must be carefully controlled to provide columnar grains with the well-defined vertical paths needed for rapid diffusion. In addition, of course, the deposition conditions must be compatible with the simultaneous formation of high-quality epitaxial silicon.

An isolation structure discussed more recently uses the familiar fine-grain polysilicon. In this trench-isolation technique [91], a groove is etched in the single-crystal silicon where isolation regions are needed. This groove is formed by anisotropic dry etching so that a narrow, deep “trench” with vertical sidewalls is obtained. The walls of the groove are then oxidized, and polysilicon is deposited to fill the groove completely. The deposition conditions are chosen to be well within the surface-reaction-limited regime so that the silicon-containing gas can readily travel down the entire length of the narrow groove. Otherwise, filling can be incomplete, leaving a void which produces a mechanically unstable structure. If the deposition on the walls of the groove is well controlled, the groove is completely filled once a layer half as thick as the width of the groove is deposited. A polysilicon layer of about the same thickness is also deposited on the oxide-covered top surface of the wafer, and this polysilicon is then removed without using a mask by anisotropic dry etching until the underlying oxide surface is exposed. After the oxide is removed from the single-crystal regions and a protective oxide is formed over the tops of the polysilicon-filled grooves, transistors are fabricated in the single-crystal regions. This technique can be used for either bipolar or MOS integrated circuits. It is very useful in dense, small-geometry CMOS integrated circuits for suppressing latchup by increasing the effective path length between adjacent devices. It can also be used to form the storage capacitor of a one-transistor, dynamic RAM cell so that a large effective storage area is obtained without occupying much of the wafer surface [92]. In a further refinement, both the access transistor and storage capacitor can be placed on the sidewalls of an etched trench [93].

As a modification of the trench-isolation technique, a selective deposition process can be employed in which silicon is deposited in the grooves, but not on the top surface of the wafer [94]. In this case, the walls of the groove are covered with oxide, but the oxide is removed from the bottom of the groove; an isolation diffusion can be added to the exposed silicon at the bottom of the groove, if desired. Silicon is then deposited selectively on the exposed silicon at the bottom of the groove, and the groove is filled from the bottom without having silicon nucleate on the top surface of the wafer. Control of the etching and deposition processes is critical for this technique to be successful.

Although the dominant use of polysilicon is for gate electrodes and interconnections in MOS integrated circuits, virtually all new, high-performance bipolar technologies employ one, or even two, layers of polysilicon to improve transistor performance and increase packing density. The availability of polysilicon has also led to its use in floating-gate memory elements and in high-value load resistors. Devices with their active elements within polysilicon layers are also being investigated. Polysilicon diodes are attractive because of their compatibility with CMOS integrated- circuit processing, and polysilicon transistors can be employed both in silicon integrated-circuit applications and for large-area circuits, such as the active matrix and peripheral circuits of displays. As integrated-circuit technology continues developing, the uses of polysilicon will undoubtedly expand, even as its use for interconnecting elements in MOS integrated circuits becomes restricted by its limited conductivity. Even for MOS integrated circuits, however, the compatibility of polysilicon with integrated-circuit processing makes its continued use attractive, and it will probably be augmented by other materials, rather than being replaced in the foreseeable future. The widespread use of polysilicon in present applications and its expansion into new applications make it a material of continuing importance in silicon integrated-circuit technology and require continued and expanded understanding of its deposition and properties [95].

4.5. Amorphous Silicon

Amorphous silicon (a-Si) is the non-crystalline allotropic form of silicon. It can be deposited in thin films at low temperatures onto a variety of substrates. It offers some unique capabilities for a variety of electronics.

Silicon is a fourfold coordinated atom that is normally tetrahedrally bonded to four neighboring silicon atoms. In crystalline silicon (c-Si) this tetrahedral structure continues over a large range, thus forming a well-ordered crystal lattice. In amorphous silicon this long range order is not present. Rather, the atoms form a continuous random network. Moreover, not all the atoms within amorphous silicon are fourfold coordinated. Due to the disordered nature of the material some atoms have a dangling bond. Physically, these dangling bonds represent defects in the continuous random network and may cause anomalous electrical behavior.

Likewise, the material can be passivated by hydrogen, which bonds to the dangling bonds and can reduce the dangling bond density by several orders of magnitude. Hydrogenated amorphous silicon (a-Si:H) has a sufficiently low amount of defects to be used within devices such as solar photovoltaic cells, particularly in the protocrystalline growth regime [96, 97]. However, hydrogenation is unfortunately associated with light-induced degradation of the material, termed the Staebler–Wronski effect [98].

While a-Si suffers from lower electronic performance compared to c-Si, it is much more flexible in its applications. For example, a-Si layers can be made thinner than c-Si, which may produce savings on silicon material cost. One further advantage is that a-Si can be deposited at very low temperatures, e.g., as low as 75 degrees Celsius. This allows for deposition on not only glass, but plastic as well, making it a candidate for a roll-to-roll processing technique. Once deposited, a-Si can be doped in a fashion similar to c-Si, to form p-type or n-type layers and ultimately to form electronic devices.

Another advantage is that a-Si can be deposited over large areas by PECVD. The design of the PECVD system has great impact on the production cost of such panel, therefore most equipment suppliers put their focus on the design of PECVD for higher throughput, that leads to lower manufacturing cost [99] particularly when the silane is recycled [10].

Amorphous silicon has become the material of choice for the active layer in thin-film transistors (TFTs), which are most widely used in large-area electronics applications, mainly for liquid-crystal displays (LCDs).

4.5.1. General Properties

Fabrication techniques, such as plasma-enhanced chemical vapor deposition (PECVD), hot wire-chemical vapor deposition, photo assisted CVD, reactive sputtering can be used to produce amorphous silicon [101]. The structural and electronic properties of a-Si has been studied extensively, with much understanding of how electronic properties relate to local structural order. By definition, amorphous silicon films containing about 4 to 40 atomic percentage of hydrogen are generally denoted as a-Si:H [101]. Unlike the crystallite silicon material, a-Si:H has only a short range order which means that all four covalent bonds of Si atoms might not be saturated [102]. Thus it has a continuous random network structure model [102], in which Si has unsaturated bonds as shown below [103] (Fig. 4.29). Si atoms with unsaturated bonds occur as it releases the strain associated with the disordered structure. Commonly known as dangling bonds, they act as recombination centers and thus lead to poor quality solar cells. Hydrogen passivation can be considered the best technique to reduce their concentration. The single electron of hydrogen atom can convert the unsaturated Si atoms to four fold coordinated ones.

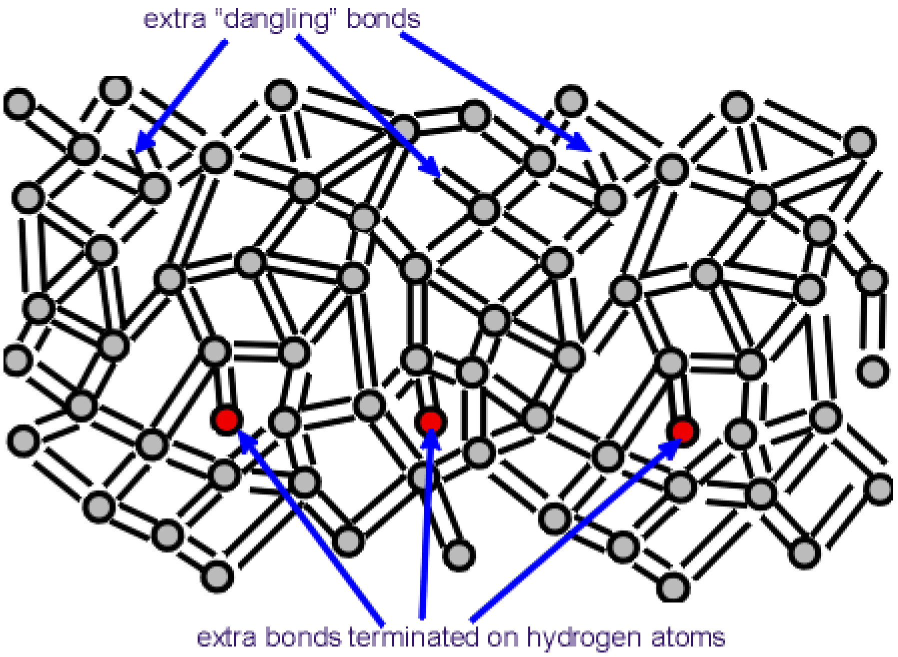


Fig. 4.29. Continuous random structure of a-Si.

Thus hydrogen passivation is successful in removing dangling bonds [104]. It has been shown by other groups that dangling bond concentration can be reduced from 10^{22} to 10^{15} cm^{-3} using this technique. For complete understanding of the working of a-Si solar cells, it is important to determine the distribution of the trap states in the bandgap. Figure 1.4 shows a density of states curve for a undoped a-Si:H film [105].

Like crystalline silicon, the existence of extended valence and conduction band states is observed. However, the presence of short range order and distortion of the covalent bonds of fourfold co-ordinated neutral silicon atoms in amorphous structure gives rise to an exponential decrease of the density of states, also known as band tails [106, 107] and are given by

$$G_C = G_{C0} \exp[(E_c - E) / E_{urc}] \tag{4.6}$$

$$G_V = G_{V0} \exp[(E - E_v) / E_{urv}] \tag{4.7}$$

where G_C and G_V are density of states for conduction band tail states and valence band tail states respectively; and E_{urv} and E_{urc} are the measure of penetration of the conduction and valence band tails into the forbidden gap respectively, called the Urbach energy. For a good quality a-Si:H solar cells, the typical values of E_{urv}= 42-45 meV [108-110] and E_{urc}=26 meV [111], which means that valence band tail extends deeper into the forbidden gap. Low Urbach energy means lesser disorder, which is a very important property of amorphous a-Si:H. It should be noted that a fraction of defect states also can be attributed to silicon atoms having three silicon neighbors and one hydrogen atom in tetrahedral configuration. These extra states change the Tauc's bandgap , depending on the hydrogen content- from about 1.5 eV for pure a-Si to almost 2.0eV when 30 at. % H content is recorded. Very close to the tail states exists several deep-level bands in the forbidden gap, believed to be related to dangling bonds [101].

The bandedge and midgap region of density of state plot for amorphous entails further deeper study. As we can see from Fig. 4.30 that a boundary exists (known as mobility edge, as suggested by Mott et al [112]), which separates two regions i.e. high mobility region (or extended states), where electrons and holes are helped by band transport from localized region or low mobility region, where carriers undergo hopping conduction. The midgap region observed in amorphous silicon region has been explained by Davis and Mott [113] by making use of "bands of compensated levels" [106] near the midgap, which can be attributed to defects such as dangling bonds present in the material [106], which can be can be empty or positively charged (D^+), neutral (D^0) or negatively (D^-) charged. The D^0 and D^- energy levels are separated by a correlation energy, U_{eff}, which is attributed to the repulsion between electrons.

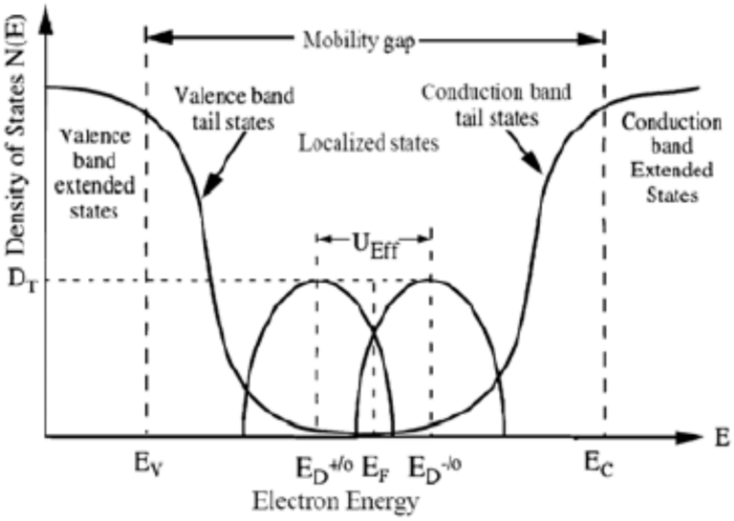


Fig. 4.30. Electronic density of states in a-Si:H [113].

Mashall and Owen's amorphous Silicon density of state model [114] highlighted the presence of bands of donors and acceptors in the upper and lower halves of the mobility gap similar to Davis and Mott model [113], which facilitated derivation of "realistic gap state models" [106]. Characteristics of these states have been assessed from such experiments like photoluminescence [115, 116], electron spin resonance (ESR) [117], and optically detected magnetic resonance (ODMR) [118]. Some defect states are obtained during the growth of the film due to hydrogen evolution from films [119]. Others seem to be related to "microstructural imperfections" [106] such as polymer chains [120] with contaminants like oxygen, carbon and other dopants [121, 122]. Other defect levels may arise from foreign atoms and neighboring dangling bonds interactions, and weak bonds between Si and other atoms [106]. The sign of U_{eff} is still controversial and is still subject of lot of debate. U_{eff} has been approximated as 0.4 eV from ESR on undoped and doped a-Si:H samples [109], assuming a positive dangling bond [106]. Based on EDS and Photothermal Deflection Spectroscopy (PDS) it was estimated as 0.2 eV [123]. Negative U_{eff} has been proposed in works [124-127], and the value of it to be about -2eV as has been established based on detailed calculation [127] and claimed that this might not be observed during the ESR measurement. In the positive U_{eff} model, it has been shown that D^0 is about 1.0 ~ 1.25eV below conduction band, while D^- is about 0.25 ~ 0.45eV higher than D^0 level [106].

4.5.2. Improvement of Stability by Chemical Annealing

The instability of an a-Si:H can be attributed to its non-equilibrium state, so it is obvious that any thermal or optical external disturbance, will cause structural changes in the material [128]. When exposed to light, electron-hole pairs are created which recombine in the film, release energy and ultimately leads to breakage of weak Si-Si and /or Si-H bonds facilitating creation of defects, which in turn deteriorates the conductivity of the film [128]. The film degradation process can be explained by these models: (1) the trap to dangling bond conversion model [129], indicates when positively correlated dangling bonds (D^0) and negatively correlated charged defect states (D^- and D^+) capture an oppositely charged carriers it gets converted into a positively correlated dangling bond state, which causes an increase in D^0 states and promotes recombination [129]. (2) Bond breaking-SJT model [130], suggests that the energy dissipated due to recombination of an electron-hole pair leads to breakage of weak Si-Si bonds and consequently increased density of midgap defect density [128]; (3) Another model [131] suggests that recombination of photogenerated carriers removes H from Si-H bonds creating concentration of Si dangling-bond defects significantly [131]. When two free H atoms combine together to form a metastable two- H, the two dangling bonds from which H is emitted also become metastable. It is seen that that during light soaking, defect density typically increases by an order from 10^{16} to about 10^{17} cm^{-3} [132]. This has been verified by Staebler and Wronski, who have also found revival of device quality being achieved by annealing at 180°C [128]. Other theories explaining this phenomenon exists as well. Some studies have identified the metastability in a-Si:H as arising from the breaking of weak silicon bonds [133], accompanied by structural rearrangements in the network, that can account for electron-spin resonance data, defect kinetics [134], charged metastable defects and hysteretic defect annealing [135], in contrast to the H-collision model [136] that has been proposed to explain metastability through the rupture of Si-H, followed by H-motion, leading to H-induced complexes in the network [136, 137]. It has been recognized that H motion can occur by cupture and formation of silicon bonds [138] in the amorphous network, and the role of H has been implicated as the origin of metastability. Local H-motion and flipping of SiH bonds also leads to a metastable state that can account for infrared absorption changes on light soaking [139], in contrast to metastable changes in H-bonding where H bonds to weak silicon bonds. For a-Si solar cells, clearly mid-gap defects created photo-generation are detrimental for the solar cell as they act as electron and hole recombination centers thereby adversely affecting their lifetime. It has also been seen to reduce the electric field in the middle region of the *i*-layer, which reduces the field assisted drift length or range and thus adversely affecting the collection of the carriers in *i*-layer [128]. With the target of improving stability of a-Si, several approaches have been tried, which includes the development of new fabrication materials, techniques and solar cell device structure. The new deposition methods that have been successful so far improving stability include remote ECR growth [140, 141], hot wire growth [142], multichamber systems [143]. It has been shown that the stability of graded bandgap *i*-layer structure solar cell is far superior when compared to that of a- Si:H solar cell with no grading [144]. Other approaches towards improving the stability of a-Si:H solar cells include use of "blocking/insulator layers, grading or p^+ multilayered, *i/n* interface grading, tailoring of the *i*-layer" [145], and tandem solar cells [146]. Researchers have extensively studied various deposition materials to improve stability - high hydrogen dilution [147], He-dilution [148], deuterium dilution [149], and SiF_4 or SiH_2Cl_2 [150] have been successful in meeting this goal.

However, chemical annealing (CA) so far has turned out to be a very attractive method for improving material stability. The basic objective of chemical annealing is to cause sufficient ion bombardment so as to remove loosely Si-H bonds, which are very vulnerable under light soaking and give rise to dangling bonds easily. Again, since the hydrogen content is reduced, it allows us to reduce bandgap as well. The success of chemical annealing in improving stability has been shown in [151] , when the authors compared the stability of *p*-*i*-*n* solar cell based on Ar CA was with standard a-Si:H solar cell fabricated from SiH_4 without H_2 dilution. Measurement found that in the beginning the standard solar cell has a higher FF, but after a light soaking of ~700 minutes, the FF of the standard one drops below that of the CA one [151]. However, the FF of CA devices was low, ~55% range, implying high defect density. So, it allows us to obtain low bandgap material which is less susceptible to light induced degradation. Thus, chemical annealing helps us in improving stability of material.

More recently, improvements in a-Si construction techniques have made them more attractive for large-area solar cell use as well. Here their lower inherent efficiency is made up, at least partially, by their thinness – higher efficiencies can be reached by stacking several thin-film cells on top of each other, each one tuned to work well at a specific frequency of light. This approach is not applicable to c-Si cells, which are thick as a result of their construction technique and are therefore largely opaque, blocking light from reaching other layers in a stack.

The main advantage of a-Si in large scale production is not efficiency, but cost. a-Si cells use approximately 1% of the silicon needed for typical c-Si cells, and the cost of the silicon is by far the largest factor in cell cost. However, the higher costs of manufacture due to the multi-layer construction have, to date, made a-Si unattractive except in roles where their thinness or flexibility is an advantage. Typically, amorphous silicon thin-film cells use a p-i-n structure. Typical panel structure includes front side glass, transparent conducting oxides, thin film silicon, back contact, polyvinyl butyral (PVB) and back side glass.

4.6. Nanocrystalline Silicon

Nanocrystalline silicon (also called microcrystalline silicon) is amorphous silicon, but also contains small crystals. It absorbs a broader spectrum of light and is flexible. It is an allotropic form of silicon with paracrystalline structure—is similar to amorphous silicon (a-Si), in that it has an amorphous phase. Where they differ, however, is that nc-Si has small grains of crystalline silicon within the amorphous phase. This is in contrast to polycrystalline silicon (poly-Si) which consists solely of crystalline silicon grains, separated by grain boundaries. The difference comes solely from the grain size of the crystalline grains. Most materials with grains in the micrometre range are actually fine-grained polysilicon, so nanocrystalline silicon is a better term. The term Nanocrystalline silicon refers to a range of materials around the transition region from amorphous to microcrystalline phase in the silicon thin film. The crystalline volume fraction (as measured from Raman spectroscopy) is another criterion to describe the materials in this transition zone.

Nc-Si has many useful advantages over a-Si, one being that if grown properly it can have higher electron mobility, due to the presence of the silicon crystallites. It also shows increased absorption in the red and infrared wavelengths, which make it an important material for use in a-Si solar cells. One of the most important advantages of nanocrystalline silicon, however, is that it has increased stability over a-Si, one of the reasons being because of its lower hydrogen concentration. Although it currently cannot attain the mobility that poly-Si can, it has the advantage over poly-Si that it is easier to fabricate, as it can be deposited using conventional low temperature a-Si deposition techniques, such as PECVD, as opposed to laser annealing or high temperature CVD processes, in the case of poly-Si.

4.7. Conclusions

Silicon is one of the most useful elements to mankind. Most is used to make alloys including aluminium-silicon and ferro-silicon (iron-silicon). These are used to make dynamo and transformer plates, engine blocks, cylinder heads and machine tools and to deoxidise steel.

The element silicon and their single-, poly-, nanocrystalline and amorphous forms are used extensively as a semiconductor in solid-state devices in the computer and microelectronics industries. For this, hyperpure silicon is needed. The silicon is selectively doped with tiny amounts of boron, gallium, phosphorus or arsenic to control its electrical properties.

The wide application of silicon is solar cells production. Solar cells based on silicon (Si) semiconductors account for nearly 90% of photovoltaic (PV) products. Annual production of Si-based PV reached more than 15 gigawatts—an order of magnitude higher than other PV technologies. Silicon in PV takes many forms, including the industry-dominant single-crystalline (c-Si) and multicrystalline wafers sawn from ingots; melt-grown ribbons; thin hydrogenated amorphous silicon (a-Si:H); and microcrystalline Si layers grown from gaseous precursors. Many improved technologies are on the horizon, including thin c-Si by epitaxial growth, crystallized polysilicon layers, and nanoscale Si.

Chapter 5. $A^{III}B^{V}$, $A^{II}B^{VI}$, and other semiconductors

5.1. Introduction

Nowadays the silicon is dominant material in semiconductor electronics and photovoltaic. But, nevertheless, semiconductor materials $A^{III}B^{V}$, $A^{II}B^{VI}$, $A^{IV}B^{IV}$ have wide applications in such fields as high frequency electronics and optoelectronics. The various applications of these materials are caused, first of all, by higher carrier mobility and direct energy band. At the moment more and more attention is paid on research, development and application of the devices based on wide bandgap materials. They allow to realize the power high frequency electronic devices and optoelectronics devices for action in shorter wavelength region of light.

GaN as a representative of its binary cousins, InN and AlN, and their ternaries along with the quaternary, is considered one of the most important semiconductors after Si. It is no wonder that it finds ample applications in lighting and displays of all kinds, lasers, detectors, and high-power amplifiers. These applications stem from the excellent optical and electrical properties of nitride semiconductors. The parameters are imperative in determining the utility and applicability of this class of materials to devices.

This Chapter is devoted the consideration and analysis of the properties and applications of wide bandgap materials and their structures.

5.2. Development and Applications of Wide Bandgap Semiconductors. Optical Devices

5.2.1. Wide Bandgap Semiconductors Indispensable for Short Wavelength Optical Devices

As the name implies, a “wide bandgap” semiconductor is one having a large bandgap energy (forbidden bandgap), which is directly related to the emission/absorption wavelength of optical devices. Typical wide bandgap semiconductors exhibit emission/absorption wavelengths in the green/blue part of the visible spectrum and on into the shorter wavelengths of violet/ultraviolet light. For example, the “blue light emitting diode” is a well known application of wide bandgap semiconductors. Thus in general, wide bandgap semiconductors can be defined as having fundamental optical absorption edges that are of shorter wavelengths than the color red [1].

Examples of optical devices include light emitting diodes, laser diodes, photodiodes, photoconductive sensors, electro-modulation devices, and optical–optical modulation devices. Semiconductor light emitting devices are ultrasmall, light weight, high efficiency, and have much longer lifetimes than other light sources. In particular, wide bandgap semiconductors have become increasingly important in the electronics industry as optical sources for full color displays, white light illumination, UV/deep UV light sources, and blue–violet laser diodes for high density DVDs. But in spite of the intense world-wide interest in the use of wide bandgap semiconductors for blue LEDs, a long time passed before technological breakthroughs led to the fabrication of *p-n* junctions in GaN. After 1990, the number of researchers involved in wide bandgap semiconductors increased dramatically following the first successful operation of the blue LED.

Wide Bandgap Semiconductor Materials

Figure 5.1 is a comparison of the bandgap energies and lattice constants (a) of wide bandgap semiconductors and other well known semiconductors including silicon (Si), GaAs and ZnSe. The three main types of wide bandgap semiconductors are:

– Group III nitrides such as GaN
– Group II oxides such as ZnO
– Group II chalcogenides such as ZnSe

Other important wide bandgap semiconductors such as silicon carbide and diamond are being studied for electronic device applications.

In Fig. 5.1, the hexagon and square symbols represent hexagonal and cubic crystal structures, respectively. Materials shown in plain and italicized text refer to direct and indirect bandgap semiconductors, respectively. Further, by multiplying the lattice constants of hexagonal crystal structures by $\sqrt{2}$, it is possible t directly compare those values to the lattice constants of cubic crystal structures in terms of the atomic bond lengths. A bandgap energy of 2.25 eV corresponds to a wavelength of 550 nm (green light emission) and semiconductors with bandgap energies larger than 2 eV are said to be “wide bandgap.” Thus it is apparent that wide bandgap semiconductors are essential for fabricating optical devices emitting visible-green, blue, and UV/deep UV wavelengths.

Direct bandgap semiconductors, such as GaN, ZnO, and ZnSe are necessary for fabricating high efficiency light emitting devices. SiC and diamond are important indirect wide bandgap semiconductors, and prior to recent developments in nitrides, SiC *p-n* junctions were used for fabricating blue LEDs operating in 460–470nm wavelength range. But efforts to commercialize such SiC LEDs were thwarted due to the poor emission efficiency which is more than two orders of magnitude less than that of nitride semiconductors.

The points in Fig. 5.1 are seen to increase toward the left hand side of the graph which indicates that semiconductors with small lattice constants exhibit large bandgap energy. The small lattice constant shows that the material exhibits strong interatomic forces, with the outermost shell electrons, that are responsible for chemical bonding, being strongly bound to the lattice thus leading to large bandgap energy.

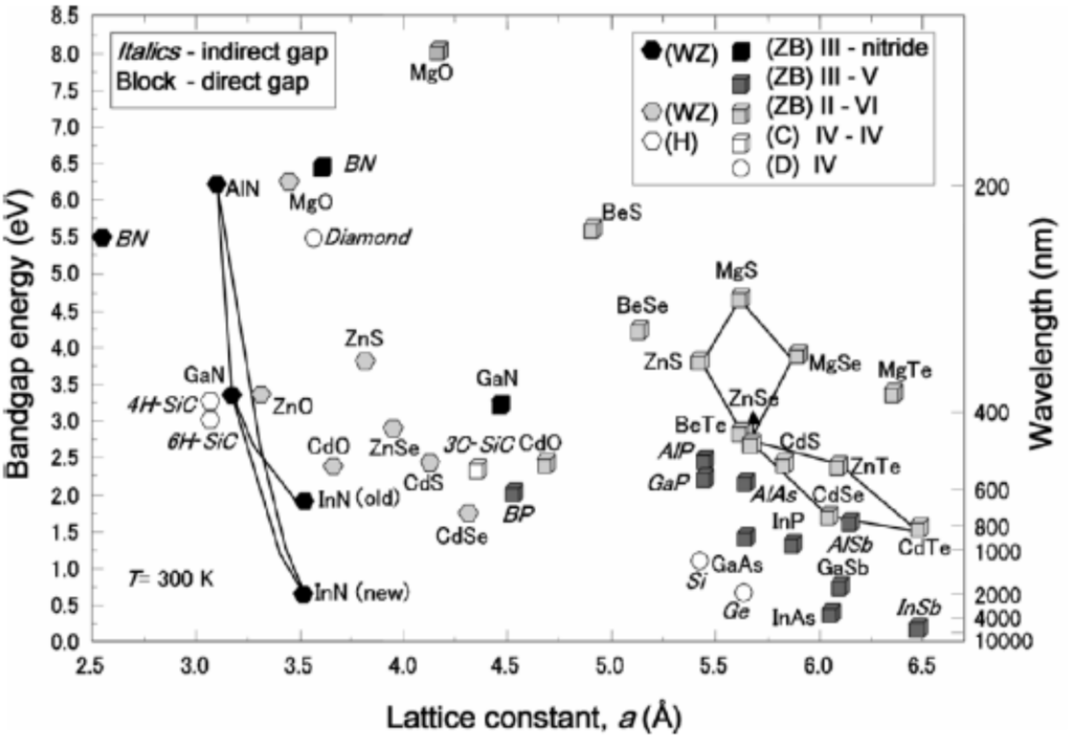


Fig. 5.1. Relationship between forbidden energy gap and lattice constant of wide bandgap semiconductors.

Compared with covalently bonded group IV elemental semiconductors, group III–V compound semiconductors exhibit slightly ionic bonding while group II–VI semiconductors exhibit greater ionic bonding.

In the case of ionic bonding, the bonding electrons are localized to the composite elements, resulting in weak bonding forces and large bandgap energies. ZnSe and ZnS are typical examples of II–VI compound semiconductors exhibiting such bonding and they have a “soft” crystalline structure that can be scratched during handling with tweezers.

Thus the bandgap energy tends to increase for compound semiconductors composed of light elements and also having high ionicity in bonding. That is, the bandgap energy increases for elements toward the top right hand region of the periodic table.

As shown in Fig. 5.1, the crystal structures of wide bandgap semiconductors are cubic (diamond and zincblende), hexagonal (wurtzite), and rock salt (NaCl). Generally, as the ionic bonding component of elements constituting compound semiconductors increases, so the interelemental attractive forces increase resulting in shorter distances between the elements with the result that crystal structures change from zincblende to wurtzite. This is the main reason that many wide bandgap semiconductors exhibit a hexagonal crystal structure. Further increases in the ionic bonding component eventually lead to rock salt crystal structures, as seen in MgO.

Overview of Optical Devices with Emission Wavelengths in the Visible Short and UV Regions.

Historically, phosphide compound semiconductors were used for fabricating short wavelength light emitting devices. For example, high efficiency LEDs emitting up to wavelengths of 580 nm can be fabricated using quaternary AlInGaP epilayers grown on GaAs or GaP substrates. For shorter wavelengths, there have been reports on nitrogen-doped GaP that has an indirect bandgap, to produce isoelectronic traps, via which bound exciton recombination paths yield green emission in GaP diode structures. However, the emission efficiency of GaN diodes at 630 nm is reduced to 1/50. Thus it is not possible to produce emission at shorter wavelengths using conventional III–V compound semiconductors.

The AlN–GaN–InN system of compound semiconductors has been extensively studied for fabricating light emitting devices for wavelengths shorter than green. As described later, it is still not possible to reproducibly produce *p*-type ZnO although there have been reports of emission from ZnO diode structures [2]. ZnSe is widely recognized as being an excellent material system for fabricating blue-green LEDs and laser diodes. In spite of the tremendous effort expended in the development of ZnSe-based devices it was found that the lifetime of ZnSe optical devices was too short for commercialization and this field of research has largely been abandoned with developments in nitride semiconductors showing greater promise. There are still groups using ZnSe for fabricating laser diodes emitting at longer wavelengths. A combination of homoepitaxial technology and new device structures are being used for the development of ZnSe-based LEDs emitting white light [3].

Figure 5.2 shows the wavelength dependence (ultraviolet to red) of the external quantum efficiency of nitride-based light emitting devices. Since blue, green, and other such semiconductor light emitting diodes are solid, robust, and highly efficient they are increasingly being used as light sources in displays and traffic signals. Further, it is also possible to fabricate short wavelength, high efficiency light sources emitting at 365 nm, corresponding to *i*-line wavelength of high pressure mercury lamps, that are widely used in industry, so the prospect of a long lifetime, solid state light source is of great industrial significance. Optical devices emitting at less than 350 nm are of interest for biological applications with increasing activity in their development. For reference, Fig. 5.2 also includes the wavelength dependence of the efficiency of AlInGaP-based LEDs. It can be seen that nitride semiconductors can be used for fabricating high efficiency light emitting devices in the short wavelength spectral range. ZnO-based light emitting devices are still in the development stage and do not compare favorably with nitrides at the present time. However, although ZnO-based devices cover almost the same wavelength range as nitrides, they are expected to enable the fabrication of lower threshold laser diode devices using excitons to operate at higher efficiencies and higher temperatures. Further advantages over nitrides include the plentiful supply of raw materials, the ease with which substrates are produced, and the wide range of device applications that could use the transparency of ZnO [4].

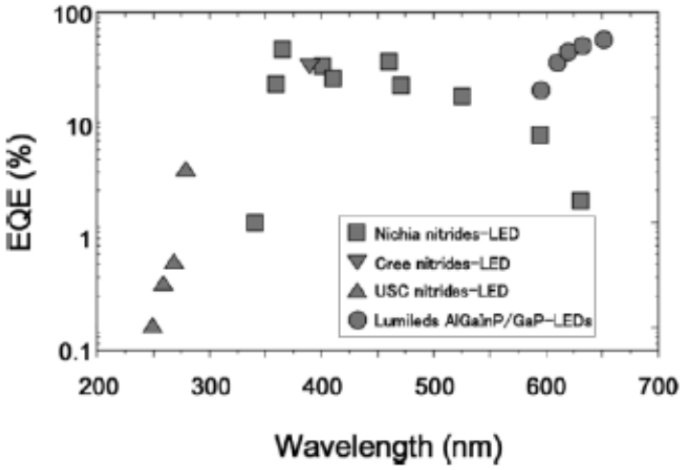


Fig. 5.2. Comparison of the wavelength dependence of the emission efficiency (external quantum efficiency) of nitride and phosphorus-based semiconductors (as of September, 2005).

Nitride-based light emitting devices are being used as white light sources in the illumination industry. The aim is to manufacture nitride devices having efficiencies higher than fluorescent phosphor light sources. The efficiencies of nitride light sources are already much higher than incandescent tungsten lamps and they are being increasingly used for localized, spot illumination and flash lights. Uniformly superimposing red, green, and blue (RGB) light sources would produce white light, but in practice it is more common to use a single light source to excite fluorescent phosphors. This is because it is easier to control the intensity of white light using a single light source than three independent RGB sources. There are two methods for producing white light using a single light source:

– Use high efficiency blue LEDs to excite fluorescent phosphors with complementary emission wavelengths with respect to blue.

– Use a UV light source to excite a fluorescent phosphor of the three primary colors.

The main issues to resolve are high efficiency and high color rendering. Nitride and ZnSe-based white diodes both have color rendering figures of merit that are similar to incandescent light sources. Nitride-based wide bandgap semiconductors are being used in high density DVD players as short wavelength laser diodes (LDs) having wavelengths in the range 405–410 nm. A range of nitride-based laser diodes have been fabricated extending continuous wave operation to wavelengths between 364 and 482 nm.

Short Wavelength Photodetectors, Modulators, and Novel Optical Devices

Ultraviolet (UV) sensors exploiting transparency in the visible region are examples of wide bandgap semiconductor photodetectors. Such solar blind UV sensors that are insensitive to visible light and respond only to UV irradiation are used as flame sensors and to detect harmful UV radiation in sunlight. High efficiency, photovoltaic type sensors have been fabricated and their simple structure has led to interest in photoconductive types of sensors as well.

Recently, the bandgap of InN has been clarified as being 0.64 eV, which corresponds to the near infra-red region and enables AlN–GaN–InN-based nitride materials to be used for fabricating devices operating from the UV to near infra-red range. These properties of nitride semiconductors are being used to develop high efficiency tandem-type solar cells.

The band offsets of wide and small bandgap semiconductor heterostructures can be several electron volts. For example, the conduction band discontinuity of AlN/GaN based heterostructures is approximately 2 eV, which is much larger than the bandgap itself of conventional semiconductors. If the thickness of the quantum well layer in such semiconductor heterostructures is grown to be only a few atomic layers, then the energy difference between the subband electron levels in the conduction band well layer is larger than the photon energy of optical communication wavelengths. The relaxation time of electrons from the excited level is extremely short at approximately 10 fs, thus enabling the possibility of fabricating inter-subband transition (ISBT) optical–optical switches and modulators.

The use of wide bandgap semiconductors in optical communication requires development of technology for the growth of superlattice structures with atomically sharp interfaces. Further, wide bandgap semiconductors could also be used for fabricating quantum cascade lasers. However, homo and heteroepitaxial growth methods for producing bulk crystals of GaN and ZnO and wide bandgap semiconductor heterostructures still need to be developed to the same level of precision and quality as for conventional group III–V compound semiconductors.

It is possible to grow wide bandgap semiconductors that are transparent in the visible spectrum, in layers of semi-insulating, semiconducting and conducting thin films. Such properties of ZnO are being studied for use as transparent displays.

5.2.2. Control of the Physical Properties of III–V Nitrides and II–VI Semiconductors

Nitride materials are environmentally friendly as well as being resistant to harsh environments. These materials contain nitrogen, a light element often participating in strong chemical bonding. Oxide materials are also composed of a light element, oxygen, and exhibit similar physical properties.

Irrespective of the choice of semiconductors, in order to fabricate high performance optical devices, it is essential to produce *p*-*n* junctions. However, it is often difficult to control both *p*- and *n*-type conduction in materials with strong chemical bonding or large bandgaps. Further, it is necessary to use multiple quantum well structures in order to improve the emission efficiency and other device characteristics, and the material properties are not always conducive to the fabrication of such structures.

Wide bandgap semiconductors have the following common physical properties and problems:

Difficulties in Controlling both *p*-Type and *n*-Type Conductivity

The development of the blue LED using wide bandgap semiconductors took a long time primarily due to difficulties encountered in making them amphoteric, i.e., controlling both the *p*- and *n*-type conduction. Wide bandgap semiconductors generally show these trends. For example, it is relatively easy to produce *n*-type GaN and ZnSe but *p*-type films were difficult to produce for a long time.

Figure 5.3 shows the theoretical models of the energy band structure of the ternary AlN–GaN–InN alloy where the Fermi level stabilizing energy, E_{fs}, is used as a reference in calculating the positions of the conduction and valence band edges [2]. The Fermi level stabilizing energy is a measure of the average dangling bond energy and shows the position where the Fermi level is pinned when defects are introduced [3]. These calculations show that controlled, *n*-type doping of GaN is easy to achieve but *p*-type is difficult. In the case of AlN, it is difficult to control both types of doping, with *p*-type being particularly challenging. Further, InN, which has recently been shown to have a bandgap of 0.64 eV, a value much smaller than the 1.9 eV previously reported, can be readily made as the *n*-type by introducing crystal defects. However, difficulties in producing the *p*-type conductivity are expected.

The results of Fig. 5.3 is also extremely important for the design of bandlineups of nitride materials. That is, the formation of heterojunctions does not depend only on bandgap differences but also on the resulting relative proportions of the conduction and valence band discontinuities. The band edge discontinuities are important parameters for forming superlattice structures.

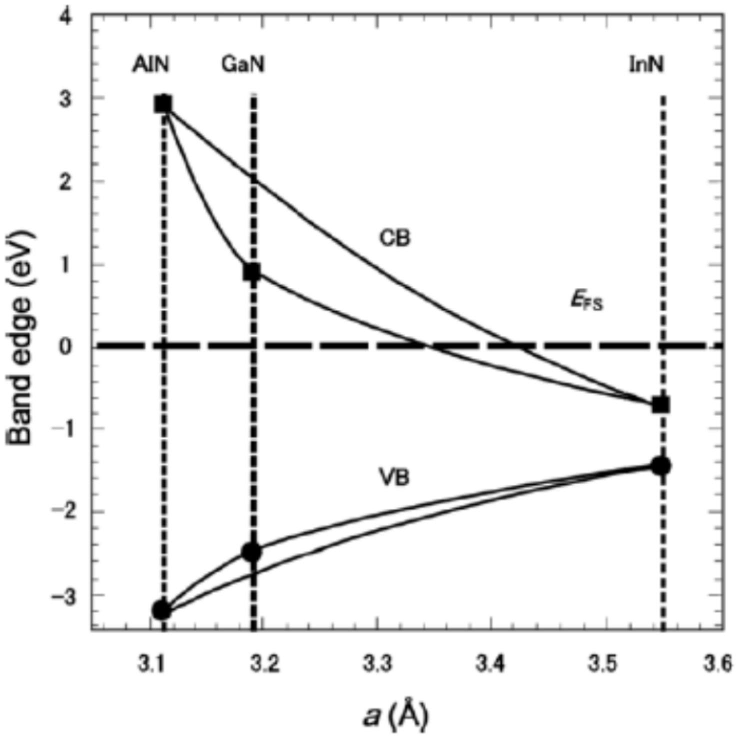


Fig. 5.3. Band line up of nitride compound semiconductors with reference to the stabilized Fermi energy of defects [2].

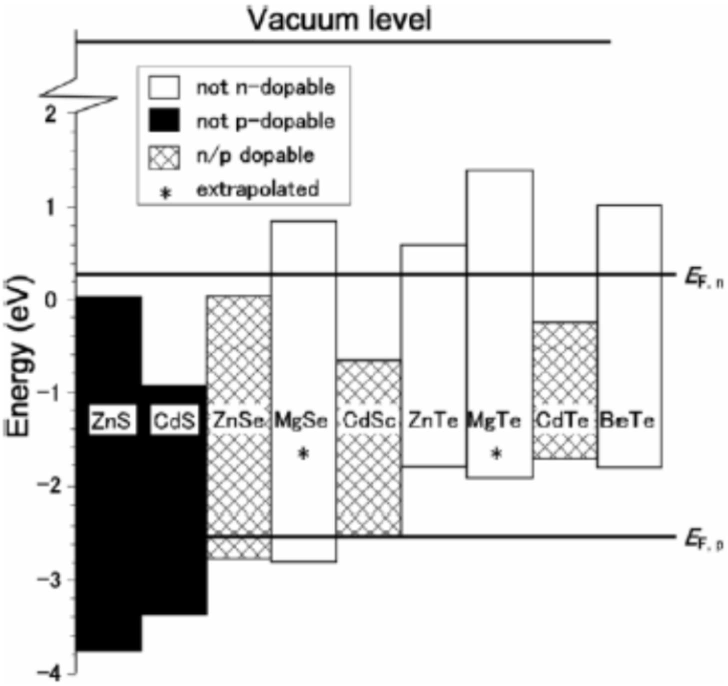


Fig. 5.4. Band line up and ease of doping of II–VI based compound semiconductors [4].

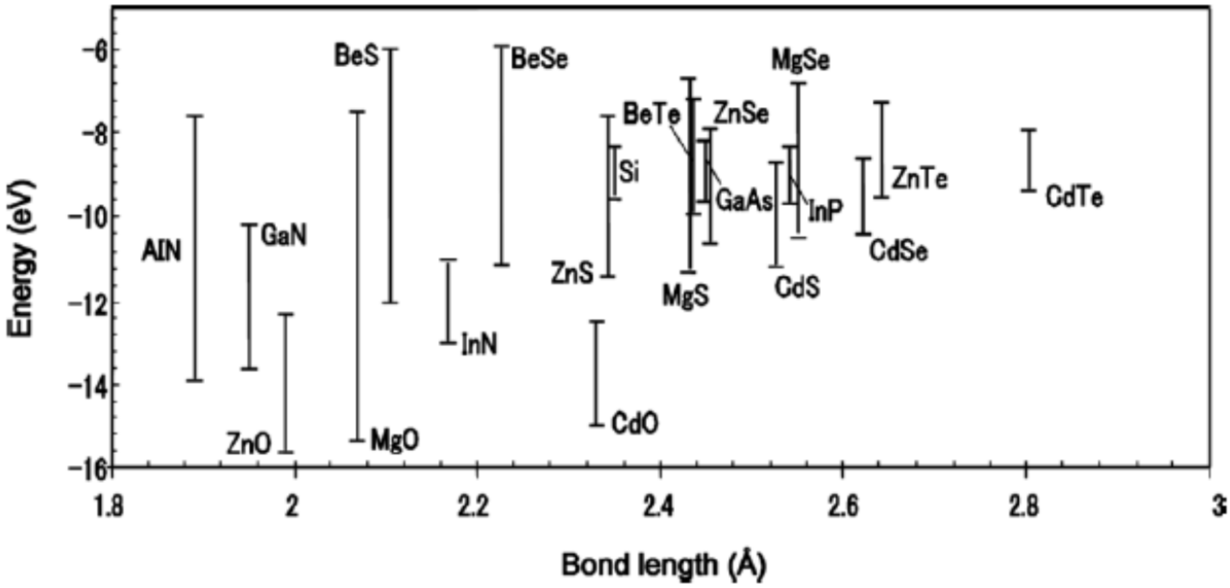


Fig. 5.5. Band line ups of a several wide bandgap semiconductors based on the Harrison model [4].

Figure 5.4 shows the band alignment and ease of doping of II–VI chalcogenide semiconductors [4], where it can be seen that it is possible to produce both *n*- and *p*-type ZnSe. The limited data for ZnO oxides are shown in Fig. 5.5 together with band alignments for a selection of other semiconductors calculated according to the Harrison model [5]. There have been many reports on the difficulties encountered in *p*-type doping of ZnO. There are also considerable hurdles to overcome in producing both *p*- and *n*-type conduction in other wide bandgap semiconductors.

Control of Material Properties and Effect of Epitaxial Crystal Structure.

The hexagonal wurtzite crystal structure is the most stable form of GaN and ZnO type compound semiconductors. This crystal structure does not have symmetry in the *c*-axis direction and is polarized in this direction. Further, nonpolar sapphire substrates are used for growing these semiconductors because of the difficulties involved in obtaining appropriate homoepitaxial substrates. Almost all device structures are epitaxially grown on *c*-plane substrates. Both GaN and ZnO grow perpendicular to the *c*-plane of the substrate where the direction of the polarity of the *c*-axis affects the processing and properties of devices. Further, as seen from Fig. 5.1, a large lattice mismatch during growth results in the generation of rotation domains (regions with rotated crystal directions) and the introduction of strain.

Effect of Crystal Polarity on Epitaxy. For the growth of GaN and ZnO on nonpolar *c*-plane sapphire substrates, the epitaxy relationship for crystalline lattice between the epilayer and substrate, in particular the polarity direction of epilayers, are uncertain. Thus a comprehensive understanding of polarity is necessary for preparing the surfaces of sapphire substrates prior to epitaxial growth. For MOCVD growth of GaN based nitrides, sapphire substrates are annealed at 1,100°C in a hydrogen atmosphere producing an excess of Al on the surface and resulting in the growth of GaN with Ga-polarity (+*c* polarity). Further, nitridation of sapphire surface results in the formation of a thin layer of AlN with a nitrogen polarity (-*c* polarity) and nitrides grown on these surfaces have N-polarity.

There are extremely large differences in the vapor pressures of elements constituting GaN-based nitride crystals. In general, +*c* polarity leads to enhanced surface diffusion of metal elements. This is used in epitaxial lateral overgrowth (ELO) technology for drastically improving crystalline quality and surface flatness. Further, polarity is also known to affect impurity incorporation mechanisms.

The epitaxial growth of ZnO-based oxides on sapphire substrates is more complicated than the growth of nitrides in spite of the presence of oxygen in both the substrate and epitaxial layer. There are several reasons for this. First, during the initial bonding of Zn or O in the ZnO layer with Al or O in the substrate, the complexity of the sapphire crystal structure and the large differences between the lattice constants of sapphire and ZnO, results in only small differences in bonding energies among several permissible epitaxy relationships.

This affects not only the polarity but also the generation of rotation domains in the ZnO. Another reason for difficulties during the growth of ZnO on sapphire is that oxide compounds are prone to grow with an amorphous structure, a trend that is more pronounced at low growth temperatures when accurate information about crystal polarity during growth becomes difficult to ascertain.

Effect of Crystal Natures and Large Lattice Mismatch on the Characteristics of Optical Devices

Several problems arise due to the large lattice mismatch between epitaxial layers and the substrates used for fabricating wide bandgap semiconductor optical devices. As a result of this extremely large lattice mismatch, GaN based blue LEDs grown on *c*-surfaces of sapphire consist of columnar crystal structures with approximately 10^{10} cm^{-2} lattice mismatch dislocations. The dislocation density in LEDs made of conventional III–V compound semiconductors is approximately five to six orders of magnitude less than that in wide bandgap semiconductors. Thus in spite of the extremely high density of dislocations, the optical emission efficiency of nitride semiconductor devices is miraculously high.

These physical properties can be explained by the short diffusion length of minority carriers. The effective mass of carriers in wide bandgap semiconductors is generally large. That is, since the diffusion length of injected minority carriers is short, it is thought that they undergo high efficiency emission transitions before reaching grain boundaries. Further, it is easy to control the electrical conduction of hexagonal wide bandgap semiconductors having even polycrystalline structures, leading to the conclusion that crystalline defects do not trap carriers in such semiconductors. Further, due to the low crystal symmetry (anisotropy), the gliding direction of the dislocation is also controllable, which enables alignment of the dislocations generated during operation in a direction parallel to the active layer. This dramatically reduces degradation of laser diodes operated at high carrier injection. The following properties enable significant improvements in device performance:

– Relatively large contribution of ionic bonding in wide bandgap semiconductors
– Low crystal symmetry
– Heavy effective mass of charge carriers
– Short diffusion length

However, crystal polarity and the large internal strain are sources of new problems in wide bandgap semiconductors. That is, an extremely large electric field arises in the quantum well of the active layer due to polarization resulting from the crystal strain. Thus the large band bending in the active layer causes spatial separation of electrons and holes, which subsequently results in a reduction of efficiency for emission, a peak shift to longer wavelengths comparing to the calculated values, and dependence of the emission wavelength on the magnitude of the current. Homoepitaxy and nonpolar growth are being studied to reduce such problems.

5.2.3. Other Characteristics and Trends of Wide Bandgap Semiconductor Optical Devices

The environment and healthcare are becoming central themes for future research and development. Conventional optical devices emitting at red and infrared wavelengths are fabricated using poisonous elements such as P and As. At the present time, there are no alternative materials for producing light sources for optical communications and information networks. On the other hand, devices made using nitride and oxide compound semiconductors are not harmful to humans and the environment, and show stable operation at elevated temperatures and under exposure to radiation. Then a number of investigations are underway for developing red LEDs and green LDs by III-nitrides. Further, research and development for solid-state UV and deep UV light source using wide bandgap semiconductors, in particular III-nitrides, are intensively undertaken all over the world now.

The large effective mass of electrons and holes in wide bandgap semiconductors could also enable realization of excitonic devices operating at room temperature if the binding energy of free excitons in the material is larger than 26 meV, the thermal energy of room temperature, 300 K. The free exciton binding energies of GaN and ZnO at room temperature are 28 meV and 60 meV, respectively, physical properties that raise expectations for room temperature exciton devices.

The physical properties of wide bandgap semiconductors are still not completely understood, thus raising the possibility for development of novel functionality devices. However, process development for fabricating nitride devices is also important. Many researchers concur that the development of growth methods for high quality and cheap GaN and ZnO bulk crystals and/or substrates will become imperative for major advances in this field.

5.3. Development and Applications of Wide Bandgap Semiconductors. Electronic Devices

5.3.1. Silicon Carbide Electronic Devices

Expectation for Wide Bandgap Semiconductors

Electronics based on semiconductors has directly contributed to information, energy and global environment [5–9]. Electronics is the core of information technology. Computers and their networks have spread all over the world, which will promote further development of our society. The semiconductors which have been used in the progress of electronics are silicon (Si), gallium arsenide (GaAs) and related III–V materials. The energy bandgaps of these semiconductors are 1.12 eV for Si and 1.43 eV for GaAs, which are mid-range values. According to the progress of electronics, strong demand to use the technology has long been continued. However, those devices from semiconductors with mid-range bandgaps cannot be used for a wide range of applications owing to the performance limit imposed by the size of the bandgap. Nowadays, wide bandgap semiconductors take strong attention for electronic devices such as high-power, high-frequency/high-power, robust devices under harsh environment such as high temperature and heavy radiation influence, optoelectronic devices for ultraviolet detection, and light-emitting devices in the short wavelength region.

Power Devices and Limits of Si Technology

The consumption of electric power has steadily been increasing by the rapid shift to the age of information. From the viewpoint of environmental protection, we should not solve the problem simply by increasing electric power supply. Cutting down the consumption of electric energy and its efficient use are urgently required, and especially highly-efficient use of electric power is very important, because it enables us to decrease the consumption of electric energy remarkably.

Many power electronic devices are used at important places such as voltage and frequency control for transmitting and converting electric power. Accordingly, reduction in loss and performance improvement of these devices will directly lead to considerable reduction in electric power consumption. Present power semiconductor devices, for example inverters, are all made of Si. The performance of Si power devices has been improved using microfabrication technology that has aided the progress of Si VLSI (very large scale integration). But now, the performance of Si power devices is approaching its limit due to physical properties, energy bandgap, and we cannot expect dramatic development in Si power devices.

Expectations of SiC

Table 5.1 shows the physical properties and figures of merit of Si, GaAs, and SiC, together with other important requirements in device processes. The large energy bandgap, breakdown field strength, saturation electron velocity and thermal conductivity (similar to Cu) of SiC enables this material to be used for fabrication of devices that can be operated at high voltages and large drive currents, in high frequency, and at high temperatures.

Conductive and semi-insulating SiC substrates can be obtained. The conductivity of SiC layers can be readily controlled over a wide range, 10^{14}–10^{19} cm^{-3} both for *p*- and *n*-types by impurity doping. Further, the surface of SiC can be covered with high-quality oxide layers by thermal oxidation, an essential factor for device fabrication.

Figure 5.6 shows a comparison of the characteristics of Si and SiC Schottky diodes, majority-carrier power devices. The upper figure is a comparison of the device length, which shows that the one order of magnitude larger breakdown field strength of SiC enables a reduction of its device length to 1/10 of that required for Si. The lower figure is a comparison of the relation between drift layer thickness and electric field strength, where the gradient of the straight line shows the impurity concentration; it is possible to introduce two orders of magnitude more impurities into SiC than Si.

Table 5.1. Properties of Si, GaAs, and SiC (4H).

	Si	GaAs	SiC (4H)
Energy gap (eV)	1.12	1.43	3.26
Electron mobility (cm^2 V^{-1} s^{-1})	1,350	8,000	1,000
Breakdown field (MV cm^{-1})	0.3	0.4	3.0
Saturation electron velocity (cm s^{-1})	1×10^7	1×10^7	2×10^7
Thermal conductivity (W $cm^{-1}K^{-1}$)	1.5	0.5	4.9
Johnson figure of merit	1	1.8	400
Baliga figure of merit	1	15	603
Conductivity control	easy	easy	easy
Oxidization	easy	difficult	easy
Conducting wafer	yes	yes	yes
Semi-insulating wafer	no	yes	yes

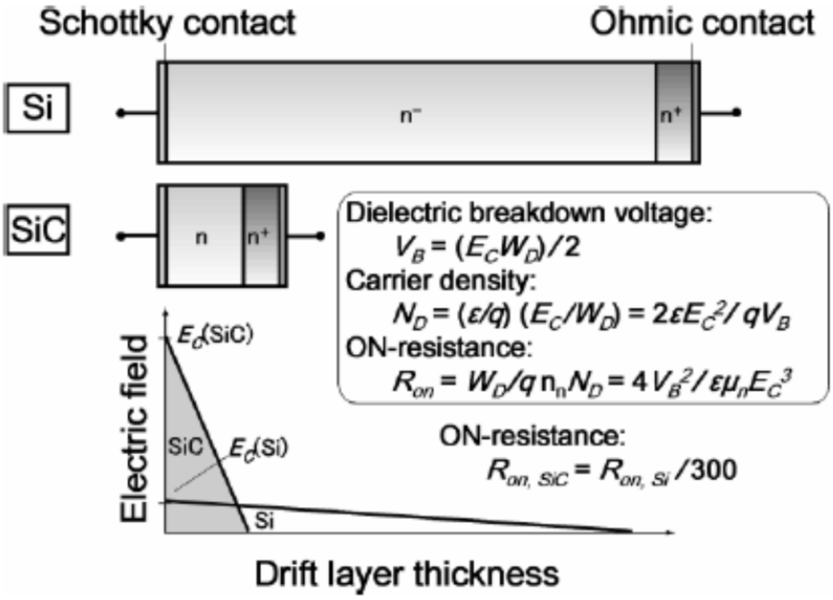


Fig. 5.6. Comparison of performance of Si and SiC devices in the case of Schottky diodes.

Because SiC has electric breakdown field strength of about ten times larger than that of Si, SiC power electronic devices can be made one tenth thinner than Si. And because handling current density is larger, the SiC devices can be smaller over all. In the case of popularly used MOSFETs (metal-oxide-semiconductor field effect transistors) for power electronic devices, the specific on-resistance that causes thermal loss during power conversion is inversely proportional to the third power of the electric breakdown field strength; consequently the on-resistance of SiC

MOSFET becomes smaller than that of Si by two or more orders of magnitude. Also, because the thermal conductivity of SiC is about three times that of Si, the equipment using SiC power electronic devices can use simple cooling systems. Thus, the future of SiC is very promising because of its smaller size, lower loss, higher efficiency, and easier heat-dissipation.

The physical properties put SiC power devices to practical use with power losses of less than 1/300 that of Si power devices. High-frequency operation of SiC devices can solve the unpleasant noise in low-frequency operation of present-day Si power devices. The upper limit of operation temperature for SiC power devices can be increased up to 300°C rather than at 150°C for Si power devices. That is, SiC offers potential for miniaturization, low power losses, high efficiency and simple cooling equipment.

Applications of SiC Power Devices

SiC Schottky diodes reported by Japanese researchers in 1993 and 1995 are commercialized in 2001, and now available as commercial products. They are slowly replacing Si *p-n* diodes in switching power supplies as fast switching diodes. Nowadays switching transistors are being widely investigated in the world. There are also signs of progress in improving electron mobility in the inversion channel of SiC MOSFETs. In a few years, normally-off MOSFETs will probably become commercially available. Devices for small power capacity with small area capable of handling several amperes are at the stage of being fabricated on 3-in. substrates for commercial production. The basic technological infrastructure for fabricating devices with larger power capacity handling several tens and above hundred amperes for power conversion (converters and inverters) is being prepared.

Power electronics encompasses a very wide range of applications including electrical and electronic equipment, household appliances, industrial production equipment, uninterrupted power supplies, tractions for trains, and high voltage DC transmission equipment. Use of SiC power electronic systems with high efficiency and simple cooling for these applications will strongly reduce electric energy consumption.

Further, there will be many major changes in the automobile industry with the introduction of hybrid electric cars based on conventional gasoline engine with electric motors and, probably in future, supported by fuel cells, which can reduce exhausted gases and reliance on fossil fuels. Considering a limited space of trunk, existence of high temperature around engine, and long-term use of batteries in automobiles, a big revolution in power electronic system is strongly expected. A new power electronic system, in which SiC power devices is the central core, is the best suited technology.

Figure 5.7 shows the main potential applications of power semiconductor devices. As shown in the figure, in the present Si power electronics, various types of power devices are used in individual applications, caused by historical step in the development of each device. SiC MOSFETs will replace Si IGBT (insulated gate bipolar transistor) technology, because the power loss in SiC MOSFETs is two orders of magnitude less compared with Si MOSFETs, which covers Si IGBT power ranges. In addition, high-frequency switching is possible with MOSFETs than IGBTs owing to the majority carrier operation. It is not a simple replacement, but it gives advantages such as miniaturization, low loss, and simple cooling system.

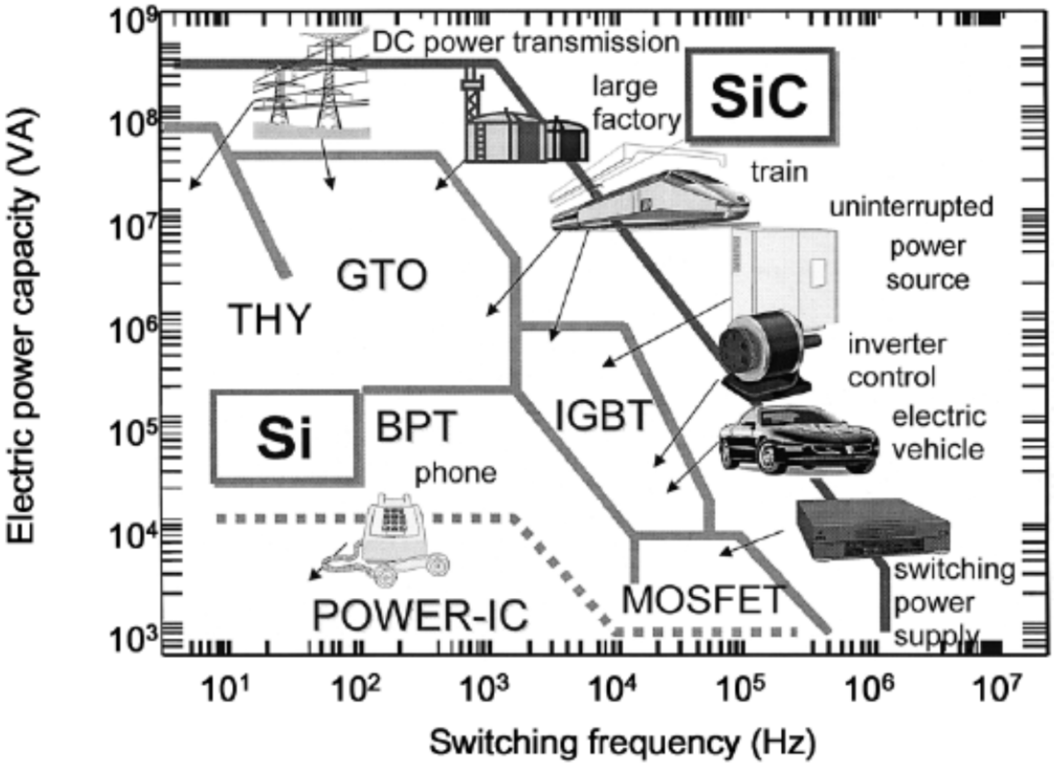


Fig. 5.7. Types of power devices

In communication, due to the generational change of cellular phones, new base stations operating at high capacity and high speed are being constructed continuously. There is also strong interest in the development of SiC MESFETs (metal-semiconductor field effect transistors) as compact, high power output microwave oscillators operating at less than 10 GHz for wireless base stations. Power outputs of several ten watts for these devices have been reported and field tested.

5.3.2. Nitride Compound Semiconductor Electron Devices

Nitride semiconductors are of interest for more effective device solutions with respect to energy and the environment, as well as in the development of the information technology industry. This section reviews applications of nitride semiconductors and highlights the potential problems and trends. GaN as a representative of its binary cousins, InN and AlN, and their ternaries along with the quaternary, is considered one of the most important semiconductors after Si. It is no wonder that it finds ample applications in lighting and displays of all kinds, lasers, detectors, and high-power amplifiers. These applications stem from the excellent optical and electrical properties of nitride semiconductors.

Applications of Nitride Semiconductors

Figure 5.8 shows the device frequencies expected to be used for portable phones, satellite communication, wireless systems, and intelligent transportation systems (ITS). With increasing frequency, f, wave propagation losses generally show an approximate $\sim f^{2.6}$ dependence. Further, the high frequency power increases linearly with increases in the volume of information. On the other hand, an extremely high degree of linearity is required in order to stand up to intense modulation and demodulation processes. These types of systems can only be realized with high frequency devices having high power output, high efficiency, good linearity, and low power consumption. To date, high frequency and low noise devices, with relatively low power output have been fabricated using conventional compound semiconductors such as GaAs and InP, but wide bandgap nitride semiconductors are required for high power output devices operating in the GHz range.

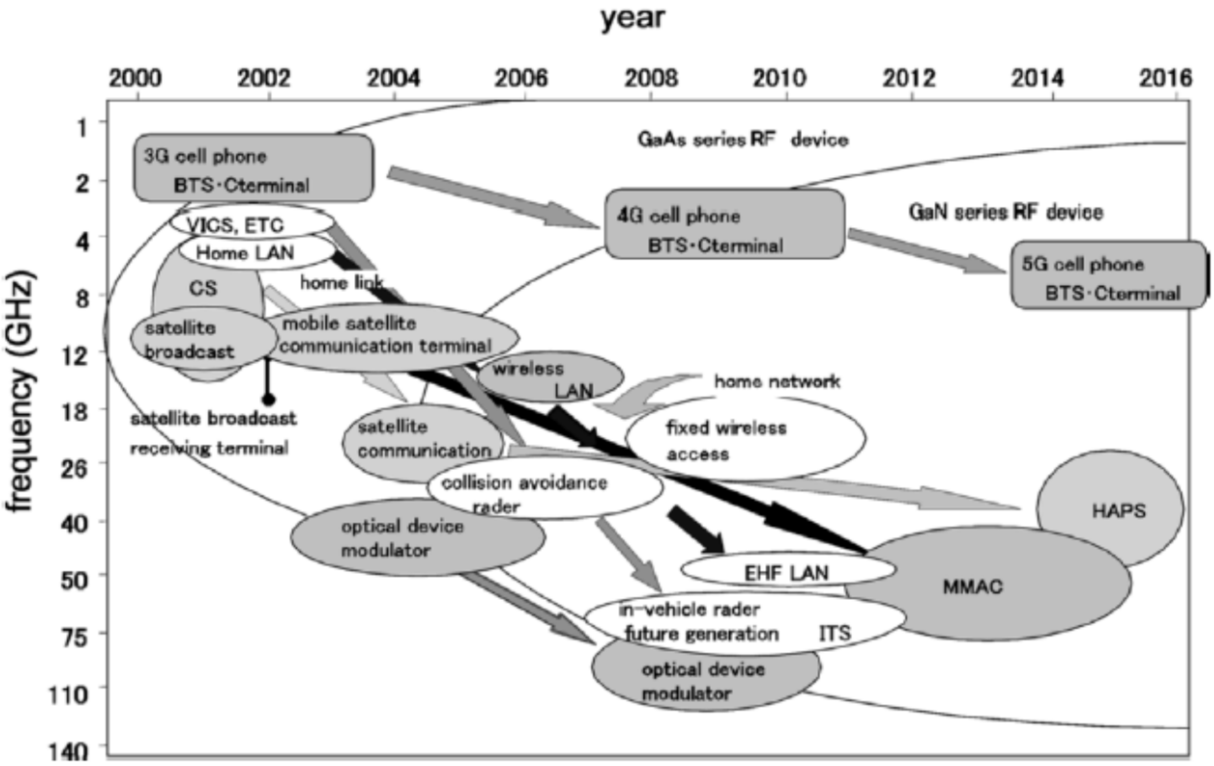


Fig. 5.8. Wireless systems and related operating frequency ranges.

Characteristics of Nitride Semiconductors and Electron Devices

The Ga–N atomic bonding is strong because GaN is composed of N, a light element. As a result the lattice constant of nitrides is smaller than conventional III–V semiconductors. Large, bulk crystals of GaN cannot be supplied as lattice matched substrates for epitaxial growth. This is a major difference compared with SiC device technology. However, the strong atomic bonding results in nitrides being thermally, chemically, and mechanically robust. Nitrides also show greater ionicity than conventional III–V semiconductors such as GaAs and exhibit spontaneous polarization due to lack of reverse symmetry along the *c*-axis of the hexagonal wurtzite crystal structure. Additional piezoelectric polarization arises when strain is introduced into the crystal structure.

Table 5.2 shows the main physical properties of commonly used wide bandgap and conventional III–V compound semiconductors. At room temperature, GaN has a large bandgap of 3.4 eV and a breakdown electric field of 3.3×10^6 V cm^{-1}, which is 8–10 times larger than GaAs. Further, the saturation drift velocity of GaN is 2.5×10^7 cm s^{-1}, approximately 1.2 times larger than that of GaAs. The thermal conductivity is a little less than that of SiC, but 4 times larger than that of GaAs.

Table 5.2. Physical properties of widely used semiconductors.

Material	Band-gap (eV)	Relative dielectric constant	Effective electron mass	electron mobility ($cm^2V^{-1}s^{-1}$)	Breakdown field (10^6 V cm^{-1})	Saturation electron velocity	Thermal conductivity	Transition *n* type (10^7 cm^{-1})
InN	0.6 - 0.7	15.0	0.04	4,000	2.0	4.2	0.8	direct
GaN	3.4	9.5	0.22	1,200	3.0	2.5	2.1	direct
AIN	6.2	8.5	0.29	–	–	2.0	2.9	direct
Si	1.1	11.8	0.19	1,500	0.3	1.0	1.5	indirect
GaAs	1.4	12.8	0.067	8,500	0.4	2.0	0.5	direct
4H–SiC	3.3	10.0	0.3	1,000	3.0	2.0	4.9	indirect
Diamond	5.5	5.5	0.2	1,800	4.0	2.5	20.9	indirect

One of the main features of nitrides when compared with SiC is the possibility of fabricating heterostructures having large band discontinuities using AlGaN and AlInGaN. A two-dimensional electron gas (2DEG) is formed at this heterointerface due to spontaneous and piezoelectric polarization. The presence of the 2-DEG leads to a large drift velocity and a relatively high electron mobility. These structures are used to fabricate heterostructural FETs (HFET) which can be driven at large current densities. The large breakdown electric field implies that a large electric field can be applied without damaging the devices. Thus the capacity to pass a large drive current and to apply large voltage enables fabrication of HFETs capable of large output power. Further, assuming a constant applied voltage, devices fabricated with shorter gate lengths would further enhance the electric field enabling ultrahigh frequency operation.

In addition to the high power output and high frequency advantage over GaAs, nitride semiconductors and those dopants contain elements that are less harmful to the environment.

Due to the large band offset and the effects of spontaneous and piezoelectric polarization at the AlGaN/GaN heterostructure interface, it is relatively easy to produce structures with sheet electron densities of $(1 - 3)\times10^{13}$ cm^{-1} for Al compositions of 30%. This electron density is 4–5 times larger than in AlGaAs/GaAs heterostructures. Also, the saturation drift velocity of GaN is 1.2 times that of GaAs. The maximum current density of HFETs is proportional to the product of the carrier density and the saturation drift velocity. A first-order calculation shows that the maximum current density that is possible per unit gate width for AlGaN/GaN HFETs is 5–6 times larger than for AlGaAs/GaAs structures. Also, it is possible to apply a voltage between the source and drain that is approximately one order of magnitude larger than in GaAs-based structures. The high frequency output power density per unit gate width of Class A amplifiers is given by

$$P_{OUT} = 1/8 \times I_{DS} \times V_{DS} \quad (5.1)$$

where, V_{DS} is the voltage applied between the source and drain and I_{DS} is the maximum current density. From this relationship, it can be seen that the P_{OUT} of an AlGaN/GaN HFET is expected to be 50–60 times larger than for AlGaAs/GaAs devices. The figure of merit defined by FM = $P_A f_{10}^2 R_L$ is also used in high frequency, high output power amplifiers. Here, P_A is the output for class A amplifiers; f_{10} is the frequency at 10 dB gain; R_L is the load resistance. For GaN, FM is expected to be 1.2×10^{23} WHz$^2\Omega$, a value 56 times that of GaAs. The performance of high output power, high frequency, low loss devices can also be defined by the Johnson and Baliga figures of merit. As shown in Table 5.2, the values of the breakdown electric field, saturation drift velocity and electron mobility of nitride semiconductors are greater than SiC, again highlighting the potential of nitrides for power electronics [10, 11]. Figure 5.9 shows the frequency and output power of electron devices fabricated using a wide selection of semiconductors.

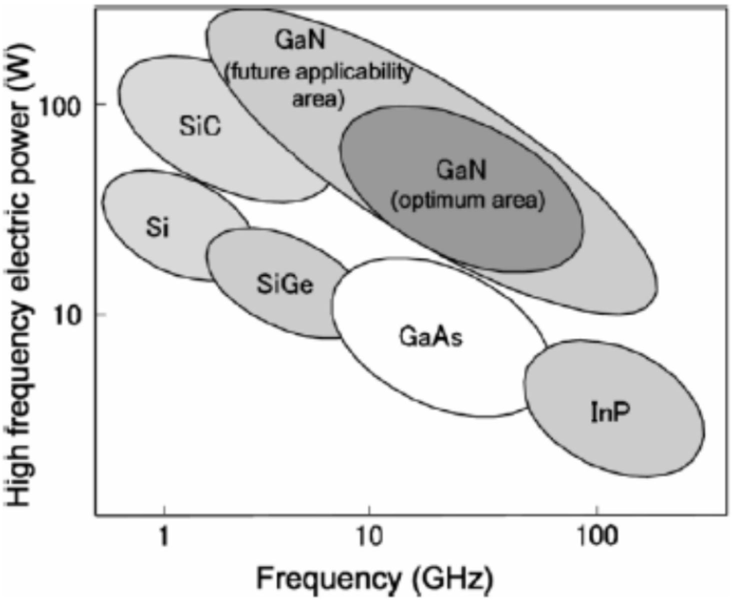


Fig. 5.9. Operating frequencies and output power of electron devices made using a range of materials.

Figure 5.10 shows the variation of output power with frequency.

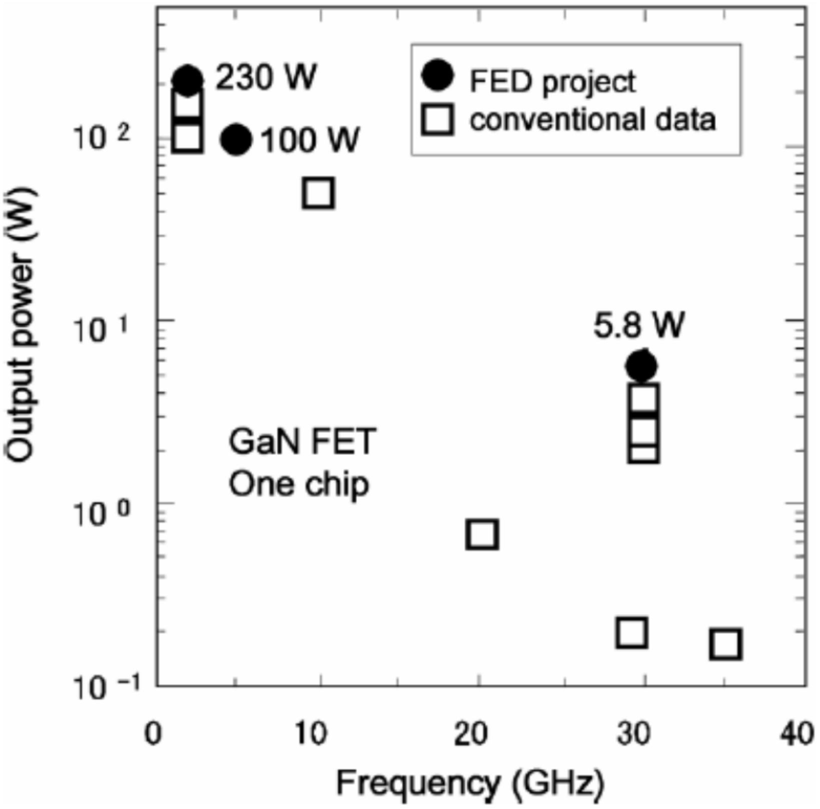


Fig. 5.10. Relationship between the R_F power output and operating frequency.

5.4. Crystal Structure of Nitrides

Group III nitrides can be of crystalline structures: the wurtzite (Wz), zinc blende (ZB), and rock salt. Under ambient conditions, the thermodynamically stable structure is wurtzite for bulk AlN, GaN, and InN. The zinc blende structure for GaN and InN has been stabilized by epitaxial growth of thin films on {011} crystal planes of cubic substrates such as Si [12], SiC [13], MgO [14], and GaAs [15]. In these cases, the intrinsic tendency to form the Wz structure is overcome by the topological compatibility. However, Wz structure could very likely be present at the extended defect sites. The rock salt form is possible only under high pressures and, therefore, is laboratory form of exercise.

Let us now discuss the space groups for the various forms of nitrides. The rock salt, or NaCl, structure (with space group $Fm\bar{3}m$ in the Hermann–Mauguin notation and O_h^5 in the Schoenflies notation) can be induced in AlN, GaN, and InN under very high pressures. The reason for this is that the reduction of the lattice dimensions causes the interionic Coulomb interaction to favor the ionicity over the covalent nature. The structural phase transition to rock salt structure was experimentally observed at the following estimated pressure values: 22.9 GPa (17 GPa in other estimates) for AlN [16], 52.2 GPa for GaN [17], and 12.1 GPa for InN [18]. Rock salt III nitrides cannot be produced by any epitaxial growth.

The space grouping for the zinc blende structure is $F\bar{4}3m$ in the Hermann– Mauguin notation and T_d^2 in the Schoenflies notation. The zinc blende structure has a cubic unit cell, containing four group III elements and four nitrogen elements. (Although the term zinc blende originated in compounds such as ZnS, which could be in cubic or hexagonal phase, it has been used ubiquitously for compound semiconductors with cubic symmetry. The correct term that should be used for the cubic phase of GaN is actually sphalerite. The position of the atoms within the unit cell is identical to the diamond crystal structure. Both structures consist of two interpenetrating face centered cubic sublattices, offset by one quarter of the distance along a body diagonal. Each atom in the structure may be viewed as positioned at the center of a tetrahedron, with its four nearest neighbors defining the four corners of the tetrahedron. The stacking sequence for the (111) close-packed planes in this structure is AaBbCc. Lowercase and uppercase letters stand for the two different kinds of constituents. The wurtzite structure has a hexagonal unit cell and thus two lattice constants, *c* and *a*. It contains six atoms of each type. The space grouping for the wurtzite structure is $P6_3mc$ in the Hermann–Mauguin notation and C_6^4 in the Schoenflies notation. The point group symmetry is *6mm* in the Hermann–Mauguin notation and C_{6v} in the Schoenflies notation. The Wz structure consists of two interpenetrating hexagonal close-packed (hcp) sublattices, each with one type of atom, offset along the c-axis by 5/8 of the cell height (5*c*/8). The wurtzite and zinc blende structures are somewhat similar and yet different. In both cases, each group III atom is coordinated by four nitrogen atoms. Conversely, each nitrogen atom is coordinated by four group III atoms. The main difference between these two structures lies in the stacking sequence of closest packed diatomic planes. The Wz structure consists of alternating biatomic close-packed (0001) planes of Ga and N pairs, thus the stacking sequence of the (0001) plane is AaBbAa in the (0001) direction.

Although the main interest is in Wz GaN as opposed to zinc blende GaN, a description of stacking sequence of both GaN polytypes with the accepted Ramsdel notation is warranted, so is the stacking order of SiC polytypes that are relevant to GaN because they are used for substrates in GaN epitaxy. The bonds describe a tetrahedron denoted by *T*, which has one atom species at each of the three corners and the other atom species in its center [19, 20]. The basal plane of this structure is defined by one face of the tetrahedron and the bond perpendicular to this plane defines the *c*-axis. A rotation of 180 ° around the *c*-axis produces a twin variant denoted by T$^{/}$ as shown in Figure 5.11a (left).

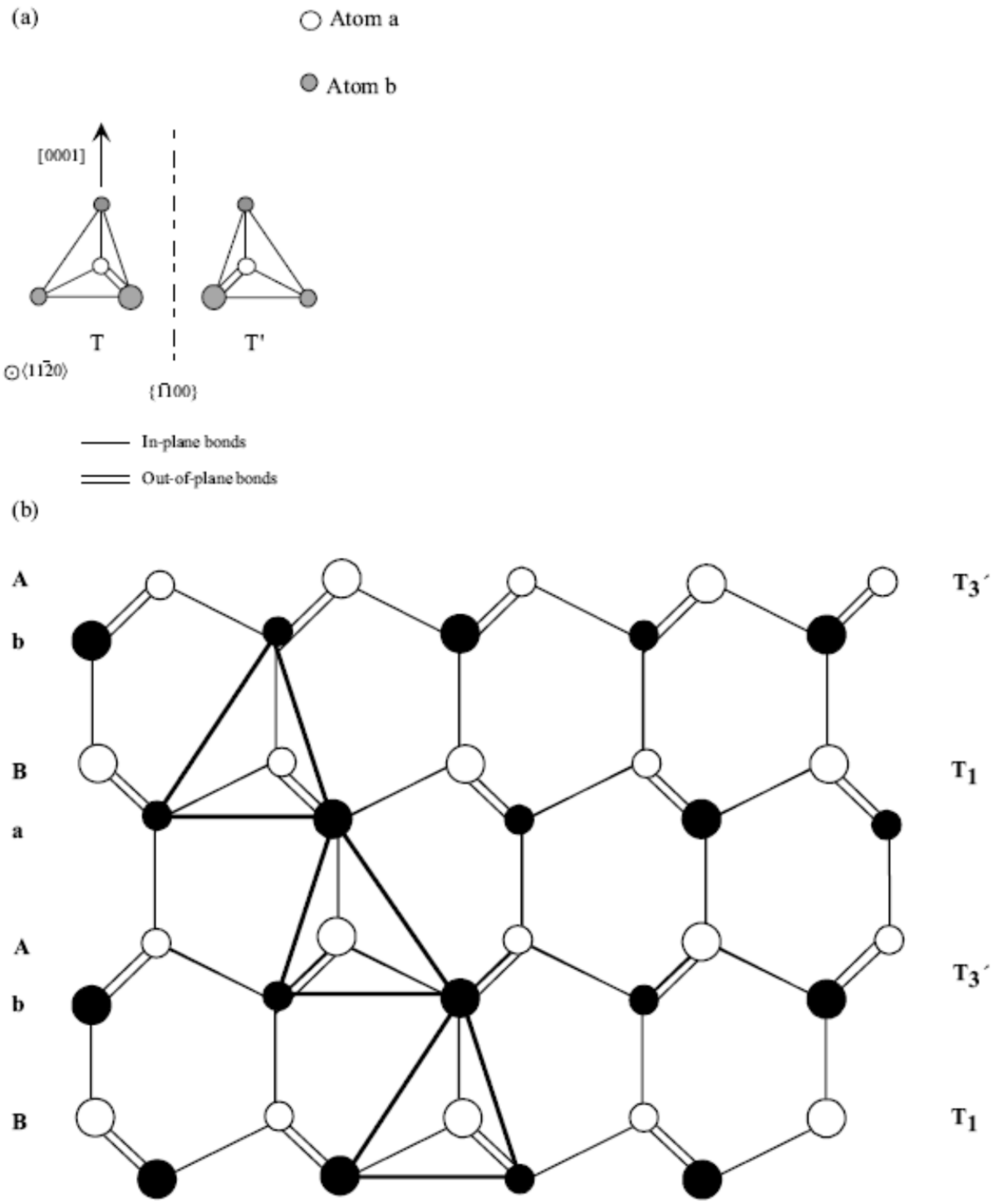


Fig. 5.11. Representation of the tetrahedrally coordinated materials in the Ramsdel notation. (a) The two possible tetrahedra. (b) The T_1, $T^{/}_3$, T_1, $T^{/}_3$, tetrahedral stacking composing the 2H sequence [20].

The two variants (twins T and $T^{/}$) are related to one another by mirror symmetry about one of the {1 1^{-} 0 0} *m*-planes. A tetrahedron can occupy one of the three possible positions in the basal plane. The representation of the tetrahedrally coordinated materials in the Ramsdel notation is shown in Figure 1.1a for two possible tetrahedra, one is the mirror image twin of the other with respect to the (1 1^{-} 0 0) *m*-plane. The single bonds are on the (1 1 2^{-} 0) plane, called the *a*-plane. The layers of the tetrahedra can then be denoted by T_1, T_2, T_3, and by T'_1, T'_2, T'_3 An example of T_1, T', T_1, T'_3 stacking order representing 2H ordering as in wurtzitic GaN is shown in Figure 1.1b. The structure of nitride semiconductors and most relevant polytypes of SiC can be completely described by a combinatorial stacking of the aforementioned six tetrahedra layers. Naturally, not all the stacking sequences must obey the following two rules to keep a corner sharing structure, as such not all stacking orders are allowed:

(*i*) A tetrahedron T can be followed by another one of the same kind with the following subscript: $T_1T_2T_3$, and inversely for the twin variant: $T'_3T'_2T'_1$.

(*ii*) A tetrahedron T1 must be followed by the twin variant of the preceding subscript: $T_1T'_3$, and inversely for its twin variant: T'_1T_2.

Recall that GaN crystallizes in the cubic structure (zinc blende or sphalerite, the latter being the correct term and the former being the one used universally) or in the more stable hexagonal structure (wurtzite). The anions (N^{3-}) form an hcp structure in which the cations (Ga^{3+}) occupy half of the tetrahedral sites. The structure of a unit cell of GaN projected along [0 0 0 1] is depicted schematically in Figure 5.12. The open symbols represent γ sites that are occupied by nitrogen atoms; the Ga atoms are in the tetrahedral sites, β. These latter sites can either be at heights (3/8)*c* above (β_1) or below (β_2) N site, depending on the crystal polarity.

A stick-and-ball representation of Ga-polarity and N-polarity Wz structure is depicted in Figure 5.13. The Wz and zinc blende structures differ only in the bond angle of the second nearest neighbor (Figure 5.14). As clearly shown, the stacking order of the Wz along the [0 0 0 1] *c*-direction is AaBb, meaning a mirror image but no in-plane rotation with the bond angles. In the zinc blende structure along the [1 1 1] direction, there is a 60° rotation that causes a stacking order of AaBbCc. The point with regard to rotation is illustrated in Figure 5.14b.

The Wz group III nitrides lack an inversion plane perpendicular to the *c*-axis; thus, nitride surfaces have either a group III element (Al, Ga, or In) polarity (referred to as Ga-polarity) with a designation of (0 0 0 1) or (0 0 0 1)A plane or a N-polarity with a designation of (0 0 0 1^{-}) or (0 0 1)B plane. The distinction between these two directions is essential in nitrides because of their implications for the polarity of the polarization charge.

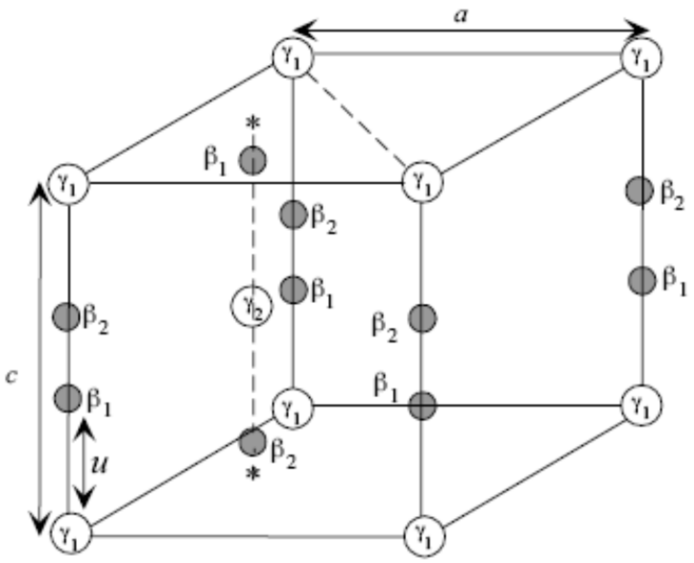


Fig. 5.12. Schematic diagram showing the β_1 and β_2 tetrahedral sites of GaN unit cell. Starting with the assumption that N occupies the γ sites, only one family of β sites can be simultaneously occupied by Ga atoms [20].

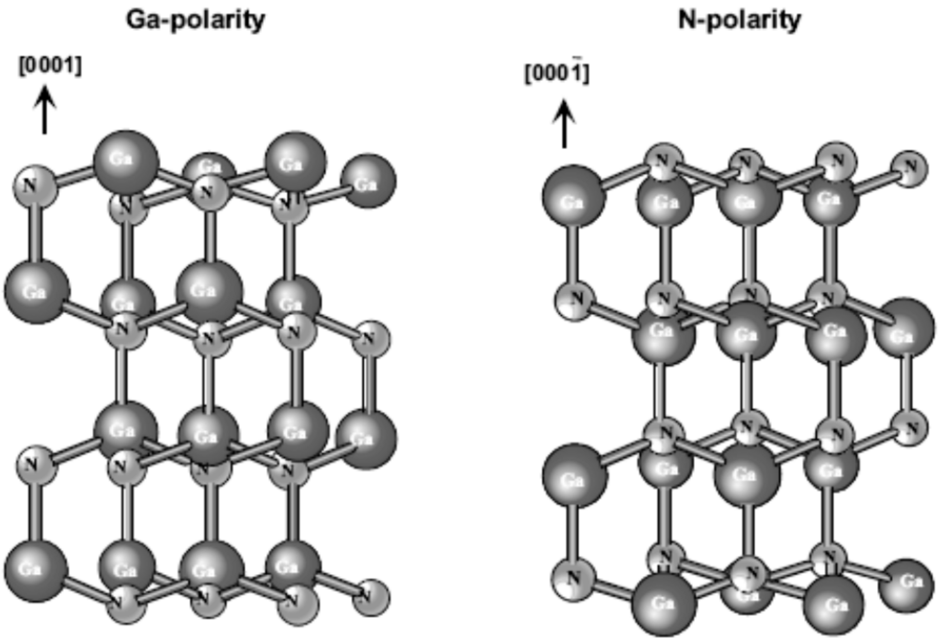


Fig. 5.13. A stick-and-ball diagram of a hexagonal structure.

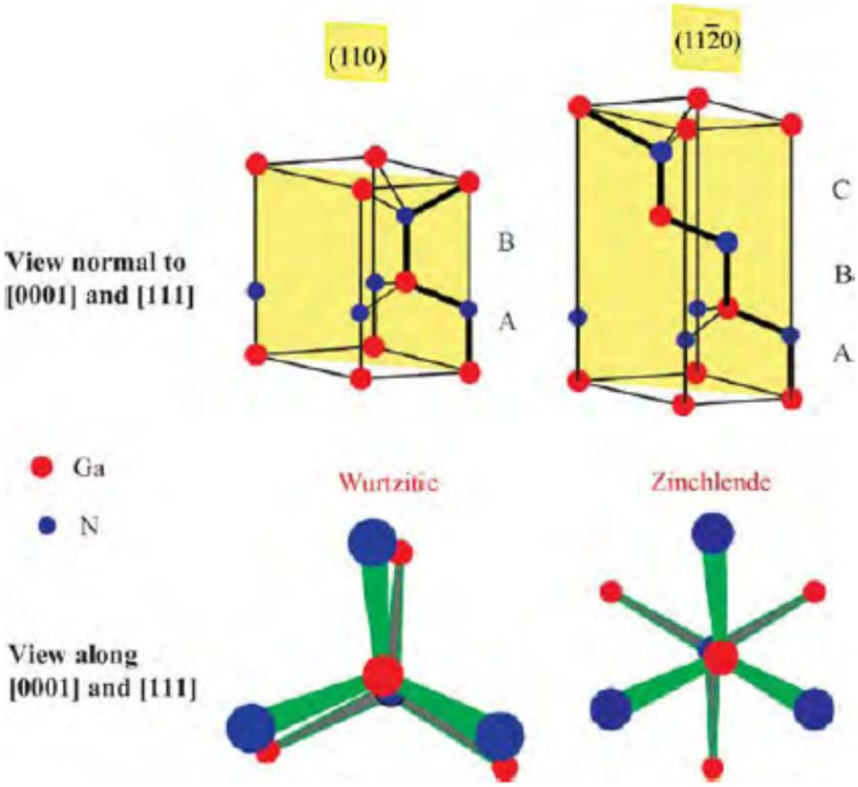


Fig. 5.14. A stick-and-ball stacking model of crystals with (a, both top and bottom) 2H wurtzitic and (b, both top and bottom) 3C zinc blende polytypes. The bonds in an A-plane (1 1 2$^-$ 0) are indicated with heavier lines to accentuate the stacking sequence. The figures on top depict the three-dimensional view. The figures at the bottom indicate the projections on the (0 0 0 1) and (1 1 1) planes for wurtzitic and cubic phases, respectively. Note the rotation in the zinc blende case along the <111> direction.

The cohesive energy per bond in the wurtzite form is 2.88 eV (63.5 kcal mol^{-1}), 2.2 eV (48.5 kcal mol^{-1}), and 1.93 eV (42.5 kcal mol^{-1}) for AlN, GaN, and InN, respectively [21]. The calculated energy difference DEW-ZB between wurtzite and zinc blende lattice is small [22]: $\Delta E_{W\text{-}ZB}$ =-18.41 meV/atom for AlN, $\Delta E_{W\text{-}ZB}$ =-11.44 meV/atom for InN, and $\Delta E_{W\text{-}ZB}$ =-9.88 meV/atom for GaN. Wurtzite form is energetically preferable for all three nitrides compared to zinc blende, although the energy difference is small.

Table 5.3 lists a comparison of measured and calculated lattice parameters reported for AlN, GaN, and InN crystallized in the wurtzite structure in more detail in terms of the specifics of the sample used for measurements and complements. The dispersion is even a greater concern in ternary and quaternaries, as compositional inhomogeneities, in addition to the aforementioned issues, cause an additional dispersion.

The wurtzite polytypes of GaN, AlN, and InN form a continuous alloy system whose direct bandgaps range, according to data that adorned the literature for years, from 1.9 eV for

InN, to 3.42 eV for GaN, and to 6.2 eV for AlN. A revisit of the InN bandgap indicates it to be about 0.78 eV [23] and the same for AlN is about 6 eV in which case the energy range covered would be about 0.7–6 eV. Thus, the III–V nitrides could potentially be fabricated into optical devices, which are active at wavelengths ranging from the red well to the ultraviolet. The bandgaps of nitrides, substrates commonly used for nitrides, and other conventional semiconductors are shown in Fig. 5.15 with respect to their lattice constants.

Table 5.3. Measured and calculated lattice constants of wurtzite AlN, GaN, and InN.

Compound	**Sample**	***a*(A°)**	***c*(A°)**
AlN	Bulk crystal [24]	3.1106	4.9795
	Powder [25]	3.1130	4.9816
	Epitaxial layer on SiC [26]	3.110	4.980
	Pseudopotential LDA [27]	3.06	4.91
	FP-LMTO LDA [28]	3.084	4.948
GaN	Bulk crystal [29]	3.189	5.1864
	Relaxed layer on sapphire [30]	3.1892	5.1850
	Powder [31]	3.1893	5.1851
	Relaxed layer on sapphire [32]	3.1878	5.1854
	GaN substrate – LEO [33]	3.1896	5.1855
	Pseudopotential LDA [27]	3.162	5.142
	FP-LMTO LDA [28]	3.17	5.13
InN	Powder [31]	3.538	5.703
	Pseudopotential LDA [27]	3.501	5.669
	FP-LMTO LDA [28]	3.53	5.54

LDA: local density approximation; FP-LMTO: full-potential linear muffin–tin orbital.

All III nitrides have partially covalent and partially ionic bonds. The concept of fractional ionic character (*FIC*) is useful in interpreting many physical phenomena in the crystals [34, 35]. The *FIC* may be defined for a binary compound AB as $FIC = |Q_A^* - Q_B^*| / |Q_A^* + Q_B^*|$, where Q_A^* and Q_B^* are effective charges on atoms A and B. The *FIC* values range from zero for a covalent compound (each atom has four electrons) to 1 for an ionic compound (all eight electrons belong to the anion). Figure 5.16 displays the charge distribution along the AB bond for all three compounds. The arrow along the bond charge indicates the atomic boundaries in the crystals that are not always at the minimum of the line charge along the bond AB. This should be expected taking into account the partial covalent bond of the compounds, because only in the ionic crystals, the atomic boundary is clearly defined. Table 5.4 lists the calculated effective radii, r_{III} and r_N, the effective charges, and FIC for AlN, GaN, and InN. The ionicity of AlN is high. This may explain the difficulties with AlN doping. It is well known that only covalent semiconductors or semiconductors with a large covalent component can form hydrogen-like shallow levels in the bandgap by substitution of a host atom with a neighbor with one more or one less electron. GaN and InN have a smaller than AlN but nearly equal ionicity. GaN was doped both *p*- and *n*-type. Thus, one can expect that InN can also be doped *n*- and *p*-type. To date, only *n*-type InN has been obtained because of high volatility of nitrogen and easiness of nitrogen vacancy formation that acts as a donor in this compound.

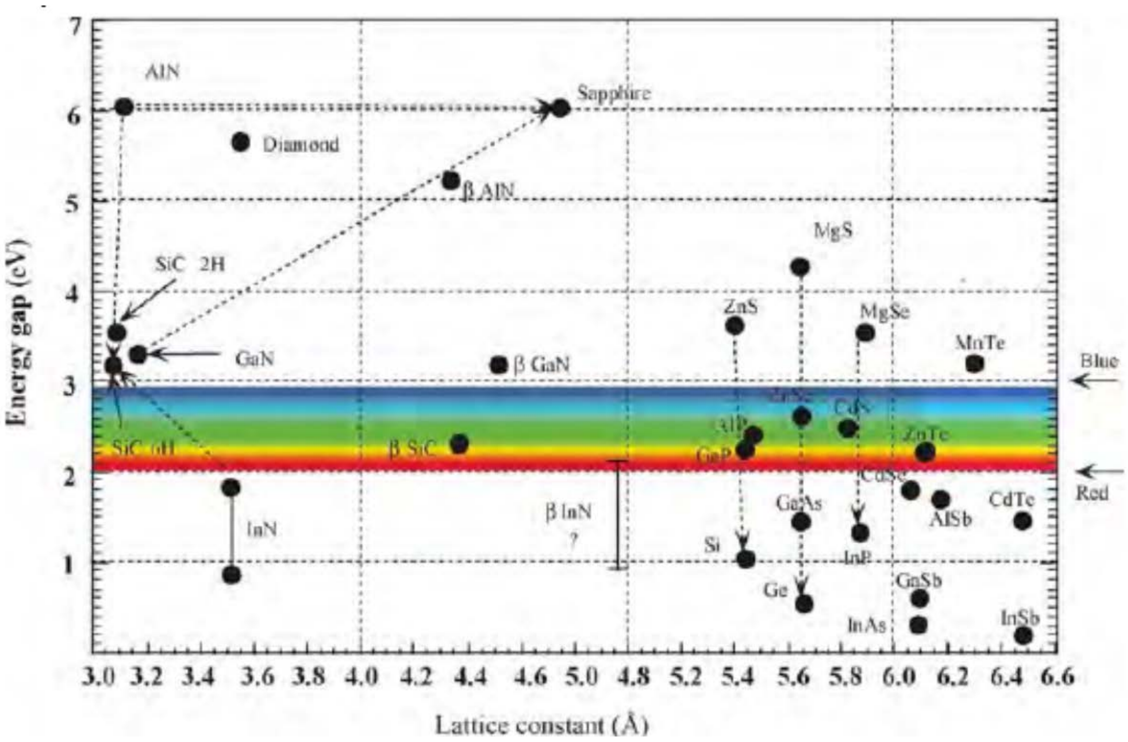


Fig. 5.15. The bandgaps of nitrides, substrates commonly used for nitrides, and other conventional semiconductors versus their lattice constants.

The III-nitrides are commonly grown on mismatched substrates because of the lack of suitable native substrates. Thus, the epitaxial layers are strained during cool down, if they are sufficiently thick for them to relax at the growth temperature. The mechanical forces related to strain dramatically change the band structure of the epitaxial layers. The pressure dependence of the bandgap energy E_g can be expressed as $E_g=E_g(0)gP+dP^2$, where $E_g(0)$ is the bandgap of stress-free semiconductor, g and d are the pressure coefficients, and P is the pressure.

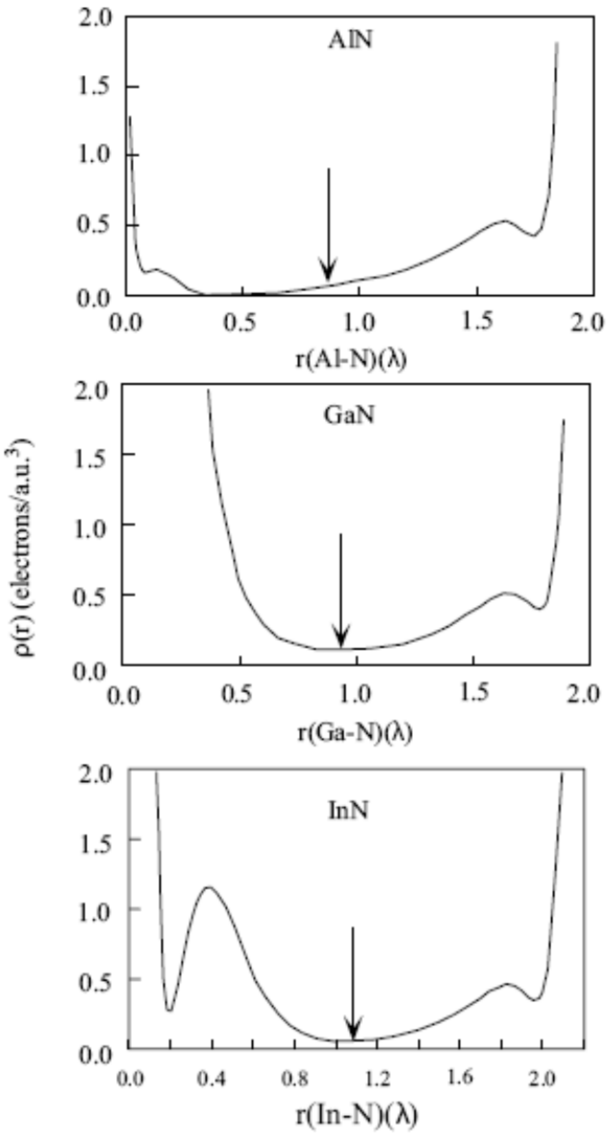


Fig. 5.16. Charge density along the III–N bond in III nitride semiconductors.

Table 5.4. Calculated ionic radii (A°), effective charges (electrons), and the fractional ionic character (FIC) for III nitrides [35].

Compound	**r_{III} (A°)**	**r_N (A°)**	**Q^*_{III} (e)**	**Q^*_N (e)**	***FIC***
AlN	0.8505	1.0523	1.12	6.88	0.72
GaN	0.9340	1.0119	1.98	6.02	0.51
InN	1.0673	1.0673	1.83	6.17	0.54

For GaN, *g* and *d* parameters are 4.2×10^{-3} and -1.8×10^{-5}, respectively [36, 37].The bandgap is in terms of eV and the pressure is in terms of kbar. The pressure dependence has, in general, a sublinear character. The variation of the GaN energy gap with pressure, both theoretical [38] and experimental [17], is shown in Figure 5.17.

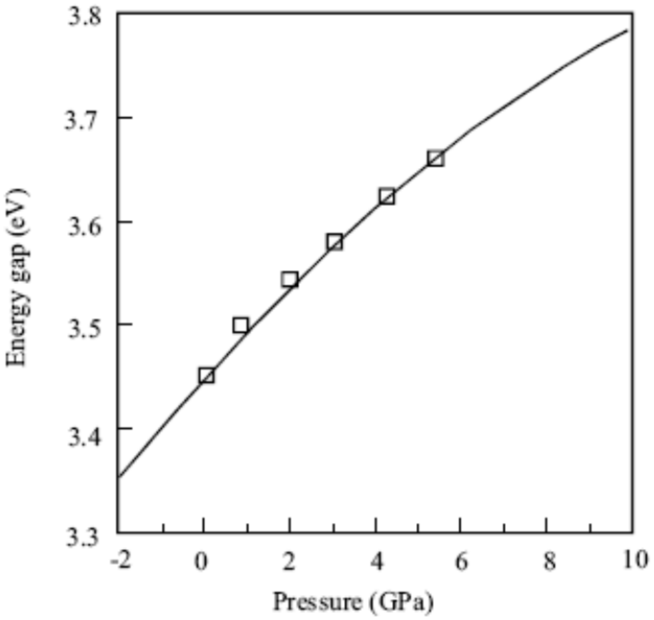


Fig. 5.17. Pressure dependence of the GaN energy gap, showing the typical sublinear character. Solid line represents the calculations [38], which have been rigidly upshifted by ~0.82 eV for a better fit with experiments, and the squares represent experimental results [17].

5.5. Remaining Issues and Trends

The performance of high power nitride devices operating in the 2GHz band is considered as being satisfactory and the main areas of concern are reliability, production yield, and cost. Improvements in yield and reductions in cost will only be possible with a stable supply of high quality 4-in. substrates. Manufacturing of semi-insulating SiC substrates is not easy and there are concerns about the costs as well. Use of n-type SiC substrates is also being investigated because they are easier for growth and less expensive. Large area Si wafers and AlN, which has a large thermal conductivity, are also being considered as substrates.

Reliability issues will require improvements in the crystalline quality, surface passivation, ohmic and gate electrodes, and efficient heat dissipation. However, compared with conventional devices fabricated using Si and GaAs, there is no doubt about the superior performance of nitride semiconductor devices, which will eventually be used for third and fourth generation mobile phone base stations.

Nitride-based devices have already surpassed the performance of GaAs for devices operating in the millimeter band range, and are expected to be used in fixed wireless access systems and for high speed wireless communication between base stations.

Figure 5.18 shows other potential applications of nitride semiconductor devices such as satellite communication, the high speed intelligent traffic systems, home networks, and sensing.

The wide bandgap of nitrides also has potential applications for devices operating at high temperature. Si is limited to operation below about 150◦C. There have been reports of nitride devices operating at 300◦C, with the possibility of operation even at 400–500◦C.

Normally-off devices are important for applications in power electronics, such as power sources.

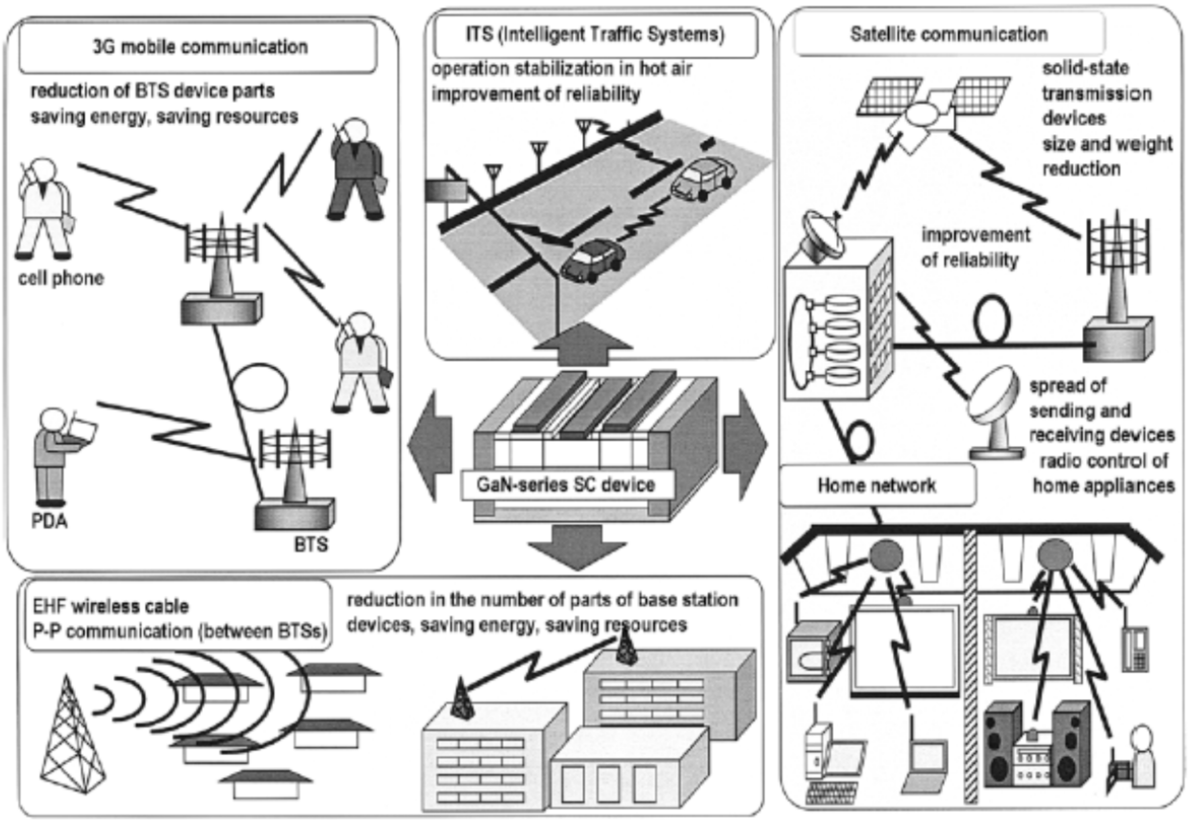


Fig. 5.18. Applications of electron devices fabricated using nitride compound semiconductors.

Chapter 6. Silicon Oxide

6.1. Introduction

Silicon dioxide (silica) is one of the most commonly encountered substances in both daily life and in electronics manufacturing. Crystalline silicon dioxide (in several forms: quartz, cristobalite, tridymite) is an important constituent of a great many minerals and gemstones, both in pure form and mixed with related oxides. Beach sand is mostly silica. The whole of planar electronics processing and the modern IC industry has been made possible by the unique properties of silicon dioxide: the only native oxide of a common semiconductor which is stable in water and at elevated temperatures, an excellent electrical insulator, a mask to common diffusing species, and capable of forming a nearly perfect electrical interface with its substrate [1]. Deposited silicon dioxide, almost always by CVD approaches, is almost as old as thermal growth on the substrate, and has been employed in various ways in IC fabrication due to its familiarity, versatility, and reliability.

SiO_2 is formed by strong, directional covalent bonds, and has a well-defined local structure: four oxygen atoms are arrayed at the corners of a tetrahedron around a central silicon atom.

Many different kinds of thin films are used to fabricate discrete devices and integrated circuits, including thermal oxides, dielectric layers, polycrystalline silicon, and metal films. An important oxide layer is the gate oxide, under which a conducting channel can be formed between the source and the drain. A related layer is the field oxide, which provides isolation from other devices. Both gate and field oxides generally are grown by a thermal oxidation process because only thermal oxidation can provide the highest-quality oxides having the lowest interface trap densities.

Semiconductors can be oxidized by some methods, including thermal oxidation and electrochemical anodization. Among these, thermal oxidation is the most important for silicon devices. It is a key process in modern silicon IC technology.

Deposited dielectric films are used mainly for insulation and passivation of discrete devices MEMS, NEMS and integrated circuits.

6.2. Silicon Dioxide for MEMS and NEMS

Silicon dioxide (SiO_2) is one of the most widely used materials in the fabrication of MEMS. In polysilicon surface micromachining, SiO_2 is used as a sacrificial material, since it can be easily dissolved using etchants that do not attack polysilicon. SiO_2 is widely used as etch mask for dry etching of thick polysilicon films, since it is chemically resistant to dry etching processes for polysilicon. SiO_2 films are also used as passivation layers on the surfaces of environmentally sensitive devices.

The most common processes used to produce SiO_2 films for polysilicon surface micromachining are thermal oxidation and LPCVD. Thermal oxidation of Si is performed at temperatures of 900 °C to 1200 °C in the presence of oxygen or steam. Since thermal oxidation is a self-limiting process, the maximum practical film thickness that can be obtained is about 2 μm, which is sufficient for many sacrificial applications. As noted by its name, thermal oxidation of Si can only be performed on Si surfaces.

SiO_2 films can be deposited on a wide variety of substrate materials by LPCVD. In general, LPCVD provides a means for depositing thick (> 2 μm) SiO_2 films at temperatures much lower than thermal oxidation. Known as low-temperature oxides, or LTO for short, these films have a higher etch rate in HF than thermal oxides, which translates to significantly faster release times when LTO films are used as sacrificial layers. Phosphosilicate glass (PSG) can be formed using nearly the same deposition process as LTO by adding a phosphorus-containing gas to the precursor flows. PSG films are useful as sacrificial layers, since they generally have higher etching rates in HF than LTO films.

PSG and LTO films are deposited in hot-wall, low pressure, fused silica furnaces in systems similar to those for polysilicon. Precursor gases include SiH_4 as a Si source, O_2 as an oxygen source, and, in the case of PSG, PH_3 as a source of phosphorus. LTO and PSG films are typically deposited at temperatures of 425 °C to 450 °C and pressures ranging from 200 mtorr to 400 mtorr. The low deposition temperatures result in LTO and PSG films that are slightly less dense than thermal oxides, due to the incorporation of hydrogen in the films. LTO films can, however, be densified by an annealing step at high temperature (1,000 °C). The low density of LTO and PSG films is partially responsible for the increased etch rate in HF.

Thermal SiO_2 and LTO are electrical insulators used in numerous MEMS applications. The dielectric constants of thermal oxide and LTO are 3.9 and 4.3, respectively. The dielectric strength of thermal SiO_2 is 1.1Ч106 V/cm, and for LTO it is about 80% of that value [2]. The stress in thermal SiO_2 is compressive with a magnitude of about 300 MPa [2]. For LTO, however, the as-deposited residual stress is tensile, with a magnitude of about 100 MPa to 400 MPa [2]. The addition of phosphorous to LTO decreases the tensile residual stress to about 10 MPa for phosphorus concentrations of 8% [3]. As with polysilicon, the properties of LTO and PSG are dependent on processing conditions.

Plasma-enhanced chemical vapor deposition (PECVD) is another common method to produce oxides of silicon. Using plasma to dissociate the gaseous precursors, the deposition temperatures needed to deposit PECVD oxide films is lower than for LPCVD films. For this reason, PECVD oxides are quite commonly used as masking, passivation, and protective layers, especially on devices that have been coated with metals.

Quartz is the crystalline form of SiO_2 and has interesting properties for MEMS. Quartz is optically transparent, piezoelectric, and electrically insulating. Like single crystal Si, quartz substrates are available as high quality, large area wafers that can be bulk micromachined using anisotropic etchants [4]. Quartz has recently become a popular substrate material for microfluidic devices due to its optical, electronic, and chemical properties.

Another SiO_2-related material that has recently found uses in MEMS is spin-on-glass (SOG). SOG is a polymeric material with a viscosity suitable for spin coating. SOG has the potential in MEMS fabrication. SOG was developed as a thick-film sacrificial molding material for thick polysilicon films [5]. The processes to deposit, polish, and etch SOG films that were 20 microns thick have been reported. The thick SOG films were patterned into molds and filled with 10 micron-thick LPCVD polysilicon films, planarized by selective CMP (Chemical Mechanical Polishing), and subsequently dissolved in a wet etchant containing HCl, HF, and H_2O to reveal the patterned polysilicon structures. The cured SOG films were completely compatible with the polysilicon deposition process. High-aspect ratio channel plate microstructures from SOG has bee also fabricated [6]. Electroplated nickel (Ni) was used as a molding material with Ni channel plate molds fabricated using a conventional LIGA process. The Ni molds were then filled with SOG, and the sacrificial Ni molds were removed in a reverse electroplating process. In this case, the fabricated SOG structures (over 100 microns tall) were micromachined glass structures fabricated using a molding material more commonly used for structural components.

6.3. Thermal Oxidation Process

The basic thermal oxidation apparatus (Fig. 6.1) consists of a resistance-heated furnace, a cylindrical fused-quartz tube containing the silicon wafers held vertically in a slotted quartz boat, and a source of either pure dry oxygen or pure water vapor. Oxidation temperature is generally in the range of 900–1200°C, and the typical gas flow rate is about 1 L/min. The oxidation system uses microprocessors to regulate the gas flow sequence, to control the automatic insertion and removal of silicon wafers, to ramp the temperature up (i.e., to increase the furnace temperature linearly) from a low temperature to the oxidation temperature, to maintain the oxidation temperature to within ±1°C, and to ramp the temperature down when oxidation is completed.

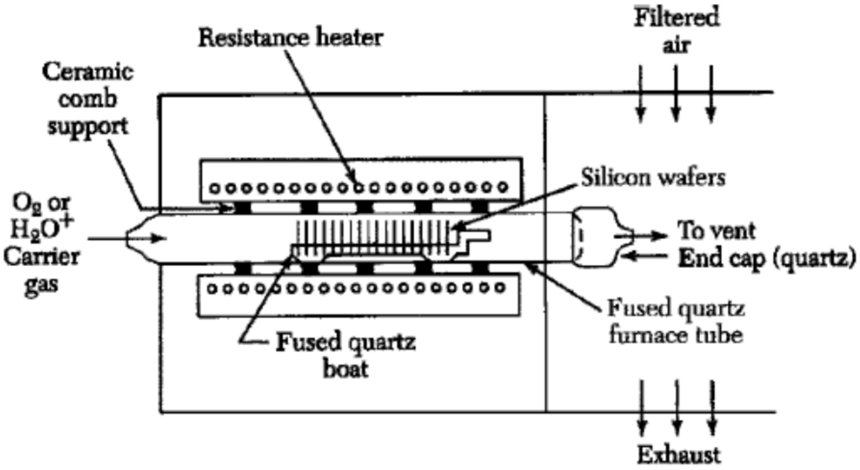


Fig. 6.1. Schematic of an oxidation furnace [1].

6.3.1. Growth Kinetics

The following chemical reactions describe the thermal oxidation of silicon in oxygen (“dry” oxidation) and water vapor (“wet” oxidation), respectively:

$$\mathrm{Si(solid)} + \mathrm{O_2(gas)} \rightarrow \mathrm{SiO_2(solid)} \tag{6.1}$$

$$\mathrm{Si(solid)} + 2\mathrm{H_2O(gas)} \rightarrow \mathrm{SiO_2(solid)} + 2\mathrm{H_2(gas)} \tag{6.2}$$

The silicon–silicon dioxide interface moves into the silicon during the oxidation process. This creates a new interface region, with surface contamination on the original silicon ending up on the oxide surface. As a result of the densities and molecular weights of silicon and silicon dioxide, growing an oxide of thickness *x* consumes a layer of silicon 0.44*x* thick (Fig. 6.2).

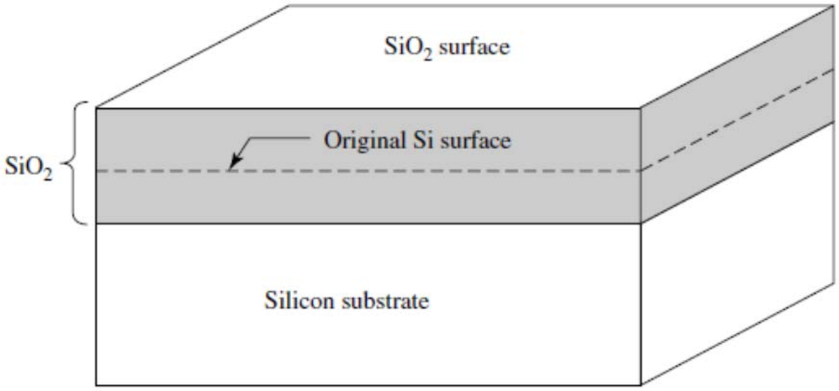


Fig. 6.2. Movement of silicon–silicon dioxide interface during oxide growth [1].

The kinetics of silicon oxidation can be described on the basis of the simple model illustrated in Fig. 6.3. A silicon slice contacts the oxidizing species (oxygen or water vapor), resulting in a surface concentration of C_0 molecules/cm^3 for these species. The magnitude of C_0 equals the equilibrium bulk concentration of the species at the oxidation temperature. The equilibrium concentration generally is proportional to the partial pressure of the oxidant adjacent to the oxide surface. At 1000°C and a pressure of 1 atm, the concentration C_0 is 5.2×10^{16} cm^{-3} for dry oxygen and 3×10^{19} cm^{-3} for water vapor.

The oxidizing species diffuses through the silicon dioxide layer, resulting in a concentration C_s at the surface of silicon. The flux F_1 can be written as

$$F_1 = D\frac{dC}{dx} \cong \frac{D(C_0 - C_s)}{x} \qquad (6.3)$$

where D is the diffusion coefficient of the oxidizing species, and x is the thickness of the oxide layer already present.

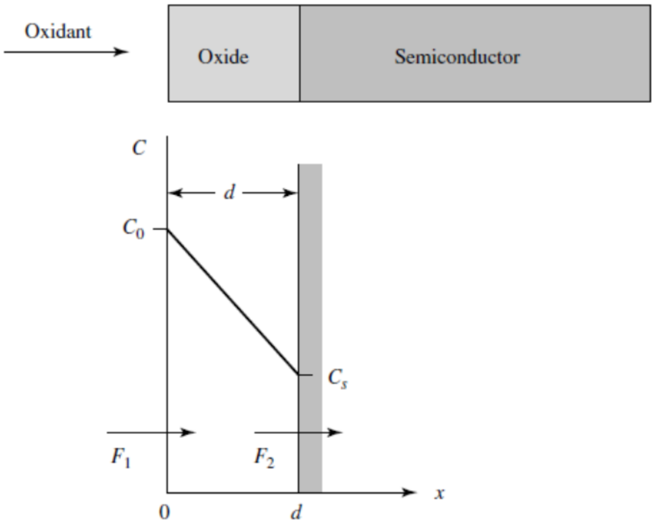


Fig. 6.3. Basic model for the thermal oxidation of silicon [1].

At the silicon surface, the oxidizing species reacts chemically with silicon. Assuming the rate of reaction to be proportional to the concentration of the species at the silicon surface, the flux F_2 is given by

$$F_2 = kC_s \tag{6.4}$$

where k is the surface reaction rate constant for oxidation. At the steady state, $F_1 = F_2 = F$. Combining Eqs. (6.3) and (6.4) gives

$$F = \frac{DC_0}{x + (D/k)} \tag{6.5}$$

The reaction of the oxidizing species with silicon forms silicon dioxide. Let C_1 be the number of molecules of the oxidizing species in a unit volume of the oxide. There are 2.2 × 10^{22} silicon dioxide molecules/cm^3 in the oxide, and one oxygen molecule (O_2) is added to each silicon dioxide molecule, whereas we add two water molecules (H_2O) to each SiO_2 molecule. Therefore, C_1 for oxidation in dry oxygen is 2.2 × 10^{22} cm^{-3}, and for oxidation in water vapor it is twice this number (4.4 × 10^{22} cm^{-3}). Thus, the growth rate of the oxide layer thickness is given by

$$\frac{dx}{dt} = \frac{F}{C_1} = \frac{DC_0/C_1}{x + (D/k)} \tag{6.6}$$

This differential equation can be solved subject to the initial condition, $x(0) = d_0$, where d_0 is the initial oxide thickness; d_0 can also be regarded as the thickness of oxide layer grown in an earlier oxidation step. Solving Eq. (6.6) yields the general relationship for the oxidation of silicon:

$$x^2 + \frac{2D}{k}x = \frac{2DC_0}{C_1}(t+\tau) \tag{6.7}$$

where $\tau \equiv (d_0^2 + 2Dd_0/k)C_1/2DC_0$, which represents a time coordinate shift to account for the initial oxide layer d_0.

The oxide thickness after an oxidizing time t is given by

$$x = \frac{D}{k}[\sqrt{1 + \frac{2C_0k^2(t+\tau)}{DC_1}} - 1] \tag{6.8}$$

For small values of t, Eq. (6.8) reduces to

$$x \cong \frac{C_0k}{C_1}(t+\tau) \tag{6.9}$$

and for larger values of t, it reduces to

$$x \cong \sqrt{\frac{2DC_0}{C_1}(t+\tau)} \tag{6.10}$$

During the early stages of oxide growth, when surface reaction is the rate limiting factor, the oxide thickness varies linearly with time. As the oxide layer becomes thicker, the oxidant must diffuse through the oxide layer to react at the silicon–silicon dioxide interface and the reaction becomes diffusion-limited. The oxide growth then becomes proportional to the square root of the oxidizing time, which results in a parabolic growth rate.

Equation (6.7) is often written in a more compact form

$$x^2 + Ax = B(t+\tau) \tag{6.11}$$

where $A = 2D/k$, $B = 2DC_0/C_1$ and $B/A = kC_0/C_1$. Using this form, Eqs. (6.9) and (6.10) can be written as

$$x = \frac{B}{A}(t+\tau) \tag{6.12}$$

for the linear region and as

$$x^2 = B(t+\tau) \tag{6.13}$$

for the parabolic region. For this reason, the term B/A is referred to as the *linear rate constant* and B is the *parabolic rate constant*. Experimentally measured results agree with the predictions of this model over a wide range of oxidation conditions. For wet oxidation, the initial oxide thickness d_0 is very small, or $\tau \cong 0$. However, for dry oxidation, the extrapolated value of d_0 at $t = 0$ is about 25 nm. Thus, the use of Eq. (6.11) for dry oxidation on bare silicon requires a value for τ that can be generated using this initial thickness. Table 6.1 lists the values of the rate constants for wet oxidation of silicon, and Table 6.2 lists the values for dry oxidation.

Table 6.1. Rate constants for wet oxidation of silicon.

Temperature (°C)	A (μm)	B (μm^2/h)	τ (h)
1200	0.05	0.72	0
1100	0.11	0.51	0
1000	0.226	0.287	0
920	0.5	0.203	0

Table 6.2. Rate constants for dry oxidation of silicon.

Temperature (°C)	A (μm)	B (μm^2/h)	τ (h)
1200	0.04	0.045	0.027
1100	0.09	0.027	0.076
1000	0.165	0.0117	0.37
920	0.235	0.0049	1.4
800	0.37	0.0011	9.0
700	-	-	81.0

The temperature dependence of the linear rate constant B/A is shown in Fig. 6.4 for both dry and wet oxidation and for (111)- and (100)-oriented silicon wafers [1, 7-11]. The linear rate constant varies as $exp(-E_a/k_BT)$, where the activation energy E_a is about 2 eV for both dry and wet oxidation. This closely agrees with the energy required to break silicon–silicon bonds, 1.83 eV/molecule. Under a given oxidation condition, the linear rate constant depends on crystal orientation. This is because the rate constant is related to the rate of incorporation of oxygen atoms into the silicon. The rate depends on the surface bond structure of silicon atoms, making it orientation-dependent. Because the density of available bonds on the (111) plane is higher than that on the (100) plane, the linear rate constant for (111) silicon is larger.

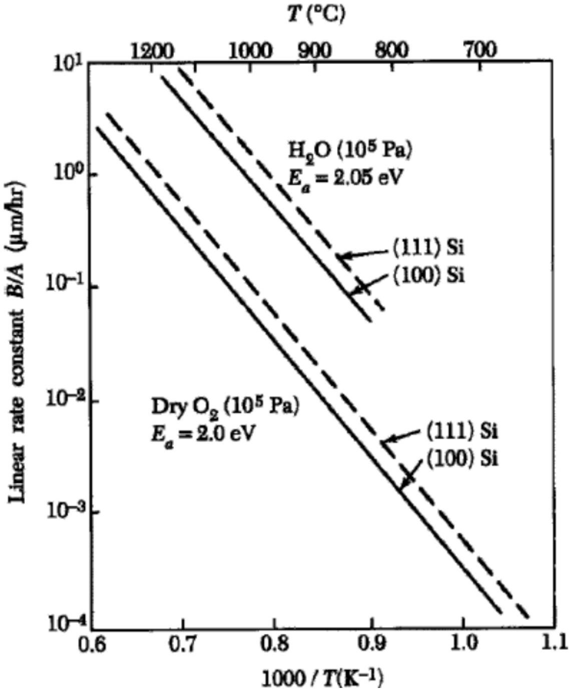


Fig. 6.4. Linear rate constant versus temperature [1].

Figure 6.5 shows the temperature dependence of the parabolic rate constant *B*, which can also be described by $exp(-E_a/kT)$. The activation energy E_a is 1.24 eV for dry oxidation. The comparable activation energy for oxygen diffusion in fused silica is 1.18 eV. The corresponding value for wet oxidation, 0.71 eV, compares favorably with the value of 0.79 eV for the activation energy of diffusion of water in fused silica. The parabolic rate constant is independent of crystal orientation. This independence is expected because it is a measure of the diffusion process of the oxidizing species through a random network layer of amorphous silica.

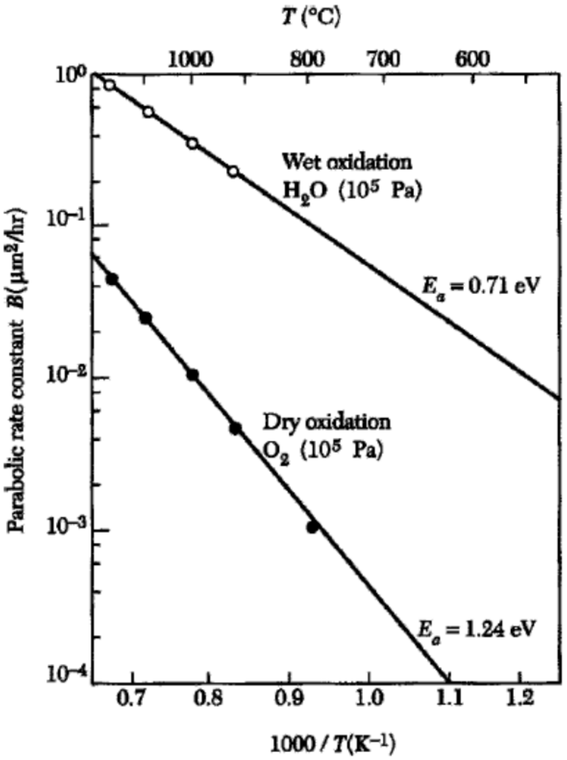


Fig. 6.5. Parabolic rate constant versus temperature [1].

Although oxides grown in dry oxygen have the best electrical properties, considerably more time is required to grow the same oxide thickness at a given temperature in dry oxygen than in water vapor. For relatively thin oxides such as the gate oxide in a MOSFET (typically ≤20 nm), dry oxidation is used. However, for thicker oxides such as field oxides (≥20 nm) in MOS integrated circuits, and for bipolar devices, oxidation in water vapor (or steam) is used to provide both adequate isolation and passivation.

Figure 6.6 shows the experimental results of silicon dioxide thickness as a function of reaction time and temperature for two substrate orientation [7]. Under a given oxidation condition, the oxide thickness grown on a (111)-substrate is larger than that grown on a (100)-substrate because of the larger linear rate constant of the (111)-orientation. Note that for a give temperature and time, the oxide film obtained using wet oxidation is about 5-10 times thicker than that using dry oxidation.

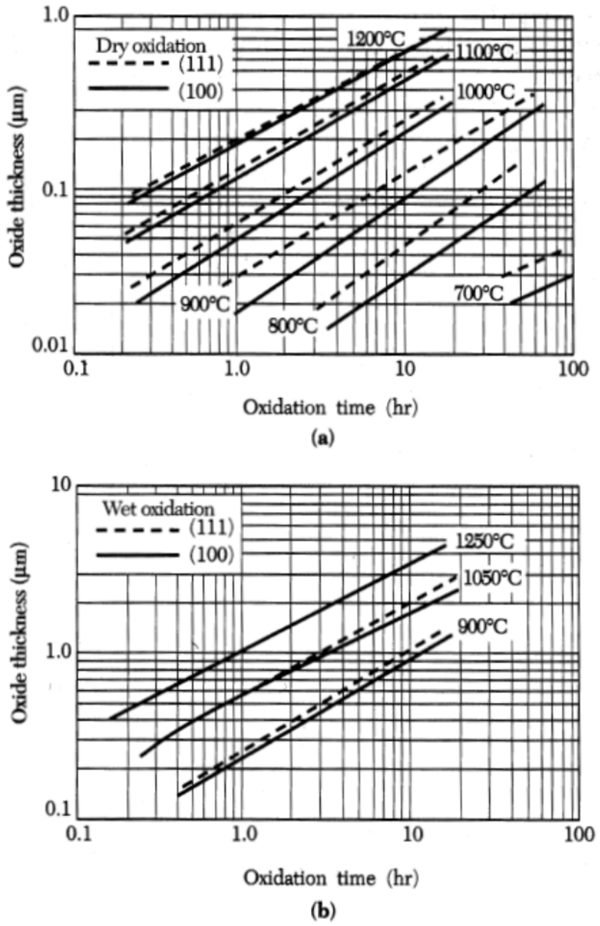


Fig. 6.6. Experimental results of silicon dioxide thickness as a function of reaction time and temperature for two substrate orientations. (*a*) Growth in dry oxygen. (*b*) Growth in steam.

6.3.2. Thin Oxide Growth

Relatively slow growth rates must be used to reproducibly grow thin oxide films of precise thickness. Approaches to achieve such slower growth rates include growth in dry O_2 at atmospheric pressure and lower temperatures (800–900°C); growth at pressures lower than atmospheric pressure; growth in a reduced partial pressures of O_2 by using a diluent inert gas, such as N_2, Ar, or He, together with the gas containing the oxidizing species; and the use of composite oxide films with the gate oxide films consisting of a layer of thermally grown SiO_2 and an overlayer of chemical vapor deposition (CVD) SiO_2. However, the mainstream approach for gate oxides 10–15 nm thick is to grow the oxide film at atmospheric pressure and lower temperatures (800–900°C). With this approach, processing using modern *vertical* oxidation furnaces can grow reproducible, high-quality 10-nm oxides to within 0.1 nm across the wafer.

It was noted earlier that for dry oxidation, there is a rapid early growth that gives rise to an initial oxide thickness d_0 of about 20 nm. Therefore, the simple model given by Eq. (6.11) is not valid for dry oxidation with an oxide thickness ≤20 nm. For ultra-large-scale integration, the ability to grow thin (5–20 nm), uniform, high-quality reproducible gate oxides has become increasingly important.

In the early stage of growth in dry oxidation, there is a large compressive stress in the oxide layer that reduces the oxygen diffusion coefficient in the oxide. As the oxide becomes thicker, the stress will be reduced due to the viscous flow of silica and the diffusion coefficient will approach its stress-free value. Therefore, for thin oxides, the value of D/k may be sufficiently small that we can neglect the term Ax in Eq. (6.11) and obtain

$$x^2 + d_0^2 = Bt \qquad (6.14)$$

where d_0 is equal to $\sqrt{2DC_0\tau/C_1}$, which is the initial oxide thickness when time is extrapolated to zero, and B is the parabolic rate constant defined previously.

6.4. Impurity Redistribution during Oxidation

Dopant impurities near the silicon surface will be redistributed during thermal oxidation The redistribution depends on several factors. When two solid phases are brought together, an impurity in one solid will redistribute between the two solids until it reaches equilibrium. The ratio of the equilibrium concentration of the impurity in the silicon to that in the silicon dioxide is called the segregation coefficient and is defined as

$$k = \frac{C_{Si}}{C_{SiO2}}, \qquad (6.15)$$

where C_{Si} is the equilibrium concentration of impurity in silicon, C_{SiO2} is the equilibrium concentration of impurity in SiO_2.

A second factor that influences impurity distribution is that the inipurity may diffuse rapidly through the silicon dioxide and escape to the gaseous ambient. If the diffusivity of the impurity in silicon dioxide is large, this factor will be important. A third factor in the redistribution process is that the oxide is growing, and thus the boundary between the silicon and the oxide is advancing into the silicon as a function of time. The relative rate of this advance compared with the diffusion rate of the impuritv through the oxide is important in determining the extent of the redistribution. Note that even if the segregation coefficient of an impurity equals unity, some redistribution of the umpurity in the silicon will still take place. As indicated in Fig. 6.2, the oxide layer will be about twice as thick as the silicon layer it replaced. Therefore, the same amount of impurity will now be distributed in a large volume, resulting in depletion of the impurity from the silicon.

Four possible redistribution processes are illustrated in Fig. 6.7 [8]. These processes can be classified into two groups. In one group, the oxide takes up the impurite (Fig. 6.7, *a* and *b* for $k<1$), and in the other the oxide rejects the impurity (Fig. 6.7, *c* and *d* for $k>1$). In each case, what happens depends on how rapidly the impurity can diffuse through the oxide. In group 1, the silicon surface is depleted of impurities; an example is boron, with k approximately equal to 0.3. Rapid diffusion of the impurity through the silicon dioxide increases the amount of depletion; an example is boron-doped silicon heated in hydrogen ambient, because hydrogen in silicon dioxide enhances the diffusivity of boron. In group 2, k is greater than unify, so the oxide rejects the impurity. If diffusion of the impurity through the silicon dioxide is relatively slow, the impurity piles up near the silicon surface; an example is phosphorus, with k approximately equal to 10. When diffusion through the silicon dioxide is rapid, so much impurity may escape from the solid to the gaseous ambient that the overall effect will be a depletion of the impurity; an example is gallium, with k approximately equal to 20.

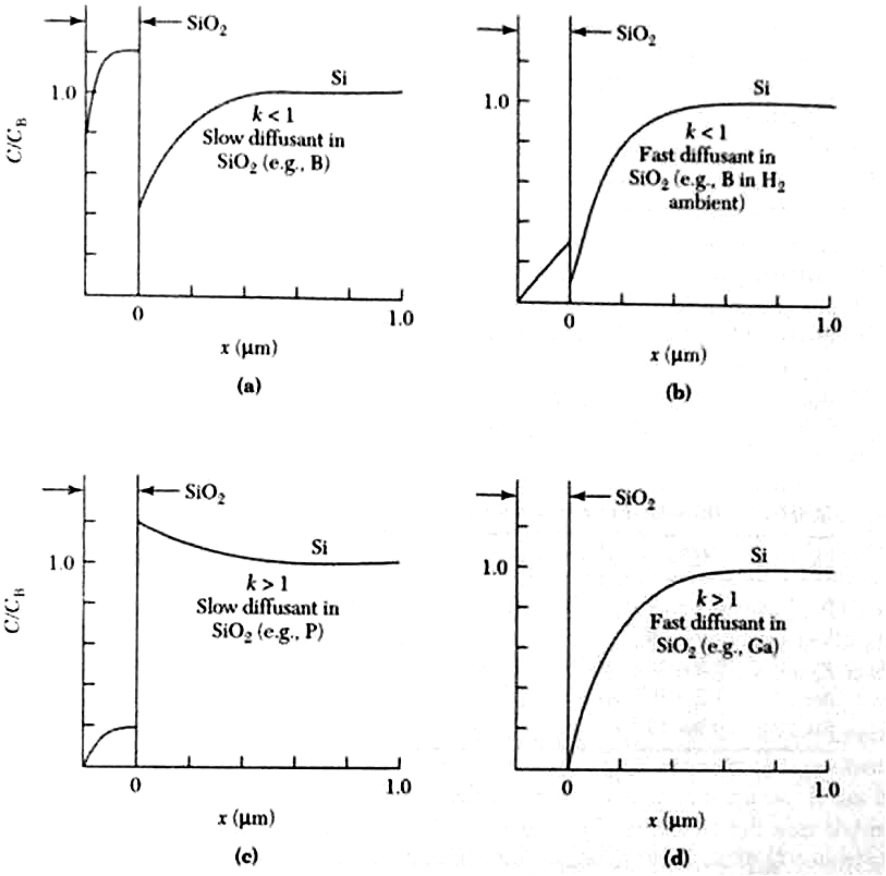


Fig. 6.7. Four different cases of impurity redistribution in silicon due to thermal oxidation.

The redistributed dopant impurities in silicon dioxide are seldom electrically active. However, redistribution in silicon has an important effect on processing and device performance. For example, nonuniform dopant distribution will modify the interpretation of the measurements of interface trap properties, and the change of the surface concentration will modify the threshold voltage and device contact resistance.

6.5. Masking properties of silicon dioxide

A silicon dioxide layer can also provide a selective mask against the diffusion of dopants at elevated temperatures, a very useful property in IC fabrication. Predeposition of dopants, whether it be by ion implantation, chemical diffusion, or spin-on techniques, typically results in a dopant source at or near the surface of the oxide. During a subsequent high-temperature drive-in step, diffusion its oxide-masked regions must be slow enough with respect to diffusion in the silicon to prevent dopants from diffusing through the oxide mask to the silicon surface. The required thickness may be determined experimentally by measuring the oxide thickness necessary to prevent the inversion of a lightly doped silicon substrate of opposite conductivity at a particular temperature and time. Typically, oxides used for masking common impurities are 0.5 to 1.0 µm thick.

The values of diffusion constants for various dopants in various dopants in SiO_2 depend on the concentration, properties, and structure of the oxide. Table 6.3 lists diffusion constants for various common dopants, and Fig. 6.8 gives the oxide thickness required to mask boron and phosphorus as a function: of diffusion time and temperature. Note that SiO_2 is much more effective for masking boron than phosphorus. Nevertheless, the diffusivities of P, Sb, As, and B in SiO_2 are all orders of magnitude less than their corresponding values in silicon, so they are all compatible with oxide masking. This is not true, however, for Ga or Al. Silicon nitride is used as an alternative masking material for these elements.

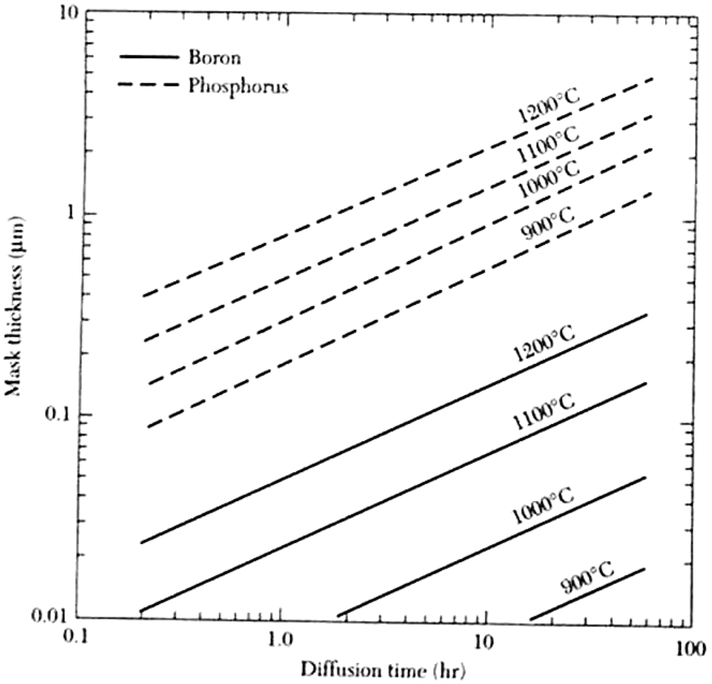


Fig. 6.8. Thickness of silicon dioxide needed to mask boron and phosphorus diffusions as a function of diffusion time and temperature.

Table 6.3. Diffusion constants in SiO_2

Dopants	Diffusion Constants at 1100 °C (cm^2/s)
B	3.4×10^{-17} to 2.0×10^{-14}
Ga	5.3×10^{-11}
P	2.9×10^{-16} to 2.0×10^{-13}
As	1.2×10^{-16} to 3.5×10^{-15}
Sb	9.9×10^{-17}

6.6. Silicon Oxide Quality

Oxides used for masking are usually grown by wet oxidation. A typical growth cycle consists of a dry–wet–dry sequence. Most of the growth in such a sequence occurs in the wet phase, since the SiO_2 growth rate is much higher when water is used as the oxidant. Dry oxidation, however, results in a higher quality oxide that is denser and has a higher breakdown voltage (5–10 MV/cm). It is for these reasons that the thin gate oxides in MOS devices are usually formed using dry oxidation.

It is well known that defects in noncrystalline SiO_2 films on silicon play an important role in determining the properties of Si/SiO_2 interface structures and, hence, of various silicon-based semiconductor devices.

MOS devices are also affected by charges in the oxide and traps at the SiO_2 –Si interface. The basic classification of these traps and charges, shown in Fig. 6.9, are interface-trapped charge, fixed-oxide charge, oxide-trapped charge, and mobile ionic charge.

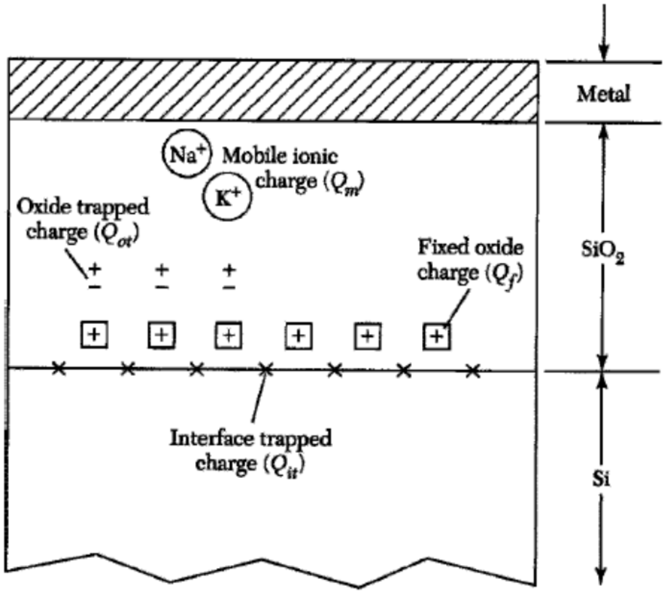


Fig. 6.9. Description of charges associated with thermal oxides [1].

Interface-trapped charges (Q_{it}) are due to the SiO_2 –Si interface properties and dependent on the chemical composition of this interface. The traps are located at the SiO_2 –Si interface with energy states in the silicon-forbidden bandgap. The interface trap density (i.e., number of interface traps per unit area and per eV) is orientation dependent. In silicon with a <100> crystal orientation, the interface trap density is about an order of magnitude smaller than that in the <111> orientation. Present-day MOS devices with thermally grown silicon dioxide on silicon have most of the interface trapped charges passivated by low-temperature (450°C) hydrogen annealing. The value of Q_{it} for <100> oriented silicon can be as low as 10^{10} cm^{-2}, which amounts to about one interface trapped charge per 10^5 surface atoms. For <111> oriented silicon, Q_{it} is about 10^{11} cm^{-2}.

The fixed charge (Q_f) is located within approximately 3 nm of the SiO_2 –Si interface. Generally, Q_f is positive and depends on oxidation and annealing conditions, as well as on the orientation of the silicon substrate. This charge is fixed and very difficult to charge or discharge. It has been suggested that when the oxidation is stopped, some ionic silicon is left near the interface. These ions, along with uncompleted silicon bonds (e.g., Si-Si or Si-O bonds) at the surface, may result in the positive interface charge. Q_f can be regarded as a charge sheet located at the SiO_2-Si interface. Typical fixed oxide charge densities for a carefully treated SiO_2-Si interface system are about 10^{10} cm^{-2} for <100> surface and about 5×10^{10} cm^{-2} for a <111> surface. Because of the lower values of Q_{it} and Q_f the <100> orientation is preferred for silicon MOSFETs.

Oxide-trapped charges (Q_{ot}) are associated with defects in the silicon dioxide. These charges can be created, for example, by X-ray radiation or high-energy electron bombardment. The traps are distributed inside the oxide layer. Most process-related Q_{ot} can be removed by low-temperature annealing.

Mobile ionic charges (Q_m), which result from contamination from sodium or other alkali ions, are mobile within the oxide under raised-temperatures (e.g., >100°C) and high-electric-field operations. Trace contamination by alkali metal ions may cause stability problems in semiconductor devices operated under high-bias and high-temperature conditions. Under these conditions mobile ionic charges can move back and forth through the oxide layer and cause threshold voltage shifts. Therefore, special attention must be paid to the elimination of mobile ions in device fabrication. For example, the effects of sodium contamination can be reduced by adding chlorine during oxidation. Chlorine immobilizes the sodium ions. A small amount (6% or less) of anhydrous HCl in the oxidizing gas can accomplish this, but the presence of chlorine during dry oxidation increases both the linear and parabolic rate constants, leading to a higher growth rate.

6.7. Silicon Oxide Structure

The basic structural unit of thermally grown silicon dioxide is a silicon atom surrounded tetrahedrally by four oxygen atoms, as illustrated1 in Fig. 6.10. The silicon-to-oxygen internuclear distance is 0.16 nm, and the oxygen-to-oxygen internuclear distance is 0.227 nm. These tetrahedra are joined together at their corners by oxygen bridges in a variety of ways to form the various phases or structures of silicon dioxide (also called silica). Silica has several crystalline structures (e.g., quartz) and an amorphous structure. When silicon is thermally oxidized, the silicon dioxide structure is amorphous. Typically amorphous silica has a density of 2.21 g/cm^3, compared with 2.65 g/cm^3 for quartz.

The basic difference between the crystalline and amorphous structures is that the former is a periodic structure, extending over many molecules, whereas the latter has no periodic structure at all. Figure 6.10, *b* is a two-dimensional schematic diagram of a quartz crystalline structure made up of rings with six silicon atoms. Figure 6.10, *c* is a two-dimensional schematic diagram of an amorphous structure for comparison. In the amorphous structure there is still a tendency to form characteristic rings with six silicon atoms. Note that the amorphous structure in Fig. 6.10, *c* is quite open because only 43% of the space is occupied by silicon dioxide molecules. The relatively open structure accounts for the lower density and allows a variety of impurities (such as sodium) to enter and diffuse readily through the silicon dioxide laver.

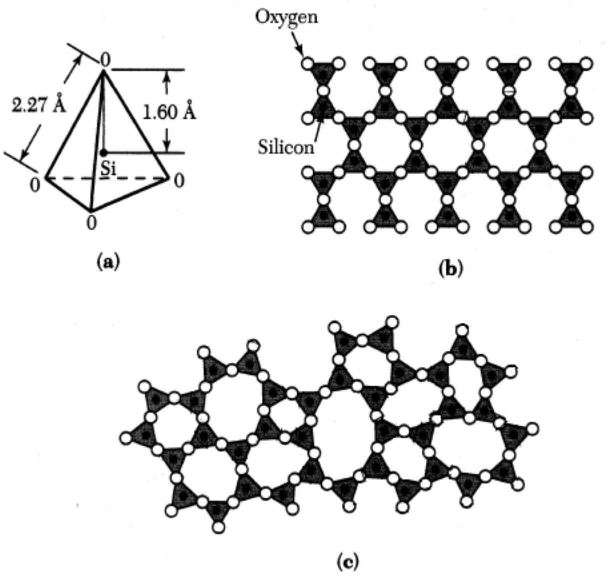


Fig. 6.10. (*a*) Basic structural unit of silicon dioxide. *(b)* Two dimensional representation of a quartz crystal lattice. *(c)* Two-dimensional representation of the amorphous structure of silicon dioxide.

The structural defects, particularly their electronic properties, are often described phenomenologically, i.e., without emphasizing their origin and relationship to the structure of noncrystalline SiO_2. Even when the immediate environment of some point defects was taken into account, for instance, in ESR studies, important features of the defect structure of the noncrystalline SiO_2 film (e.g., ordered regions termed channels and chemical interactions between defects) were usually not considered. A different approach was based on the localized properties of chemical bonds: It was suggested that a unique defect in noncrystalline SiO_2 film is the structural channel [12, 13] and that network defects in vitreous SiO_2 are significantly different from point defects in crystals, particularly with respect to irradiation behavior. Numerous observations have been made which can be interpreted in a self-consistent manner within the framework of these ideas.

The structure of essentially all crystalline and non-crystalline polymorphs of SiO_2 is based on the tetrahedrical configuration of the oxygen atoms around the silicon: that is, the Si-O coordination is 4:2 as expressed by the $SiO_{4/2}$ formula. The exception is stishovite in which the silicon is surrounded by six oxygen atoms in an octahedral configuration. Stishovite can be written as $SiO_{6/3}$ indicating the 6:3 coordination. The properties of stishovite are drastically different from those of the 4:2 coordinated crystalline polymorphs and vitreous SiO_2;

however, they illustrate in an extreme manner some trends exhibited by the polymorphs of 4:2 coordination.

Polymorphs based on $SiO_{4/2}$ tetrahedra.

One of the many remarkable features of SiO_2 is that there are nine polymorphs, including vitreous SiO_2, based on essentially identical $SiO_{4/2}$ tetrahedra; that is, the Si-O bond length varies only from 1.60 to 1.63 A° and the O-Si-O bond angle is ≈109° in all of them. The uniformity in the short-range order (SRO) is responsible for the very small difference (about 1%) in the standard free enthalpies of formation of the most stable and the least stable forms of silica , i.e., α-quartz and vitreous SiO_2 (the values are -197.2 kcal mole^{-1} and -195.3 kcal mole^{-1}, respectively). Another manifestation of the same SRO is the fact that the optical properties of vitreous and crystalline SiO_2 are identical in the first approximation, whereas the behavior of most of the noncrystalline solids indicates some disorder (reduction in the average coordinate matrix as in amorphous Si or reduced second neighbour interaction as in amorphous Se).

In contrast, the crystal structure and density of SiO_2 polymorphs exhibit significant variation, for example, the densities of quartz and vitreous silica are 2.65 and 2.20 g cm^{-3} respectively (this corresponds to a change of 17%). Other properties also exhibit some variation, e.g., the bond overlap population changes by ≈3% from quartz to vitreous silica. The main cause of the change in density is the variation in the packing arrangement (i.e. topology) of the SiO_2 tetrahedra. The Si-O-Si bond angle, which links the tetrahedra plays an important role m determining certain properties. This angle may exhibit a significant distribution even for some crystalline polymorphs, for instance, tridymite from 139.7° to 173.2° in the large unit cell comprising 320 $SiO_4/_2$ units. The histogram of the Si-O-Si bond angle distribution is shown in Fig. 6.11. These observations indicate that the *variation among the $SiO_{4/2}$ polymorphs is essentially of conformational nature*, that is, significant changes in the long-range order (LRO) are associated with little change in the energy content. This fact has important implications for the structure and properties of noncrystalline (vitreous) SiO_2.

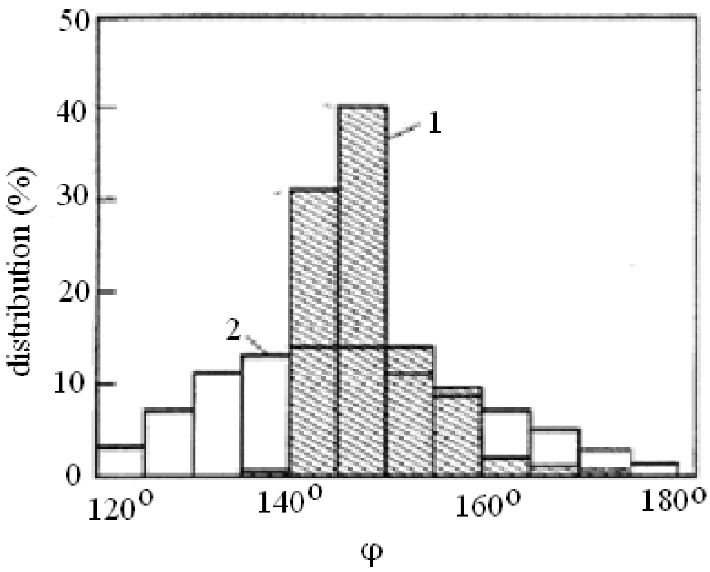


Fig. 6.11. Histogram of Si-O-Si bond angle distribution. (1) Tridymite (triclinic), (2) vitreous SiO_2.

Structure of noncrystalline SiO_2.

From the viewpoint of X-ray or electron diffraction analysis, noncrystalline SiO_2 films obtained by thermal or anodic oxidation of silicon are identical to fused vitreous silica. Therefore, the structure of fused silica will be considered first. The diffraction analysis of fused SiO_2 indicates that the shortest interatomic distances (i.e. the first Si-O, O-O, and Si-Si distances) correspond to those present in the crystalline phase, and that the Si-O-Si bond angles have a very wide distribution extending from 120° to 180°; the distribution curve peaks at 144°. According to a refined analysis, the mean value of the Si-O-Si bond angle is 152°. The histogram of the bond angle distribution in vitreous silica is shown in Fig. 6.11.

The results of the diffraction analysis are often interpreted in terms of random network model. A recent model based on -Si-O-Si-loops consisting of 4, 5, 6, 7, and 8 Si atoms resulted in 153° as the mean value of the Si-O-Si bond angle; therefore, it appears to be a realistic description of the structure of vitreous SiO_2. It is clear from Fig. 6.11 that the range of the Si-O-Si bond angles is larger for noncrystalline SiO_2 than for tridymite, particularly in the region below 140°. The reason is that, in addition to the distorted 6-member loops in tridymite, vitreous silica contains 4- and 5-member loops. The Si-O-Si bond angle decreases as the loop size decreases, and puckering of the loops (i.e., deviation from nonplanaritv) also tends to decrease the bond angle. Despite the difference in the range of the Si-O-Si bond angle distribution, the essential structural feature of crystalline tridymite and vitreous silica is common: the Si-O-Si bond angles exhibit an unusually large variation.

A interpretation of the diffraction analysis of vitreous SiO_2 suggested that it consists of tridymite-like regions up to at least 20 A° in size, which are bonded together similarly to twinned crystals but in such a manner that isotropic properties would be present. It is important to realize that the size of these tridymite-like regions is smaller than that of the unit cell of the tridymite crystal (the *c*-axis is 82 A°); therefore, these regions cannot be considered as microcrystals. Rather, they represent a form of *structural ordering* in the sense that the bonding topologies of SiO_2 glass and tridymite show a distinct resemblance. In contrast, crystallographic ordering, which refers to the translational symmetry associated with the periodicity of the crystal is lacking in vitreous silica. From the mechanical and thermal properties of SiO_2 glass it has also been inferred that pre-ordered regions exist in silica, which exhibit some similarities to particular crystalline polymorphs. Similar conclusions were reached from the polarizability behavior of silica glass during pressure densification.

The flexibility of the SiO_2 structure, as manifested in the wide distribution of the Si-O-Si bond angles in noncrystalline SiO_2 and in crystalline tridymite, is the reason that despite the lack of LRO in noncrystalline SiO_2, the SRO is still very high and there is *no need to introduce defects* in significant density to obtain the noncrystalline structure. The model of vitreous SiO_2 based on 4- to 8-member loops does not contain broken Si-O bonds (i.e., nonbridging oxygens) except on the surface. The essentially identical SRO in vitreous and crystalline SiO_2 as well as the lack of defects in vitreous SiO_2 are the reasons that the configurational entropy of vitreous silica is very low, ≈ 0.9 cal K^{-1}. For these reasons, fused SiO_2 and thermally or anodically grown SiO_2 films are considered *vitreous* rather than amorphous.

Despite the structural similarity between fused silica and thermally or anodically grown SiO_2 films on silicon, there is a significant difference in their crystallization behavior. The crystallization product of fused silica is practically always cristobalite, whereas thermally grown SiO_2 films on silicon may crystallize to cristobalite, quartz, or tridymite. Anodic oxidation of silicon may even result in a thin single crystal quartz film beneath the noncrystalline oxide. The fact that the unique crystallization behavior of SiO_2 films on silicon is very likely related to the presence of ordered regions (channels).

***π*-Bonding in SiO_2.**

The structure and chemistry of siloxane (silicone) polymers, whose framework consists of Si-O-Si groups, is usually discussed in terms of the properties of the Si-O bond. Similar considerations have also been applied to the structural chemistry of SiO_2 poly morphs and silicates. The concept of $d\pi$-$p\pi$ bonding arising from the overlap between the 2*p* orbitals of the oxygen containing the lone pair electrons and the 3*d* orbitals of silicon plays an important role in these considerations. An important feature of the Si-O bond is that π-bonding increases and the ionicity of the Si-O bond decreases with the Si-O-Si bond angle. Also, π-bonding can be influenced by other bonds, e.g., OH and O^-Na^+. The concept of π-bonding was also employed to explain the bond polarizability and other properties of SiO_2 polymorphs.

In contrast, the optical spectra of SiO_2 can usually be adequately interpreted without considering the Si 3d orbitals because, based on siloxane chemistry, $d\pi$-$p\pi$ bonding does not have an appreciable effect on the bond energy; however, it has a great effect on structure and chemical behavior. This is demonstrated by the small (0.14%) difference in the total energy of H_4SiO_4 calculated with and without involving the Si 3d orbitals. In contrast, the difference in molecular orbital energies is much larger for those orbitals which are affected by including the Si 3d orbitals than for those which are not affected by these orbitals. This is particularly true for that orbital which corresponds to π-bonding, that is, the orbital energy is reduced by 48% when the Si 3d orbitals are considered. The total population of electrons in the molecular orbitals which involve the Si 3d orbitals is 1.43 (0.36 per Si-O bond). These considerations of the H_4SiO_4 molecule are very relevant to SiO_2 since the X-ray fluorescense spectrum of silica could be interpreted on this basis. The 92.3 and 94.5 eV peaks in the $L_{2,3}$ spectrum are products of transitions (to Si 2*p*) from orbitals winch contain significant 3d contributions. The X-ray spectra of silicate and SiO_2 crystals have been also interpreted in terms of Si 3*d* admixture into the oxygen lone pair bands. It was pointed out that, although this admixture, is relatively small, its effect can be decisive in the structure and chemistry of SiO_2.

This effect π-bonding arises from the fact that the Si-O bond is mixed ionic- covalent. The σ component of the covalent bond (overlap between Si $3sp^3$ hybrid and O 2*p* orbitals) is responsible for the almost invariant nature of the $SiO_{4/2}$ in SiO_2 polymorphs. According to molecular orbital calculations the overlap between oxygen orbitals, i.e., the bond overlap population, expressed as *n*(Si-O), increases with increasing Si-O-Si bond angle (see Fig. 6.12). The Si-O bond overlap population increases even when the Si 3d orbitals are not included in the calculations. However, the confidence level of the correlation between Si-O bond length and *n*(Si-O) is better when the Si 3*d* orbitals are included, and with respect to *n*(Si-O), the $Si\text{-}O_{nb}$ conation is separated from the Si-O-Si configuration. This separation reflects the large difference in the chemical behavior of bridging and non-bridging oxygens. This overlap population is higher than the value of ≈0.36 per bond for H_4SiO_4 because the hydrogen in the OH group acts as an electron acceptor relative to silicon and, hence, decreases the π–component of the Si-O bond. These n(Si-O) values represent trends rather than absolute numbers and, hence they should be only considered as means for grouping and classifying. Concomitant with the increase in π-bonding, the ionicity of the bond decreases with angle so that the overall bond strength remains approximatety unchanged. Consequently, many properties of SiO_2 polymorphs, such as optical and X-ray spectra, as well as the free enthalpy of formation, are practically structure invariant. In contrast, several important properties (e.g., bond polarizability, infrared absorption and etching), depend on the extent of π-bonding and, hence, display a trend as a function of the Si-O-Si bond angle as shown in Fig. 6.12.

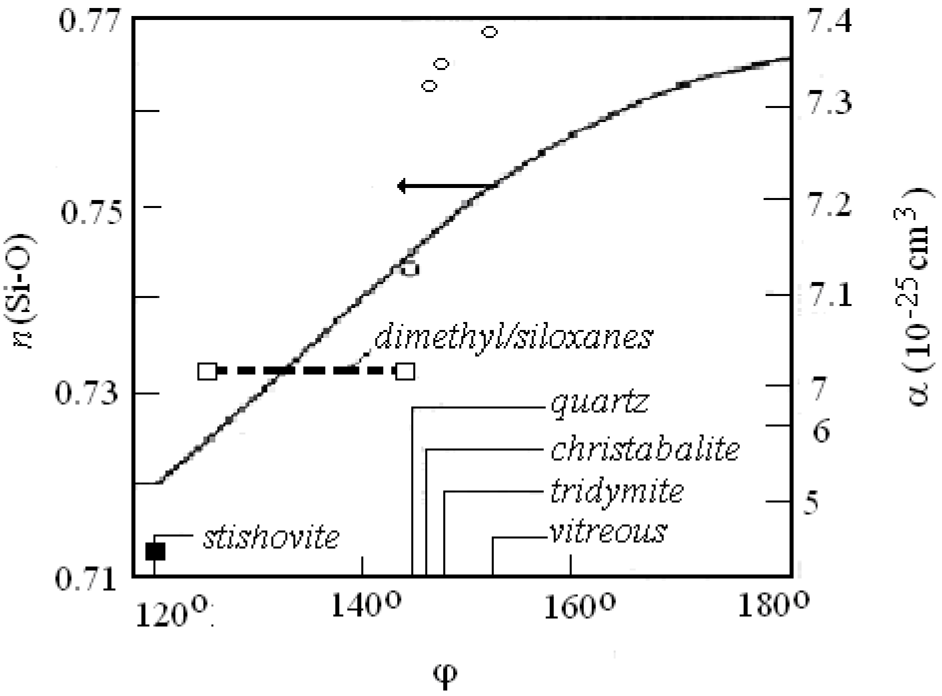


Fig. 6.12. Various properties of the Si-O bond as a function of the Si-O-Si bond angle. The solid line is the bond overlap population, n(Si-O) calculated from the Si(*spd*) basis. This curve does not hold for stishovite because the coordination of oxygen is three; i.e., it is not a bridging oxygen. The full and open symbols represent the Si-O bond polarizability, α, and refer to 6:3 and 4:2 Si-O coordinations, respectively. The bond angles representing tridymite and vitreous silica are the mean values of the corresponding angle distributions. Note the break in the α-scale.

Ionicity of the Si-O bond.

Figure 6.12 shows that the bond polarizability of stishovite, 4.32×10^{-25} cm^3 is much smaller than that of the other SiO_2 polymorphs and siloxanes ≈7.0 to $\approx7.4\times10^{-25}$ cm^3. Also, the Si-O stretching vibration frequency (not shown in Fig. 6.12) is much less for stishovite than for the other polymorphs, the respective values being 885 cm^{-1} and 1077 to 1106 cm^{-1}. Furthermore, the Si-O bond length is larger in stishovite (1.77 A°) than in $Si0_{4/2}$ polymorphs (1.59 to 1.63 A°). These observations indicate that Si-O bond is more ionic in stishovite than in the $SiO_{4/2}$ polymorphs. This conclusion is further strengthened by arguments based on the electronic dielectric properties The value of *f*, a coefficient associated with first and second neighbour delocalization, is 3.8 for stishovite; this value is characteristic of ionic crystals in which the anions are in close contact with each other as in rutile (TiO_2). In contrast, the *f* value for $SiO_{4/2}$ polymorphs are in the range of 5.0 to 5.1, which is typical of covalent solids The conclusion that the Si-O bond is more ionic in stishovite than m the $SiO_{4/2}$ polymorphs is contrary to the ionicity values 0.41 for stishovite and 0.53 to 0.65 for the $Si0_{4/2}$ polymorphs. The implications of the difference between the two interpretations are discussed below. Another great difference in the character of the Si-O bond in stishovite and $SiO_{4/2}$ polymorphs is the lack of π-bonding in the former: this is partially responsible for the fact that stishovite is soluble in H_2O but insoluble in HF, while the $SiO_{4/2}$ polymorphs behave in the opposite manner.

Since the SRO in the various $SiO_{4/2}$ polymorphs and siloxane polymers is essentially the same, the variation in their properties is not as great as the difference between $SiO_{4/2}$ polymorphs and stishovite. Nevertheless, this difference and the trend shown in Fig. 6.12, together with the increase in the ratio of the force constants of the Si-O-Si bending and Si-O stretching vibration with increasing Si-O-Si bond angle, demonstrate that *the ionicity of the Si-O bond decreases as the Si-O-Si angle increases.*

This conclusion is at variance with recent suggestions concerning the ionicity of the Si-O bond. The difference between the two interpretations, which is very pertinent to the chemical behavior of defects and other properties of SiO_2 films, is briefly discussed. One of the important conclusions is that the ionicity, F_i increases and the covalent character $(1 – F_i)$ of the Si-O bond decreases with increasing Si-O-Si bond angle, φ. This conclusion was reached on the basis of a relationship between the electronic dielectric constant, $\varepsilon(0)$, and F_i: $\varepsilon(0) = 1 + 1.26d^2(l - F_i)$, where d is the Si-O bond length in angstroms. The fact that this relationship is incorrect in this case is clearly demonstrated by the wrong order of ionicity values for stishovite and $SiO_{4/2}$

polymorphs as discussed above. This gross discrepancy exists because the dielectric ionicity concept which underlies this relationship cannot be considered as a quantitative guide to the correlation between bond and dielectric properties; the first-neighbour delocalization (covalency) and second-neighbour delocalization (anion-anion contact) as expressed by the structural parameter, *f*.

This point can be further illustrated by the internal inconsistency of using the above relationship and another closely related equation for the ionicity,

$$F_i = \frac{C^2}{E_h^2 + C^2} \qquad (6.16)$$

where

$$E_h^2 + C^2 = E_g^2 \qquad (6.17)$$

In this equation, E_g is the average bandgap, C and E_h are its ionic (coulombic) and covalent components, respectively. The values of F_i are 0.57 for quartz and 0.65 for vitreous $Si0_2$, representing a change of 14%. According to (6.16), the corresponding change in the value of $[C^2/(C^2 + E_h^2)]^{1/2}$ should be ≈4%. However, both the experimental and theoretical studies indicate that the optical bandgap of the $Si0_{4/2}$ polymorphs is essentially structure-invariant.

The F_l value of stishovite is ~ 30% less than that of quartz. Hence, the optical bandgap of stishovite should be significantly less than that of quartz. However, based on the similarity in the packing of oxygen atoms in rutile and stishovite, it can be easily shown that the optical bandgap of stishovite is essentially identical to that of quartz; the refractive index of stishovite calculated by this assumption (1.82) is practically identical to the experimental value (1.83). The 14-% variation in F_i, between quartz and vitreous SiO_2 is much larger than the corresponding change in the bond overlap population, ≈3% and is in the opposite direction. This discrepancy exists for the same reason as the discrepancy concerning the ionicity of stishovite discussed above.

The ionicity, F_i was represented as a monotonically increasing function of φ. Subsequently, this relationship was modified on the basis of a qualitative argument involving repulsion between adjacent Si atoms; the pertinent curves are shown in Fig. 6.13. However, the points in Fig. 6.13 clearly demonstrate that the F_i values simply exhibit a random scatter, particularly if the correct values for d and φ are used. Hence, the ionicity of the Si-O bond increases and the covalency decreases with the Si-O-Si bond angle is unjustified, even if the theoretical considerations were correct (however, they are not). Thus, not withstanding the rejection of the conceptual framework based on π-bonding as an "unclear point, the properties of SiO_2 polymorphs can be interpreted in a consistent manner on this basis. This was apparently recognized.

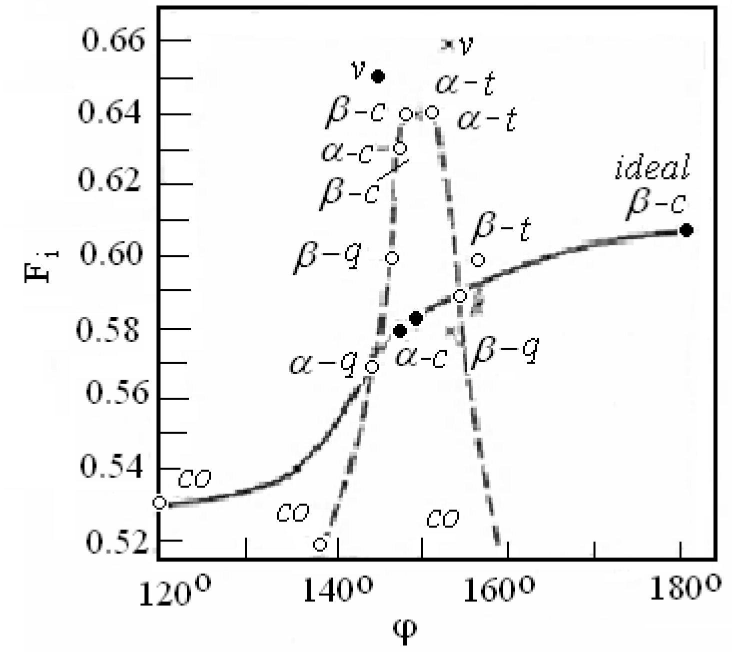


Fig. 6.13. Ionicity, Fi, of the Si-O bond as a function of the Si-O-Si bond angle φ. The abbreviations *v, t, c, q, k*, and *co* represent vitreous, tridymite, cristobalite, quartz, keatite, and coesite, respectively. Ideal β-cristobalite refers to a *hypothetical* structure with $\varphi = 180°$ The crosses represent Fi values calculated using up-to-data values for the Si-O bond length and plotted against up-to-date φ. Average *d* and φ values were used in these calculations for vitreous SiO_2, α and β tridymite, and coesite.

Si-O bond and defect structure of vitreous SiO_2.

With respect to the defect structure of vitreous SiO_2, an important effect is the decrease of the Si-O bond length with increasing Si-O bond overlap, i.e., with increasing Si-O bond angle as shown in Fig. 6.14. This means that the bond/structural flexibility of vitreous SiO_2 arising from the wide distribution of the Si-O-Si bond angles is further increased. Figure 6.14 also shows that for a given bond length the bond overlap is larger for the $Si-O_{nb}$ (O_{nb}= non-bridging oxygen) than the Si-O-Si configuration (containing a bridging oxygen) As will be discussed below, this increased bond overlap in the $Si-O_{nb}$ configuration has significant consequences with respect to network defects in vitreous SiO_2.

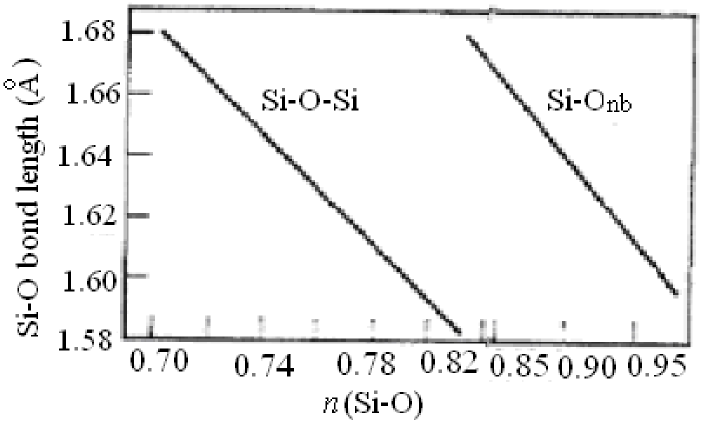


Fig. 6.14. Relationship between Si-O bond length and bond overlap population. Note the break in the abscissa.

Based on the behavior of bond polarizability, infrared spectra, etc., it has been suggested that vitreous SiO_2 has the highest π-bond order among SiO_2 polymorphs, excluding the unstable melanophlogite (see Fig. 6.12). This idea has been strengthened by a model of the α-β phase transformation of quartz crystal. As the temperature increases from 25 to 600 °C, the Si-O-Si bond angle increases from 144° to 153° and the Si-O bond length decreases from 1.607 to 1.59 A°; this has been interpreted as an increase in π-bonding with temperature. It is well known from crystallography that the high temperature polymorph is always characterized by increased symmetry relative to the low temperature polymorph. Vitreous SiO_2, which is isotropic, represents the highest symmetry among SiO_2 polymorphs and, hence, is expected to have the highest π-bond order.

Another feature of the α-β phase transformation is that increasing π-bond order tends to pull the oxygen atoms to the plane perpendicular to the c-axis; this explains the negative thermal expansion coefficient in the *c*-direction in β-quartz. The atomic arrangement along the *c*-axis is such that Si and O atoms form a helix along the *c*-axis where each loop consists of six Si atoms. This rearrangement of the oxygen atoms indicates that π-bonding increases preferentially along the axis of the structural channel as the temperature increases.

A similar phenomenon is that those five Si-O-Si bond angles in tridymite, which are larger than 160°, are between layers (consisting of Si-O-Si loops) rather than within a layer. The layers are connected in such a manner that structural channels are present along the c-axis. These observations demonstrate that structural channels in crystalline SiO_2 polymorphs represent increased π-bonding between the Si and O atoms forming these channels. Thus, the earlier suggestion that channel defects in vitreous SiO_2 are characterized by increased π-bonding appears to be substantiated.

6.8. Oxidation of Polycrystalline Silicon

In most integrated circuits, the polysilicon is electrically isolated from overlying conductors (either metal or additional layers of polysilicon) by silicon dioxide, which may be formed by thermal oxidation of the polysilicon or by chemical vapor deposition. In most integrated circuits these oxide layers must simply be highly insulating. However, specialized devices, such as the electrically erasable, programmable read-only memories (EEPROMs) used with increasing frequency in VLSI circuits require a thin oxide with well-controlled conductivity above the polysilicon.

Numerous studies have shown that the oxidation rate of polysilicon can differ substantially from that of single-crystal silicon and also that the electrical properties of the oxide grown on polysilicon are different from those of a similar thickness of oxide grown on single-crystal silicon. In this section, the differences in the oxidation rates of polysilicon and single-crystal silicon will be first examined and these differences will be related to the structure of the polysilicon. In some applications the differences in the oxidation rate can be used constructively, while in other applications, they complicate integrated-circuit fabrication.

6.8.1. Oxide Growth on Polysilicon

Undoped Films. Substantial differences can exist between the thickness of oxide grown on polysilicon and that of oxide simultaneously grown on single-crystal silicon. For lightly doped films one of the dominant factors leading to this difference is the presence of grains with different orientations in the polysilicon. To demonstrate the importance of differently oriented grains, in one study [14] thick layers of polysilicon were polished so that the resulting smooth surface contained regions with different crystal orientations. When oxidized, each differently oriented grain oxidizes at a rate characteristic of that particular orientation of crystalline silicon. Grains with a given orientation exhibit the same oxide color as seen on the similarly oriented, single-crystal control wafers. These differences are accentuated under surface-reaction-limited oxidation conditions and reduced when the oxidation is performed under diffusion-limited conditions. Although polishing the samples removed the faceted surface structure which causes the exposed crystal planes to differ from the grain orientation, the study did show that the macroscopically measured oxide thickness can be expected to be a suitable average of the oxide thicknesses grown on differently oriented grains.

A similar trend is seen in oxides grown on the fine grain polysilicon typically used in integrated circuits. The oxide thicknesses grown on thin polysilicon films deposited either in an atmospheric-pressure reactor at 960°C or in a low-pressure reactor at 625°C are between the oxide thicknesses grown on rapidly oxidizing (111)-oriented, single-crystal silicon and slowly oxidizing (100)-oriented silicon [11]. Because the oxide thickness grown depends on the dominant crystal orientations in the film being oxidized, it can differ significantly for polysilicon deposited under different conditions, necessitating process modification when the polysilicon deposition conditions change significantly. The crystalline texture of polysilicon films depends strongly on the deposition temperature and the oxide thickness should vary correspondingly.

Heavily Doped Films. Although oxidation of lightly doped polysilicon is similar to that of lightly doped single crystal silicon, marked differences are seen between the oxidation of heavily doped polysilicon and similarly doped single-crystal silicon. Heavily doped single-crystal silicon oxidizes much more rapidly than does lightly doped single-crystal silicon because of the excess point defects present in the heavily doped material. As the influence of the added point defects begins to dominate the oxidation process, the differences between the oxidation rates of the different orientations of silicon decrease significantly. However, the oxidation-rate enhancement on polysilicon is usually much less than that on single-crystal silicon doped at the same time [15].

Figure 6.15 [15] shows the thickness of oxide grown on two different orientations of single-crystal silicon and on polysilicon deposited under two different deposition conditions. Varying amounts of phosphorus were added by gaseous diffusion from a $POC1_3$ source to sets of samples each containing all four types of material, and then all samples were simultaneously oxidized under surface-reaction-limited conditions. The resulting oxide thicknesses are shown as functions of the sheet resistance measured on the (100)-oriented single-crystal silicon contained in each set of samples. As the dopant concentration increases (decreasing sheet resistance), the oxide thickness grown on the polysilicon becomes a smaller fraction of that grown on single-crystal silicon, even though the amount of dopant added to the polysilicon is expected to be at least as great as that added to the single-crystal silicon.

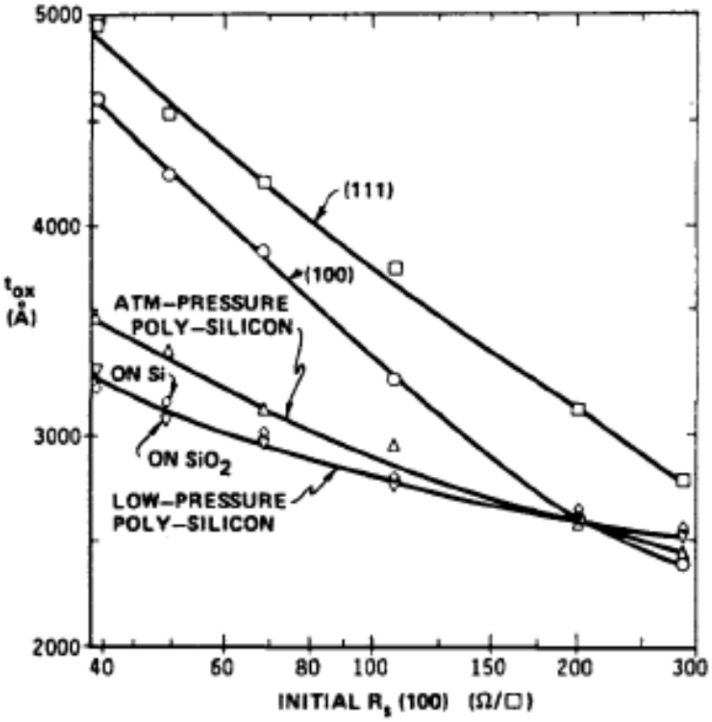


Fig. 6.15. Oxide thickness grown on phosphorus-doped single-crystal and polycrystalline silicon during a 150 min, 850°C, pyrogenic steam oxidation as a function of the sheet resistance R_S measured on the (100)-oriented, single-crystal silicon wafer in each set.

This apparently anomalous behavior is best understood by considering the interaction of diffusion and oxidation. The impurities diffuse much more rapidly in polysilicon than in single-crystal silicon [16]. During doping and oxidation, therefore, the phosphorus added near the surface can diffuse toward the back of the silicon film more readily in polysilicon than in single-crystal silicon. The surface concentration is lower; and, consequently, a thinner oxide is grown. The differences between the oxide thicknesses grown on the different types of polysilicon depend on the ease of diffusion in each sample, which is governed by the detailed grain structure.

An alternate explanation suggests that the difference between the oxide thicknesses grown on polysilicon and on single-crystal silicon is dominated by the different electrical activity of the dopant in each type of material. Because the dopant is less active in polysilicon, the Fermi level is closer to the intrinsic Fermi level during oxidation; the charged point defects which enhance the oxidation are less numerous; and the oxide grown is thinner. The different behavior of the point defects in single-crystal silicon and in polysilicon may also affect the oxidation rate. Grain boundaries and other structural defects in polysilicon can act as recombination sites for point defects, increasing their concentration gradients near the surface. If the silicon interstitials injected by the oxidation process can diffuse away from the oxidizing surface more readily, the oxidation rate should increase, rather than decrease.

Differences in the dopant-diffusion rates between polysilicon and single-crystal silicon can also explain the *more rapid* oxidation of heavily doped polysilicon than single-crystal silicon sometimes seen. We can understand this behavior by considering the finite thickness of the polysilicon. Table 6.4 [15] shows that the oxide thickness grown on heavily doped polysilicon also depends on the thickness of the polysilicon film and is greater on thinner films than on thicker films. In thick films, the dopant can readily diffuse away from the polysilicon surface, reducing the surface dopant concentration and the oxidation rate, as discussed above. In thinner polysilicon films, however, the finite thickness confines the dopant atoms; as dopant atoms approach the back surface, the increasing concentration of dopant there decreases the concentration gradient driving the diffusion. The surface concentration remains high, and the oxide grown is, consequently, thicker.

Table 6.4. Oxide thicknesses grown on n^+ polysilicon films of different thicknesses during a 75 min, 850°C, pyrogenic steam oxidation.

Polysilicon thickness (μm)	Oxide thickness (nm)	
	LP	AP
0.5	393	378
1.0	280	301
1.5	276	310
n^+ (100) single crystal	515	

The increase in oxide thickness with increasing dopant concentration does not continue indefinitely, however. It saturates at a value approximately corresponding to the solid solubility of phosphorus in silicon at the oxidation temperature. As shown in Fig. 6.16 [17], when oxidation occurs at 750°C, the oxide growth rate remains constant for phosphorus chemical concentrations greater than about 10^{21} cm^{-3}, which is close to the solid solubility of phosphorus in silicon at this temperature, especially if segregation of the dopant at grain boundaries is considered. Thus, the oxide thickness depends on the phosphorus concentration from the concentration at which the Fermi level at the oxidation temperature departs from the intrinsic Fermi level up to the concentration corresponding to the solid solubility of the dopant in silicon.

The considerably greater oxide thickness grown on heavily doped polysilicon than on lightly doped, single-crystal silicon can be used advantageously in the fabrication of many integrated circuits. In a typical silicon-gate integrated circuit, the heavily doped, *n*-type polysilicon gate is adjacent to the lightly doped single-crystal silicon regions which subsequently form the source and drain of an MOS transistor. After definition of the polysilicon, however, these single-crystal regions are still lightly doped. Oxidation under suitable conditions produces a thick oxide on the polysilicon, while only a thin oxide is grown on the single-crystal silicon. The thick oxide on the polysilicon can serve as an implant mask during processing, or it can be used to reduce capacitance in the finished circuit. The differential oxidation rate is also useful in reducing critical mask alignment; the thin oxide on the single-crystal silicon can be removed while not exposing the heavily doped polysilicon, which is covered by a much thicker oxide.

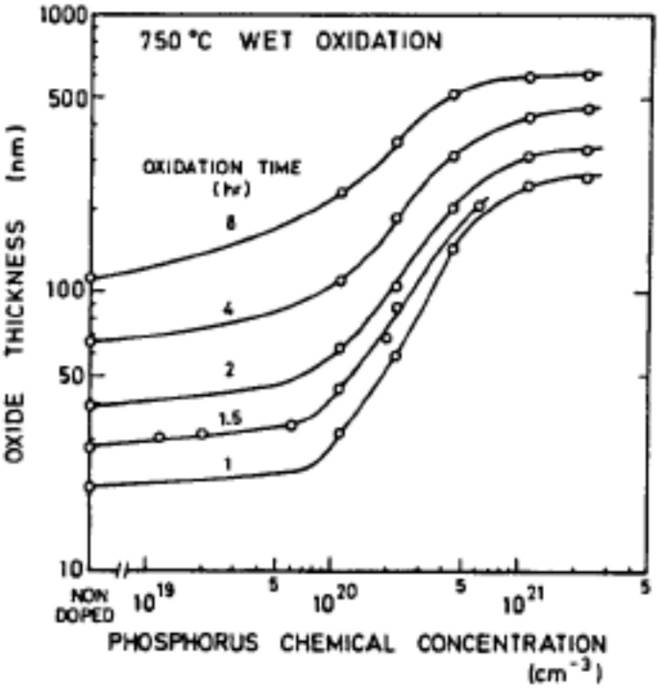


Fig. 6.16. Oxide thickness grown on heavily phosphorus doped polysilicon during a 750°C, wet-oxygen oxidation [17].

Grain-Boundary Oxidation. The disordered structure near the grain boundaries might be expected to oxidize more rapidly than the crystalline structure near the center of the grains. In the study of polished samples of undoped, large grain polysilicon, deposited at a high temperature, careful examination of the grain-boundary regions did not reveal any enhanced oxidation at the grain boundaries. Because grain boundaries occupy a larger fraction of fine-grain polysilicon, grain-boundary effects might be expected to play a more important role in the oxidation of fine-grain material, especially in films deposited at lower temperatures, which might be expected to be less ordered. If the grain boundaries oxidize rapidly, the oxide formed there should cause considerable compressive stress in the oxidized film. However, no tendency toward stress is caused by oxidation of *undoped* films [18]. In fact, a slight increase of tensile stress is found, suggesting that the heat treatment orders the structure near the grain boundaries. On the other hand, in phosphorus-doped films oxidation causes compressive stress [18]. In addition, examination of the local film thickness near the grain boundaries as oxidation proceeds shows that the entire thickness of the polysilicon film is consumed first at the grain boundaries while unoxidized silicon still remains near the centers of the grains [18]. Although differences in the thickness of the polysilicon film near the center of the grains and near the grain boundaries could also cause the polysilicon to be completely consumed by oxidation near the grain boundaries first, these observations suggest that the region near the grain boundaries oxidizes more rapidly than does the silicon away from the grain boundaries. Because the grain boundaries are the lowest portions of the polysilicon surface before oxidation and they oxidize most rapidly, the surface roughness of the polysilicon is expected to increase as oxidation proceeds.

Other studies have confirmed the more rapid oxidation near grain boundaries. High resolution, cross-section transmission electron microscopy shows that Si-P precipitates can form at grain boundaries when the phosphorus concentration exceeds its solid solubility at the oxidation temperature [14]. Because this phase oxidizes more rapidly than does silicon, the oxide near the grain boundaries can be considerably thicker than that over the centers of the grains. Consider a polysilicon film doped to solid solubility at an intermediate temperature. When it is oxidized at a higher temperature, the phosphorus concentration is below solid solubility at the oxidation temperature, and no enhanced oxidation is expected at the grain boundaries. At lower oxidation temperatures the phosphorus concentration is greater than solid solubility; excess phosphorus concentrates at the grain boundaries; and the grain-boundary regions oxidize more rapidly than do the centers of the grains. In addition to forming more rapidly, the oxide grown over grain-boundary precipitates appears to contain a high phosphorus concentration and, therefore, etches more rapidly than does SiO_2. As the oxide grown on polysilicon is etched, narrow grooves can be left in the polysilicon surface. In the extreme case, the indentation can extend through the entire thickness of the polysilicon film so that the underlying oxide (e.g., the gate oxide) can be attacked.

6.8.2. Oxide-Thickness Evaluation

Although measuring the oxide thickness on single-crystal silicon nondestructively by optical techniques is straight forward using an ellipsometer or a spectrophotometer, measuring the oxide thickness grown on polysilicon is more complex because of the multilayer structure on which the oxide is grown. At the wavelengths typically used for oxide-thickness measurements (e.g. λ = 628 nm for ellipsometry or λ = 400-800 nm for spectrophotometry), polysilicon is transparent, and the reflected signal being analyzed is influenced by reflections from the underlying interfaces, as well as from the top and bottom interfaces of the oxide layer grown on the polysilicon. In theory, the reflectance of the total multilayer structure can be analyzed, and the oxide thickness of the polysilicon can be extracted. However, in most cases variations in the thicknesses of the underlying layers make the indicated thickness of the oxide grown on the polysilicon layer too uncertain for practical use.

The optical techniques can, however, be adapted to the ultraviolet wavelength range, in which polysilicon is opaque. In this case, the reflected signal is dominated by interference at the top and bottom of the oxide layer grown on the polysilicon. The wavelength range from 200 to 400 nm is suitable for this type of measurement. However, near the lower end of this wavelength range, the surface roughness of the polysilicon also affects the reflected signal, and at λ =280 and 370 nm, structural bands of crystalline silicon can influence the reflected signal. The interference of light reflected from the two surfaces of the top oxide is strong, however, and the technique is a useful, nondestructive, method of measuring the oxide thickness grown on polysilicon. Figure 6.17 shows the interference signal obtained over the wavelength range from 200 to 500 nm, with interference in the top oxide below about 400 nm and interference in the multilayer structure at higher wavelengths. Ellipsometry using ultraviolet light can also be employed to determine the oxide thickness, again taking advantage of the fact that silicon is opaque to ultraviolet light.

Of course, the oxide thickness can be determined destructively by etching a step and measuring the step height with a surface profilometer. Although this technique is the most straightforward, it can be more time consuming than the optical techniques, its resolution is limited by the surface roughness of the polysilicon, and it cannot readily be used on device wafers.

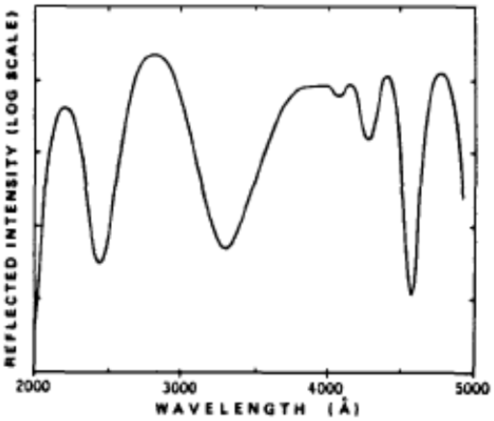


Fig. 6.17. The reflectance in the wavelength range from 200 to 400 nm arises from interference in the oxide above the polysilicon, while the signal at longer wavelengths is influenced by reflection at interfaces beneath the polysilicon as well.

6.9. Silicon Dioxide Deposition

There are three commonly used deposition methods: atmosphericpressure CVD, low-pressure CVD (LP CVD), and plasma-enhanced chemical vapor deposition (PE CVD, or plasma deposition). PE CVD is an energy-enhanced CVD method, in which plasma energy is added to the thermal energy of a conventional CVD system. Considerations in selecting a deposition process are the substrate temperature, the deposition rate and film uniformity, the morphology, the electrical and mechanical properties, and the chemical composition of the dielectric films.

The reactor for atmospheric-pressure CVD is similar to the one for oxidation**,** except that different gases are used at the gas inlet. In a hot-wall, reduced-pressure reactor as shown in Fig. 6.18 *a,* the quartz tube is heated by a three-zone furnace, and gas is introduced at one end and pumped out at the opposite end. The semiconductor wafers are held vertically in a slotted quartz boats [19]. The quartz tube wall is hot because it is adjacent to the furnace, in contrast to a cold-wall reactor such as the horizontal epitaxial reactor that uses radio frequency (rf) heating.

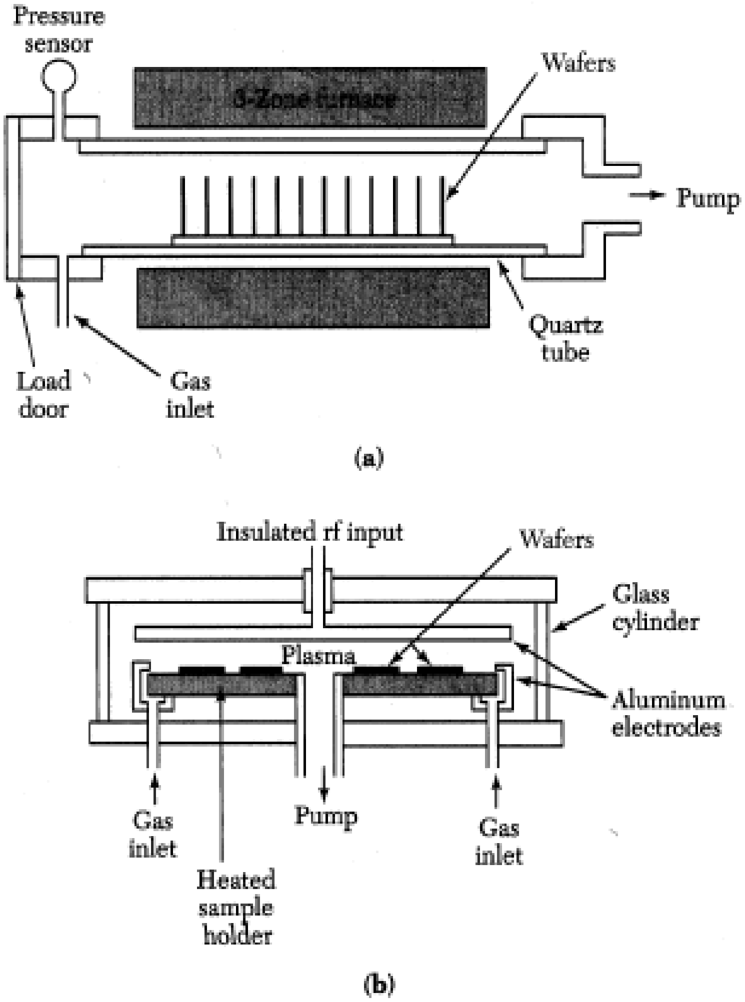


Fig. 6.18. Schematic diagrams of chemical-vapor deposition reactors. (a) Hot-wall, reduced-pressure reactor. (b) Parallel-plate plasma deposition reactor [19].

The parallel-plate, radial-flow, PE CVD reactor shown in Fig. 6.18,b consists of a cylindrical glass or aluminum chamber sealed with aluminum endplates. Inside are two parallel aluminum electrodes. An rf voltage is applied to the upper electrode, whereas the lower electrode is grounded, The rf voltage causes a plasma discharge between the electrodes. Wafers are placed on the lower electrode, which is heated between 100°C and 400°C by resistance heaters. The reaction gases flow through the discharge from outlets located along the circumference of the lower electrode. The main advantage of this reactor is its low deposition temperature. However, its capacity is limited, especially for large-diameter wafers, and the wafers may become contaminated if loosely adhering deposits fall on them.

CVD silicon dioxide does not replace thermally grown oxides because the best electrical properties are obtained with thermally grown films. CVD oxides are used instead to complement the thermal oxides. A layer of undoped silicon dioxide is used to insulate multilevel metallization, to mask ion implantation and diffusion, and to increase the thickness of thermally grown field oxides. Phosphorus-doped silicon dioxide is used both as an insulator between metal layers and as a final passivation layer over devices. Oxides doped with phosphorus, arsenic, or boron are used occasionally as diffusion sources.

6.9.1. Deposition Methods

Silicon dioxide films can be deposited by several methods. For low-temperature deposition (300°C - 500°C), the films are formed by reacting silane, dopant, and oxygen. The chemical reactions for phosphorus-doped oxides are

$$SiH_4 + O_2 \rightarrow^{(450°C)} SiO_2 + 2H_2, \qquad (6.18)$$

$$4PH_3 + 5O_2 \rightarrow^{(450°C)} 2P_2O_5 + 6H_2. \qquad (6.19)$$

The deposition process can be performed either at atmospheric pressure in a CVD reactor or at reduced pressure in an LP CVD reactor (Fig. 6.18a). The low deposition temperature of the silane-oxygen reaction makes it a suitable process when fdms must be deposited over a layer of aluminum.

For intermediate-temperature deposition (500°C-800°C), silicon dioxide can be formed by decomposing tetraethylorthosilicate, $Si(OC_2H_5)_4$ in an LP CVD reactor. The compound, abbreviated TEOS, is vaporized from a liquid source. The TEOS compound decomposes as follows:

$$Si(OC_2H_5)_4 \rightarrow^{(700°C)} SiO_2 + \text{by-products}, \qquad (6.20)$$

forming both SiO_2 and a mixture of organic and organosilicon by-products. Although the higher temperature required for the reaction prevents its use over aluminum, it is suitable for polysilicon gates requiring a uniform insulating layer with good step coverage. The good step coverage is a result of enhanced surface mobility at higher temperatures. The oxides can be doped by adding small amounts of the dopant hydrides (phosphines, arsine, or diborane), similar to the process in epitaxial growth.

The deposition rate as a function of temperature varies as $exp(-E_a/k_BT)$, here E_a, is the activation energy, The E_a of the silane-oxygen reaction is quite low: about 0.6 eV for undoped oxides and almost zero for phosphorus doped oxide. In contrast, E_a for the TEOS reaction is much higher: about 1.9 eV for undoped oxide and 1.4 eV when phosphorus doping compounds are present. The dependence of the deposition rate on TEOS partial pressure is proportional to $(1- e^{-P/P0})$, where P is the TEOS partial pressure and P_0 is about 30 Pa. At low TEOS partial pressures, the deposition rate is determined by the rate of the surface reaction. At high partial pressures, the surface becomes nearly saturated with adsorbed TEOS and the deposition rate becomes essentially independent of TEOS pressure [19].

The atmospheric-pressure and low-temperature CVD processes using TEOS and ozone (O_3) have been proposed, as shown in Fig. 6.19 [20]. This CVD technology produces oxide films with high conformality and low viscosity under low deposition temperature. In addition, the shrinkage of oxide film during annealing is also a function of ozone concentration, as shown in Fig. 6.20. Because of their porosity, O_3-TEOS CVD oxides are often accompanied by plasma-assisted oxides to permit planarization in ULSI processing.

For high-temperature deposition (900°C), silicon dioxide is formed by reactin dichlorosilane, $SiCl_2H_2$, with nitrous oxide at reduced pressure:

$$SiCl_2H_2 + 2N_2O \rightarrow^{(900°C)} SiO_2 + 2N_2 + 2HCl. \qquad (6.21)$$

This deposition gives excellent film uniformity and is sometimes used to deposit insulating layers over polysilicon.

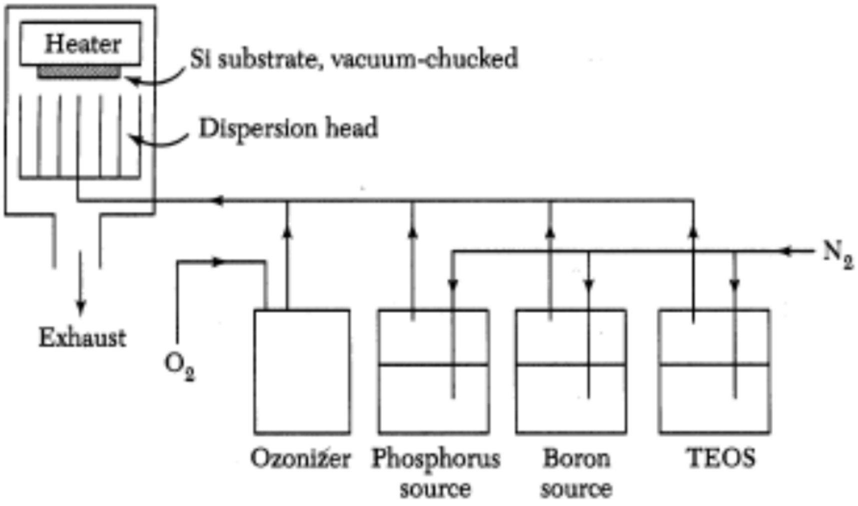


Fig. 6.19. Experimental apparatus for the O_3-TEOS chemical-vapor deposition (CVD) system.

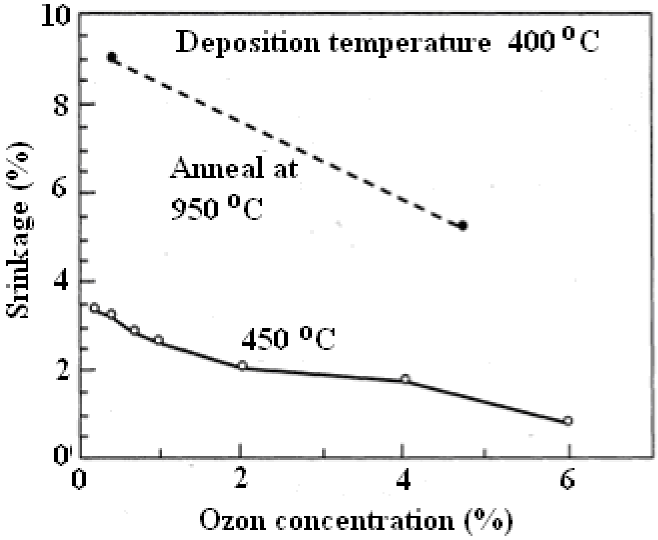


Fig. 6.20. Dependence of the shrinkage of the O_3-TEOS CVD film on ozone concentration using annealing.

6.9.2. Properties of Silicon Dioxide

Deposition methods and properties of silicon dioxide films are listed in Table 6.5 [19]. In general, there is a direct correlation between deposition temperature and film quality. At higher temperatures, deposited oxide films are structurally similar to silicon dioxide that has been thermally grown.

Table 6.5. Properties of SiO_2 films.

Property	Thermally grown at 1000°C	$SiH_4 + O_2$ at 450°C	TEOS at 700°C	$SiCl_2H_2$ + N_2O at 900°C
Composition	SiO_2	SiO_2(H)	SiO_2	SiO_2(Cl)
Density (g/cm^3)	2.2	2.1	2.2	2.2
Refractive index	1.46	1.44	1.46	1.46
Dielectric strength (10^6 V/cm)	>10	8	10	10
Etch rate	30	60	30	30

(A°/min) (100:1 H_2O:HF				
Etch rate (A°/min) (buffered HF)	440	1200	450	450
Step coverage	-	Nonconformal	Conformal	Conformal

The lower densities occur in films deposited below 500°C. Heating deposited silicon dioxide at temperatures between 600°C and 1000°C causes densification, during which the oxide thickness decreases, whereas the density increases to 2.2 g/cm^3. The refractive index of silicon dioxide is 1.46 at a wavelength of 0.6328 μm. Oxides with lower indices are porous, such as the oxide from the silane-oxygen deposition, which has a refract index of 1.44. The porous nature of the oxide also is responsible for the lower dielectric strength, which is the applied electric field that will cause a high current to flow in the oxide film. The etch rates of oxides in a hydrofluoric acid solution depend on deposition temperature, annealing history, and dopant concentration. Usually higher-quality oxides are etched at lower rates.

6.10. Silicon Oxide Nanoparticles

Oxide nanoparticles are of particular interest in the fabrication of electronic devices based on the ultra-large scale integration (ULSI). Integrated circuits become more and more complex with increasing layers of metallization and interlayer dielectrics. This has led to the search for a method to planarize the wafers between the deposition and the processing steps. The method adopted is chemical mechanical polishing (CMP), which involves pressing the Si wafers against a polishing pad in the presence of colloidal slurries of oxides such as SiO_2, Al_2O_3, CeO_2, or ZrO_2. The size, shape, surface chemistry, and stability of the oxide particles govern the efficiency of the process. The CMP process has turned out to be an important step in the semiconductor industry [21, 22].

Prior to the synthesis of a nanophase material, the size and the dimensional features of the material to be prepared have to be defined. Accordingly, a suitable preparative method can be adopted. This is very crucial because the particle size, dimensionality, and composition govern the resultant properties of the nanostructured materials that are assembled from nanoparticles as building units in order to achieve certain desired properties. For instance, the fascinating electronic, electrical, optical, catalytic, thermal, and mechanical properties of the nanoparticles and the nanostructured materials have triggered enormous multipronged research activities in recent years in this area.

Several methods have been developed over the years for the synthesis of nanomaterials. The various routes include synthesis using liquid, solid, or gas phase precursors, which come under physical or chemical processing. The major issues for the synthesis of nanoparticles are (*i*) the control of the particle size and composition, and (*ii*) the control of the interfaces and the distributions of the nanobuilding blocks within the fully formed nanostructured materials.

The oxide nanomaterials are of interest in a variety of structural and functional applications. The advantages of equiaxed, dispersive, submicron ceramic oxide particles with a narrow size distribution have been intensively discussed [23]. Typical nanoparticles are agglomerates of several primary particles. The agglomerates are termed as secondary particle. The secondary particle size is obtained by scanning electron microscopy (SEM), whereas the X-ray line broadening helps to estimate the primary particle size or crystallite size. Minimizing the agglomeration and deriving the properties exclusive to the primary particles are the important objectives in the synthesis of oxide nanoparticles. Let's consider some important methods for the preparation of oxide nanoparticles are described.

Gas Phase Condensation Methods

The production of nanoparticles by condensation of gaseous precursor molecules generally comes under the category of aerosol processes [24, 25]. For the synthesis of nanoparticles or ultra-fine particulates (particle size <100 nm), the gas phase condensation is a suitable method. The initial step in this process is the formation of gaseous precursor molecules by a suitable physical or chemical method in aerosol reactors [26 - 30]. The precursor molecules then react in the vapor phase to form tiny nuclei of the desired phase (gas-to-particle conversion). The size of a typical nucleus thus formed has dimensions comparable to that of a molecule of a refractory oxide. Subsequently, the primary particles undergo collision and coalescence to form aggregates, which in turn form agglomerates held together by the weak van der Waals forces [31, 32]. The method is suitable for the production of nanoparticles in laboratories as well as in industries. The mechanism involved in the gas phase condensation process allows one to link the properties of gaseous precursor to the nanoparticles formed. The important parameters involved in the processing are time, temperature, and the amount of particles produced per unit volume. The decisive characteristics of the prepared nanoparticles include the size distribution of the primary particles, the grain boundaries, the pore sizes, the defect concentrations, and crystallinity. Apart from these characteristics, the fractal dimension of the aggregates and the particle bond energies are also of interest for investigation. The various methods used for the gas phase condensation of oxide nanoparticles are summarized as a block diagram in Fig. 6.21 [33]. Let's consider an important preparative route—flame processing related techniques such as sputtering for oxide nanoparticles. The laser ablation and microwave plasma-assisted chemical reactions in the gas phase are other interesting methods.

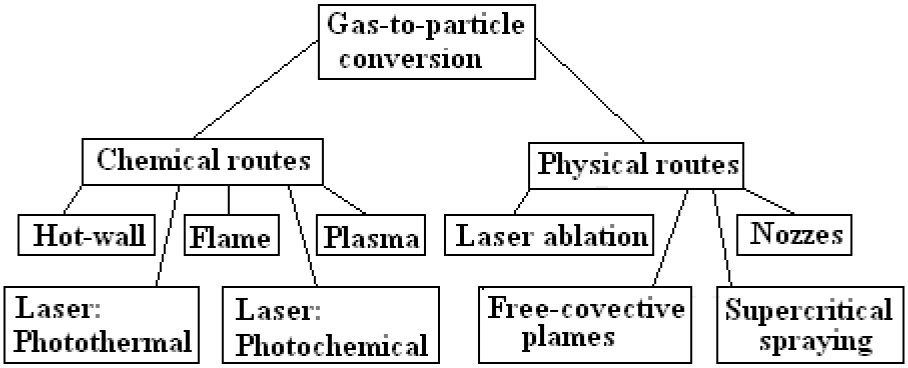


Fig. 6.21. Methods used for gas phase condensation of nanoparticles [33].

Flame Processing.

Oxide nanoparticles are produced commercially on a large scale in a flame reactor [29, 30]. The precursor in vapor phase is fed into a reactor in the presence of oxygen and ignited. The burning step may also take place in other gaseous atmospheres such as inert gas, hydrogen, or methane. The schematic diagram for the particle formation in a flame reactor is shown in Fig. 6.22 [33], which illustrates the eventual formation of nanoparticles through collision, coalescence, and agglomeration in an aerosol reactor. Temperatures as high as 1200–2500 K can be achieved in a typical flame reactor. The method is commercially used for the production of industrially useful silica (SiO_2) from silicon tetrachloride vapor. It is also important to correlate the particle size of the material produced to the processing conditions and the material properties in order to standardize the production method for the nanoparticles. The particle size and the surface area of the oxide can be tuned by adjusting the process temperature. The lower the process temperature, the smaller the particle size and larger the surface area. The sizes of the oxide nanoparticles that are produced using the aerosol technique fall in the range of 7 to 27 nm.

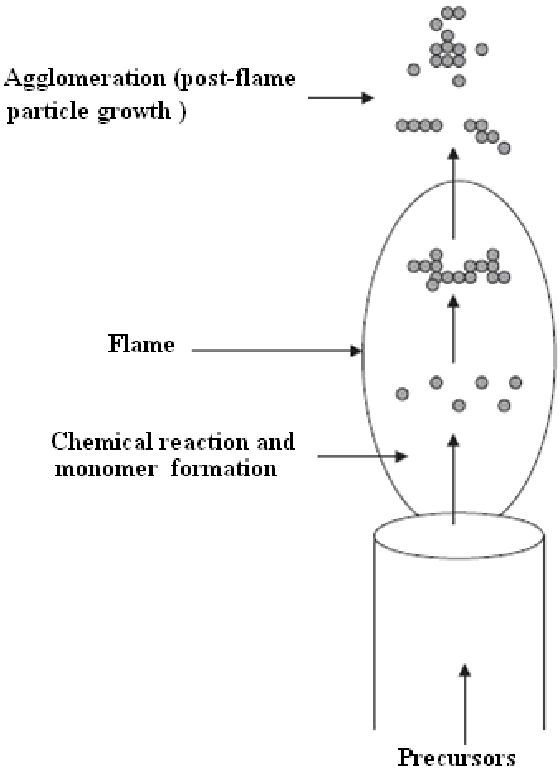


Fig. 6.22. Schematic representation of particle formation in a flame reactor [33].

Electric Dispersion Reaction

The electric dispersion reaction is a precipitation reaction that is carried out in the presence of a pulsed electric field to synthesize ultrafine precursor powders of advanced ceramic materials. The technique involves subjecting the reactor liquid (metal-alkoxide solution) to a dc electric field (3 to 10 kV/cm at pulsing frequencies in the range 1 to 3 kHz). Under the applied electric field, the sol is shattered to micron-sized droplets, termed as microreactors that contain hydrous precursor precipitate. The formed precursor powders can be thermally processed to obtain oxide nanoparticles. This technique has been developed for the synthesis of various ultrafine ceramic powders from the corresponding metal alkoxide solution [34-37]. A typical microreactor droplet for the formation of nanosized oxide particles is schematically represented in Fig. 6.23. Metal alkoxide and metal salt solutions have been used for the synthesis of ceramic precursor powders (including SiO_2, Al_2O_3, ZrO_2, and 1:2:3 yttrium-barium-copper oxide) with chemical homogeneity at submicrometer level

[34]. The method also allows control and modification of the particle morphology. Moreover, the method has been shown to be a viable means for the production of microspheres and porous spherical shells. A scanning electron micrograph of hydrous 1:2:3 Y-Ba-Cu oxide particles synthesized by an electric dispersion reaction from the corresponding metal nitrates revealed that the particle sizes were in the micrometer range (~5 µm), which on further thermal processing form submicrometer powders [34].

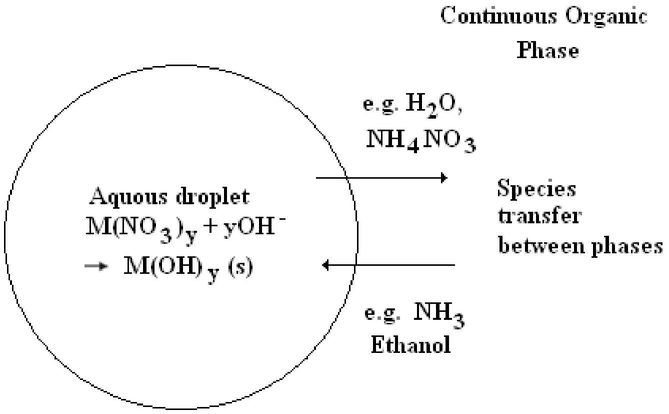


Fig. 6.23. Typical microreactor droplet for the formation of nanosized oxide particles during an electric dispersion reaction [130].

Sol–Gel (Colloidal) Processing

Sol–gel processing is a popular processing route for the synthesis of a wide variety of materials in desired shapes (particles, fibers, or films). The formation of a sol by dissolving the metal alkoxide, metal-organic, or metal-inorganic salt precursors in a suitable solvent is the primary step in a sol–gel in which the solvent molecules are trapped inside a solid (gel) [38, 39]. Subsequent drying of the gel followed by calcination and sintering leads to the final ceramic product. As the reacting species are homogenized at the atomic level in a sol–gel process, the diffusion distances are considerably reduced compared to a conventional solid-state reaction; thereby the product forms at much lower temperatures.

Depending on the nature of the precursor, which can be an aqueous solution of an inorganic salt or metal-organic compound, the species involved in the intermediate steps of the sol–gel process differ. The nature and composition of the intermediate species formed also depend on the oxidation state, the pH, or the concentration of the solution. These factors are crucial in the formation of a colloid and its stability. Metal alkoxides are noted for the very high reactivity toward hydrolysis and are therefore being used as a common starting material for the sol–gel synthesis [40]. Electronegativity, valence states, and the coordination number of the metal affect the rate of hydrolysis and hence the nature of the colloidal solution. The resultant properties of the synthesized powder depend on the colloidal aggregate formed as a result of the hydrolysis step. This has been demonstrated in the case of many ceramics, as the chemical tuning of the precursor sol and the nature of the colloid formed have affected the final product [41,42]. The preparation of powders with the desired microstructure from colloids depends on the tuning of the interparticle forces in the colloid, which differ depending on the nature of the suspension - well-dispersed, weakly flocculated, or strongly flocculated. The total interparticle potential energy can be expressed as

$$V_{total} = V_{vdW} + V_{elect} + V_{steric} + V_{structural} \tag{6.22}$$

where V_{vdW} is the attractive potential energy due to long range van der Waals interaction between particles, V_{elect} is the electrostatic repulsive potential energy between likecharged particles, V_{steric} is the potential energy arising out of the steric interactions among bulky groups present on the particle surface, and $V_{structural}$ is the potential energy due to various nonadsorbed species in solution [43]. The use of sol–gel (colloidal) method for the synthesis of advanced ceramic oxide such as silica (SiO_2) has been demonstrated [44-47]. The grain sizes in the nanometer range were obtained when samples were sintered from nanoparticles with a uniform green microstructure prepared through colloidal processing. Recently, it has been demonstrated that the grain sizes of high permittivity thin films can be suitably tuned by varying the sol–gel precursor concentration as well as the subsequent annealing temperatures of the films [48].

Silica (SiO_2) is a potential material in optical, electrical, and thermal applications. To understand the densification and sintering characteristics of silica powders, monodisperse, spherical hydrous silica (SiO_2) was synthesized through a colloidal sol–gel process and the resultant powder preparative step involved the addition of tetraethylorthosilicate ($Si(OC_2H_5)_4$, TEOS) to a reagent-grade ethanol and concentrated NH_4OH solution with constant stirring at room temperature. The TEOS undergoes hydrolysis/condensation reactions, forming polysilicic acid, which is then followed by the precipitation of silica particles. The reactant concentration controlled the average diameter of the particles that are formed during the synthesis. The precipitate was washed with distilled water, dried, and subsequently calcined in loose stack in the temperature range of 200 °C to 1050 °C. The diameters of the spherical particles ranged from 200 to 600 nm. However, by tuning the reactant concentration and the thermal processing conditions, the particles can be crystallized within the nanograin size range. The sample calcined at 200 °C was suspended in distilled water using ultrasonication and pH adjustments. The powder was then allowed to settle down in plastic tube under gravity to form a green compact. The SEM characterization of the green compact formed by the particle sedimentation showed the two-dimensional, hexagonal close-packed array of particles with an ordered array extending to a few micrometers. Figure 6.24 illustrates the hexagonal pore formation in SiO_2. The hcp layers are stacked one over the other to form a self-organized three-dimensional ordered structure. The striking feature of these investigations was the tuning of the diameter of the nano-sized pores in the compact, which was prepared by dispersion and flocculation techniques. The green compact of the dispersed particles had smaller pores compared to that of flocculated particles. The pore size distributions were narrower and bimodal in nature because of the presence of three-particle and four-particle pore channels. The study showed that by appropriate tuning of the sol preparation, colloid formation, and dispersion conditions, three-dimensional ordered SiO_2 powder compact can be formed.

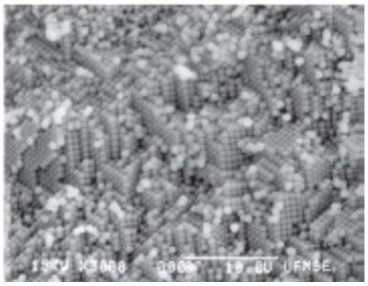


Fig. 6.24. SEM picture illustrating the formation of hexagonal and cubic ordered array in SiO_2.

6.11. Conclusions

Silicon dioxide is a high-quality insulator that can be thermally grown on silicon wafers. It can also serve as a barrier layer during impurity diffusion or implantation, and it is a key component of MOS devices and circuits. These factors have contributed significantly to silicon's current status as the dominant semiconductor material in use today.

This chapter described the mechanisms of thermal oxidation of silicon and polycrystalline silicon and presented a kinetic model of oxide growth. This model accurately predicts oxide growth rate for a wide range of process conditions. The chapter also discussed dopant redistribution and the masking properties of oxides. The CVD method of SiO_2 film deposition have been also described and analysed. Oxide characterization methods and oxide quality were discussed as well. Many important properties of silicon dioxide such as masking properties, oxide structure and quality have been considered.

The special attention has been paid on technologies of silicon oxide nanoparticles.

Chapter 7. Metallic Thin Films

7.1. Introduction

Thin metal films have been studied for decades and are widely employed in a myriad of applications. Recently, especially metallic nano- and microstructures have attracted substantial attention from the researchers due to their unique properties. Nevertherless the metals have been studied for more than a hundred years ago, the development of numerical methods, electromagnetic simulation codes, and computers have speeded up research into metallic films, micro- and nanostructures. Furthermore, the development of nanofabrication techniques has progressed hand in hand with computational techniques, providing the tools for designing, fabricating and analyzing the properties of metallic films and structures.

One of the fascinating fields of optics is plasmonics. It studies the interaction of the conduction electrons and electromagnetic field at metallic interfaces or in small metallic nanostructures. It relies on the free essence of the conduction electrons that may be induced to oscillate collectively and longitudinally. Plasmonics has led to a wide range of applications such as biosensors [1, 2], nanoscale light guiding [3, 4], and labeling of molecular objects [5-7]. The list will undoubtedly continue to expand as plasmonics is integrated into new areas.

Recent advances in preparing atomically uniform films have transformed this field of research; measurements can now be performed with precisely known film thicknesses in terms of the number of atomic layers. It is shown that the physical properties of ultra thin films can deviate substantially from the bulk counterparts, and such differences are related to the electronic structure that can be determined directly from angleresolved photoemission. Specifically, the film property variations as a function of film thickness and boundary conditions can be largely understood in terms of a “one-dimensional shell effect” akin to the periodic property variations of the elements in the periodic table. More complicated electronic effects can arise from coherent coupling of the electrons in the film to the substrate electronic states and from diffraction of the electrons by the substrate atomic structure.

7.2. Electron Theory of Metals

The response of matter to an electromagnetic field is characterized by the behavior of electrons in the applied field. In the case of metals, the electrons are essentially free, which gives rise to unique metallic features, both from the chemical and optical point of view. The behavior of electrons may be approached using classical or quantum–mechanical models, but in general the combination of the two provides the deepest understanding of the optical properties.

7.2.1. Origin of Optical Properties

Let us consider a molecule that is illuminated by an electromagnetic wave. The electric charges in the molecule are set into oscillatory motion by the electric field of the incident wave. Accelerated electric charges radiate electromagnetic energy. In the case of a system of molecules, each molecule is affected not only by the incident field but also by the resultant of the secondary fields of all the other molecules. Therefore, inside a medium the secondary waves superpose on each other and on the incident wave [8, 9]. This process is the origin of the optical properties of the medium.

The interaction of the field with electric charges leads to reduction in the propagation speed of the field because the response to the external field is not immediate. In addition, for an oblique incident angle the propagation direction changes. These properties are observed at a macroscopic level and are described by the real part of the refractive index, n, which is defined by the relationship between the speed of light in a vacuum, c_0, and the speed of light in the medium, c

$$n = \frac{c_0}{c}. \tag{7.1}$$

The wavelength of the electromagnetic field in the medium also experiences a change to $\lambda = \lambda_0/n$, where λ_0 is the wavelength of the incident field in the vacuum.

In addition to reradiating electromagnetic energy, the excited elementary charges may transform part of the incident electromagnetic energy into other forms, thermal energy, for example. This process is called absorbtion. The amplitude of the electromagnetic field decreases as it propagates through an absorbing medium. The imaginary part, k, of the complex refractive index [8],

$$n^*(\omega) = n(\omega) + ik(\omega), \tag{7.2}$$

determines the rate at which the wave is attenuated. In Eq. (7.2) i represents the imaginary unit, ω is the angular frequency of the incident field, and n is defined by Eq. (7.1). It is evident from Eq. (7.2) that the refractive index, and therefore the motion of the electrons, is dependent on the frequency of the electromagnetic wave.

An external electric field polarizes material so that the positive atom nucleus and negative electrons move in opposite directions. Susceptibility, χ, is a measure of how easily the material is polarized. Susceptibility is actually part of another optical constant, electric permittivity, defined as

$$\varepsilon^*(\omega) = \varepsilon_0[1+\chi(\omega)] + i\frac{\sigma(\omega)}{\omega}, \tag{7.3}$$

where ε_0 is the permittivity in a vacuum and σ is the electric conductivity of the material. The conductivity is non-zero only for materials with electrons that are not bound to the atom nucleus, so-called free electrons. The magnitude of the oscillatory motion of free electrons in the external field is proportional to the electric conductivity. The imaginary part of the permittivity,

$$\mathrm{Im}\{\varepsilon(\omega)\} = \mathrm{Im}\{\chi(\omega)\} + R\{\sigma(\omega)\} \tag{7.4}$$

represents the absorption of the field in the medium. $Im\{\cdot\}$ denotes the imaginary part and $R\{\cdot\}$ the real part. The imaginary part of the susceptibility is associated with the absorption of electrons that are bound to the nucleus and the real part of the conductivity is associated with absorption by the free electrons.

The relative permittivity is related to the refractive index for a non-magnetic medium by

$$\varepsilon_r^* = \frac{\varepsilon^*(\omega)}{\varepsilon_0} = \varepsilon'(\omega) + i\varepsilon''(\omega) = n^{*2}(\omega) \tag{7.5}$$

where ε_0 is the permittivity in the vacuum and $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ are both real. The relative permittivity is also called dielectric function.

The real and imaginary parts of the complex optical constants, such as refractive index, susceptibility, and permittivity, are connected by integral relations. For relative permittivity the relations may be written as [10–12]

$$\varepsilon'(\omega') - 1 = \frac{2}{\pi} P\int_0^\infty \frac{\omega\varepsilon''(\omega)}{\omega^2 - \omega'^2}\, d\omega, \tag{7.6}$$

$$\varepsilon''(\omega') - 1 = \frac{2\omega'}{\pi} P\int_0^\infty \frac{[\varepsilon'(\omega) - 1]}{\omega^2 - \omega'^2}\, d\omega + \frac{\sigma_0}{\varepsilon_0\omega'}, \tag{7.7}$$

where P denotes the Cauchy principal value of the integral, σ_0 the dc conductivity, and ω' is an artificial pole on the real axis of a complex angular frequency. The equations (7.6) and (7.7) are called Kramers–Kronig relations. They are a direct consequence of the causality principle and they relate the change in phase of an electromagnetic wave to an absorption process [13].

The quantum–mechanical approach to optical properties is based on discretesizing energy into packets called quanta. The energy level diagram of isolated atoms, illustrated in Fig. 7.1, consists of a series of states with discrete energies. A result of the periodicity of the crystal lattice of a medium is that the energy levels are grouped into bands. Optical transitions between these levels lead to absorption and emission of electromagnetic radiation. For example, the absorption of a photon leads to the transition of an electron from a low energy level to the one above it.

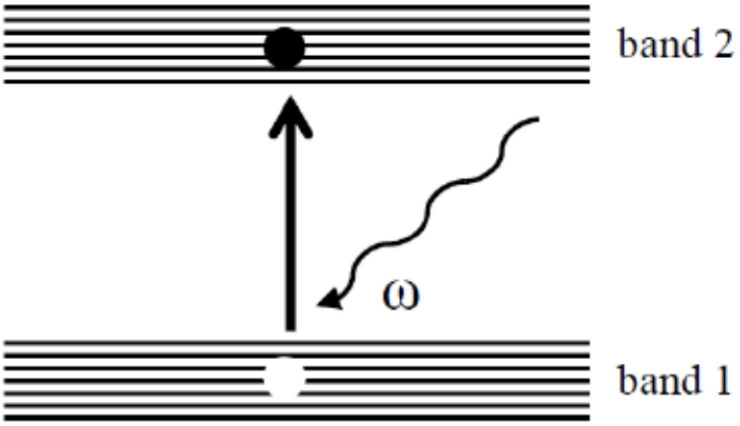


Fig. 7.1. Interband absorption of light with angular frequency ω in the energy level diagram. Bands 1 and 2 consist of several energy levels.

Energy bands are divided into three groups. The band representing the core electrons, the valence band and the conduction band. For insulators and semiconductors the conduction band is empty; for conductors it is partly filled. The size of the gap between the valence and conduction band determines whether the material is an insulator or a semiconductor. For semiconductors the band gap is less than 3 eV. For conductors the two bands overlap.

In metals the electrons in the conduction band can be excited into adjacent unoccupied states by applying an electric field, which results in an electric current. This availability of vacant electron states in the same energy band provides a mechanism, intraband absorption, for absorption of low-energy photons. Absorption in nonconductors, interband absorption, is only likely for photon energies greater than the band gap [14].

The classical model treats electrons as if they were attached to the nucleus by a spring, as shown in Fig. 7.2. From the equation of motion for a small mass attached to a large mass by a spring we obtain the dipole moment and, consequently, an expression for the relative permittivity,

$$\varepsilon_r^*(\omega) = 1 + \frac{Ne^2}{m\varepsilon_0}\frac{1}{\omega_0^2 - \omega - i\gamma\omega}, \tag{7.8}$$

where e is the magnitude of the electric charge, ω_0 the resonance frequency of the spring, N the number of oscillators per unit volume, and γ the damping factor of the springs. Equation (7.8) is called the Lorentz harmonic oscillator model. The quantum–mechanical analog for the resonance frequency ω0 is the transition frequency of an electron between two energy levels. The damping factor relates in the quantum–mechanical analog to the probability of absorption processes such as transitions to all other atomic states [15]. In general, an optical medium will have many characteristic resonance frequencies. The permittivity may then be described by taking the sum of the Lorentz oscillators with different natural resonance frequencies [14].

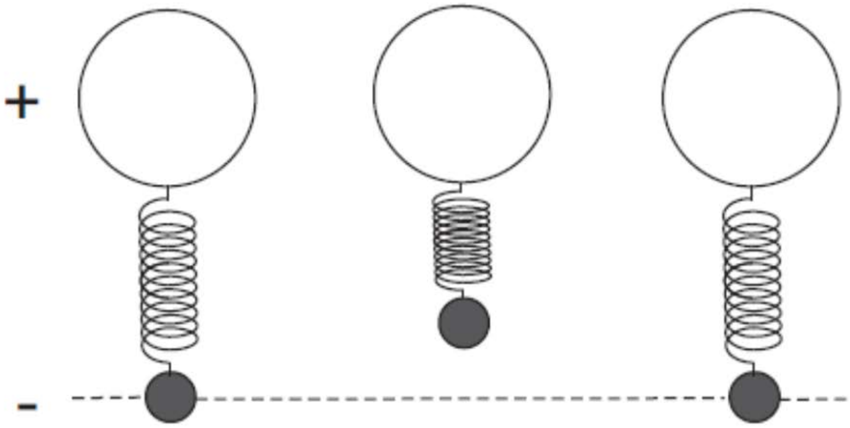


Fig. 7.2. Lorentz harmonic oscillator consisting of a heavy positive charge and a light negative mass. Electric field *E* causes displacement from the equilibrium position.

The optical response for free electrons can be obtained from the Lorentz harmonic oscillator model demonstrated in Fig. 7.2 by ignoring the springs, that is, by setting the spring constant in Eq. (7.8) to zero. Then, we obtain the expression for the relative permittivity [16],

$$\varepsilon_r^*(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - i\gamma\omega}, \tag{7.9}$$

where

$$\omega_p = \sqrt{\frac{Ne^2}{m_0\varepsilon_0}} \tag{7.10}$$

is the plasma frequency. Now, the constant γ in Eq. (7.9) represents the damping due to the scattering of electrons associated with electrical resistivity [15]. It may be written $\gamma = 1/\tau$, where τ is the mean free time between collisions. The distance between the collisions is called the mean free path. If we ignore the damping, that is $\gamma = 0$, the relative permittivity becomes real and may be written

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \tag{7.11}$$

By substituting Eq. (7.11) into Eq. (7.5) we notice that the refractive index is purely imaginary when $\omega < \omega_p$, real when $\omega > \omega_p$, and zero when $\omega = \omega_p$.

Drude theory alone does not accurately describe the optical characteristics of many metals. Metals usually exhibit some freeelectron type of behavior, which can be treated with the Drude theory, but they also have a substantial bound-electron component, due to interband transitions. For example, the colors of gold and copper are a result of interband transitions at the visible spectral range [8]. A more accurate way to describe the permittivity of metals is to combine multiple Lorentz oscillators with the Drude model.

The relative permittivity of aluminum has been calculated with Drude model and the combined Lorentz and Drude model in Fig. 7.3. The calculated values are compared to measured values taken from [17]. At wavelengths below the plasma wavelength, 82 nm, the low-damping approximation given in Eq. (7.11) is valid and the permittivity is real. The real part approaches unity as the wavelength decreases because the electrons are no longer capable of responding to the driving field. As the wavelength increases, the intraband transitions take place, metal becomes conductive, and the imaginary part grows along with the real part according to the Kramers–Kronig relations. However, there is a disturbance in the smoothly behaving permittivity around 800 nm, which is due to interband transitions. This is the reason the Drude model fails in this region.

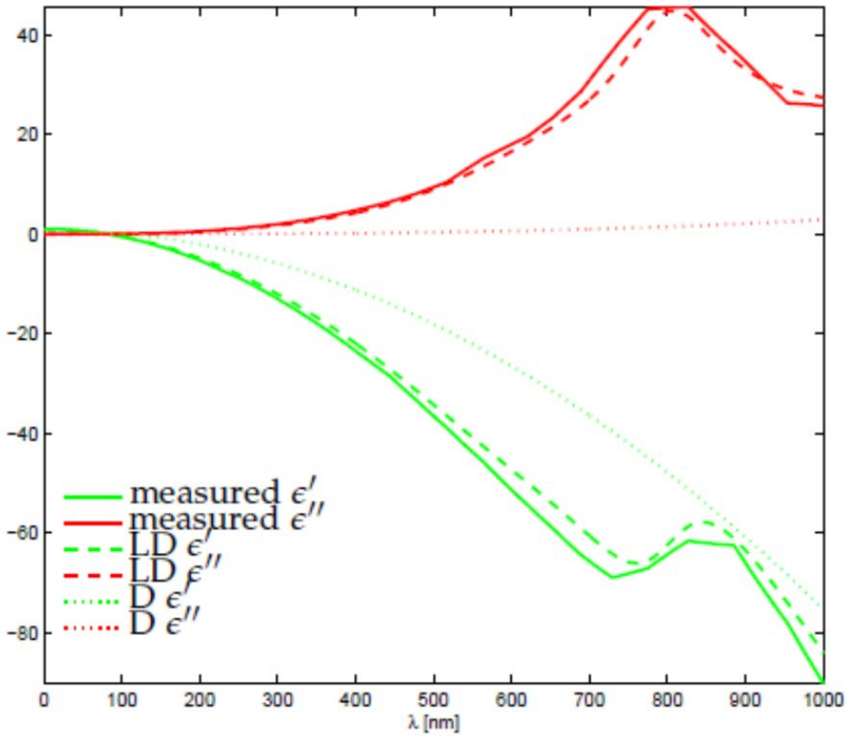


Fig. 7.3. Real part ε′ and imaginary part ε″ of relative permittivity for aluminum. LD represents the combined Lorentz–Drude model, D the Drude model. The measured values have been taken from [17].

7.2.2. Electronic Resonances

One property of metals that arises from the free electrons is that light may excite electronic resonances in them. A common feature of electronic resonance is that the oscillation amplitude of the local electric field overcomes the excitation amplitude by orders of magnitude [18]. The resonant oscillation also results in charge density oscillations that are either collective or longitudinal or both, so that the charge density has an oscillatory time dependence $exp(-i\omega t)$, where t is the time. The charge density oscillation is also referred to as plasma oscillation and its quantum is plasmon [19]. Plasmons may be divided into three groups, which are described in the following sections.

Bulk plasmons

If the damping in Drude's model Eq. (7.9) is assumed to be zero, the relative permittivity vanishes at the plasma frequency. This creates longitudinal electron oscillations called bulk plasma or volume plasma oscillation. Since the relative permittivity is zero, so is refractive index according to Eq. (7.5). The wavelength of the electromagnetic field then approaches infinity inside the medium and the electrons start to move collectively in a phase. This produces a displacement of the whole free electron gas. The fixed ion lattice will exert a restorative force to counteract this displacement of electrons, which causes the electrons to move back to the original direction and consequently again form a restorative force in the opposite direction and so on. The result is that the whole electron gas oscillates at the plasma frequency backwards and forwards with respect to the fixed ions [14, 15].

True plasma oscillation persists after the external field is removed and can be excited only by a beam of charged particles, such as electrons [15]. Therefore, in optical studies the bulk plasma oscillation refers to the forced oscillations generated by light. There exist some studies where the bulk plasma oscillations have been excited by illuminating a metallic surface at an oblique angle of incidence by light having an electric field component normal to the surface [20].

Surface plasmons

Surface plasmon resonance is a charge density oscillation existing at the interface of the dielectric and metal. The oscillation carries energy along the surface while the field in a direction normal to the surface is evanescent, as illustrated in Fig. 7.4. Because these oscillations may be excited by light, their quantum is called surface plasmon polaritons or surface plasmons.

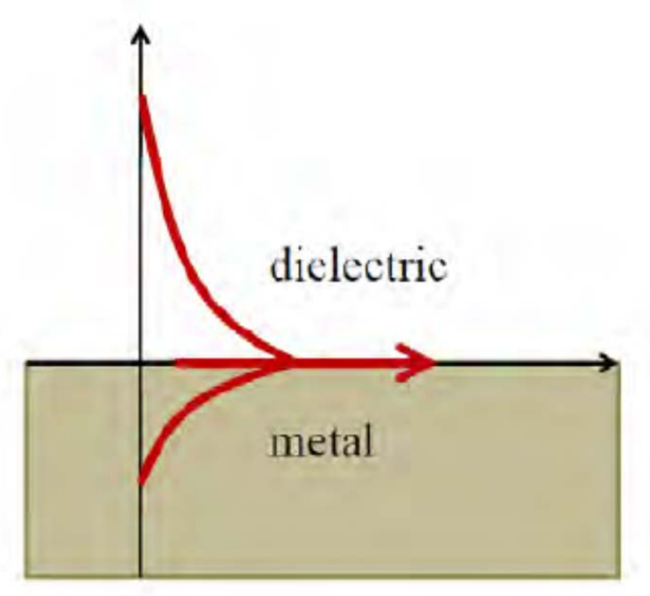


Fig. 7.4. Propagation direction of the surface plasmon polariton is indicated by the red arrow. The field in the direction normal to the surface is evanescent in the dielectric and metal, which is indicated by the red lines.

Let us define the wave number as $k = 2\pi/\lambda$. The wave number is the length of the wave vector which points in the propagating direction of the wave. The propagation constant that is the wave vector component parallel to the interface, for a surface plasmon polariton is given by [21]

$$k^{//}_{SP} = k_0 \sqrt{\frac{\varepsilon^*_2 \varepsilon_1}{\varepsilon^*_2 + \varepsilon_1}}, \qquad (7.12)$$

and the component normal to the surface [22]

$$k^{\perp}_{SP} = k_0 \sqrt{\frac{\varepsilon^2_q}{\varepsilon^*_2 + \varepsilon_1}}, \qquad (7.13)$$

where ε_1 and $\varepsilon^*_2 = \varepsilon'_2 + i\varepsilon''_2$ are the permittivities of the dielectric and metal, respectively, $k_0 = 2\pi/\lambda_0$ is the wave number in a vacuum, and ε_q with $q = 1, 2$ refer to either of the materials. Confinement to the surface demands that the field in the direction normal to the surface is evanescent. This is possible when the incident light has an electric field component normal to the interface and $\varepsilon'_2 < 0$ if $\varepsilon_1 > 0$. At optical wavelengths, this condition is fulfilled by several metals, gold and silver being the most commonly used [1].

If we approximate the permittivity with Eq. (7.11) and if $\varepsilon'_2 < -\varepsilon_1$, the wave vector component $k^{//}_{SP}$ becomes real and the surface plasmon propagates without attenuation. This is, however, approximately true for only some metals and only in the ultraviolet frequency range and frequencies below it. In general, the propagation constant is complex and the field in the propagation direction is attenuated. The real part of the propagation constant determines the surface plasmon wavelength

$$\lambda_{SP} = \frac{2\pi}{R\{k_{SP}^{//}\}}, \qquad (7.14)$$

and the imaginary part determines the propagation *1/e* decay length

$$l_{SP} = \frac{1}{2\,\mathrm{Im}\{k_{SP}^{//}\}}. \qquad (7.15)$$

The wave vector component parallel to the surface for light incident from the dielectric material is

$$k^{//} = \sqrt{\varepsilon_1}\sin\theta_1 \frac{2\pi}{\lambda_0}, \qquad (7.16)$$

where θ_1 is the angle if incidence. In order to excite surface plasmons, the real parts of the propagation constant components given by Eq. (7.12) and (7.16) should be equal. Since $\varepsilon'_2 < 0$ and $\varepsilon_1 > 0$, the real part of the propagation constant of the surface plasmon $k^{//}_{SP}$ is greater than that of the field in the dielectric medium $k^{//}$. This problem can be overcome, for example, using the famous Kretschmann configuration [23], which includes a prism placed on the metal surface. The light incident from the dielectric side to the prism experiences total internal reflection from the prism-metal interface and the generated evanescent wave couples with the surface plasmon mode.

If the surface is corrugated with a period d, the wave vector component parallel to the surface is [22]

$$k_c^{//} = k^{//} + m\frac{2\pi}{d}, \qquad (7.17)$$

where $k^{//}$ is defined by Eq. (7.16) and m is the diffraction order. The $k^{//}_c$ may be tuned to match the wave vector given by Eq. (7.12). It should be noted that the dispersion relation in Eq. (7.12) hardly changes if the metal surface has a shallow corrugation.

If we assume that the corrugated surface lies on the *xy*-plane, the wave vector for a two-dimensional corrugated surface is

$$k_c^{//} = k^{//} + m\frac{2\pi}{d_x}x^* + n\frac{2\pi}{d_y}y^*, \qquad (7.18)$$

where d_x and d_y are the grating periods in the *x*- and *y*-directions, respectively, $k^{//}$ the parallel wave vector of the incident light, and *m* and *n* the diffraction orders. It follows from the generation of diffraction orders that surface plasmons may be excited with normally incident light in contrast to a non-corrugated surface. The same applies also in the case of one-dimensional corrugated surfaces.

Localized surface plasmons

For a small particle, with a size in the range of the penetration depth of the electromagnetic field into the metal, the clear distinction between surface and bulk plasmons vanishes. Let us consider a spherical metal particle with a diameter much smaller than the wavelength of light embedded in the dielectric material. From the Mie theory [24], by taking a first-order approximation we find the polarizability, a measure for how easily individual particle is polarized, inside the particle is

$$\alpha = 4\pi a^3 \frac{\varepsilon_2^* - \varepsilon_1}{\varepsilon_2^* + 2\varepsilon_1}, \qquad (7.19)$$

where a is the radius of the particle, ε_1 the permittivity of the surrounding material and ε^*_2 the permittivity of the metal. If we assume that $\varepsilon^{//}_2$ varies slowly along the frequency, the polarizability reaches its maximum when [21]

$$\varepsilon_2^{/} = -2\varepsilon_1. \qquad (7.20)$$

The oscillation mode associated with this condition, which is in fact the lowest-order surface mode, has become known as the Frohlich mode [8], which is illustrated in Fig. 7.5. If $\gamma << \omega^2_p$, which is usually true for most metals at room temperature, it follows from Eq. (7.9) that the frequency that satisfies Eq. (7.20) is

$$\omega_F = \frac{\omega_P}{\sqrt{1+2\varepsilon_1}}, \qquad (7.21)$$

which in air $\varepsilon_1 = 1$ reduces to

$$\omega_F = \frac{\omega_P}{\sqrt{3}}. \qquad (7.22)$$

The frequency ω_F is called the Frohlich frequency.

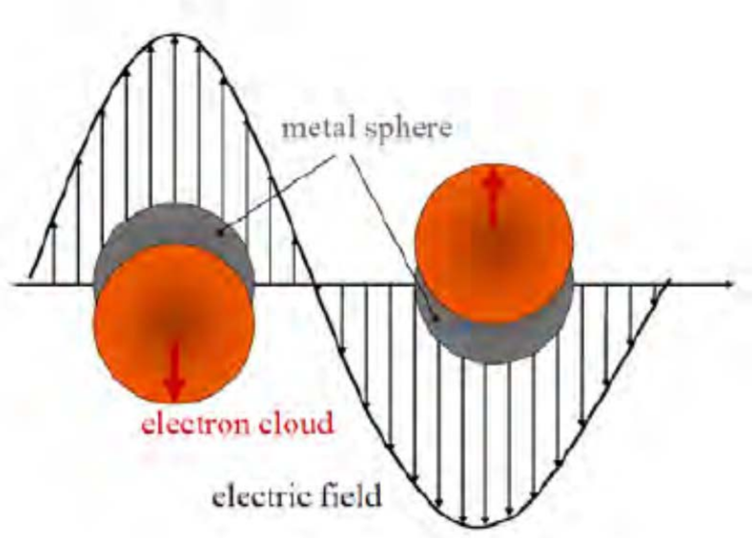


Fig. 7.5. Electron oscillations in the Fr¨ohlich mode corresponding to the dipole particle plasmon resonance.

The field inside the metal particle for which $a << \lambda$ is homogenous and therefore drives the electrons in a collective oscillative motion. When Eq. (7.20) is satisfied, the electric field is highly localized in the metallic particle. The quantum of the electron oscillation in the case of a highly localized electric field is called the localized surface plasmon or particle plasmon.

We may estimate the effect of the particle size by taking second order approximation from Mie's theory. The resonance condition for the Frohlich mode is then [8]

$$\varepsilon_2' = -(2 + \frac{12}{5}x^2)\varepsilon_1, \tag{7.23}$$

where x is the so-called size parameter defined by

$$x = \frac{2\pi\sqrt{\varepsilon_1}a}{\lambda}. \tag{7.24}$$

It may be concluded from Eq. (7.23) that the resonance frequency shifts to lower frequencies when the size of the sphere increases.

The Frohlich mode is also known as the dipole mode. For larger particles higher-order modes also appear at different frequencies. The second lowest mode is the quadrupole mode, in which half of the electron cloud inside the metal moves parallel to the applied field and half moves antiparallel [25].

For a sphere, which is a symmetrical particle, the localized plasmons may be excited independently on the polarization of the applied field. However, in the case of an infinitely long cylinder, only a field with a component normal to the surface is capable of exciting plasmons.

The resonance condition corresponding to Eq. (7.20) in an infinitely long cylinder illuminated by light polarized perpendicular to the cylinder axis is given by [26]

$$\frac{\varepsilon_2'}{\varepsilon_1} = -1. \tag{7.25}$$

It is evident by comparing the resonance condition given by Eq. (7.20) and Eq. (7.25) that the shape of the particle has a significant influence on the resonance.

Let us consider a spherical dielectric inclusion in a metallic body. This is opposite to the problem that was solved by Eq. (7.20) and we may consider it simply by switching the permittivities, $\varepsilon_2' \rightarrow \varepsilon_1$ and $\varepsilon_1 \rightarrow \varepsilon_2'$. Then we obtain the resonance condition for a spherical void, or cavity, from [21]

$$\varepsilon_2' = -\frac{1}{2}\varepsilon_1. \tag{7.26}$$

The resonance associated with Eq. (7.26) is called cavity plasmon resonance. The resonance condition for other shapes of cavities is obtained with a similar switch of permittivities.

7.2.3. Oxidation

Valence electrons are responsible for chemical reactions such as oxidation. The natural tendency of two elements in the oxidation process is to obtain an octet, a state where the bonding has a total of eight valence electrons and does not undergo any further chemical reactions. Since metals have only 1–3 valence electrons, they react easily with oxygen, except this is not the case for the most noble metals.

The bonding between oxygen and the metal is defined by the electronegativity difference of the elements. Usually the resulting bonding is mostly ionic in nature, which means that the oxygen receives electrons from the metal and it is held together by electrostatic forces. Some of the most common metal oxides are titanium oxide TiO_2, aluminum oxide Al_2O_3, chromium oxide Cr_2O_3, which is formed, for example, on stainless steel, and iron oxide Fe_2O_3, also known as rust.

7.3. Fabrication of Metallic Films and Nanostructures

Fabrication techniques for nano- and micro-optical devices have improved significantly since 1876, when David Rittenhouse made the first grating by wrapping fine wire between a pair of 0.25 mm pitch screws [27]. The feature size has decreased down to ten nanometers using electron beam lithography [28, 29] and techniques such as DNA self-assembling [30] seem to reduce it even further.

7.3.1. Thin Film Deposition

Thin film deposition is an important step in nano- and microstructure fabrication. Two widely used thin film deposition techniques are atomic layer deposition [31] and thermal evaporation [32, 33]. Due to their different physical mechanisms to grow the film, they result in different physical structures. The physical structure of the film is of great importance since it affects the optical properties and further processability of the film (see Paper I), [34].

Chemical-Vapor Deposition

CVD is attractive for metallization because it offers coatings that are conformal, has good step coverage, and can coat a large number of wafers at a time. The basic CVD setup is the same as that used for deposition of dielectrics and polysilicon (see Fig. 7.6). Low-pressure CVD (LPCVD) is capable of producing conformal step coverage over a wide range of topographical profiles, often with lower electrical resistivity than that from PVD. One of the major new applications of CVD metal deposition for integrated circuit production is in the area of refractory-metal deposition. For example, tungsten's low electrical resistivity (5.3 μΩ-cm) and its refractory nature make it a desirable metal for use in integrated circuit fabrication,

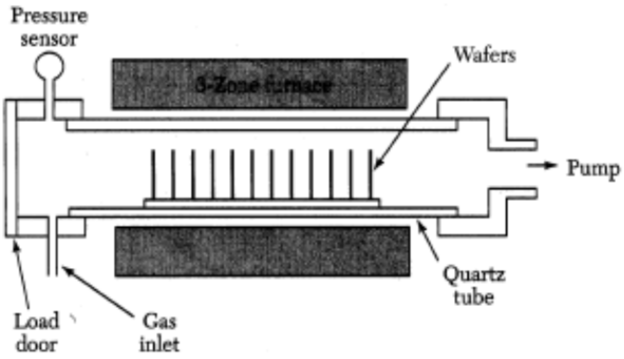


Fig. 7.6. Schematic diagrams of chemical-vapor deposition reactors. Hot-wall, reduced-pressure reactor.

CVD-W

Tungsten is used both as a contact plug and as a first-level metal. Tungsten can be deposited by using WF, as the W source gas, since it is a liquid that boils at room temperature. WF, can be reduced by silicon, hydrogen, or silane. The basic chemistry for CVD-W is as follows:

$$WF_6 + 3H_2 \rightarrow W + 6HF \text{ (hydrogen reduction),} \qquad (7.27)$$

$$2WF_6 + 3Si \rightarrow 2W + 3SiF_4 \text{ (silicon reduction),} \qquad (7.28)$$

$$2WF_6 + 3SiH_4 \rightarrow 2W + 3SiF_4 + 6H_2 \text{ (silane reduction).} \qquad (7.29)$$

On a Si contact, the selective process starts from a silicon reduction process. This process provides a nucleation layer of W grown on Si but not on SiO_2. The hydrogen reduction process can deposit W rapidly on the nucleation layer, forming the plug. The hydrogen reduction process provides excellent conformal coverage of the topography. This process, however, does not have perfect selectivity, and the HF gas by-product of the reaction is responsible for the encroachment of the oxide, as well as for the rough surface of deposited W films.

The silane reduction process gives a high deposition rate and much smaller W grain size than that obtained with the hydrogen reduction process. In addition, the problems of encroachment and rough W surface are eliminated because there is no HF by-product generation. Usually, a silane reduction process is used as the first step in blanket W deposition to serve as a nucleation layer and to reduce junction damage. After the silane reduction, hydrogen reduction is used to grow the blanket W layer.

CVD TiN

TiN is widely used as a diffusion barrier-metal layer in metallization and can be deposited by sputtering from a compound target or by CVD. The CVD TiN can provide better step coverage than PVD methods in deep submicron technology. CVD TiN can be deposited [36-38] using $TiC1_4$ with NH_3, H_2/N_2, or NH_3/H_2:

$$6TiC1_4 + 8NH_3 \rightarrow 6TiN + 24HC1 + N_2, \qquad (7.30)$$

$$2TiC1_4 + N_2 + 4H_2 \rightarrow 2TiN + 8HC1, \qquad (7.31)$$

$$2TiC1_4 + 2NH_3 + H_2 \rightarrow 2TiN + 8HC1. \qquad (7.32)$$

The deposition temperature is about 400°-700°C for NH_3 reduction and is higher than 700°C for the N_2/H_2 reaction. The higher the deposition temperature, the better the TiN film and the less C1 incorporated in TiN (-5%).

Atomic Layer Deposition

Atomic layer deposition (ALD) may be employed to grow oxide, nitride, and metal films [39]. In the process two or more precursors react chemically with each other and form one monolayer of the desired material. The precursors are led into a vacuum chamber in turns and after each precursor pulse the chamber is purged by an inert gas. These steps are repeated until the desired thickness is achieved.

The temperature, pressure of the chamber, and other process parameters are adjusted according to the process chemistry and targeted film properties. For example, with increasing temperature the reactions become more complete and the impurity content decreases, which leads to higher refractive indexes [31, 40, 41].

ALD is a very precise method for growing thin films because it allows us to adjust the film thickness by one atomic layer at a time. Depending on the process, one cycle produces a 0.1–3 °A film thickness [41]. Due to this atomic scale precise control, the microstructure of the deposited film is dense and conformal.

Physical-Vapor Deposition

The most common methods of physical-vapor deposition (PVD) of metals are evaporation, e-beam evaporation, plasma spray deposition, and sputtering. Metals and metal compounds such as Ti, Al, Cu, TiN, and TaN can be deposited by PVD. Evaporation occurs when a source material is heated above its melting point in an evacuated chamber. The evaporated atoms then travel at high velocity in straight-line trajectories. The source can be molten by resistance heating, by rf heating, or with a focused electron beam. Evaporation and e-beam evaporation were used extensively in earlier generations of integrated circuits, but they have been replaced by sputtering for ULSI circuits.

In ion-beam sputtering, a source of ions is accelerated toward the target and impinged on its surface. Figure 7.7 shows the standard sputtering system. The sputtered material deposits on a wafer that is placed facing the target. The ion current and energy can be independently adjusted. Since the target and wafer are placed in a chamber that has lower pressure, more target material and less contamination are transferred to the wafer.

One method to increase the ion densi and, hence, the sputter-deosition rate is to use a third electrode that provides more electrons for ionization. Another method is to use a magnetic field, such as electron cyclotron resonance (ECR), to capture and spiral electrons, increasing their ionizing efficiency in the vicinity of the sputtering target. This technique, referred to as magnetron sputtering, has found widespread applications for the deposition of aluminum and its alloys at a rate that can approach 1 μm/min.

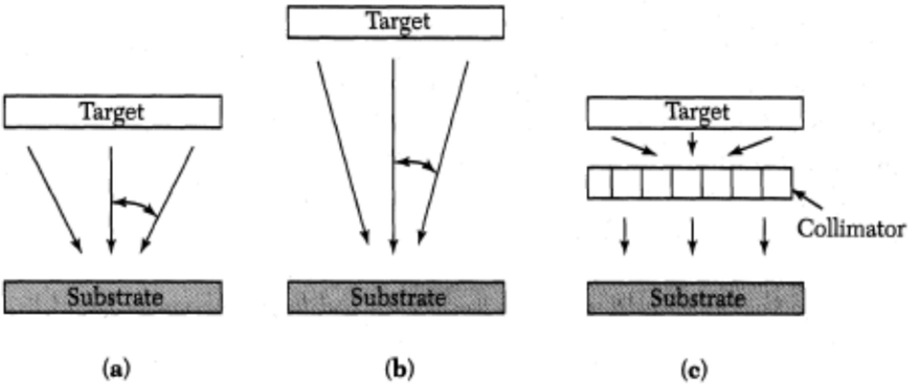


Fig. 7.7. (a) Standard sputtering, (b) long-throw sputtering, and (c) sputtering with a collimator.

Long-throw sputtering is another technique used to control the angular distribution. Figure 7.7b shows the long throw-sputtering system. In standard sputtering configurations there are two primary reasons for a wide angular distribution of incident flux at the surface: the use of a small target to substrate separation, d_{ts}, and scattering of the flux by the working gas as the flux travels from the target to the substrate. These two factors are linked because a small d_{ts} is needed to achieve good throughput, uniformity, and film properties when there is substantial gas scattering. A solution to this problem is to sputter at very low pressures, a capability that has been developed using a variety of systems, which can sustain the magnetron plasma under more rarefied conditions. These systems allow for sputtering at working pressures of less than 0.1 Pa. At these pressures, gas scattering is less important, and the target-to-substrate distance can be greatly increased. From a simple geometrical argument, this allows the angular distribution to be greatly narrowed, which permits more deposition at the bottom of high-aspect features such as contact holes.

Contact holes with large aspect ratio are difficult to fill with material, mainly because scattering events cause the top opening of the hole to seal before appreciable material has deposited on its floor. This problem can be overcome by collimating the sputtered atoms by placing an array of collimatin tubes just above the wafer to restrict the depositing flux to normal ±5°. Sputtering wit f a collimator is shown in Fig. 7.7c. Atoms whose trajectory is more than 5° from normal are deposited on the inner surface of the collimators.

Thermal Evaporation

The objective of evaporation is to transfer atoms from a heated target to a substrate. The target material is heated to the evaporation point and the thermal energy is transferred to the kinetic energy of the evaporated atoms that impinge on the substrate [42]. The target may be heated by an electron beam or resistively in a boat by passing an electric current through it.

One significant difference between ALD and thermal evaporation is that the microstructure of evaporated film changes significantly along the film thickness. The structural evolution for evaporated films starts with the adsorption of atoms on the surface and is followed by the condensation of the adsorbed atoms, adatoms. The condensation of the adatoms is also referred to as nucleation. The nuclei grow into islands until they coalesce into each other and form a continuous network of grains [43, 44].

The development of the grain structure is determined mainly by the surface mobility of the adatoms relative to the deposition rate and the substrate temperature relative to the melting temperature of the film material. As a rule, higher substrate temperatures and lower deposition rates lead to larger grain size [45]. In face-centered-cubic metals such as aluminum, silver, gold, copper and iridium, the surface mobility of the adatoms is high and the structure continues to evolve during the thickening of the film. The final grain size is typically on the order of the film thickness. Another type of material includes refractory metals and elements with a diamond-cubic crystal structure such as silicon, chromium and iron. They have a lower surface mobility and the resulting microstructure consists of vertically long pillars [46].

The resulting thin film for both material types consists of grains which are separated by grain boundaries. The space between the grain boundaries is filled with impurities such as nitrogen and oxygen that originate from the vacuum chamber. Therefore, the amount of space between the grain boundaries and the quality of the vacuum impact on the optical properties of the evaporated thin film.

Aluminum Metallization

Aluminum and its alloys are used extensively for metallization in integrated circuits, The A1 film can be deposited by a PVD or CVD method. Because aluminum and its alloys have low resistivities, 2.7 μΩ-cm for A1 and up to 3.5 μΩ-cm for its alloys, these metals satisfy the low-resistance requirements. Aluminium also adheres well to silicon dioxide. However, the use of aluminum in integrated circuits with shallow junctions often creates problems, such as spiking and eletromigration.

Junction Spiking

Figure 7.8 shows the phase diagram of the A1-Si system at 1 atm. [47]. The phase diagram relates these two components as a function of temperature. The A1-Si system exhibits eutectic characteristics; that is, the addition of either component lowers the system's melting point below that of either metal. Here, the minimim melting temperature, called eutectic temperature, is 577°C, corresponding to a 11.3% Si and 88.7% Al composition. The melting points of silicon are 660°C and 1412°C, respectively. Because of the iutectic characteristics, during aluminum deposition the temherature on the silicon substrate must be limited to less than 577°C.

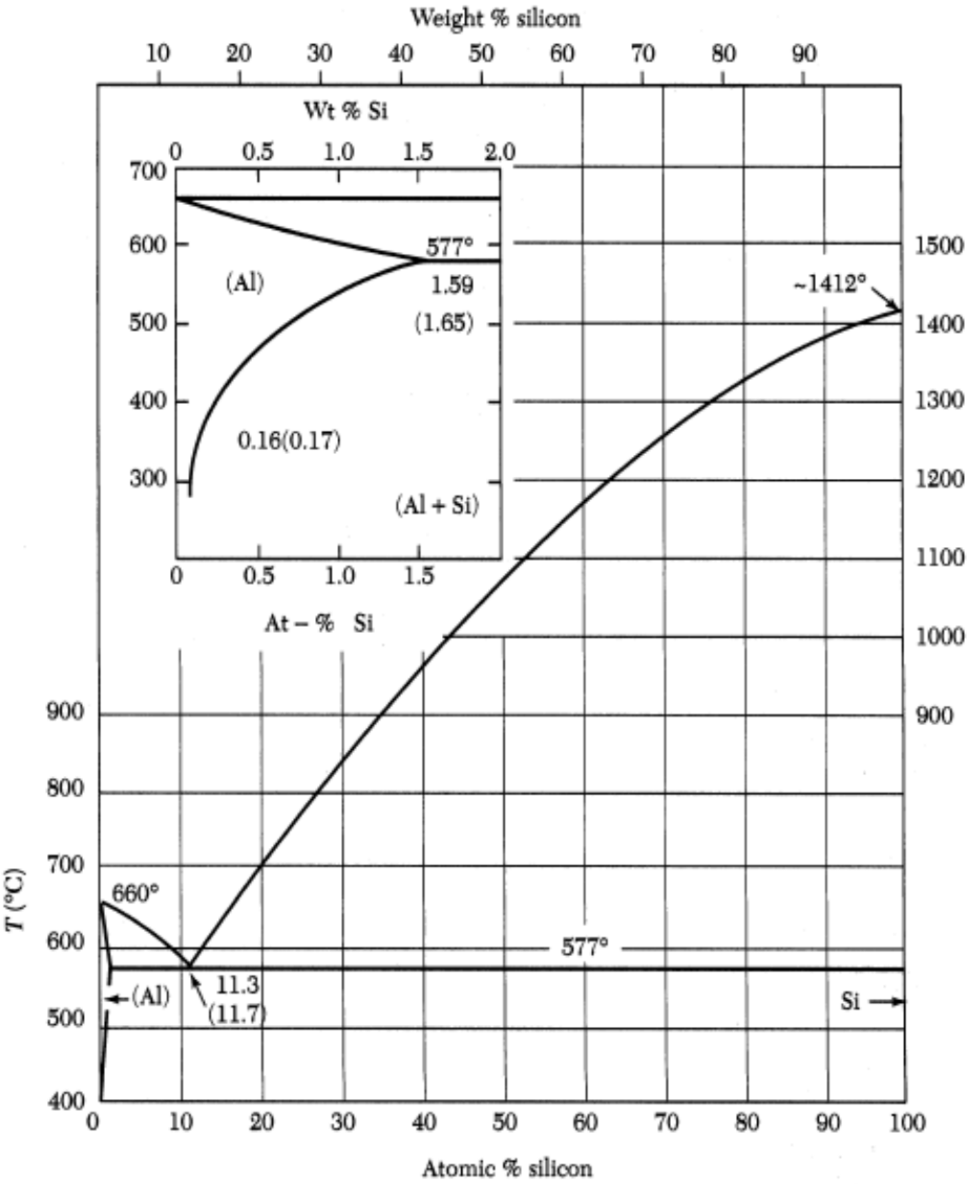


Fig. 7.8. Phase diagram of the aluminum –silicon system [47].

The inset of Fig. 7.8 also shows the solid solubility of silicon in aluminum. For example, the solubility of silicon in aluminum is 0.25 wt% at 400°C, 0.5 wt% at 450°C, and 0.8 wt% at 500°C. Therefore, wherever aluminum contacts silicon, the silicon will dissolve into the aluminum during annealing. The amount of silicon dissolved will depend not only on the solubility at the annealing temperature but also on the volume of aluminum to be saturated with silicon. Consider a long aluminum metal line in contact with an area ZL of silicon as shown in

Fig. 7.9. After an annealing time t, the silicon will diffuse a distance of approximately $\sqrt{Dt}$ along the aluminum line from the edge of the contact, where D is the diffusion coefficient given by $4\times10^{-2}\exp(-0.92/kT)$ for silicon diffusion in deposited aluminum films. Assuming that this length of aluminum is completely saturated with silicon, the volume of silicon consumed is then

$$Vol = 2\sqrt{Dt}(HZ)S(\frac{\rho_{Al}}{\rho_{Si}}), \tag{7.33}$$

where ρ_{Al} and ρ_{Si} are the densities of aluminum and silicon, respectively, and S is the solubility of silicon in aluminum at the annealing temperature [48]. If the consumption takes place uniformly over the contact area A (where $A = ZL$ for uniform dissolution), the depth to which silicon would be consumed is

$$b = 2\sqrt{Dt}(\frac{HZ}{A})S(\frac{\rho_{Al}}{\rho_{Si}}). \tag{7.34}$$

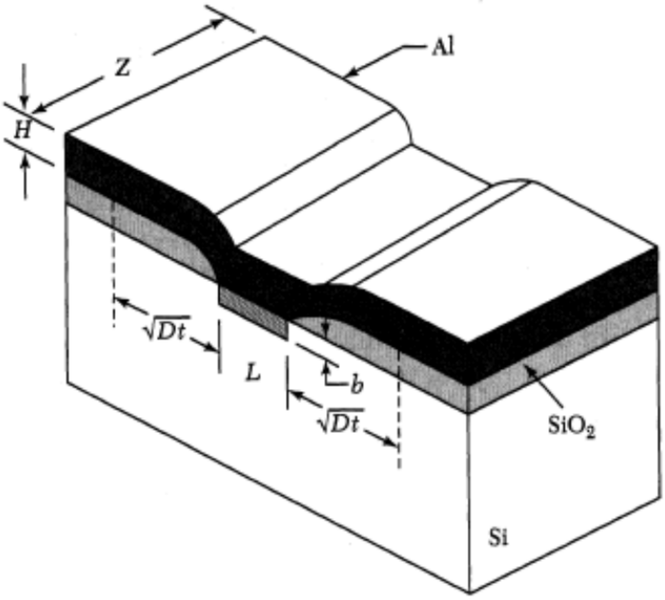


Fig. 7.9. Diffusion of silicon in aluminum metallization [48].

In a practical situation, the dissolution of silicon does not take place uniformly but rather at only a few points. The effective area in Eq. 7.34 is less than the actual contact area; hence b is much larger. Figure 7.10 illustrates the actual situation in the p - n junction area of aluminum penetrating the silicon at only the few points where spikes are formed. One way to minimize aluminum spiking is to add silicon to the aluminum by coevaporation until the amount of silicon contained by the alloy satisfies the solubility requirement. Another method is to introduce a barrier metal layer between the aluminum and the silicon substrate (Fig. 7.11). This barrier metal layer must meet the following requirements: it forms low contact resistance with silicon, it will not react with aluminum, and its deposition and formation are compatible with the overall process. Barrier metals such as titanium nitride (TiN) have been evaluated and found to be stable for contact annealing temperatures up to 550°C for 30 min.

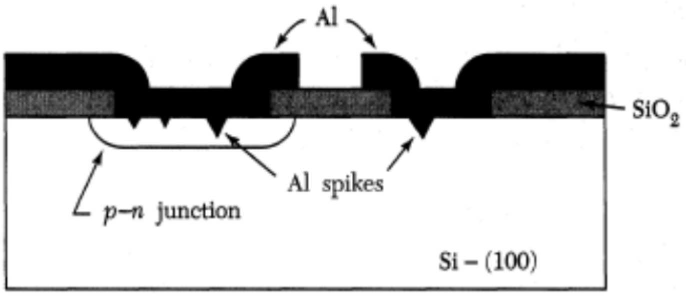


Fig. 7.10. Schematic view of aluminum films contacting silicon. Note the aluminum spiking in the silicon.

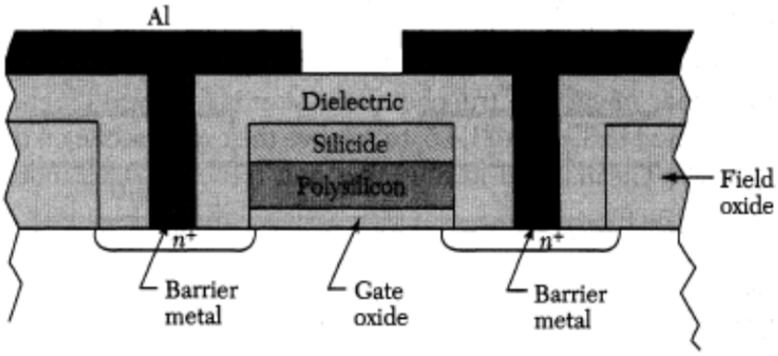


Fig. 7.11. Cross-sectional view of a MOSFET with a barrier metal between the aluminum and silicon and a composite gate electrode of silicide and polysilicon.

Electromigration

As the device becomes smaller, the corresponding current density becomes larger. High current densities can cause device failure due to electromigration. The term electromigration refers to the transport of mass (i.e., atoms) in metals under the influence of current. It occurs by the transfer of momentum from the electrons to the positive metal ions. When a high current passes through thin metal conductors in integrated circuits, metal ions in some regions will pile up and voids will form in other regions. The pileup can short-circuit adjacent conductors, whereas the voids can result in an open circuit.

The mean time to failure (MTF) of a conductor due to electromigration can be related to the current density J and the activation energy E_a by

$$MTF \approx \frac{1}{J^2}\exp\left(\frac{E_a}{kT}\right). \qquad (7.35)$$

Experimentally, a value of E_a = 0.5 eV is obtained for deposited aluminum. This indicates that low-temperature, grain-boundary diffusion is the primary vehicle of material transport, since E_a = 1.4 eV would characterize the self-diffusion of single-crystal aluminum. The electromigration resistance of aluminum conductors can be increased by using several techniques. These techniques include alloying with copper (e.g., A1 with 0.5% Cu), encapsulating the conductor in a dielectric, or incorporating oxygen during film deposition.

Copper Metallization

It is well known that both high conductivity wiring and low-dielectric-constant insulators are required to lower the RC time delay of the interconnect network. Copper is the obvious choice for a new interconnection metallization because it has higher conductivity and higher electromigration resistance than aluminum. Copper can be deposited by PVD, CVD, and electrochemical methods. However, the use of Cu as an alternative material to A1 in ULSI circuits has drawbacks, such as its tendency to corrode under standard chip manufacture conditions, its lack of a feasible dry-etching method or a stable selfpassivating oxide similar Al_2O_3 on Al, and its poor adhesion to dielectric materials, such as SiO_2 and low-*k* polymers.

Several different techniques for fabrication of multilevel Cu interconnects have been reported [49, 50]. The first method is a conventional method to pattern the metal lines followed by dielectric deposition. The second method is to pattern the dielectric layer first and fill copper metal into trenches. This step is followed by chemical mechanical polishing to remove the excess metal on the top surface of dielectric and leave Cu material in the holes and trenches. This method is also known as a damascene process.

Damascene Technology

The approach for fabricating a copper/low-*k* dielectric interconnect structure is by the "damascene" or "dual damascene." Figure 7.12 shows the dual damascene sequence for an advanced Cu interconnection structure. For a typical damascene structure, trenches for metal lines are defined and etched in the interlayer dielectric (ILD) and then followed by metal deposition of TaN/Cu . The TaN layer serves as a diffusion barrier layer and prevents copper from penetrating the low-k dielectric. The excess copper metal on the surface is removed to obtain a planar structure with metal inlays in the dielectric.

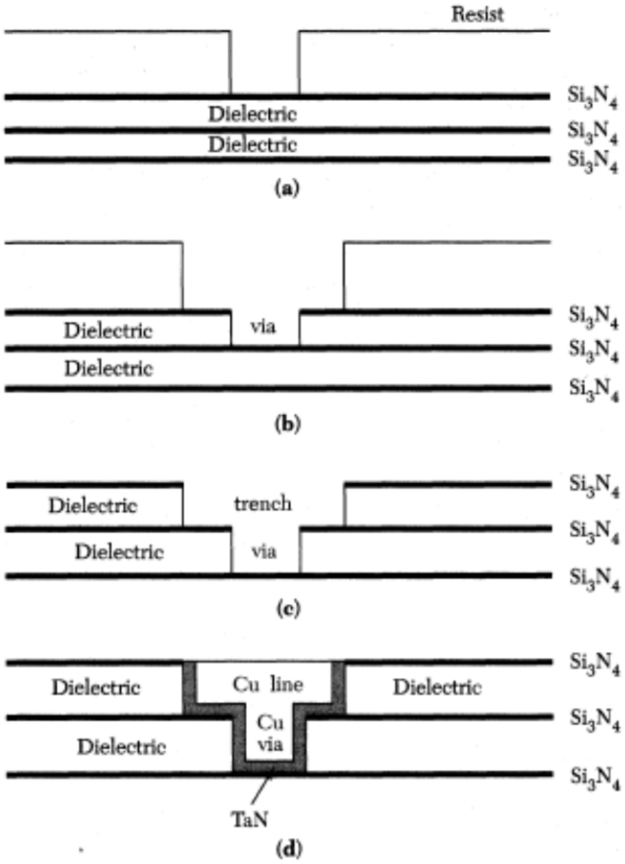


Fig. 7.12. Process sequence used to fabricate a Cu line-stud structure using dual damascene. (a) Resist stencil applied; (b) reactive ion etching dielectric and resist patterning; (c) trench and via definition; and (d) Cu depositions followed by chemical-mechanical polishing (CMP).

For the dual damascene process, the vias and trenches in the dielectric are defined using two lithography and reactive ion etching (RIE) steps before depositing the Cu metal (Fig. 7.12a-c). Then a Cu chemical-mechanical polishing process is used to remove the metal on the top surface, leaving the planarized wiring and via imbedded in the insulator [51]. One special benefit of dual damascene is that the via plug is now of the same material as the metal line and the risk of via electromigration failure is reduced.

7.4. Characterization of Metallic Nanostructures

7.4.1. Ellipsometry

An ellipsometer, according to its name, measures the ellipticity of the polarization state of light reflected from a measured surface. More accurately, the measured values ψ and δ are defined by

$$\tan\psi e^{i\delta} = \frac{r^*_{TM}}{r^*_{TE}}, \tag{7.36}$$

where r^*_{TM} and r^*_{TE} are the Fresnel reflection coefficients for the sample. If the measured sample is thin film, the Fresnel coefficients for beams that reflect a number of times from the back and top surface of the film. In the case of bulk material only one reflection needs to be considered. [52] An ellipsometer may be employed to determine the refractive index, thickness and reflectance and transmittance of the sample.

The evaluated refractive indexes must satisfy the ellipsometric equation Eq. (7.36). When dielectric material is measured over a certain wavelength range, we are usually interested in finding refractive indexes with dispersion described by the Cauchy's relation

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}, \tag{7.37}$$

where A, B, and C are the adjusted parameters. The parametric model for metals consists of the sum of the Drude and Lorentz oscillators defined by Eqs. (7.8) and (7.9). In addition, the Kramers-Kronig relation between the imaginary and real part of the refractive index may be applied to determine the solution.

If no parametric model is used, the refractive indexes are evaluated by using an initial estimate from the refraction indexes of an already known material. Then, the solution is given by the refraction indexes that are closest to the initial estimate and satisfy the ellipsometric equation Eq. (7.36).

The correspondence of ψ and δ generated by the calculated refractive indexes between the measured ψ and δ is evaluated by the root mean square error (RMSE). Refraction indexes resulting in the smallest error are the solution for Eq. (7.36). The RMSE is calculated with numerical iterative Levenberg–Marquardt algorithm [53].

The ellipsometer may also be used to measure change in the polarization state for transmitted light. If the light incident on the sample is linearly polarized, the ellipticity, χ, and polarization rotation angle, θ, for transmitted light, defined in Fig. 7.13, can be obtained from simple equations [54]

$$\tan[2(45^\circ - \theta)] = \tan 2\psi \cos\delta \tag{7.38}$$

and

$$\sin 2\chi = \sin 2\psi \cos\delta, \tag{7.39}$$

where ψ and δ are defined as in Eq. (7.36) except that the Fresnel reflection coefficients are replaced by transmission coefficients.

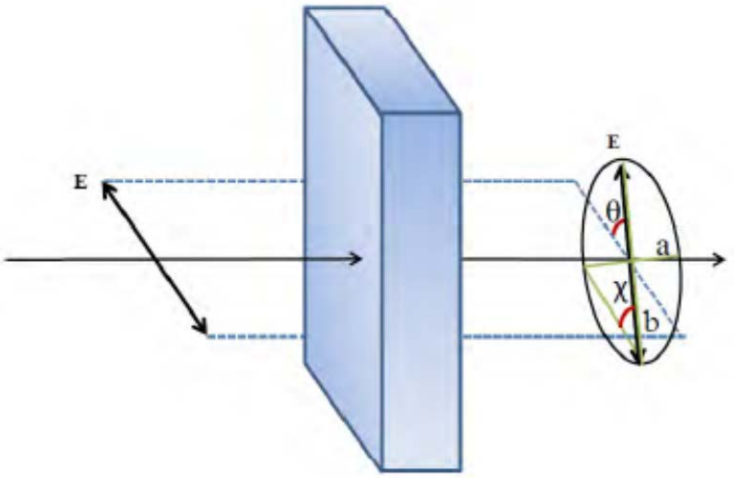


Fig. 7.13. Ellipticity, χ, defined by $\arctan \chi = a/b$, where a and b are the minor and major axes of the ellipse, and polarization rotation, θ, of the transmitted light. E is the electric field vector.

7.5. Thin Metal and Metal Oxide Films

It was revealed that the microstructure of thin films is affected by the deposition method and the process parameters. In this chapter we review how these factors contribute to the macroscopic optical properties. More precisely, we examine how the complex refractive index of a thin film is dependent on thickness and the deposition technique.

7.5.1. Changes in Refractive Index in Metal Films

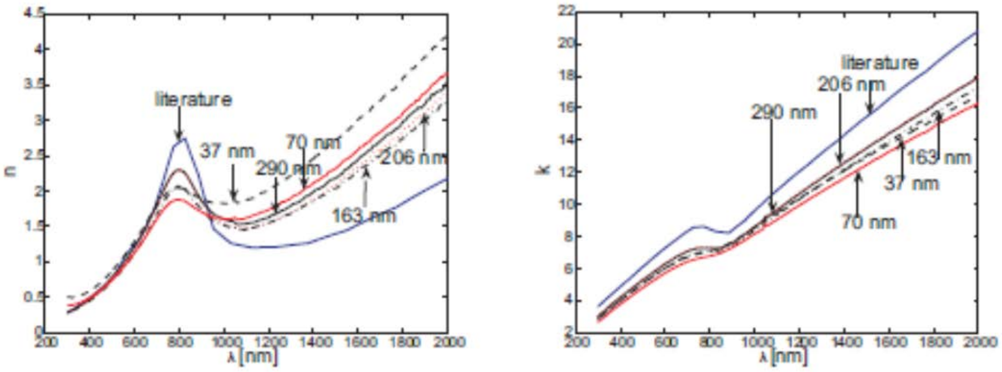
Retrieving refractive indexes from the optical constants database, such as those represented by the references [17, 55], is not a task without concern. The refractive indexes tabulated in these databases are measured for several differently fabricated and processed materials. The materials may have been fabricated by evaporation, sputtering or arc-zone melting, and some of them are bulk. The process parameters, such as temperature and pressure for evaporated or sputtered samples, are certainly not equal. Furthermore, post-processing of the samples may or may not include, for example, polishing, annealing, and rinsing [56–62].

In order to design optical devices, the refractive index used for the theoretical calculations should be known as precisely as possible. Especially the optical response of resonant elements, such as waveguide resonators and plasmon resonators, is very sensitive to changes in the refractive index. Even modification in the second decimal can shift the resonance peak out of the operational range and alter the spectral intensity.

In case of thin films, in addition to the fabrication technique and process parameters, the thickness of the film also affects the refractive index in many different ways. We have demonstrated how the complex refraction index changes between films of different thickness and different deposition techniques. The techniques that were used for the deposition of the films included atomiclayer deposition (ALD) and electron-beam gun evaporation. The deposited materials were aluminum and iridium. Iridium was deposited by both methods but aluminum was deposited only by evaporation since the ALD of aluminum is extremely challenging. The thickness of aluminum films ranged between 37–290 nm; for iridium it was between 30–144 nm.

The samples were cleaned [35] to remove contaminants that could affect the nucleation and further film growth. The process parameters, including evaporation rate, vacuum chamber pressure, and temperature, were kept constant for each material. The refractive indexes were determined by a variable angle spectroscopic ellipsometer, VASE, provided by J. A. Woollam Co. Refractive indexes were calculated using the initial estimate approach, and the combined Lorentz–Drude model given by Eq. (7.8) and Eq. (7.9). In the case of aluminum, we also added a 3-nm-thick oxide layer on top of the metal layer.

The aluminum oxide thickness used in the modeling has a significant impact on the refractive indexes we obtain for the aluminum film in the modeling. The 3 nm thickness is not necessarily the accurate value for the oxide layer. However, since the process parameters in evaporation were chosen to be such that they would minimize the surface roughness and the samples were measured at the same age, within one day, we could assume that the thickness of the oxide film was the same for all aluminum samples. As a result, the refractive indexes represented in Fig. 7.14 (a) and (b) might scale up or down but in similar manner for all samples, permitting the thickness dependence to remain.

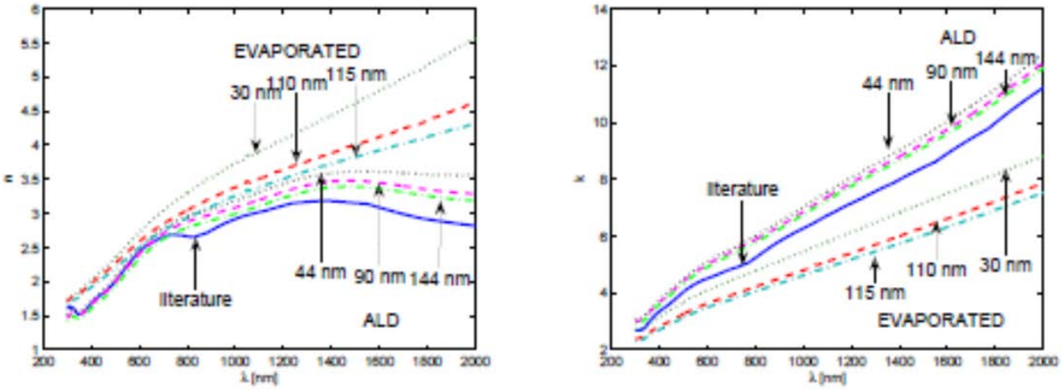


(a) (b)

Fig. 7.14. Real part (a) and imaginary part (b) of the refractive index for 37 nm, 70 nm, 163 nm, 206 nm, and 290 nm thick electron-beam-gun-evaporated aluminum films. The literature values are for ultra-high-vacuum-evaporated film [17, 62].

For iridium the analysis of thickness dependence is more straightforward since no native oxide layer is formed on it at room temperature. Figure 7.15 indicates that the real part of the refractive index approaches the literature values, taken for a crystal rod of iridium, when the thickness of the film increases. For ALD iridium the change is not as significant as for the evaporated films. This is not surprising if we consider the well-controlled microstructure that ALD produces. The same applies to the imaginary part of the refractive index, except for the fact that the imaginary part of evaporated films is closer to the literature values for thinner films, but not the thicker.

Figure 7.15 shows that the imaginary and real part of the ALD refractive index drops when the thickness increases. This could be due to an increase in the amount of impurities and defects occurring in the ALD process when more monolayers are deposited. It is a well-known fact that these kinds of defects make the refractive index more effective-like [64–66]. Moreover, when the impurities consist of air, both the real and imaginary part of the effective refractive index decreases compared to the value of the pure metal. For evaporated iridium the thickness effect is mostly dominated by cracks on the surface that increase in size for thicker films. The cracks were caused by the different thermal expansion coefficients of iridium and silicon dioxide, which was used as a substrate.



(a) (b)

Fig. 7.15. Real part (a) and imaginary part (b) of the refractive index for 44 nm, 90 nm, and 144 nm thick atomic-layer deposited iridium compared to 30 nm, 110 nm, and 115 nm thick evaporated iridium. The literature values are for a polycrystalline iridium rod [17, 63].

For evaporated aluminum the behavior of the imaginary part of the refractive index is opposite to that of ALD films. The imaginary part decreases for thinner films. Since the surface of the aluminum samples was smooth and homogenous, the cracks were not an issue. Instead, the decrease could be explained by grain structure evolution. The size of the evaporated grain for aluminum tends to be the same as the thickness of the film. Consequently, for the thinnest aluminum layers the grain size could have a limiting effect on the electron mean free path, which is about 15 nm at room temperature [19, 67]. If the mean free path is reduced, the conductivity drops and, as does the imaginary part of the refractive index. In addition, the proportion of impurities, such as oxygen, might be smaller for thicker films. The oxygen content in the vacuum chamber is consumed by the oxidation of aluminum. Therefore, as the film gets thicker, the oxygen content gets smaller and smaller. The increased proportion of air and other impurities for thinner films would be consistent with the fact that the interband transitions of aluminum at 800 nm are not as strong for thin films as they are for thicker ones. This is shown in Fig. 7.14(b), where the curve is smoother at the interband transition region for the thinner films.

Some inconsistencies between the behavior of the complex refractive indexes also exist. For example, why does the film with a thickness of 70 nm have a smaller extinction coefficient than the 37 nm thick film? A factor that might contribute to this is the fluctuating evaporation rate that ranged between 1.8–2.2 °A/s for aluminum. The evaporation rate affects grain growth and therefore the optical properties as well. Another unexplained feature is the high real refractive index of evaporated iridium films. We might also assume that the real part would be smaller than that of the ALD film due to the more effective-like optical properties, but obviously this is not the case.

7.6. Metals in Micro- and Nanoelectromechanical Systems

It can be argued that of all the material categories associated with MEMS, metals may be among the most enabling, since metallic thin films are used in many different capacities ranging from etch masks used in device fabrication to interconnects and structural elements in microsensors and microactuators. Metallic thin films can be deposited using a wide range of techniques, including evaporation, sputtering, CVD, and electroplating. Since a complete review of the metals used in MEMS is far beyond the scope of this chapter, the examples presented in this section were selected to represent a broad cross section of where metals have found uses in MEMS.

Aluminum (Al) and gold (Au) are among the most widely employed metals in microfabricated electronic and electromechanical devices, as a result of their use as innerconnect and packaging materials. In addition to these critical electrical functions, Al and Au are also desirable as electromechanical materials. One such example is the use of Au micromechanical switches for RF MEMS. For conventional RF applications, chip level switching is currently performed using FET- and PIN diode-based solid state devices fabricated from gallium arsenide (GaAs) substrates. Unfortunately, these devices suffer from insertion losses and poor electrical isolation. In an effort to develop replacements for GaAs-based solid state switches, the development of an electrostatically actuated, cantilever-based micromechanical switch fabricated on GaAs substrates have been reported [68]. The device consisted of a silicon nitride-encased Au cantilever constructed on a sacrificial silicon dioxide layer. The silicon nitride and silicon dioxide layers were deposited by PECVD, and the Au beam was electroplated from a sodium sulfite solution inside a photoresist mold. A thin multilayer of Ti and Au was sputter deposited in the mold prior to electroplating. The trilayer cantilever structure was chosen to minimize the deleterious effects of thermal and process-related stress gradients in order to produce unbent and thermally stable beams. After deposition and pattering, the cantilevers were released in HF. The processing steps proved to be completely compatible with GaAs substrates. The released cantilevers demonstrated switching speeds better than 50μs at 25V with contact lifetimes exceeding 10^9 cycles.

In a second example from RF MEMS, the fabrication of an Al-based micromachined switch as an alternative to GaAs FETs and PIN diodes has been reported [69]. In contrast to the work [68], this switch utilizes the differences in the residual stresses in Al and Cr thin films to create bent cantilever switches that capitalize on the stress differences in the materials. Each switch is comprised of a series of linked bimorph cantilevers designed in such a way that the resulting structure bends significantly out of the plane of the wafer due to the stress differences in the bimorph. The switch is drawn closed by electrostatic attraction. The bimorph consists of metals that can easily be processed with GaAs wafers, thus making integration with GaAs devices possible. The released switches were relatively slow, at 10 ms, but an actuation voltage of only 26 V was needed to close the switch.

Thin-film metallic alloys that exhibit the shapememory effect are of particular interest to the MEMS community for their potential in microactuators. The shape-memory effect relies on the reversible transformation from a ductile martensite phase to a stiff austenite phase in the material with the application of heat. The reversible phase change allows the shape-memory effect to be used as an actuation mechanism, since the material changes shape during the transition. It has been found that high forces and strains can be generated from shape-memory thin films at reasonable power inputs, thus enabling shape-memory actuation to be used in MEMS-based microfluidic devices, such as microvalves and micropumps. Titanium-nickel (TiNi) is among the most popular of the shape-memory alloys, owing to its high actuation work density (50 MJ/m^3) and large bandwidth (up to 0.1 kHz) [70]. TiNi is also attractive because conventional sputtering techniques can be employed to deposit thin films, as detailed in a recent report [70]. In this study, TiNi films were deposited by co-sputtering elemental Ti and Ni targets, and a co-sputtering TiNi alloy and elemental Ti targets. It was reported that co-sputtering from TiNi and Ti targets produced better films due to process variations related to the roughening of the Ni target in the case of Ti and Ni co-sputtering. The TiNi/Ti co-sputtering process has been used to produce shape-memory material for a silicon spring-based microvalve [71].

Use of thin-film metal alloys in magnetic actuator systems is another example of the versatility of metallic materials inMEMS.Magnetic actuation in microdevices generally requires the magnetic layers to be relatively thick (tens to hundreds of microns) to generate magnetic fields of sufficient strength to generate the desired actuation. To this end, magnetic materials are often deposited by thick film methods, such as electroplating. The thickness of these layers exceeds what can feasibly be patterned by etching, so plating is often performed in microfabricated molds made from materials such as polymethylmethacrylate (PMMA). The PMMA mold thickness can exceed several hundred microns, so X-rays are used as the exposure source during the patterning steps. When necessary, a metallic thin film seed layer is deposited prior to plating. After plating, the mold is dissolved, which frees the metallic component. Known as LIGA (short for Lithography, Galvanoforming and Abformung), this process has been used to produce a wide variety of high-aspect ratio structures from plateable materials, such as nickel-iron (NiFe) magnetic alloys [72] and Ni [73].

In addition to elemental metals and simple compound alloys, more complex metallic alloys commonly used in commerical macroscopic applications are finding their way into MEMS applications. One such example is an alloy of titanium known as Ti-6Al-4V. Composed of 88% titanium, 6% aluminum and 4% vanadium, this alloy is widely used in commercial aviation due to its weight, strength, and temperature tolerance. The use of this alloy in the manufacture of MEMS based winged structures for micro aerial vehicles have been explored [74]. The authors considered this alloy not only because of its weight and strength, but also because of its ductility and its etching rate at room temperature. The designs for the wing prototype were modelled after the wings of bats and various flying insects. For this application, Ti-alloy structures patterned from bulk (250 μm-thick) material by an $HF/HNO_3/H_2O$ etching solution were used, rather than thin films. Parylene-C was deposited on the patterned alloy to serve as the wing membrane. The miniature micromachined wings were integrated into a test setup, and several prototypes actually demonstrated short duration flight.

7.7. Quantum Properties of Thin Metal Films

Surfaces, thin films, and self-organized surface structures can exhibit interesting and useful properties markedly different from their bulk counterparts due to symmetry reduction, geometric confinement of electrons, and boundary effects. The underlying quantum physics is a cornerstone for nanoscale science [75] which is a broadly based interdisciplinary enterprise highly relevant to the advancement of materials, devices, and technologies. Thin films, in particular, are of fundamental interest; the simple geometry facilitates a detailed exploration of the connection between atomistic details and macroscopic physical and chemical properties. Thin films also provide a research path bridging surfaces and bulk materials. Through systematic studies of films with varying thicknesses, effects pertaining to the surface, the interface, and the bulk can be identified and characterized in detail. A thorough understanding of the basic scientific issues in such systems enables advanced materials concepts through artificial layering, interfacial engineering, layer alloying, and doping. Furthermore, preparation and processing of thin films by deposition and annealing can lead to novel self-assembled and self-organized structures that function as quantum dots, wires, and stripes. A detailed investigation of thin film effects and phenomena is essential for establishing a basis for understanding the evolution, kinetics, energetics, and properties of these nanoscale systems. The vast opportunities for scientific and technological advances have fostered a strong interest in the basic physics of thin metal films.

A key issue in thin film physics is quantum confinement of the electrons and the resulting quantization of the electronic wave vector along the direction perpendicular to the film surface [76]. The continuum states characteristic of the bulk are thus replaced by a discrete set of quantum well states or subbands with their energies dependent on the boundary conditions. The resulting modifications to the electronic structure affect the physical properties, leading to atomic-layer-by-atomic-layer variations that can be quite dramatic [77]. Experimental results on property variations reported in the literature include studies of surface energy, thermal stability, work function, surface adsorption, electron-phonon coupling, superconducting transition temperature, etc. The subband electronic structure in thin films can be mapped directly by angleresolved photoemission spectroscopy. Recent advances in thin film preparation have made it possible to create atomically uniform films, thus facilitating highly precise measurements of the electronic energies as a function of system configuration including the film thickness and interfacial structure. To a first approximation, the electronic structure of a thin film can be described in terms of the standard model – a particle confined in a quantum box – as commonly described in quantum mechanics textbooks. As the film thickness increases, the number of occupied subbands (below the Fermi level) increases. Each time a subband crosses the Fermi level, the system properties must change in response to the new electronic configuration. Such Fermi level crossing happens periodically as a function of the film thickness. The result is a damped oscillatory modulation of properties as a function of film thickness, akin to the periodic property variations of the elements in the periodic table. This “one-dimensional shell effect” has been observed in a number of cases.

While the particle-in-a-box model works well in most cases, there are situations where this model does not adequately describe the experimental results for a number of reasons. For example, the detailed atomic structure of the substrate can be important, and the interfacial atomic potential can diffract or scatter the electrons in the film. Coherent coupling of the electronic states in the film to the substrate states can also be important, leading to partial confinement or hybridization. This area of research is progressing rapidly.

7.7.1. Spectroscopy of the Electronic Structure of Films

Quantum well states in films appear as discrete peaks in angle-resolved photoemission spectra for electron emission along the surface normal, the direction of confinement [76, 78-84]. Shown in Fig. 7.16 is a schematic diagram for the normal-emission geometry. Experimental results taken from Ag films grown on Fe(100) are shown in Fig. 7.17 [85, 86]. The discrete peaks in the spectra correspond to quantum well states, or standing waves of electrons formed by coherent multiple reflections between the two boundaries of the film. The bottom spectrum is for a film with a thickness of 38 monolayers (ML). Adding 0.5 ML to this film yields the middle spectrum, which exhibits two sets of quantum well peaks. One set is at the same positions as the 38 ML case, and the other corresponds to a thickness of 39 ML. Adding another 0.5 ML to this film for a total coverage of 39 ML (top spectrum spectrum), the peaks corresponding to 38 ML are completely suppressed, and only the 39 ML peaks remain. This discrete layer behavior, observed over a wide range of thicknesses to over 100 ML, establishes that the film is uniform on an atomic scale within the region probed by photoemission (~1 mm).

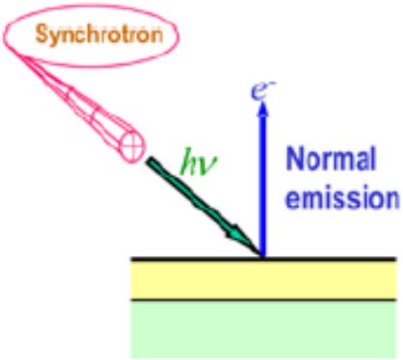


Fig. 7.16. Schematic diagram for photoemission from a film along the surface normal direction.

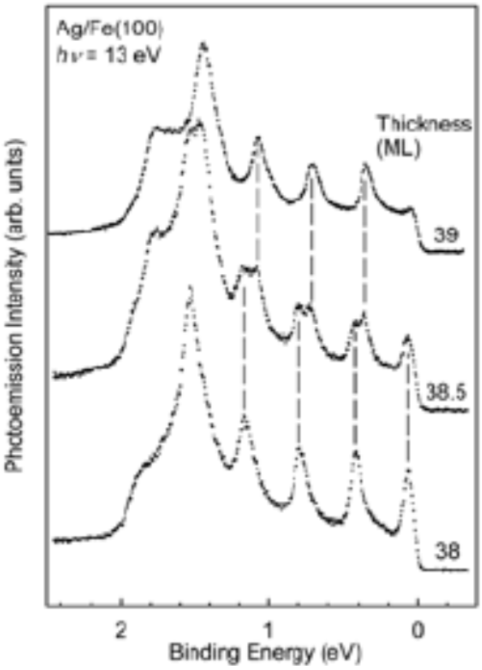


Fig. 7.17. Normal-emission spectra taken from 38, 38.5, and 39 ML of Ag on Fe(100). The 38.5 ML spectrum shows two sets of peaks, one set at the 38 ML positions, and the other at the 39 ML positions.

Such atomically uniform films had been thought to be impossible. Film growth by deposition is inherently a random process, which tends to yield roughness that usually scales with the film thickness. Creating atomically uniform films requires tailoring the growth process to follow a pathway that favors the formation of a uniform thickness. Determining the absolute thickness of a thin film can be problematic, but with uniform films, quantum-well spectroscopy provides a solution. Fig. 7.18 presents normal-emission spectra for Ag coverages of N = 1-15 ML [87]. The peaks move discontinuously, and the spectrum for thickness N shows no emission from peaks associated with thicknesses of $N\pm1$. This discrete atomic layer resolution allows absolute calibration of film thickness by atomic layer counting.

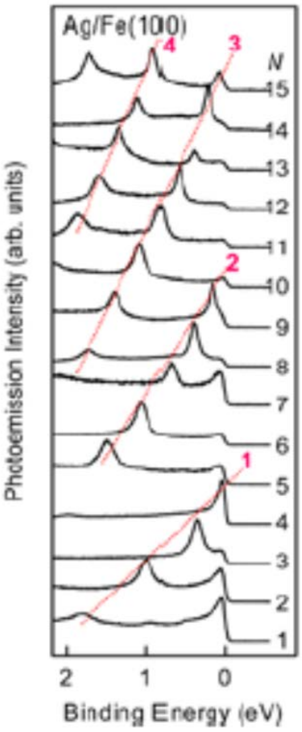


Fig. 7.18. Normal-emission spectra of Ag films on Fe(100) with thicknesses N = 1-15. The quantum numbers (1-4) are indicated.

The energies of the quantum well states are determined by the Bohr-Sommerfeld quantization rule:

$$2kNt + \Phi = 2n\pi, \tag{7.40}$$

where k is the magnitude of the wave vector of the Bloch electron along the surface normal direction, t is the monolayer thickness, Φ is the sum of the phase shifts at the two film boundaries, and n is a quantum number. The quantum numbers n = 1-4 are labeled for the quantum well peaks in Fig. 7.18. Both the wave vector and the phase shift depend on energy. With quantum well peak positions determined for a number of diffeent film thicknesses, one could solve the above equation to yeild the bulk band structure $E(k)$. This is a very accurate method for band structure determination [76, 87].

Fig. 7.19 presents another case in which atomically uniform films have been successfully prepared [88, 89]. The system is Pb films grown on Si(111). Unlike the case of Ag on Fe(100) in which there is a very good lattice match, Pb and Si have very different lattice constants. Nevertheless, the films are atomically uniform as verified by the discrete peak evolution in the normal-emission phtoelectron spectra. The in-plane crystallographic directions of the Pb film are parallel to those of the Si substrate, but the lattices are otherwise incommensurate. Pb is a free-electron-like metal. Its electronic structure can be viewed, to first order, as a jellium (or a Fermi sea of electrons). It costs very little energy to move the electrons around. The dangling bonds on the Si(111) surface can be easily terminated by the mobile electrons in the Pb jellium, and so epitaxial constraint is unimportant for this type of interfacial electronic bonding. As a result, the Pb film simply adopts its own natural lattice constant to minimize the strain energy in the film [90].

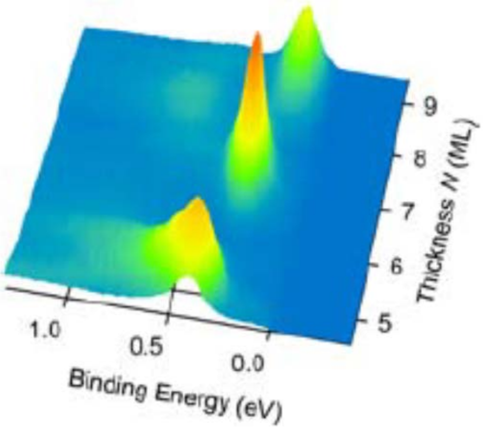


Fig. 7.19. Photoelectron intensity at normal emission as a function of film thickness and binding energy for Pb films prepared on a Pb terminated Si(111) surface.

7.7.2. Surface Properties of Thin Films – Work Function and Chemisorption

Per density functional theory, the ground state of a system is a unique functional of the electron density. Each time a subband edge crosses the Fermi level as the film thickness varies, the electron density function changes, and the physical properties should change correspondingly. Such changes generally follow a damped oscillatory behavior as a function of N with a functional form resembling Friedel oscillations. The system should become bulklike in the limit of a very large film thickness. The oscillation period can be found from Eq. (7.40). Taking the difference between two consecutive crossings ($\Delta n = 1$) at the Fermi level yields

$$\Delta N t = \frac{\pi}{k_F} = \frac{\lambda_F}{2}. \qquad (7.41)$$

Thus, the oscillation period is just one half of the Fermi wavelength. This is the same oscillation period for the giant magnetoresistance (GMR) effect in certain multilayer systems [91-93]. For Ag(100), the period is 5.8 ML. This is a dominant contribution to the variations in physical properties, but there can be others.

A measurement of the work function of Ag/Fe(100) for *N* = 0-15 reveals such quantum oscillations, as shown in Fig. 7.20 [94]. The upper panel of the figure shows the energy positions of the quantum well states deduced from a fit to the experimental data. As photoemission measures only occupied states, the points above the Fermi level are deduced from this model fit. The fit is excellent for the states below the Fermi level. Crossings of the Fermi level for the different subbands (*n* = 1-4) are marked by arrows. Based on fairly general arguments, the work function should exhibit a dip (or cusp) at each crossing [95, 96]. This prediction corresponds closely to the experimentally observed work function variations. Also shown are results from first-principles calculations of the work function. A complication is that the Ag and Fe lattices are slightly mismatched. The figure shows two calculations. The one labeled “unstrained” was for a system in which the Fe substrate was slightly strained to conform to an unstrained Ag film. The results are in fairly good agreement with the experiment; specifically, the dips at the first two crossings are well reproduced. One possible source for the discrepancy at *N* = 2 and below is strain effects. Films this thin are likely strained to conform to the substrate lattice, while thicker films are strain relieved. The theoretical results labeled “strained” were for a system in which the Ag film was slightly strained to conform to an unstrained Fe substrate. Ignoring an overall shift, the point at *N* = 2 becomes much lower than the point at N = 3, with a difference very close to the experiment. Thus the good agreement between theory and experiment is extended down to *N* = 2.

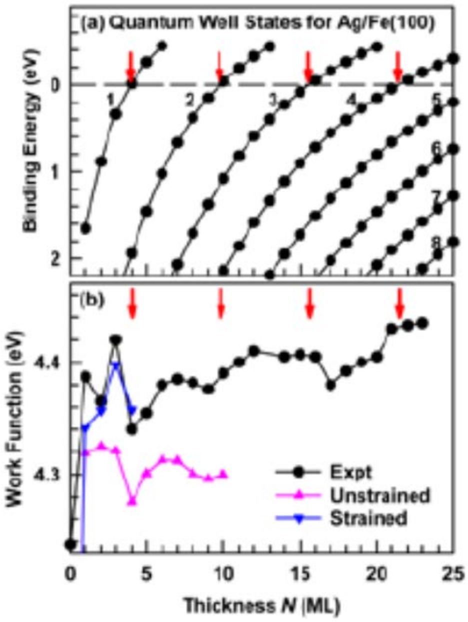


Fig. 7.20. (a) Energies of quantum well states at normal emission as a function of N for Ag on Fe, deduced from a fit to the experimental quantum well peak energies below the Fermi level. The arrows indicate Fermi level crossings of subbands. (b) Measured and computed work functions as a function of N.

The work function is just one of the many surface properties that are closely coupled to the electronic structure. By implication, chemisorption and catalytic properties of thin film surfaces can also exhibit quantum oscillations. These effects have been reported in the literature [97]. Nanoscale engineering of catalytic materials has been a topic of intense interest. Uniform thin films are not necessarily fit for catalytic applications, but they do provide a basis for understanding the general phenomenon of property modulation by varying the system dimensions.

7.7.3. Thermal Stability and Morphological Evolution

Thermal stability is an important practical issue for thin film applications. Since atomic motion associated with thermal instability is generally thermally activated with an exponential dependence on the energy difference between different configurations, a slight difference in electronic energy can have a large effect on the stability. For Ag films on Fe(100), the measured maximum stability temperatures for different thicknesses are presented in Fig. 7.21(a) [98]. In the experiment, each Ag/Fe film was ramped up in temperature until its morphology changed. Films with thick nesses of N = 1, 2 and 5 ML were stable to temperatures over 800 K, while other films for N up to 15 began to bifurcate at $T \sim 400$ K into adjacent-integer-monolayer thicknesses $N \pm 1$. Of special interest is the case of N = 5. This thickness was so stable that the film survived the highest annealing temperature available during the experiment, at which the sample was observed to glow in the chamber. Changing the thickness by just one monolayer to N = 4 or 6 made the film unstable at about room temperature. The effect is very dramatic.

Each quantum well state of Ag/Fe as seen in Fig. 7.18 corresponds to a subband which disperses as a function of the inplane momentum $k_{||}$. From photoemission results, we can compute the total electronic energy $A(N)$ of the system by summing over the occupied states. Cutoff at the Fermi level of subband occupancy gives rise to monolayer-by-monolayer variations in the total electronic energy, thus affecting the thermal stability. The quantity relevant to stability against $N \rightarrow N \pm 1$ bifurcation is the energy difference

$$\Delta(N) = \frac{1}{2}[A(N+1) + A(N-1)] - A(N). \qquad (7.42)$$

This is proportional to the discrete second derivative of $A(N)$. A large positive $\Delta(N)$ corresponds to a stable film thickness. Fig. 7.21(b) shows the results of a calculation based on photoemission measurements (not applicable for $N = 1$). Indeed, $N = 2$ and 5 should be particularly stable, in agreement with the experiment. A firstprinciples total-energy calculation has confirmed the finding [99].

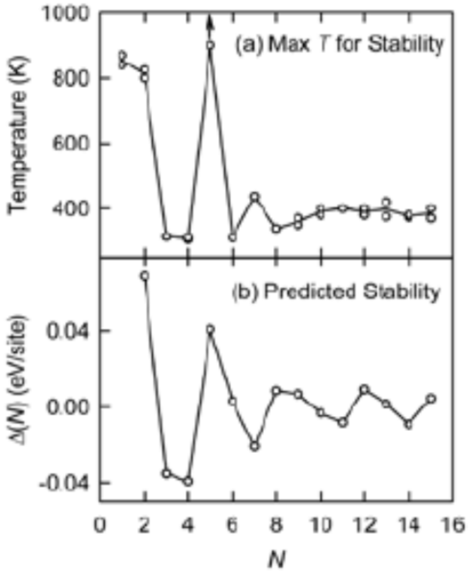


Fig. 7.21. (a) Temperature T at which a Ag film on Fe with an initial thickness of N becomes unstable. (b) Calculated energy difference $\Delta(N)$ against bifurcation.

The free energy as a function of film thickness can be readily surveyed by measuring the roughness of a film that has been annealed to high temperatures, as demonstrated in a study of Pb films deposited on Si(111) [100, 101]. From Eq. (7.41) and taking into account the discrete atomic ayer structure of films, Pb films should have a quantum oscillation period of 2.2 ML, which implies a nearly bilayer modulation of properties. Thus, films with even N are expected to be markedly different from films with odd N. However, over a sufficiently wide range of N, the phase of the even-odd oscillations can reverse because the period is not exactly 2 ML. The result is a beating pattern with a period of 9 ML superimposed on the bilayer oscillations.

Fig. 7.22 shows the annealing behavior of a Pb film on Si(111) with an initial coverage of 11 ML. Plotted is the film thickness distribution p_N expressed in terms of a percentage of the surface coverage deduced from synchrotron x-ray reflectivity data. At the base deposition temperature of 110 K, the initial film thickness distribution is narrow, but the film is not atomically uniform. It turns out that 11 ML is an unstable thickness (compared to the neighboring thicknesses 10 and 12 ML); so it is difficult to prepare an atomically uniform film at this particular thickness. As the sample is progressively annealed to higher temperatures as indicated in the figure, the film thickness distribution goes through several stages. The first stage is "bifurcation," and the result is a film largely made of the more stable thicknesses 10 and 12 ML. The next stage is the "uniform-height-island" stage with the surface dominated by islands 12-ML high separated by a wetting layer. This comes about because of phase separation of the system into a state corresponding to a local minimum in the surface energy at thickness N = 12 and a state corresponding to the global minimum at N = 1 (wetting layer). The thickness 10 ML is actually more stable than 12 ML. However, the system is prevented from forming uniform 10 ML islands because the deep minimum in the surface energy at N = 1 favors the formation of taller islands to increase the area covered by the wetting layer. Similar uniform-height islands have been observed by STM and x-ray diffraction for growth at intermediate temperatures at which the surface mobility is sufficiently high for self organization [102, 103].

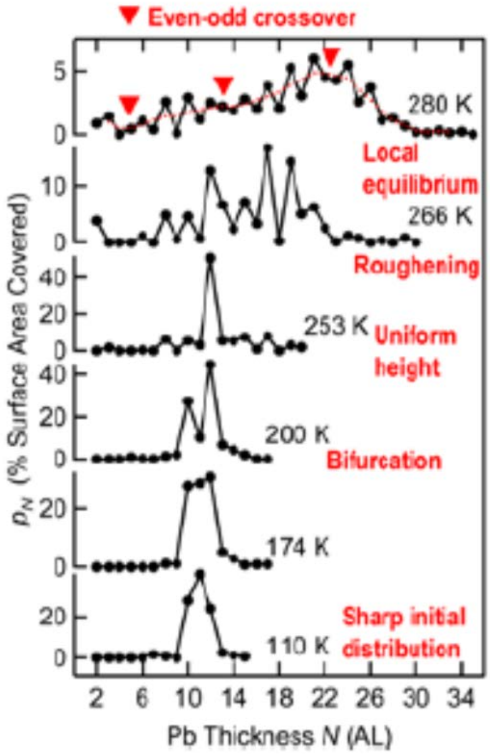


Fig. 7.22. Island height distribution for a Pb film with an initial coverage of 11 ML after annealing to various temperatures. The final distribution shows bilayer oscillations with even-odd crossovers at 9 layer intervals.

Annealing to higher temperatures results in a broadening of the thickness distribution. At 280 K, the residual preference for 12 ML from the uniform height phase has disappeared, and the film has reached local equilibrium. The resulting broad distribution is traditionally described in terms of a roughness. However, an inspection of the data reveals a structure within this thickness distribution. Superimposed on the dotted "background roughness" related to entropy effects are bilayer oscillations with even-odd crossovers occur by the triangles. The relative population differences can be related to a Boltzmann factor. From the measurements, the surface energy for different film thicknesses can be extracted. The results are well described by a functional form based on a free electron model:

$$E_S(N) = B\frac{\sin(2k_F Nt + \phi)}{N^\alpha} + C, \qquad (7.43)$$

where k_F is the Fermi wave vector, t is the monolayer thickness, B is an amplitude parameter, ϕ is a phase shift that depends on the interface properties, α is a decay exponent, and C is a constant offset. A plot of this function with the parameters chosen for a best fit to the Pb/Si data is shown in Fig. 7.23 (the constant offset is ignored).

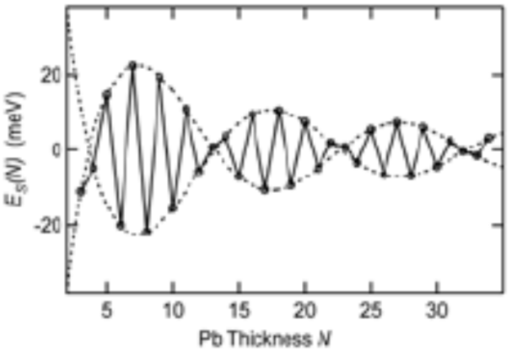


Fig. 7.23. Surface energy for Pb films on Si(111) deduced from an x-ray analysis of film roughness.

7.7.4. Structural Relaxation of Thin Films

An issue of great interest for smooth films is the internal layer structure. The neighboring atomic layer spacings could deviate from the bulk value in response to the modified electronic structure near the boundaries [104, 105]. Specifically, confinement leads to Friedel-like charge oscillations within the film with a period of 2.2 atomic layers. This period, given by one half of the Fermi wavelength, is the same as that governing the film property variations as a function of film thickness. A model calculation for a freestanding 7-ML Pb film is shown in Fig. 7.24(a) as an illustration. The electronic density wave exerts a force on the atomic planes, leading to lattice distortions. The force can be calculated, to first order, from either the electrostatic field or the local charge gradient, as shown in Fig. 7.24(b). The two methods of calculation yield very similar answers within the film. This approximately bilayer relaxation effect is quite large for Pb films on Si(111), as verified by synchrotron x-ray diffraction experiments. Results from first-principles calculations are in good accord [106].

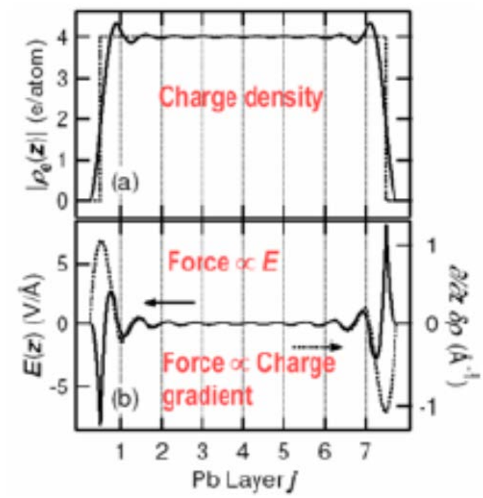


Fig. 7.24. (a) Calculated charge density of a 7-ML Pb film. (b) Force on the atomic layers calculated from the electrostatic field or charge gradient. The two calculations agree closely within the film, but differ outside.

7.7.5. Electron-Phonon Coupling and Superconductivity

Electron-phonon coupling plays a central role in many important and useful physical effects and phenomena, including superconductivity, charge density waves, and structural phase transitions [106, 107]. Angle-resolved photoemission has emerged as a powerful tool for measuring this quantity [108, 109]. At sufficiently high temperatures and at energies not too close to the Fermi level, the lifetime width ΔE of a given electronic state as measured by photoemission shows a linear dependence on temperature T caused by phonon scattering. The slope of this linear dependence is related to the electron-phonon coupling strength. The so-called electron-phonon mass enhancement parameter λ is determined by the following relationship:

$$\lambda = \frac{1}{2\pi k_B}\frac{d\Delta E}{dT}, \tag{5}$$

For thin films, this quantity shows oscillatory variations as a function of film thickness for the same reason as discussed above – a one-dimensional shell effect.

Measurements of λ have been carried out for Ag(100) films grown on Fe(100) [110, 111]. The results are summarized in Fig. 7.25. The top panel shows the binding energies of the quantum well states of interest. The middle panel shows the corresponding λ determined from photoemission. The bottom panel shows the results from a simple model calculation, which agree well with the experiment. As expected, λ exhibits oscillatory variations as a function of *N*. Furthermore, there is a *~1/N* decay pattern that overlays the oscillations, leading to an enhancement of λ at small N. This enhancement can be attributed to interface effects. A large λ is often associated with a high superconducting transition temperature within the BCS model. The results suggest an interesting possibility of enhanced or novel superconducting behavior in thin films. Ag in the bulk form is not superconducting. For Ag on Fe, the chances of finding superconductivity are probably slim because Fe is ferromagnetic. Other substrates may be better candidates.

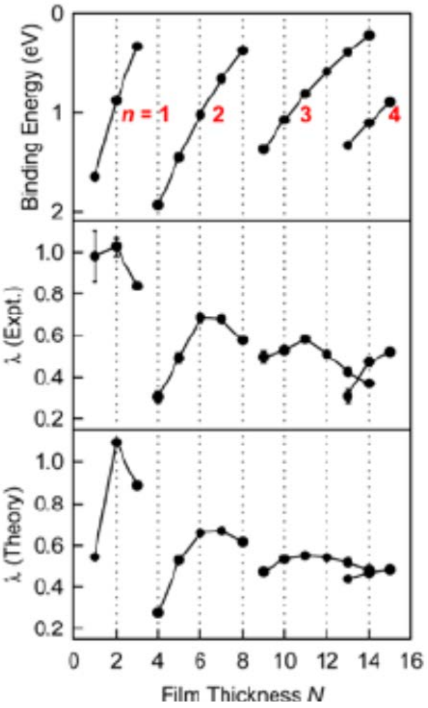


Fig. 7.25. Top, energies of quantum well states for Ag on Fe as a function of N. Middle and bottom, measured and calculated electronphonon coupling parameters.

While enhanced or novel superconductivity is yet to be discovered, oscillatory variations in the superconducting transition temperature TC has been found in Pb films deposited on Si(111) [112-114]. The results from an experiment are shown in Fig. 7.26 [112], where bilayer oscillations are evident. The samples used in this experiment were Pb films deposited on Si(111), then coated with a Au overlayer for protection in order to transfer the sample from the growth chamber through air to the low temperature transport measurement equipment. The Au overlayer could introduce a phase shift, affecting the oscillation pattern.

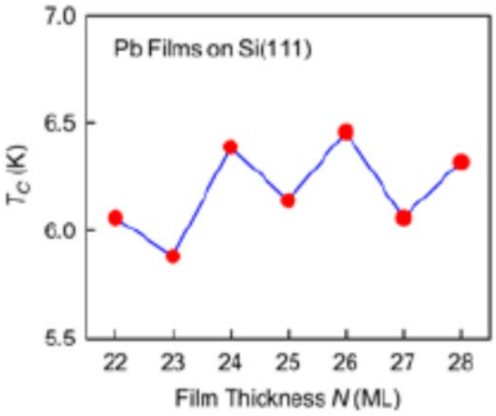


Fig. 7.26. Measured superconducting transition temperatures of uniform Pb films of various thicknesses on Si(111).

7.7.6. Beyond the Simple Particle-in-a-Box Model: Substrate Effects

The simplest picture for a thin film quantum well is that of a pair of parallel electron mirrors that reflect electrons back and forth to form standing waves in a manner similar to the optical modes in a Fabry-Pérot interferometer [86]. The actual electronic structure of a thin film can be much richer. The main differences are: (1) electrons in the substrate can couple to those in the film, resulting in a coupled structure that can be difficult or impractical to describe in terms of a single-particle phase shift analysis, as given by Eq. (7.40) [106, 115, 116]; (2) the film can support surface states, which interact with the rest of the system to create a complicated spectral weight function [117]; and (3) the corrugation potential at the film-substrate interface can lead to "multi-beam" mixing for incommensurate interfaces, resulting in complex electronic effects [106, 118].

7.8. Conclusions

Some phenomena in metallic films, nano- and microstructures have been considered. It was shown that the refractive index of metallic thin films depends on the thickness and fabrication method. As it was concluded the main factors contributing to the refractive index were grain size, impurities, surface deformation, oxidation, and free path limitations. Overall, it became evident that instead of using literature values, it is most advisable to measure the refractive indexes for home-made thin film and use those values in designing optical elements. Especially for metals that react easily with oxygen, such as aluminum, the oxide inside and on the film has a significant effect on the refractive indexes. It is probable that metallic nano- and microstructures are even more prone to surface deformations and oxidation due to several air–

metal interfaces, and the refractive indexes are drifted further away from the literature values representing thin films or bulk materials.

Overall, metallic nano- and microstructures possess great potential for several significant applications. One interesting possibility for metallic nanostructures lies in metamaterials that are artificial materials with optical properties different from anything in nature or capable of producing optical effects much stronger than natural materials. In the future, instead of today’s fabrication techniques for photonic structures, chemical methods, such as self-organization of molecules, might be used as the first-choice fabrication method to produce truly nano-sized structures leading to the creation of next generation materials.

Nanoscale phenomena are of fundamental scientific interest and technological importance. Ultrathin films, with one nanoscale dimension that can be controlled with atomic layer precision, provide an excellent model platform for experimenting with size and boundary effects. The basic quantum physics of thin films have been illustrated. A central issue is quantum confinement, which leads to discretization of the electronic states. This in turn leads to a one-dimensional shell effect that modulates the physical properties of thin films as a function of thickness. Property variations in thermal stability, work function, electron-phonon coupling, superconducting transition temperature, etc. have been demonstrated.

These effects also provide opportunities of novel device concepts. There is indeed a great potential for scientific and technological advances.

Chapter 8. Metal oxides and ceramics

8.1 Introduction

Metal oxides are an important class of materials: from both scientific and technological perspectives they present interesting opportunities for research. Thin films are particularly attractive owing to their relevance in devices and also for the ability to pursue structure–property relations studies using controlled microstructures. The inherent compositional complexity (due to the presence of ionic species) leads to rich set of properties, while in several cases, coupling of structural complexity with dynamic electronic properties leads to unexpected interfacial phenomena. Oxide semiconductors are gaining interest as new materials that may challenge the supremacy of silicon. A further recent area of research is in understanding interfaces in oxides and how they influence carrier transport. Thin film oxides are extensively used to probe strong electronic correlations.

8.2 Routes to achieve thick films of metal oxides

8.2.1 Vapor deposition

Spanning both atomic and gaseous deposition routes, vapor deposition processes rely on the transport of material in the vapor phase at the atomic or molecular level. This can be either a cloud of atoms (as in physical vapor deposition (PVD) techniques) or a precursor gas (as in chemical vapor deposition (CVD)). The names themselves hint at the mechanism by which films are formed.

PVD entails the creation of an atomic vapor that is transported to, and then condenses on, the surface of a substrate. The condensation on a bathroom mirror is a fairly good model to consider when trying to understand the process. In the bathroom, water is turned to steam through the application of energy (heat). The steam is emitted from the source (target) and is transported across the bathroom (chamber) until it condenses on a cold surface such as the mirror (substrate). The deposition process can be affected either by fixed obstacles (i.e. you keep the shower door closed) or by deflecting the steam (you open the window or switch on the extractor fan) away from the mirror. Clearly, in a PVD system, the conditions are more closely controlled, but the principle is broadly the same. To maximize the deposition rate you should avoid obstacles between the target and substrate therefore, preventing the vapor from being deflected. In a PVD system this means operating the deposition process under a high vacuum because even the molecules in the air will block the atomic vapor. Operating under a vacuum has the added advantage that it becomes possible to deposit reactive species, such as metals, without fear that they will react with the oxygen in the atmosphere. The careful control of oxygen in the chamber can be used to purposefully react with the metal being deposited in order to create a ceramic (metal oxide) film.

Returning briefly to the bathroom scenario, heat was used to create the water vapor. Similarly, heat can be used in PVD systems to vaporize the different target materials, although there are other energy sources that are also used – particularly with high-melting-point materials such as ceramics. The way in which the energy is delivered differentiates the various deposition techniques.

1. Thermal evaporation – Resistive heating of a conductive tray or holder is used to melt and evaporate the target material. This is most often used for low-melting-point metals such as gold.
2. Sputtering – High-energy charged ions are bombarded onto the surface of target material, causing atoms from the target to be expelled; this can be described as cannon balls (charged ions) hitting the castle wall (target) causing bricks (atoms) to explode outward.
3. Electron beam deposition (EBD) – An incident electron beam is used to locally heat the surface to a very high temperature, vaporizing the target material.
4. Ion beam deposition (IBD) – An ion beam is used to knock surface atoms off the surface of the target.
5. Pulsed laser deposition (PLD) – High-power laser light is used to vaporize the surface of the target.

A chemical vapor deposition (CVD) system is slightly more complex in that a chemical reaction also needs to occur. This can be between two reactive gases or between the reactant and the substrate. A bathroom analogy could be the reaction between the water and exposed ironwork resulting in the production of rust (iron oxide). As with PVD, there are many variants to CVD that arise depending on the heat source:

- Atmospheric pressure chemical vapor deposition (APCVD)
- Low-pressure chemical vapor deposition (LPCVD)
- Metal organic chemical vapor deposition (MOCVD)
- Plasma-assisted/enhanced chemical vapor deposition (PACVD/PECVD)
- Laser chemical vapor deposition (LCVD)
- Photochemical chemical vapor deposition (PCVD)
- Chemical beam epitaxy (CBE)

As the deposition processes essentially rely on the small-scale transport of matter, they are not well suited to creating very thick films. Faster deposition rates are achieved with liquid deposition routes, without the need for expensive vacuum equipment.

8.2.2 Liquid deposition

Chemical solution deposition (CSD) is an overarching term used to describe any technique whereby a chemical precursor solution is used to create a film. Unlike suspensions of material, CSD systems do not contain the actual material that is to be deposited. Instead, precursor chemicals are either dissolved or suspended in a solvent that, when deposited and heated, will result in the formation of the material of interest.

The simplest form of CSD is metal-organic decomposition (MOD) which, as the name suggests, consists of a process where metal-organic precursors decompose to produce the required film. These precursors (long-chain alkoxides, acetates, or salts) are dissolved in a carrier fluid (alcohol or water). The MOD precursor solution can then be deposited and heat-treated to remove the carrier and solvents. Further heat treatment (200–500 ^{o}C) will then cause the precursors to decompose, leading to a ceramic material that can be crystallized at even higher temperatures. MOD is a very simple process and well suited to simple ceramic materials. It is less suited to producing complex multi-ion ceramics such as lead zirconate titanate (PZT) because of the difficulty in controlling the decomposition of the precursors. If one precursor decomposes ahead of the others, the resultant ceramic film will not be homogeneous. For example, if the Pb, Ti, and Zr precursors for PZT all decompose at very different temperatures, the resultant film will consist of PbO, TiO2, and ZrO2 phases and the only way to achieve a PZT film would be to heat the film to high temperatures in order to encourage atomic diffusion. This could be acceptable for bulk systems, but for the creation of films, the application of higher temperatures can result in many unwanted effects.

Sol–gel processing is different from MOD processes in that the metal-organic precursors are reacted together to produce molecular clusters in which the ratio of the different metal atoms is fixed and homogeneously distributed throughout the system. Reactive metal-organic compounds such as metal acetates, propoxides, and butoxides (short-chain metal alkoxides) are used as starting reagents along with solvents and stabilizing compounds that extend the lifetime of the sol. Metal alkoxides are relatively larger molecules consisting of a metal ion surrounded by a number of organic alkoxide groups. During sol synthesis, the metal-organic precursors undergo hydrolysis and condensation reactions to build up the 3D nanoscale metal-organic clusters. During hydrolysis reactions, the metal alkoxides become partially hydrolyzed by reacting with water to replace one or more of the alkoxide side groups. Condensation reaction can then take place, leading to the polymerization of these partially hydrolyzed metal alkoxides and the release of water and alcohol.

The polymeric network consists of metal–oxygen–metal connections where the metal atoms can be any of the precursor elements. This ensures that there is atomic-level mixing.

While a significant number of metal-oxygen–metal (M–O–M) bonds are developed, a significant number of the unreacted alkoxide side groups still remain. Therefore, while the stoichiometry of the metal-organic cluster is approaching that of the desired ceramic, it cannot be considered a true ceramic due to the organic residue.

Water plays a key role in the synthesis of sol–gel material. Under optimal conditions, the initial hydrolysis stage occurs in a controlled manner to allow the condensation reaction to proceed and the polymer networks to form. If the hydrolysis proceeds in an uncontrolled manner, or is allowed to proceed over an extended period of time, unstable oxide particles are produced which precipitate out of the solution. Sol–gel systems are therefore typically prepared under dry conditions (inert and/or dry atmosphere) and also exhibit a finite shelf life which can have an influence on repeatability if used for film production.

While sol–gel is very effective for simple ceramics such as ZrO2, it comes into its element for complex ceramics such as Pb(Zr,Ti)O3-based piezoelectric materials where the individual metal atoms, including the minor dopants, need to be incorporated homogenously.

As with MOD processing, the sol–gel system is converted to an oxide ceramic through a pyrolysis stage where heat is used to decompose the organic component of the sol–gel to leave behind only the ceramic material. Unlike the MOD system, because the sol–gel consists of nanoscale clusters of pre-reacted metal-organic compounds there is no danger that one species will decompose prior to the other. Initial heating results in the removal of the solvents to leave behind a gel network of the nanoscale clusters. On further heating (200–450 ^{o}C), the M–O–M organic networks begin to break down to produce amorphous metal oxide particles which then crystallize on further heating (450–600 oC).

Sol–gel can be used to make films up to 5 µm thick. However, because of the significant amount of volume reduction on removing the solvent and organic components, as well as densifying the gel network, the control of internal stresses poses a significant challenge. So, although 5-µm-thick films have been produced this required a complex and well-controlled processing route. The majority of sol– gel films produced are still below 1 mm in thickness.

CSD coatings have been deposited using a variety of techniques including spray, spin, and dip coating. Once the wet films have been deposited, they are dried and then pyrolyzed to produce a stable film. Due to the high levels of stress in the film (caused by the volume reduction) it is usual to use a multilayer approach to achieve the final desired film thickness. Each layer is typically between 50 and 150 nm in thickness and once pyrolyzed can have another layer deposited on top of it. If the layer is not pyrolyzed first, the second layer can redissolve (partially or fully) the first layer and the combined shrinkage of both films on pyrolysis can cause the system to crack. The final crystallization stage can be conducted at intermittent stages or at the end of the process. Films at the upper limit of processing will typically have been exposed to multiple crystallization stages as these high-temperature stages help to anneal out some of the stored stresses.

To achieve the best results, it is important to select an appropriately matched substrate. In the majority of cases, the amorphous ceramic film will crystallize through homogenous nucleation resulting in a randomly orientated polycrystalline structure (see chapter 4 to see how this affects the properties). However, by selecting a suitable substrate with a crystal structure closely matched to the final film, it is possible to select for heterogeneous nucleation on the substrate to lead to the formation of aligned columnar grains. This alignment results in an enhancement in the functional properties of many films. For PZT-type materials (111), orientated platinum serves as an excellent substrate for use in a variety of piezoelectric MEMS devices. This form of preferred orientation, generated through heterogeneous nucleation, does not occur in thick films due to the presence of the powder particles.

8.2.3 Thick-film powder-based routes

Processing

At the heart of the thick-film deposition route is the powder. Routes by which this powder can be deposited. Here, we will explore the processing of the green-body films and the mechanisms that occur during this processing. Central to the processing of ceramic thick films is the need to transform the porous ceramic powder bed into a continuous high-density film in order to maximize the properties and increase the strength of the film. This is achieved through a number of processes including: causing the particles to fuse together under the action heat (sintering), binding the particles together using a second phase (composite), or consolidating the particles under the action of high-speed impact (particles joined) or bulk material (particles fully fused), then the energy of the system can be greatly reduced.

Sintering

Sintering describes the process whereby solid powder particles fuse together to eliminate the porosity between them, thereby increasing the density of the system. The driving force for this densification is the reduction in energy associated with the system, which is given by the sum of the bulk, surface, and grain boundary energies. As surface energy (that associated with the free surface of a material) is typically much higher than grain boundary energy (that associated with the boundary between two grains), which again is higher than the bulk energy, it is possible to reduce the energy of the system as a whole by eliminating as much surface as possible. In the case of a powder compact, the total surface energy will be related to the total surface area of all the particles. If these surfaces can be eliminated through transforming them into either grain boundaries (fig.1).

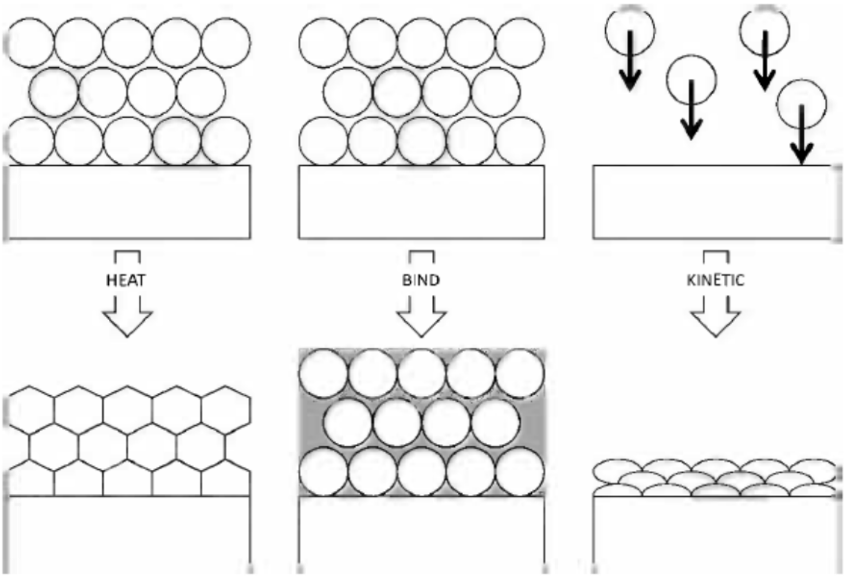


FIGURE 1 Schematic illustrating different routes for creating continuous high-density thick films. Films can be sintered through the action of heat, held together with the addition of a binder, or consolidated due to high kinetic energy impacts.

Composite sol–gel

Composite sol–gel combines the thick-film-forming capabilities of powder-based systems with the low temperature processing of sol–gel. The composite sol–gel route is a combination of powder-based processing and CSD. It combines the advantages of thick-film processing offered by the presence of the powder with the advantages of low-temperature processing offered by CSD. By mixing the sol–gel precursor and the powder together to create an ink (known as a composite ink), it is possible to create a system where the sol serves a dual purpose of carrier and binder (Figure 2). In this way, the composite sol–gel approach combines the infiltration and power processing stages into a single stage to enhance sintering. In addition, it then becomes possible to compositionally match the binder to the powder so that there is significantly less inactive material present in the final system Alternatively, the powder and sol-derived material can be different, allowing unique properties to be obtained from these composite systems. This approach opens up the possibility to create multifunctional systems.

As the composite sol–gel system is essentially a mixture of sol and powder, it is possible to produce inks with very different fluidic properties by altering either the sol concentration or the ratio of sol to powder. This allows a wide range of inks, with very different rheological properties, to be produced. In turn, this means that the composite sol–gel approach can be applied to a variety of different deposition techniques.

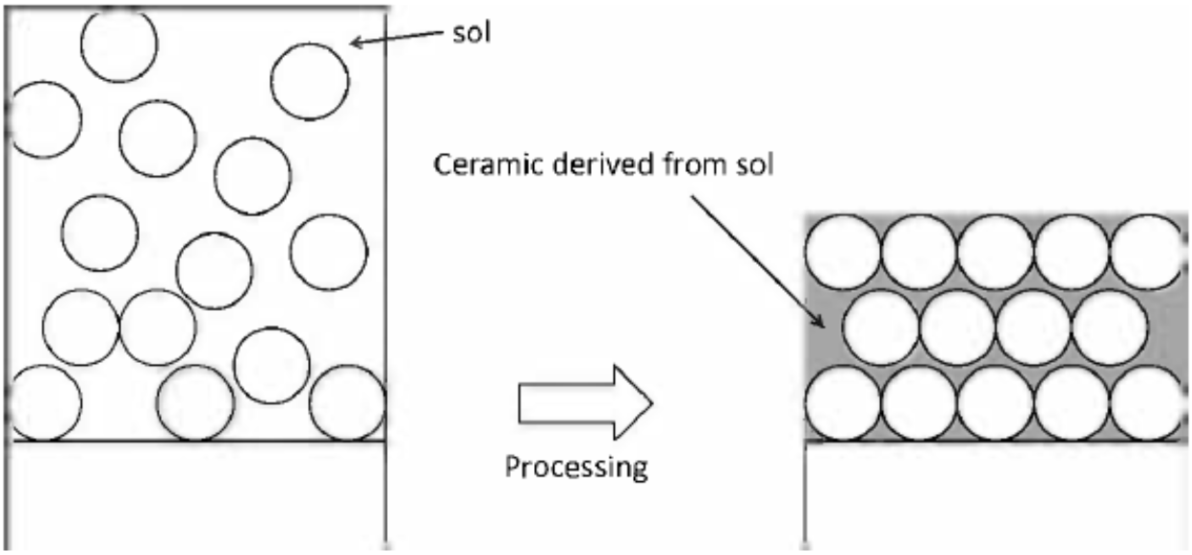


FIGURE 2 Schematic of composite sol–gel systems. During deposition the sol acts as a carrier fluid and, after processing, the sol–gel-derived ceramic binds the ceramic particles together and to the substrate.

Irrespective of the route used to deposit the film, the underlying process of working with a composite sol–gel system remains the same and can be broken down into a number of key processing stages that need to be considered when creating thick films. Central to successful processing is the control of the reduction in volume as volatile components are driven off. The sol–gel phase is composed of the metal-organic precursors along with the solvent and stabilizing agents. At the end of the process, all that will remain will be the ceramic phase that is derived from the metal-organic precursor. This means that a significant amount of material needs to be removed during processing which, in turn, means that the system will experience a significant volume reduction. This volume reduction needs to be carefully controlled in order to minimize the stresses within the system. The reduction in volume occurs in each of the processing stages but is most significant during drying and pyrolysis.

Drying

During the drying stage, the low-boiling-point carrier materials and stabilizing species are removed. Within this stage, the ink remains fluid for a significant period of time, allowing reorganization of the solid matter with the ink. If the drying stage is conducted too quickly, then the solid phase may become ‘locked’ into position early on in the process, resulting in a significant amount of residual porosity and possibly high levels of stress (fig.3).

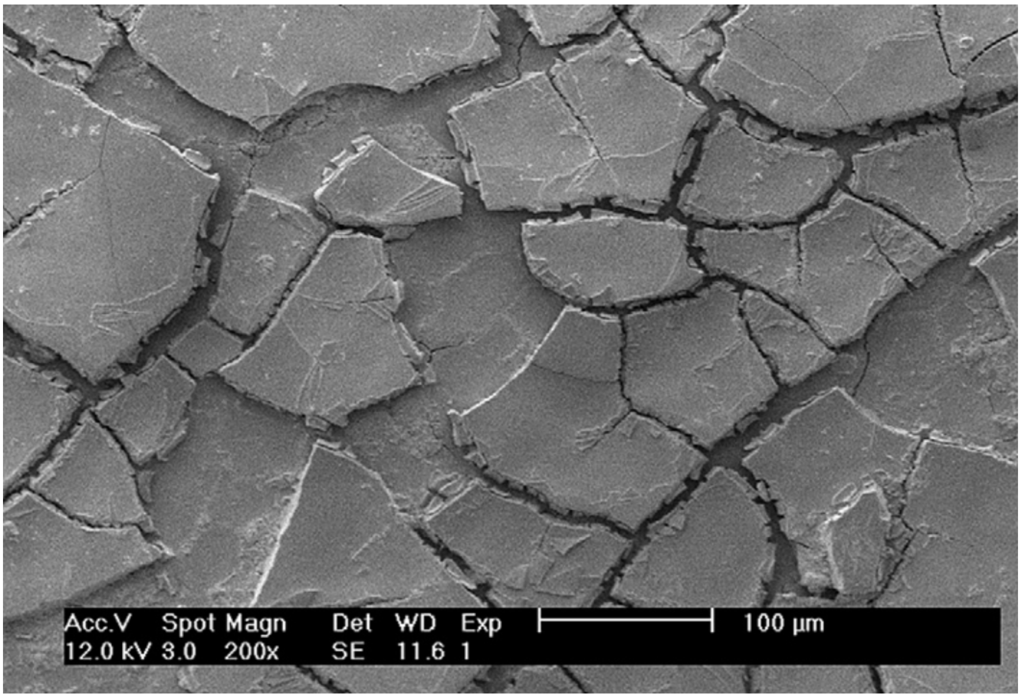


FIGURE 3

Scanning electron micrograph illustrating drying-induced cracking in a thick film.

Pyrolysis

Pyrolysis typically occurs in the temperature range of 250–550 ^{o}C where further volume reduction occurs as the organic component of the sol is removed. Once the lowvolatile compounds have been removed from the system during drying, the material that remains is composed of solid material (i.e. powder) and metal-organic precursor compounds. These precursor compounds are composed of metal ions bonded in an organic network consisting of C, H, and O atoms.When heated above a critical temperature, these precursor compounds decompose to yield volatile compounds such as H_2O and CO_2 and an amorphous metal oxide material. The removal of these volatile compounds results in a further significant loss in mass, which needs to be carefully controlled. This is particularly demanding as the ability of particles to rearrange and relive stress is greatly diminished in relation to the drying stage. Rapid changes in strength and increases in stress can also lead to cracking.

Crystallization

The final stage in the process is the conversion of the amorphous metal oxide material into a crystalline metal oxide material. It is this crystallization stage that develops the desirable functional properties in the material. While this stage does not exhibit the massive loss of mass of the previous stages, there can still be a slight volume change brought about by the reordering of the atoms on crystallization.

Whether a volume increase or decrease occurs will depend on the relative density of the crystalline phase with respect to the amorphous phase.

Unlike conventionally produced ceramic thick films, the powder sol–gel films are composed of two ceramic phases, which means that it is possible to create a range of microstructures depending on the ratio of sol to powder in the starting ink. The microstructure and properties of the film depend on the ratios used as well as the thickness of the film.

At one extreme is the pure sol system which will yield very thin films and, if processed correctly, a high-density film. The addition of small amounts of powder to the sol causes the viscosity of the fluid to increase resulting in a thicker deposit. The resultant microstructure will then consist of a ‘sea’ of sol–gel-derived material with isolated powder particles embedded in it – similar to individual grapes encased in jelly (Figure 4). Depending on the concentration of the sol and powder loading, you may end up with a flat film or one where the profiles of individual powder particles are visible on the surface. This second case occurs when there is insufficient solid material derived from the sol (usually caused by a low concentration sol) to fill the gaps between the particles while at the same time the initial powder loading was too low to form a continuous coverage of the substrate. The first case is favored with higher sol concentrations and low powder loadings.

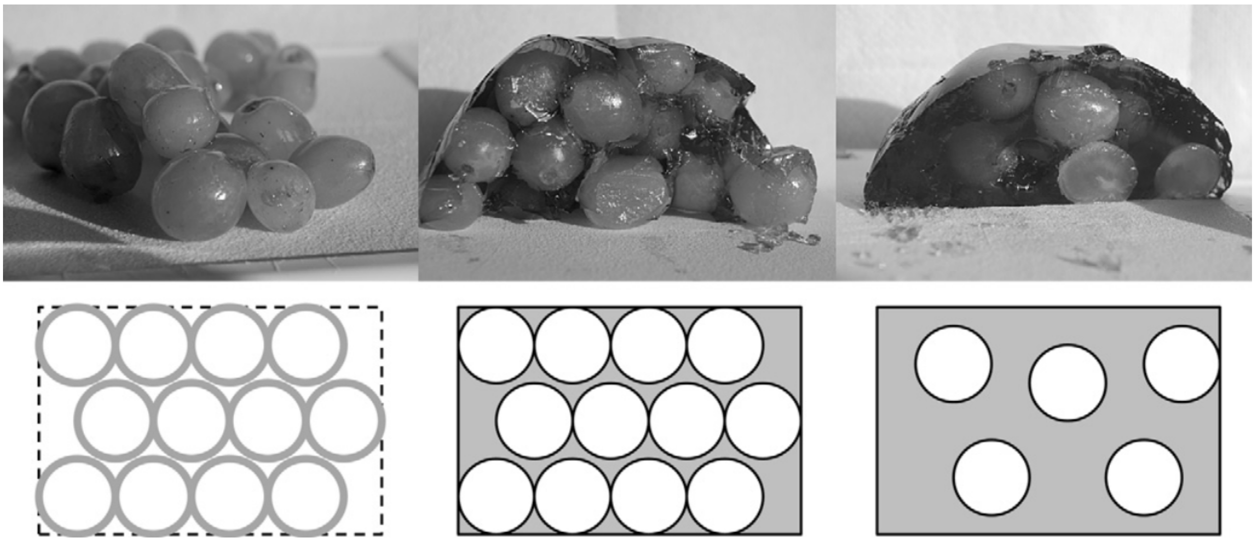


FIGURE 4 Grape/jelly analogies of composite sol–gel processing. Particles (grapes) are held together by thin coating of sol–gel material (or jelly); well-packed particles are encased in sol–gel-derived phase or; isolated particles are suspended in a sea of sol–gel-derived material.

As the quantity of powder is increased, the viscosity of the fluid increases still further and thicker films are produced. In addition, because of the higher powder loading, the particles cease to remain isolated in the final film. Initially, a loose network of particles will be formed, but as more powder is added to the system this will tend toward a conventional packed powder bed, much like the jelly part in a fruit trifle (Figure 4). This can be considered the tipping point – where the volume of solid derived from the sol fills the pores between the particles. Adding more powder beyond this stage will result in an open porous structure being produced; one where the individual powder particles are still held together by the sol but where there are voids between the particles (Figure 4).

In the case of powder-rich systems, with pores between the particles, subsequent sintering stages may be employed to increase the density of the film if required. This produces films with a microstructure where high density is achieved with less sol–gel-derived material. Such approaches may be appropriate where the properties of the powder are more important than, or superior to, those of the sol–gel-derived material.

8.2.4 Powder manufacturing routes

There are many producers of ceramic powders; so, often, the easiest route to thick-film fabrication is simply to obtain your powder from such a supplier. There will be times when this is not suitable: the powder may not be available, it may not be of the right purity or size, or you may simply wish to control the production yourself. In these instances, it will be necessary to make your own powder. The following information provides an overview of different routes for making ceramic powders (Figure 5).

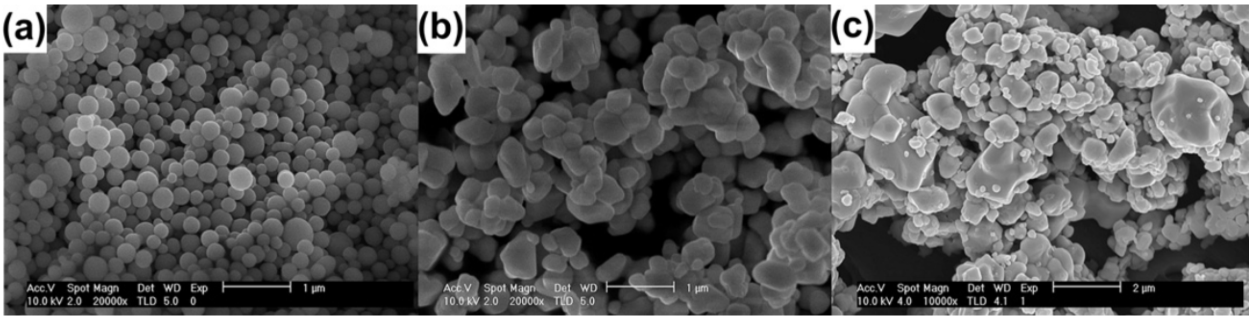


FIGURE 5

Electron micrographs showing the comparison of powder prepared by spray dried (a), molten salt (b), and mixed oxide (c) methods.

Solid state

The most common way to manufacture powders is through the solid-state route, sometimes called mixed oxide route, where the starting reagents (usually oxide materials, but carbides can be also used) are mixed together to ensure a homogeneous distribution of components. This mixing can be done either dry or wet, provided the starting reagents do not react with the liquid used. The powder mix is then heated to relatively high temperatures (ca 600–1000 oC) to cause decomposition and/or atomic diffusion while at the same time minimizing coarsening or sintering.

Nonoxide precursors (such as carbides) decompose to produce the oxide and carbon dioxide. For complex mixed oxide ceramics, it is the atomic interdiffusion between the different reagents that is critical for the formation of the desired powder. This reliance on atomic diffusion explains the need for high temperatures to be used and also explains why small particle sizes are critical. At the appropriate processing temperature, atomic interdiffusion will occur between the different materials, first resulting in the creation of core–shell structures (Figure 6) and finally the desired product. In cases where atomic diffusion of all species is approximately equal, the different species will diffuse equally so that all particles gradually develop into the desired phase. In other cases, there will be species that do not diffuse readily (e.g. in TiO2) and the desired product will only form as the other species diffuse into the stable material.

Molten salt

Similar to the mixed oxide route, molten salt synthesis is also a diffusion-based process. Unlike the mixed oxide route, molten salt synthesis relies on a liquid-phase diffusion mechanism which has the potential of enabling lower temperature synthesis compared to the mixed oxide route. The starting reagents, along with a low-melting-point salt, are mixed to produce a homogeneous distribution of reagents. On heating, the salt (or mix of salts) melts and suspends the starting reagents

Two mechanisms have been identified that can occur depending on the solubility of the different reagents. In one case, all reagents are soluble in the molten salt, while, in the other, only some of the reagents are soluble (Figure 7).

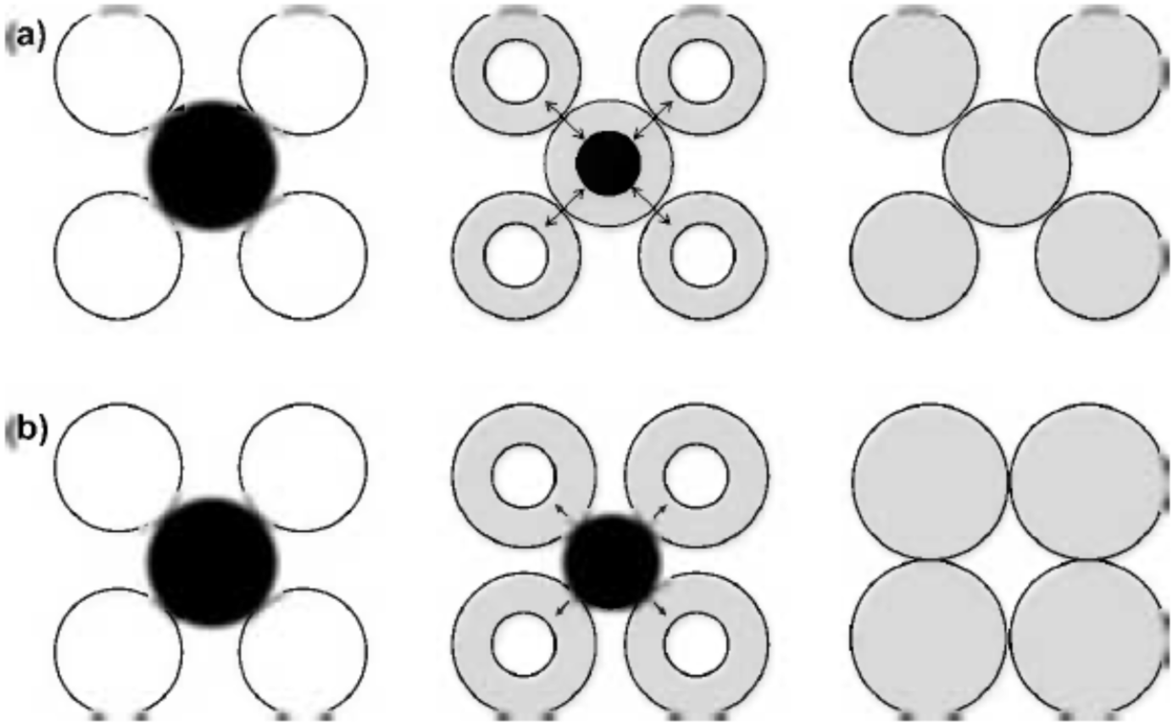


FIGURE 6 Illustration of the evolution of powder during solid-state synthesis, where a) both starting reagents have equal diffusivity or b) the diffusivity of the white reagent is negligible.

When all reagents are soluble, they all dissolve completely in the molten salt where they react with each other to produce an insoluble product (ideally the desired ceramic). For instance, PbO and ZrO_2 are both soluble in a NaCl/KCl mix and react to produce $PbZrO_3$. On cooling, the salt solidifies and encases the product. Subsequent washing can then be used to dissolve the salt and reveal the final product.

The second mechanism occurs when not all the species are soluble in the salt. Where one of the species is not soluble, the other reagents dissolve in the salt and react with the surface of the undissolved species. At this stage, the soluble species diffuse (by solid-state diffusion) into the undissolved material, creating a core–shell structure in the initial stages. As time progresses, the core is eliminated through continued diffusion. Due to the reliance on solid-state diffusion, this second mechanism may take longer to reach completion compared to the first case where all species are soluble. The production of $PbTiO_3$ from PbO and TiO_2 is an example of this second type of reaction where the TiO_2 does not dissolve in the molten salt. When more than one species is not soluble in the molten salt, the reaction cannot proceed to completion without large-scale solid-state diffusion. In this case, it may not be possible/practical to use molten salt synthesis.

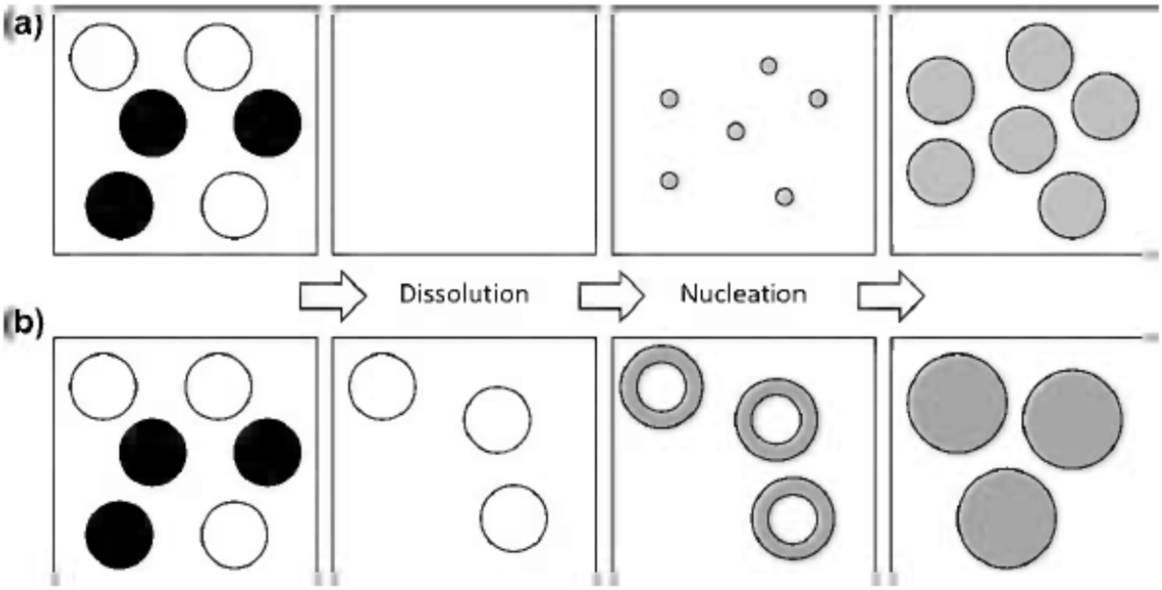


FIGURE 7 Diagram showing the evolution of ceramic material during molten salt synthesis, where a) both starting reagents dissolve in salt solution where they react and nucleate the desired ceramic particles which continue to grow or b) one reagent does not dissolve – the formation of desired powder proceeds through diffusion of the dissolved species into undissolved particles.

8.3 Application of metax oxides for MEMS

8.3.1 Piezoelectric devices

When considering functional ceramics, piezoelectric materials are probably used in the most diverserange of devices. Piezoelectric materials are those that change shape when an electrical signal isapplied to them and produce an electrical signal when subjected to a mechanical force. The operationof piezoelectric devices can be considered in light of the three main operation modes of piezoelectric materials: thickness (d33), lateral (d31), and shear (d51). The rest, as they say, is packaging. Consider a piezoelectric material with a net polarization in the 3 direction (Figure 8). With electrodes applied to the two faces perpendicular to the polarization direction, both thickness and lateral modes can be operated. When an electric field is applied to the electrodes, the sample will either increase or decrease in thickness depending on the sign of the field applied; this is the thickness mode.

At the same time, the sample will also increase or decrease in width (and depth); the lateral mode of operation. The shear mode is activated by applying the electrodes to the sides of the sample. When activated, the piezoelectric material undergoes a shearing motion around the 2 direction. When the piezoelectric material is used in the form of a film, the d33 and d31 modes are predominantly used due to the ease of arranging the electrodes and poling the material. Compared to the bulk system, the way in which piezoelectric films are used is very different due to the very nature of a film – thinness and the presence of the substrate. The thinness of the film means that any movement in the through-thickness direction or through-shear deflection is very small (Figure 8). Furthermore, the presence of the substrate introduces a bending motion in the system, which provides a new mode ofactuation not available to bulk materials, as well as reducing the degree of piezoelectric activity due toclamping.

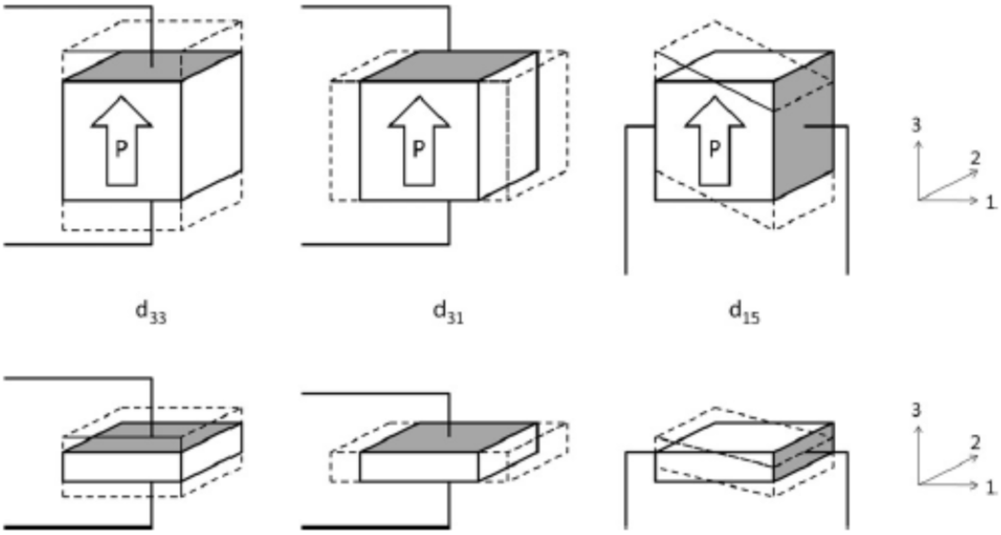


FIGURE 8

Piezoelectric operating modes d33 (thickness), d31 (lateral), and d15 (shear). Comparison between top and bottom rows shows that only the d31 mode exhibits the same degree of deflection as the thickness of the material is reduced.

8.3.2 Actuators, speakers, and pumps

The application of an electric field causes the piezoelectric material to deform and act upon the environment. All devices in this section rely on the production of the greatest deflection for maximum effect:

-Actuator – large movement equates to a large range of motion

-Speaker – large movement results in a large pressure wave that equates to a loud sound-Pump – large movement gives greater volume of fluid pumped

To achieve large ranges of motion cantilever, bridge or diaphragm structures are required (Figure 9) because of the small displacements achievable with shear and thickness mode excitation.

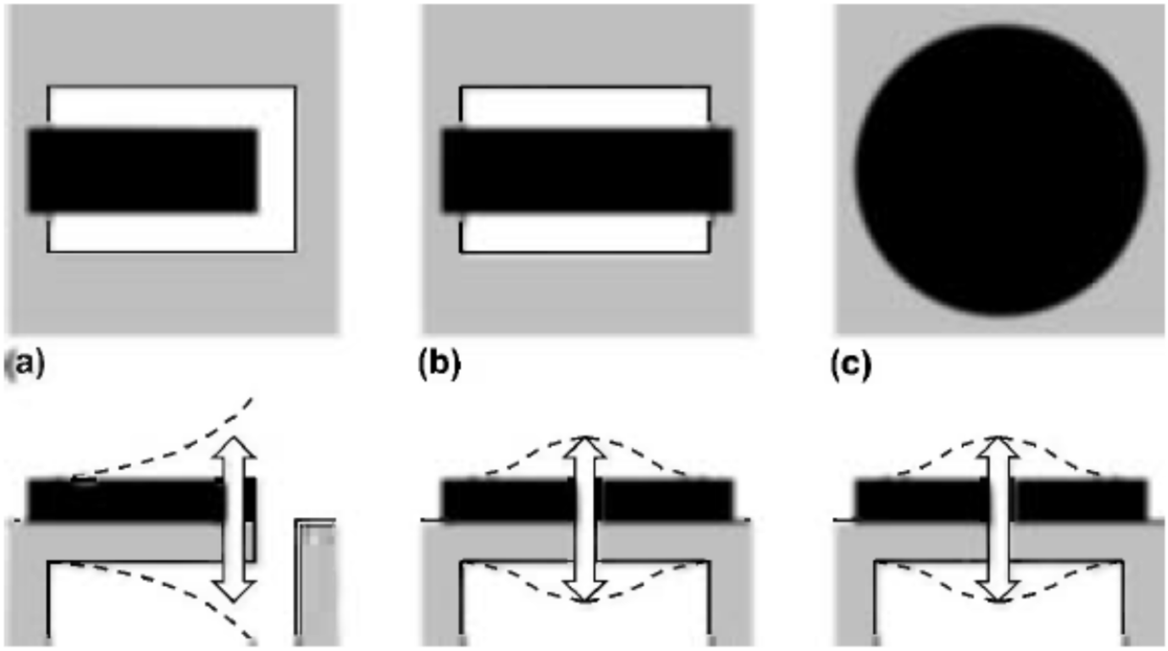


FIGURE 9 Schematics showing a) cantilever, b) bridge, and c) diaphragm structures used in micro-actuators, speakers, and pumps.

8.3.3 Sound and acoustic emission sensors

Sound is a pressure wave that is detected by the ear when the pressure wave caused the hairs in the cochlea to vibrate and nerve signals to be generated. In an analogous way, sound waves can also cause a piezoelectric material to vibrate resulting in the generation of an electrical signal. Both bending-mode and thickness-mode devices can be used to detect the sound. The choice of which one to use depends on the application. The ability of the piezoelectric material to detect the sound depends on the transmission of that sound energy from the medium (i.e. air, water, or solid) to the sensors. This degree of transmission, in turn, depends on the acoustic matching, or coupling, of the medium and device. If the acoustic impedances of the medium and the device are not similar, a significant amount of energy will not be transmitted from one to the other and the sensor will lack sensitivity.

In thickness mode, the impedance of the device is dominated by the acoustic impedance of the piezoelectric materials (Z_0 = 30–35 MPa s/m). When compared to the acoustic impedance of air (Z_0 =4.28 Pa s/m), it is clear that a thickness-mode device would be very insensitive in an air environment. For water (Z_0 = 1.5 MPa s/m), the matching is better, and better still for steel (Z_0 = 28.5 MPa s/m). Thickness-mode devices are therefore very well suited for sensing sound in metal (and water to lesser extent), i.e. for detecting cracking in metal structures via acoustic emission sensing.

In bending mode, the impedance of the device is dominated by the elastic stiffness of the materials used and the dimensions (thickness and diameter) of the diaphragm. Large, thin diaphragms made with low-stiffness materials will exhibit a low acoustic impedance. In practice, the choice of materials is limited, so the stiffness range is relatively small. Hence, acoustic impedance is mainly controlled by adjusting the dimensions of the diaphragm. Bending-mode diaphragm structures can be engineered to have a range of acoustic impedances but are best suited for air applications and occasionally, water applications.

Complicating the matter further, for both thickness- and bending-mode devices, the sensitivity of microphones increases significantly at the resonant frequency of the device. The resonant frequency of the device is also dictated by the material properties and dimensions of the device. This can create a conflict between the desired resonant frequency and acoustic impedance, which greatly restricts the design of the device. If a broad frequency range of sound is to be detected, nonresonant devices are favored as each sound frequency will receive approximately equal weighting in the resultant signal.

8.3.4 Ultrasound transducers

Transducers combine both the actuation and sensing aspects of the piezoelectric device in one. As such, they can act upon their surrounds and also sense the response from the environment. Sonar is one example where bulk piezoelectric materials are used. On the smaller scale, the same technology is used to create ultrasound devices. Ultrasound is sound with a frequency greater that 20 kHz (above the range of human hearing).

In ultrasound applications, the transducers first act as a transmitter sending an ultrasound signal outward. When the sound wave encounters an object with a different density (e.g. bone), a proportion of sound will be reflected backward toward the transducer (Equation) as with an echo. When the reflected sound wave reaches the transducer, an electrical signal is generated. By measuring the time difference between transmission and reception of the signal, it is possible to calculate the distance of the object from the transducer. Scanning the ultrasound beam physically or electronically allows a 2D or 3D map of the environment to be constructed:

$$R_A = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2$$

where Z1 and Z2 are the acoustic impedances (density of medium/speed of sound in medium) of medium 1 and 2, respectively, and RA is the reflection coefficient describing the fraction of acoustic energy reflected off the second object.

Macroscale ultrasound transducers are made using bulk ceramic material a few millimeters thick. These typically work in the 1–3-MHz range and are used for deep penetrative scans such as abdominal investigation (e.g. pregnancy) in medicine. The operating frequency of the transducer is dictated by the resonant frequency of the device (for maximum power and sensitivity). To operate at higher frequencies in order to achieve higher-resolution images, the device must be reduced in size. To achieve resonance at tens or hundreds of MHz requires piezoelectric elements a few tens of micrometers thick – the realm of thick films (Figure 10). Alternatively, bending-mode transducers can be created using thick films that resonate at a few MHz as withconventional bulk transducers. The advantages of using resonating bending-mode structures include better acoustic matching and the potential to make transducer arrays, which allow the acoustic beam to be scanned electronically instead of mechanically, thereby increasing reliability and accuracy. Bending-mode transducers offer better sensitivity at the expense of lower transmission power.

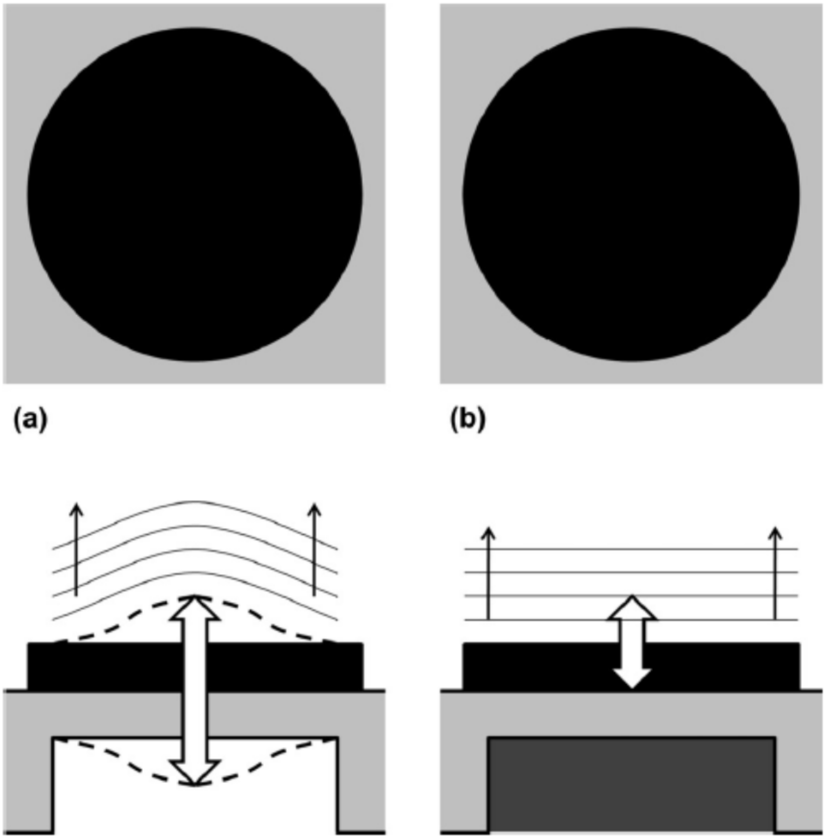


FIGURE 10 Bending-mode (a) and thickness-mode (b) ultrasound transducers. Thickness-mode transducers are more suited to operating in high acoustic impedance environments, while bending-mode transducers are more suited to low acoustic impedance environments. Note that despite the two devices having the same dimensions it is unlikely that the resonant (operating) frequency will be the same.

8.3.5Acceleration sensors

In addition to responding to pressure waves, piezoelectric materials will also produce an electrical response when deformed as a result of the application of a mechanical force. By connecting a piezoelectric element to a body and a large mass (proof mass), this effect can be used to create sensors capable of detecting acceleration (Figure 11). Under the action of acceleration, the body will move while inertia will tend to keep the proof mass stationary, creating a differential movement between the body and the proof mass.

The force experienced by a body is equal to the sum of its mass and the acceleration experienced. As the response of a piezoelectric film is proportional to the force applied, the response of the piezoelectric device can be used to determine the magnitude of the acceleration. In addition, through the appropriate design of device (i.e. restricting motion to one direction only), it is also possible to determine the direction of acceleration. These are the foundations of inertial positioning devices that can monitor the movement and position of different objects or people. In the majority of applications, thickness- and shear-mode thick-film accelerators lack the required sensitivity, thus favoring bending mode devices.

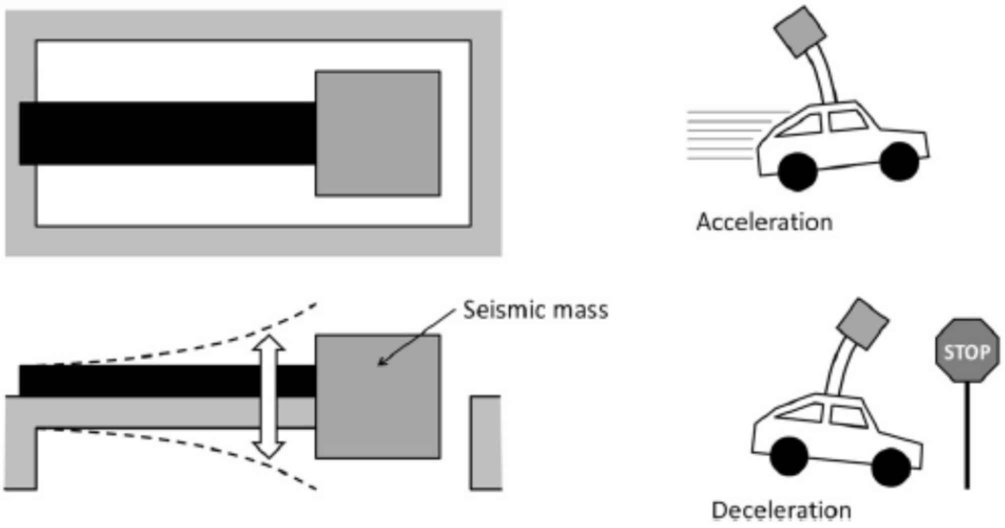


FIGURE 11 Schematic of bending-mode piezoelectric accelerometer. Acceleration and deceleration will cause the accelerometer to deform in different directions giving rise to a different signal. Greater acceleration/deceleration will result in greater deflection and a larger signal. Steady-state driving will not give rise to a signal

8.3.6 Pressure and chemical sensors

One of the features of resonating structures is that their behavior is governed by the material properties of the different parts of the device as well as the environmental conditions in which the device operates; change either of these and the response of the device will also change.

For example, a simple cantilever will have a characteristic resonant frequency which will change if its environment (e.g. viscosity or density) is altered. However, changes in fluid characteristics of air at different pressures are relatively small and result in very small changes in resonant frequency. If, instead of using a cantilever, a sealed diaphragm is used, changes in the pressure on one side of the diaphragm relative to the other side will stiffen the structure and result in a larger shift in resonant frequency. If the characteristics of the diaphragm device are known, then a given change in resonant frequency can be converted to a pressure reading (Figure 12).

The resonant frequency will also change if the mass of the system increases (more mass, lower resonant frequency) or if the stiffness of the system changes (decreased stiffness, lower resonant frequency). Therefore, when an adsorbent layer is incorporated into the structure, the resonant frequency will change when material is adsorbed. By using a selective adsorbent layer, the device will only respond when subjected to specific species, thus allowing specific stimuli to be detected. Many types of resonating structure can be used for these applications with higher resonating frequencies giving greater sensitivity. In addition, for thick-film systems, any adsorbed material will give rise to a greater relative mass change than in a bulk material system. Hence, film-based resonating structures, operating at high frequencies, offer the potential for high-sensitivity chemical and gas sensors.

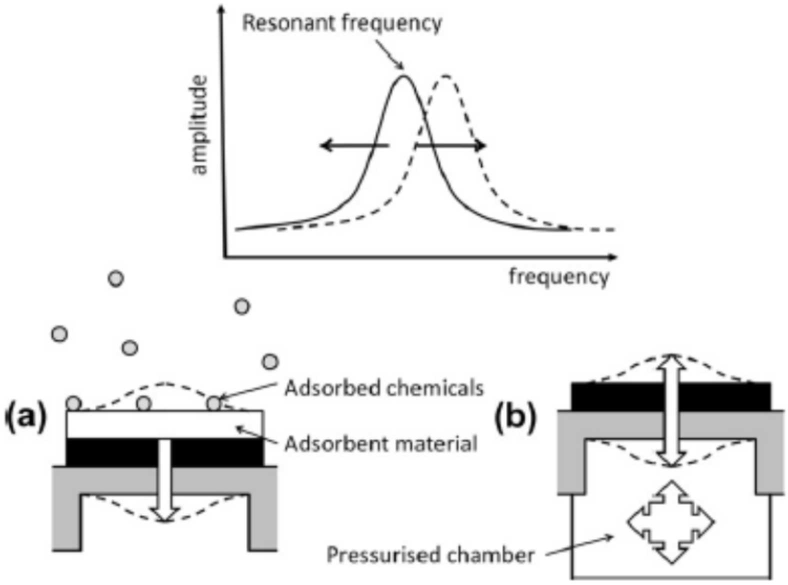


FIGURE 12 Schematic of resonating chemical (a) and pressure (b) sensor structures and an idealized resonance response. Increased mass due to an adsorbed chemical species will decrease the resonant frequency of the structure (a) while increased gas pressure in the chamber will increase the resonant frequency (b).

8.3.7 Transformers

Related to transducers, transformers also combine both actuation and sensing components; however, instead of stimulating the environment and sensing a response, the actuator part of the transformer directly stimulates the sensor part in order to transmit energy across the two parts of the device. The voltage generated by a sensor is related to its structure (i.e. electrode spacing and configuration) and the amplitude of the mechanical signal. Likewise, the amplitude of the mechanical vibrations created will depend on the device design as well as the input signal amplitude. Therefore, voltage transformation can be accomplished by simply having appropriate structures within the transmission and reception parts of the device. The configuration of the two parts dictates whether voltage is increased or decreased through this action (Figure 13).

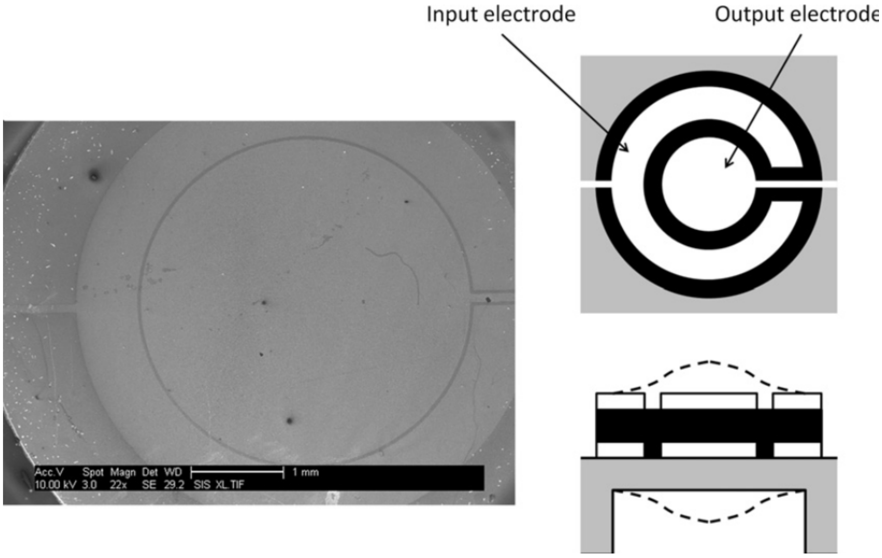


FIGURE 13 Scanning electron micrograph and related schematic of a thick-film electrical transformer. The diaphragm to deformed by applying a voltage to the input electrodes. This caused piezoelectric material at the center of thediaphragm to be also strained which results in a (different) voltage being generated at the output electrodes

8.3.8 Energy transmitters

When energy or signals need to be transmitted through a solid obstacle, the most common solution is to drill a hole and pass a wire through the hole before sealing the gap. This is not always practical, for example, in pressure vessels, containment chambers, or water or airtight compartments on sea vessels and aircraft. In such situations, a matching pair of transducers either side of the bulkhead can be used to transmit and receive signals and energy in the form of sound waves. This is similar to the transformer, but now the focus is on ensuring maximum energy transmission, with minimum dispersion/loss.

Voltage transformation is possible in such systems, but the most effective transmission of energy will be achieved when both transmitter and receiver are tuned to the same resonant frequency. Incorporating a voltage transformer component would affect this matching and reduce efficiency of transmission. Essentially, the system consists of a focused ultrasound transmitter matched to a corresponding receiver that reconverts the sound energy into electrical energy.

8.3.9 Energy harvesters

The principle behind the operation of piezoelectric energy harvesters is the conversion of mechanical oscillation into electrical energy. The majority of available ‘waste’ energy is low frequency and to achieve the best energy conversion, the harvester should operate at resonance. This can be achieved either by designing bending-mode harvesters that are sufficiently large that their resonant frequency is within the appropriate range or by using a mechanical frequency transformer to increase the frequency of the available oscillation, much in the same way that a guitarist converts their slowly oscillating hand into a high-speed oscillating string by plucking the string with a plectrum.

The maximum conversion efficiency occurs when the piezoelectric device operates at resonance; however, the available ‘waste’ vibration energy is not always available at a convenient single frequency. The two approaches adopted to maximize the amount of energy harvested using thick-film devices are: to use an array of harvesters each tuned to different resonant frequencies or, to make a single harvester that has a broad range over which it resonates (broadband resonator). The second approach entails creating devices with different cross-sectional profiles and thicknesses. While this creates a broadband device, for any given frequency, the amount of energy harvested is less than it would be for a single narrow-band harvester. The effectiveness of the device, therefore, depends on the characteristics of the individual environments in which it is used. It is for this reason that the majority of piezoelectric harvesters are designed specifically for individual applications. The size of the harvester directly influences the amount of energy that can be harvested. For thick-film devices, typical energy harvesting levels are in the mW range.

8.3.10 Pyroelectric devices

Related to piezoelectric devices are pyroelectric devices, which share a common family of materials. This is because the piezoelectric materials used in thick-film devices are actually ferroelectric materials. In addition to being piezoelectric, these ferroelectric materials are pyroelectric. Pyroelectric materials generate an electrical signal when subjected to changes in temperature.

Pyroelectric sensors

Responding to changes in temperature, pyroelectric materials are used in a range of thermal-sensing devices including intruder alarms, infrared thermal imagers, and gas sensors. The advantage of using films for these applications stems from their low mass, which means that they are more responsive to temperature fluctuations. In addition, the use of films – and, in particular, structured films – means that imaging or diagnostic systems can be created. Just like a digital camera consists of an array of lightsensitive elements, so does a thermal imaging camera. Unlike piezoelectric systems, pyroelectric materials are operated using just one mode; the same as that used for piezoelectric thickness mode with the electrodes perpendicular to the poling direction.

Pyroelectric energy harvester

While the pyroelectric effect can be used to harvest energy from temperature fluctuations, such fluctuations are not of significant magnitude or frequency to enable significant levels of energy to be captured (pW). Higher levels of energy can be captured by making use of the bending motion induced by thermal expansion mismatches between the pyroelectric material and the substrate. While driven by a changing temperature, power is generated via the piezoelectric effect.

8.3.11 Other functional devices

There are a host of other functional ceramics that can be incorporated into microsystems in a similar way to piezoelectric materials. The structures adopted are often based on the same basic structure with electrical connections to the functional ceramic, in order to transmit the resultant signal or power, which is placed on a supporting substrate.

Solid oxide fuel cells

Solid oxide fuel cells (SOFCs) are complex multilayer systems where all layers may be ceramic thick films (Figure 1.14). Each of the layers serves a different purpose and requires very different structures. A basic SOFC structure consists of an electrolyte sandwiched between two electrodes, one of which may be the supporting substrate. Fuel enters on one side of the electrolyte (anode) and air on the other (cathode). The oxygen becomes ionized at the cathode/electrolyte interface and oxygen ions diffuse through the electrolyte to combine with the fuel at the anode/electrolyte interface. Electrons flow around the external electrical circuit from the anode to the cathode.

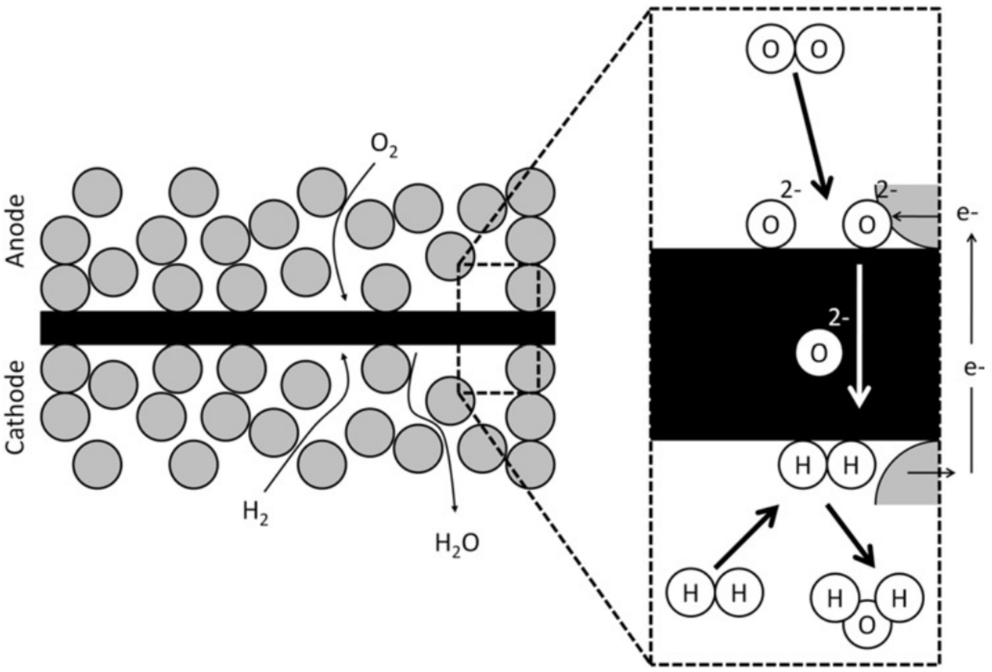


FIGURE 14 Schematic of solid oxide fuel cell structure showing the porous anode and cathode and the high-density electrolyte. The insert shows the transport of oxygen ions through the electrolyte, after which they can combine with hydrogen atoms to produce water vapor. Intimate contact is required between the electrodes, electrolyte, and gas phases for power generation.

The electrolyte is an ionic conductor (e.g. oxygen ions) that needs to be gas tight to prevent the fuel and oxygen from mixing. To maximize the current generated, it should have a large area and small thickness, besides possessing a high ionic conductivity. It should also have a high electrical resistance to prevent backward current leakage. Yttrium-stabilized zirconia (YSZ) is frequently used for such applications.

In contrast to the electrolyte, both the anode (fuel side) and the cathode (oxygen side) need to be porous to provide a high amount of surface area for the fuel and air to interact with. In addition to having a high specific surface area, the electrodes should provide a large number of triple points where the gas phase, electrode, and electrolyte materials are in close proximity with each other to facilitate interaction between the gases, ions, and electrons. Both electrodes should have a high electronic conductivity to minimize ohmic losses. The anode needs to be stable under reducing conditions, while the cathode needs to be stable under oxidizing condition. A Ni/YSZ composite is frequently used for the anode, while La0.8Sr0.2MnO3 is used as the cathode. To increase the ionic conductivity in the YSZ, the SOFCs are operated at elevated temperatures (500–800 oC), which places additional demands on the materials and structure as they must remain stable for prolonged periods of operation.

Electrochemical gas senor

Consisting of a high-density thick film of ionic conducting ceramic material (e.g. YSZ) with a porous platinum electrode either side, a potentiostatic gas sensor works by generating a potential difference across a ceramic film because of the difference in partial pressure of oxygen in the combustion gas zone and atmosphere (Equation). The ceramic film needs to be gas tight with both sides exposed to the gaseous environments:

$$E = \frac{RT}{|z|F} \ln \left(\frac{p(O_2)_A}{p(O_2)_B} \right)$$

where F is the Faraday constant (9.648x 10^4 C/mol), z is the electron charge of the system (4 for two O^{2-} ions), and p(O2) is the oxygen partial pressure (A and B representing the environment either side of the ion conducting film).

Resistive gas senor

Ceramic thick films are also used to detect gases via changes in electrical resistance. Here, the thick film is porous to allow the gases to permeate the structure and interact with the maximum amount of surface. Unlike the potentiometer sensor, which detects a potential difference across the ceramic, resistive gas sensors detect a change in resistance of the film. Due to the low conductivity at room temperature, such sensors are often combined with micro-resistance heaters. Under normal operating conditions, oxygen is adsorbed onto the surface of the ceramic materials (e.g. SnO2, WO3, and ZnO), creating negative ions with the effect that electrons are removed from the conduction band of the ceramic near to the surface of the particles. This results in an increase in the electrical resistance of the film as conduction is predominantly governed by surface/grain boundary effects. When a combustible gas is introduced, it also adsorbs on to the surface of the ceramic and reacts with the oxygen ions. In the process, electrons are released back into the conduction band, resulting in an increase in the electrical conductivity (Figure 15).

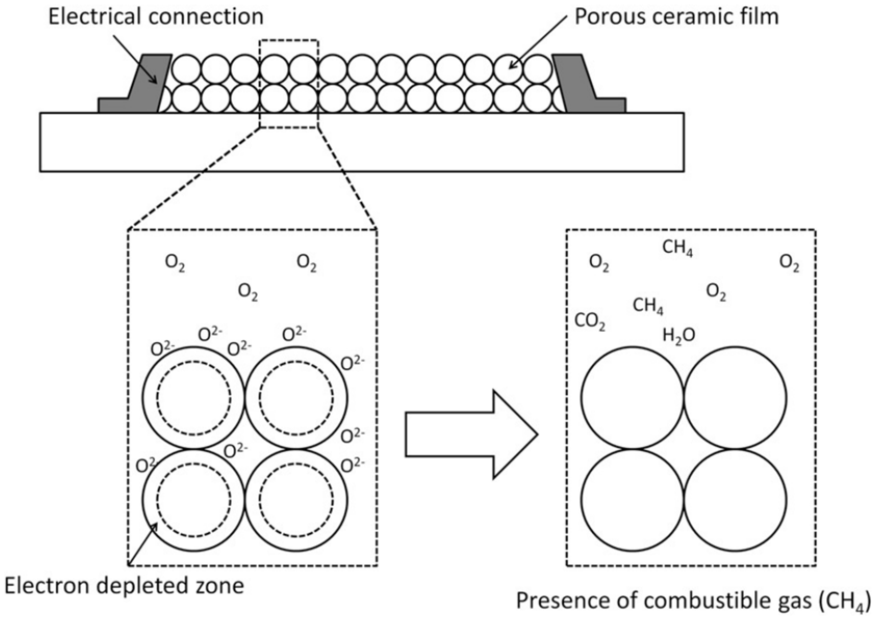


FIGURE15

Schematic of SnO2 resistive gas sensor. In air, oxygen adsorbs onto surface forming O2– ions, depleting the surface zone of conduction band electrons. When a combustible gas is introduced, it reacts with the oxygen ions, allowing the surface zone to become more conductive

Thermoelectric generator

The majority of commercially available thermoelectric generators are based on metal alloy systems (e.g. Bi2Te3). However, there are a number of ceramic materials that also exhibit thermoelectric characteristics (e.g. Ca3Co4O9) and offer potential for high-temperature thermoelectric generation where the metallic materials cease to function.

The basic structure adopted is the same in both cases and consists of pillars of p- and n-type semiconducting ceramics positioned alternately and connected in series electrically (Figure 16). In a thick-film device, the ceramics can be printed on an electrically insulating substrate and then capped with a second electrically insulating coating to prevent electrical short circuits or leakage. A temperature gradient is established between the top face and the bottom face resulting in the generation of electrical current.

Micro hotplate

Electrically conducting ceramics can be used as high-temperature heating elements. If they are nonoxides, such as titanium nitride (TiN), they may need to be sufficiently protected from the atmosphere to prevent oxidation. In such cases, a multilayer structure consisting of the TiN sandwiched between two gas tight layers (e.g. alumina) provides the basis for such a design.

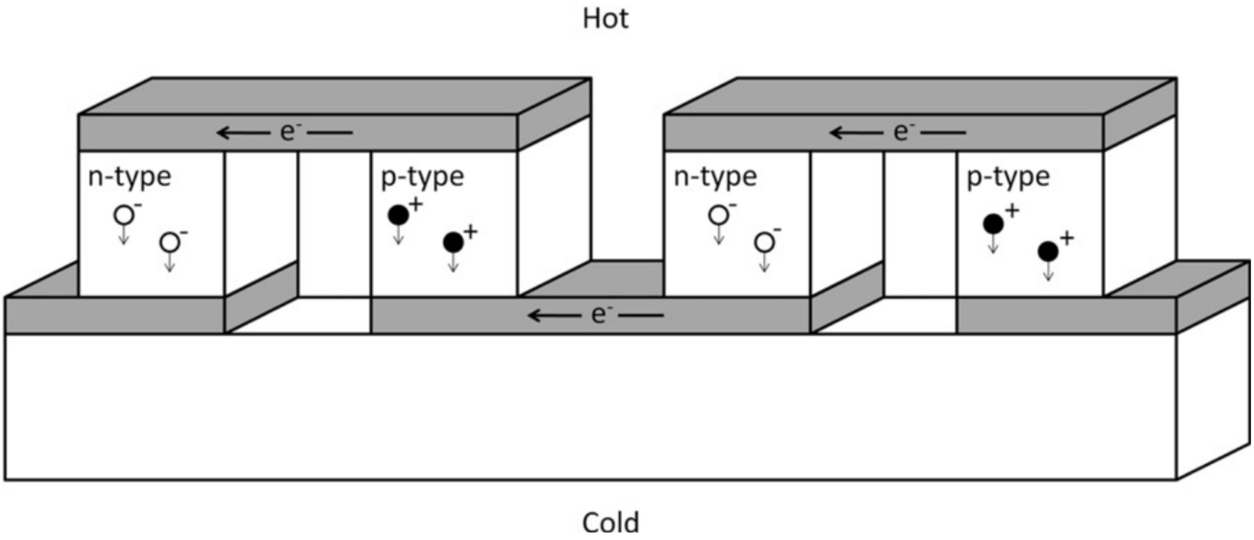


FIGURE 16 Schematic of thermoelectric device. Thick-film semiconducting ceramic structures are connected in series to generate electric current from a temperature difference.

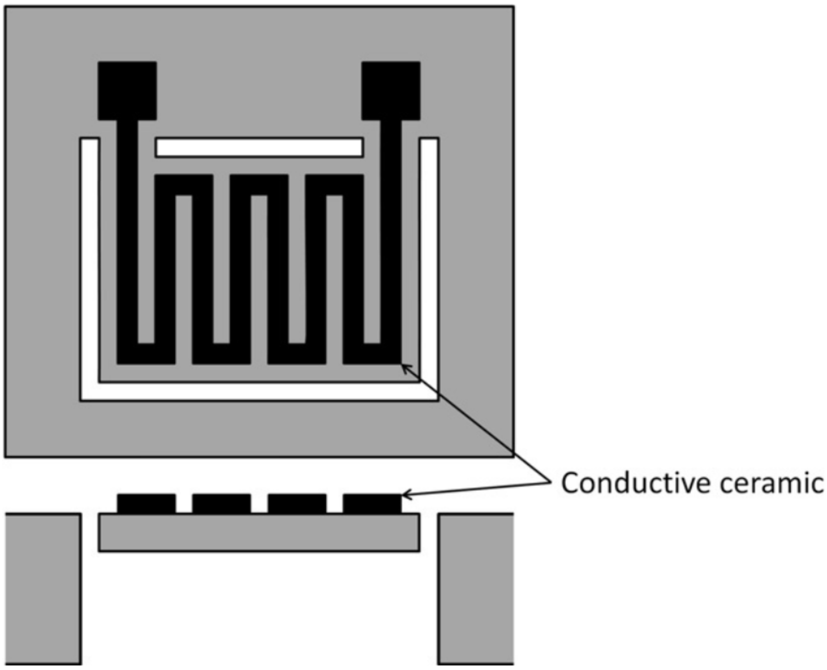


FIGURE 17 Schematic of micro-hotplate based on conductive ceramic. If the conductive ceramic is a nonoxide material, anadditional protective layer can be applied on top of the heating elements.

By thinning the underlying substrate (or removing it completely) and isolating the heating zone from the surrounding material, the responsivity and stability of the device can be improved as there is less mass for the heating element to heat and thermal losses due to conduction to the substrate are reduced considerably (Figure 17). Micro-hotplates find applications in a range of devices where localized heating is required. They are frequently found in gas sensors, where they are used to heat the ceramic gas sensor either to increase conductivity or to drive off adsorbed species to refresh the sensor, requiring the integration of two separate devices.

8.4 Functional metal oxides for adaptive electronics

We survey functional oxide materials that may have application in adaptive electronic devices. A list of oxides that will be discussed subsequently is given in Fig. 18 with corresponding crystal structures and band gaps.

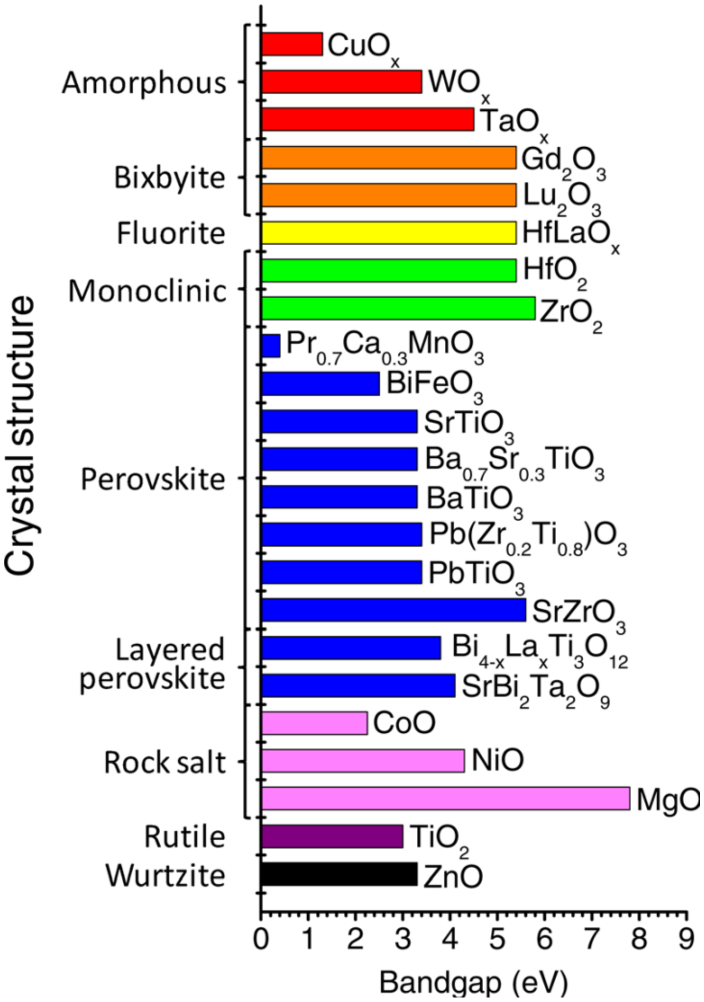


Fig. 18 List of oxides discussed in this review organized by crystal structure and experimentally reported band gaps

An essential requirement of an adaptive electronic component is a multiply-valued internal state that can be tuned in a non-volatile or quasi-stable manner. The particular internal state can be resistivity, permittivity, polarization, magnetization, optical transmission, or other properties (see Fig. 19).

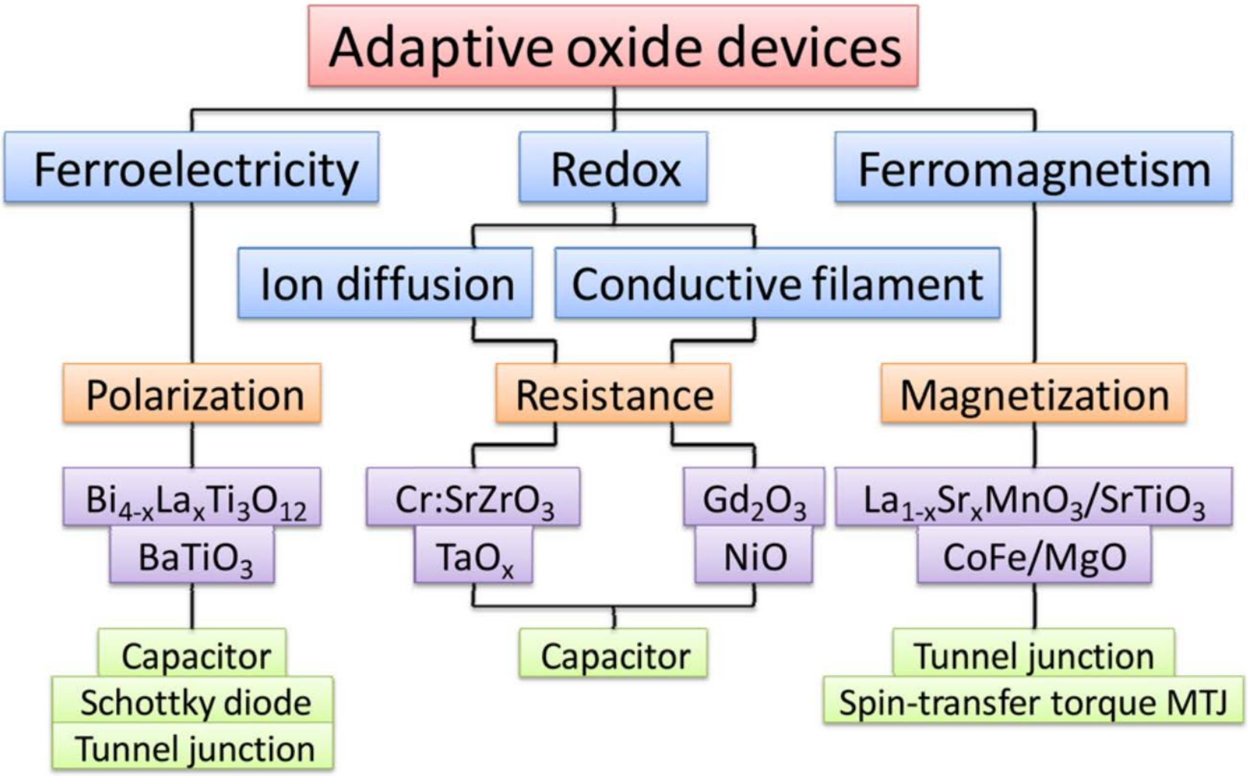


Figure 19 Diagram of adaptive oxide devices

Moreover, the perturbation which alters the internal state can be electric, magnetic, thermal, optical, or another external source. The degree to which a device can be considered multiply-valued or non-volatile is not well defined. In principle, it may be possible to modify the internal state of a device with infinite precision between two extremes, but if the intermediate states are not distinguishable within the capabilities of the peripheral circuitry, then they have little utility. In this regard, maximization of the range between internal state extrema will likely, although not necessarily, lead to greater distinguishability between intermediate states. If, in the absence of stimulus, the intermediate states fluctuate or degrade significantly over time, then the states may no longer be distinguishable and the device cannot be considered non-volatile.

The time scale over which the internal states of the device remain stable and distinguishable is referred to as the retentiontime, which is generally a function of temperature. A related metric is the endurance or fatigue factor, which is a measure of how many times the internal state can be modified before distinguishability is lost. Maximizing both endurance and retention time are key to device performance. There are several device properties that are not essential for adaptive electronics, but merit consideration for practical realization. For example, the implementation of twoterminal device architectures, as opposed to three-terminal architectures, is more suitable for various applications, such as synaptic connections in hardware neural networks and high-density crossbar arrays. Although there are applications of three-terminal adaptive electronic devices, the bulk of this review will focus on oxides implemented in two terminal structures. Other device parameters that are important are density, operating temperature range, switching speed, and CMOS processing compatibility. Additionally, there are properties that will become increasingly significant as the field matures, such as the ability to intuitively implement learning algorithms in hardware neural networks. Fig.20 shows the example of application of metal oxide natrix for one electron device.

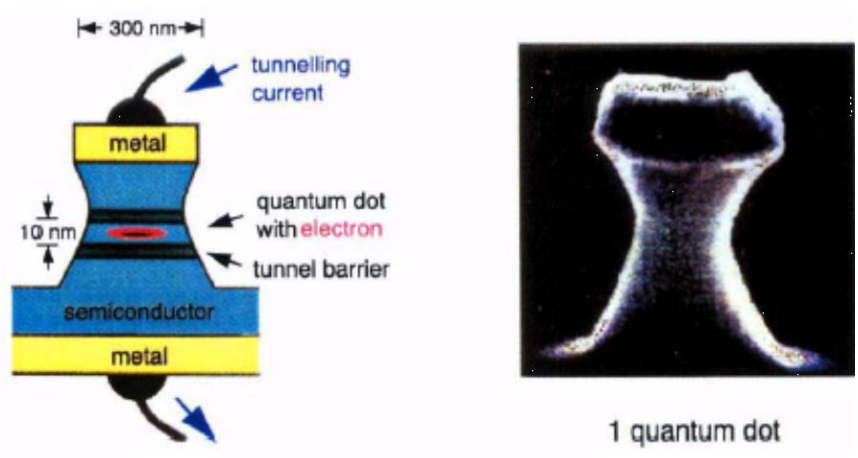


Fig.20 Quantim dots as artificial single electron device

Chapter 9 Polymeric Materials for Micromachining

9.1 INTRODUCTION

The word polymer originates in the Greek language. It is composed of two words, namely, *poly*=many and *meros*=part. Each individual part of this multitude is referred to as a *monomer*. A monomer is a compound composed of atoms or molecules capable of forming one or more characteristic and for each compound an unchangeable unit. Linking the monomers chemically produces a uniformly built polymer, which is also called a *macromolecule*. These compounds can be either manufactured synthetically or obtained or transformed from natural products.

With the development of synthetic polymers materials they are serving as conductors, insulators, semiconductors, protective layers, materials for MEMS etc., at extremely low and high temperatures, at a wide range of frequencies, in adverse environments. There are special polymers or their compounds that exhibit other specialized properties (Table 9.1).

9.2 TYPES OF POLYMERS FOR MICRO AND NANOELECTRONICS

Essentially, polymers are organic materials, prepared by processes in which simple organic molecules (*monomers*) are used as building blocks to produce large molecules (*polymers or macromolecules*). The number of monomer repeat units in a macromolecule is so large that the characteristics of a polymer are not significantly modified when removing or adding a single monomer repeat unit. Commercially available polymers are comprised of 1000 monomer repeat units that result in molecular weights ranging from 10^4 to 10^7. The number of monomeric units determines the *degree of polymerization* usually denoted by the symbol N. The structure of most macromolecules consists most frequently of one or two types of monomer repeat units, defined as either *homopolymer* or *copolymer*, respectively. Figure 9.1 shows an example of a monomer (ethylene) and polymer (polyethylene).

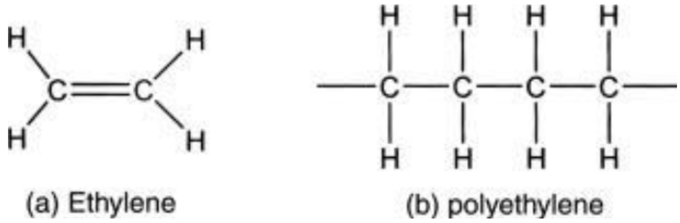
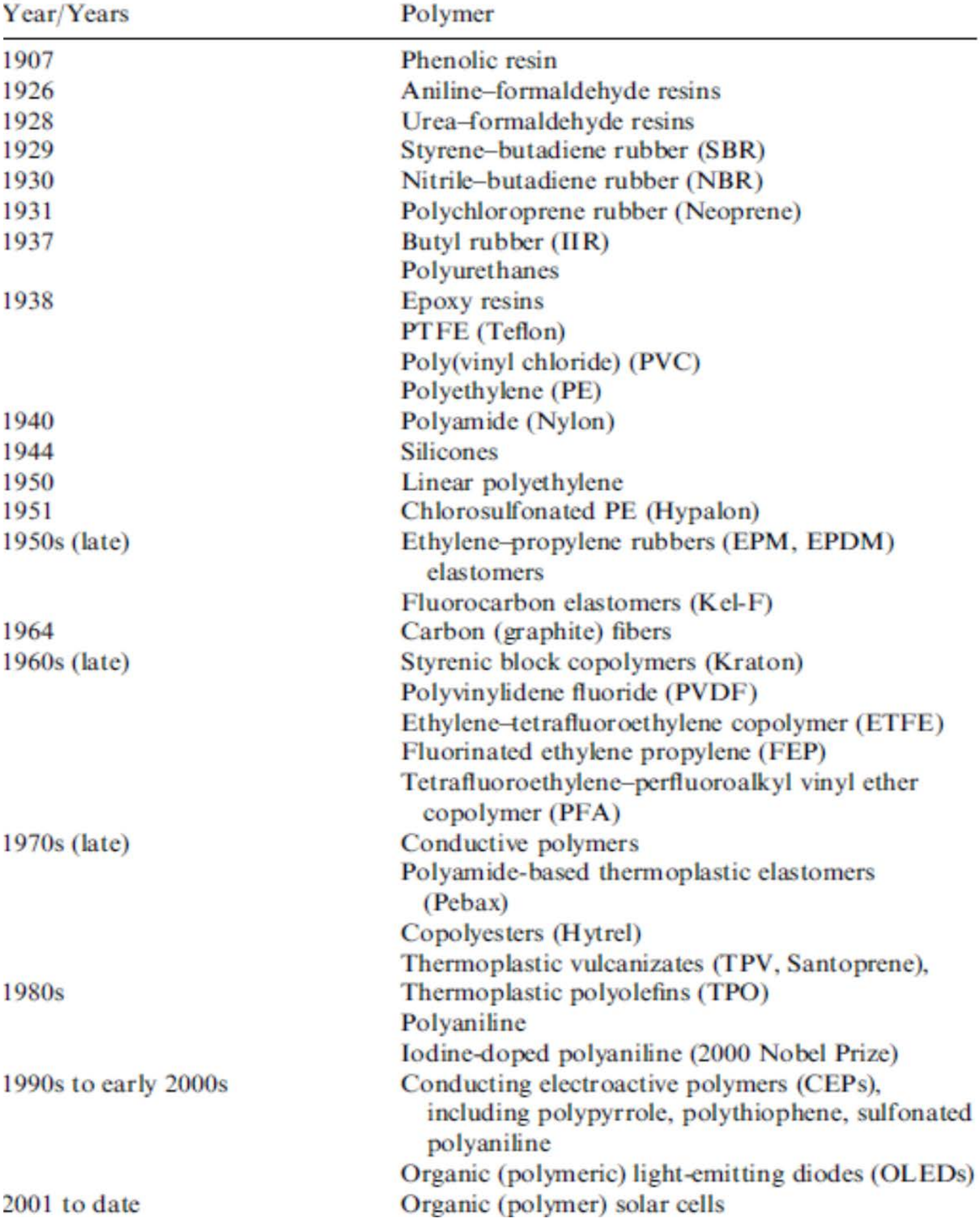

(a) Ethylene
(b) polyethylene

FIGURE 9.1 (a) Ethylene (monomer) and (b) polyethylene (polymer); simplified formulas.

TABLE 9.1 Milestones in the Development of Synthetic Polymers for Electrical and Electronic Applications



Year/Years	Polymer
1907	Phenolic resin
1926	Aniline–formaldehyde resins
1928	Urea–formaldehyde resins
1929	Styrene–butadiene rubber (SBR)
1930	Nitrile–butadiene rubber (NBR)
1931	Polychloroprene rubber (Neoprene)
1937	Butyl rubber (IIR)
	Polyurethanes
1938	Epoxy resins
	PTFE (Teflon)
	Poly(vinyl chloride) (PVC)
	Polyethylene (PE)
1940	Polyamide (Nylon)
1944	Silicones
1950	Linear polyethylene
1951	Chlorosulfonated PE (Hypalon)
1950s (late)	Ethylene–propylene rubbers (EPM, EPDM) elastomers
	Fluorocarbon elastomers (Kel-F)
1964	Carbon (graphite) fibers
1960s (late)	Styrenic block copolymers (Kraton)
	Polyvinylidene fluoride (PVDF)
	Ethylene–tetrafluoroethylene copolymer (ETFE)
	Fluorinated ethylene propylene (FEP)
	Tetrafluoroethylene–perfluoroalkyl vinyl ether copolymer (PFA)
1970s (late)	Conductive polymers
	Polyamide-based thermoplastic elastomers (Pebax)
	Copolyesters (Hytrel)
	Thermoplastic vulcanizates (TPV, Santoprene),
1980s	Thermoplastic polyolefins (TPO)
	Polyaniline
	Iodine-doped polyaniline (2000 Nobel Prize)
1990s to early 2000s	Conducting electroactive polymers (CEPs), including polypyrrole, polythiophene, sulfonated polyaniline
	Organic (polymeric) light-emitting diodes (OLEDs)
2001 to date	Organic (polymer) solar cells

Substitution of individual atoms by different substituents can occur on the backbone of a linear macromolecule, which has a more or less regular structure. Such substituents are commonly composed of atoms or atomic groups, which are chemically different from the monomer. An example, shown in Fig. 9.2, is polystyrene, a polymer in which phenyl groups are attached as side groups to the C-C backbone.

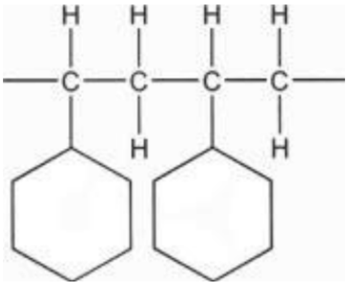


FIGURE 9.2 Polystyrene.

Macromolecules can be linked together by either physical or chemical bonds. Chemically linked macromolecules, are known as cross-linked macromolecules. Cross-linked macromolecules form a *network*. Physical bonds can develop between the individual macromolecular segments in addition to the existing chemical bonds to strengthen the entire network. The individual macromolecules prefer the coiled state and penetrate each other, thus forming an entangled chain structure. Entanglements will hold the macromolecules together and as a result the physical bonds become mobile.

Compared to the existence of physical bonds, chemical bonds are up to 10^3 times stronger. One significant difference between physical and chemical bonds is that physical bonds are reversible when subjected to heat, solvents, or mechanical forces, whereas chemical bonds are irreversible. The chemical and physical structures that illustrate normal linear, branched, cross-linked, and entangled macromolecules are shown schematically in Fig. 9.3.

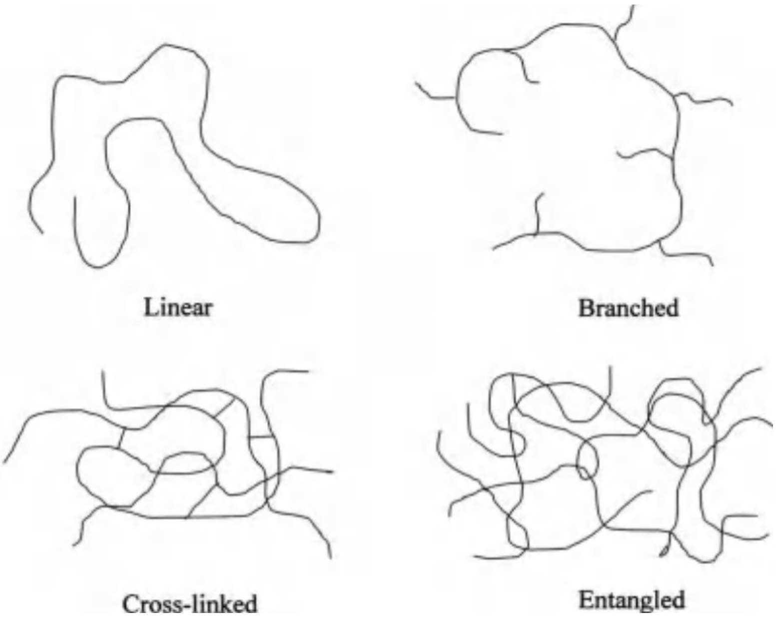


FIGURE 9.3 Schematic of linear, branched, cross-linked, and entangled macromolecular chains.

A selection of conventional polymers for microelectronics is shown in Table 9.2. These polymers are electrically neutral. If the polymeric chains are built from monomers that contain an ionizable group, that is, a group that can dissociate into a chain-fixed cation and anion and a mobile counterion bearing the opposite charge, a *polyelectrolyte* is obtained. Table 9.3 collects a few typical examples. The first three compounds are synthetic polymers, the other two samples are biopolymers; cellulose and starch in the form of derivatives which include ionizable substitutes.

TABLE 9.2 Conventional Polymers, Part A

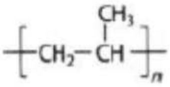
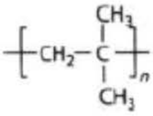
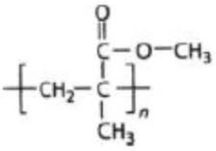
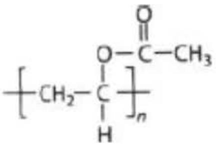
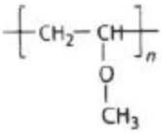
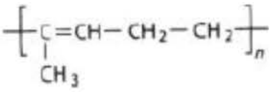
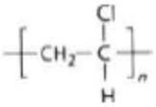
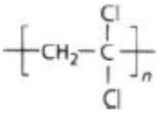
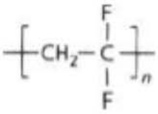
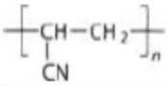
Structure	Name (abbrevation)
 $\left[CH_2-\overset{CH_3}{\overset{\vert}{CH}} \right]_n$	Polypropylene (PP)
 $\left[CH_2-\underset{CH_3}{\underset{\vert}{\overset{CH_3}{\overset{\vert}{C}}}} \right]_n$	Polyisobutylene (PIB)
 $\left[CH_2-\underset{CH_3}{\underset{\vert}{\overset{\overset{O}{\Vert}\atop{C-O-CH_3}}{\overset{\vert}{C}}}} \right]_n$	Poly(methylmethacrylate) PMMA
 $\left[CH_2-\underset{H}{\underset{\vert}{\overset{O-\overset{O}{\overset{\Vert}{C}}-CH_3}{\overset{\vert}{C}}}} \right]_n$	Poly(vinylacetate) (PVAc)
 $\left[CH_2-\underset{\underset{\underset{CH_3}{\vert}}{O}}{\underset{\vert}{CH}} \right]_n$	Poly(vinylmethylether) (PVME)
$\left[CH{=}CH-CH_2-CH_2 \right]_n$	Polybutadiene (PB)
 $\left[\underset{CH_3}{\underset{\vert}{C}}{=}CH-CH_2-CH_2 \right]_n$	Polyisoprene (PI)
 $\left[CH_2-\underset{H}{\underset{\vert}{\overset{Cl}{\overset{\vert}{C}}}} \right]_n$	Poly(vinyl chloride) (PVC)
 $\left[CH_2-\underset{Cl}{\underset{\vert}{\overset{Cl}{\overset{\vert}{C}}}} \right]_n$	Poly(vinylidene chloride) (PVDC)
 $\left[CH_2-\underset{F}{\underset{\vert}{\overset{F}{\overset{\vert}{C}}}} \right]_n$	Poly(vinylidene fluoride) (PVDF)
$\left[CF_2-CF_2 \right]_n$	Poly(tetrafluoroethylene) (PTFE)
 $\left[\underset{CN}{\underset{\vert}{CH}}-CH_2 \right]_n$	Poly(acrylonitrile) (PAN)

TABLE 9.2 Conventional Polymers, Part B

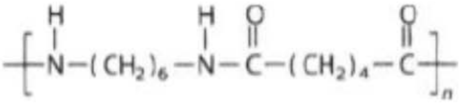
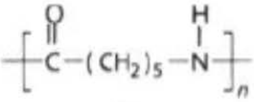
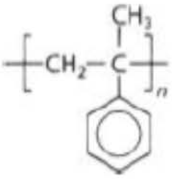
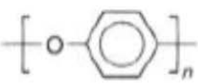
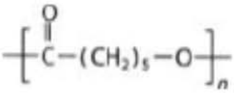
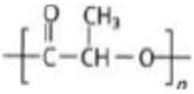
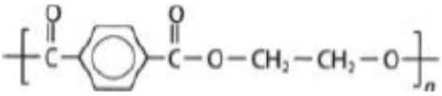
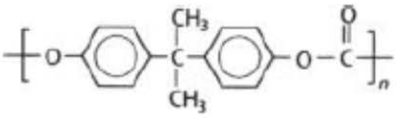
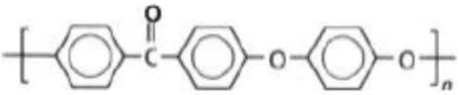
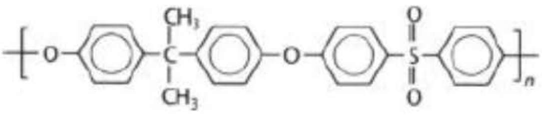
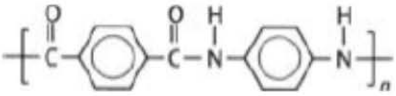
Structure	Name (abbrevation)
$\left[O-CH_2 \right]_n$	Poly(oxymethylene) (POM)
$\left[O-(CH_2)_2 \right]_n$	Poly(ethyleneoxide) (PEO)
 $\left[\overset{H}{N}-(CH_2)_6-\overset{H}{N}-\overset{O}{C}-(CH_2)_4-\overset{O}{C} \right]_n$	Poly(hexamethylene adipamide) (nylon 6,6)
 $\left[\overset{O}{C}-(CH_2)_5-\overset{H}{N} \right]_n$	Poly(*ε*-caprolactam)(nylon 6)
	Poly(*α*-methylstyrene)
	Poly(*α*-phenylene oxide) (PPO)
 $\left[\overset{O}{C}-(CH_2)_5-O \right]_n$	Poly(*ε*-caprolactone) (P*ε*CL)
 $\left[\overset{O}{C}-\overset{CH_3}{CH}-O \right]_n$	Poly(L-lactic acid) (PLLA)
	Poly(ethylene terephthalate) (PET)
	Polycarbonate (PC)
	Poly(ether ether ketone) (PEEK)
	Polysulfone (PSF)
	Poly(*p*-phenylene-terephthalamide) (Kevlar)

TABLE 9.2 Conventional Polymers, Part C

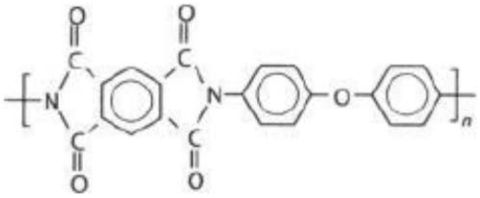
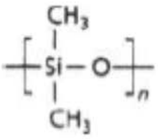
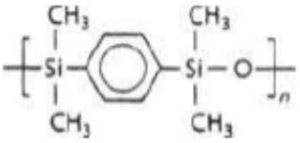
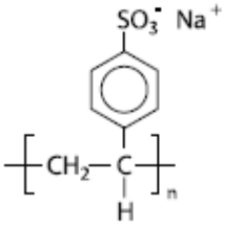
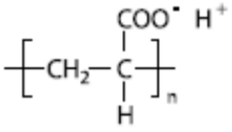
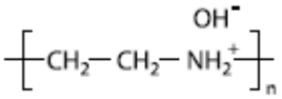
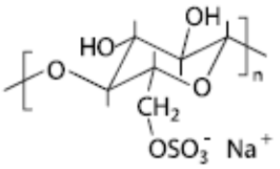
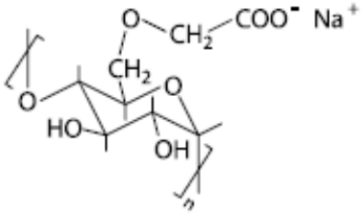
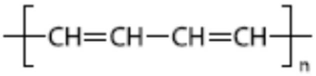
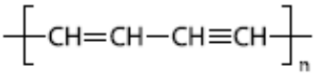
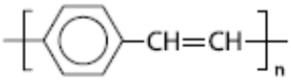
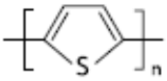
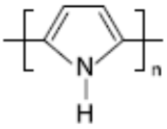
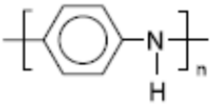
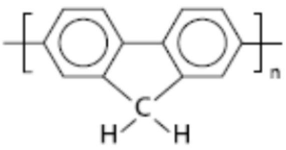
Structure	Name (abbrevation)
	Polyimide
	Poly(dimethylsiloxane) (PDMS)
	Poly(tetramethyl-*p*-silphenylene siloxane) (TMPS)

Table 9.3. A selection of polyelectrolytes

	sodium poly(styrene sulfonate) NaPSS
	poly(acrylic acid) PAA
	poly(ethylene imine)
	sodium cellulose sulfate
	sodium carboxymethyl amylose

Charges on a chain can also be created by doping processes. For *conjugated polymers*, i.e., chains with conjugated C–C double bonds, this is particularly easy. Even more importantly, the produced charges are mobile and thus provide electrical conductivity. Table 9.4 compiles some of these special materials.

Table 9.4. A selection of conjugated polymers

	polyacetylene PAc
	polydiacetylene PdAc
	poly(p-phenylenevinylene) PPhV
	poly(p-phenylene) PPh
	polythiophene PTh
	polypyrrole PPy
	polyaniline PAn
	polyfluorene

9.3 POLYMERIZATION

The number of monomer units in a macromolecule is referred to as the *degree of polymerization.* Typical degrees of polymerization are

Polycarbonate (PC): 80–160

Polyamide (PA) 6: 100–300

PVC: 800

Polystyrene (PS): 1500–4000

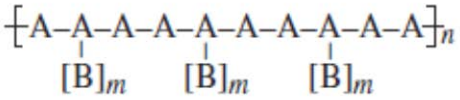
PP: 3500–35000

Ultrahigh Molecular Weight Polyethylene (PE–UHMW): 200,000

If only one type of monomer is involved in the polymerization reaction, the resulting product is a *homopolymer*:

$$\left[\text{A–A–A} \cdots \text{A–A–A} \right]_n$$

If two or more monomers are involved, the resulting product is a *copolymer*. For the coupling of the two monomeric units in the chain, two limited cases exist. In the first, the coupling is statistical and determined by the probabilities of attachment of the two monomers on the growing chain. Chains of this kind are called *statistical copolymers*. They exhibit a short-range order with preferred sequences, and thus are different from a *random mixing*, but possess no order in the chemical composition over the long range. *Block copolymers* are the second limiting case, which are formed by coupling long macromolecular sequences of uniform composition, and depending on the number of sequences, *di-, tri-, and multiblock copolymers* may be prepared. The remaining structures, which are most frequently represented, are *random and alternating copolymers*. The typical copolymer structures are shown below:

$\left[\text{A–B–B–A–A–B–A–B–A–A} \right]_n$	Random copolymer
$\left[\text{A–B–A–B–B} \cdots \text{A–B–A–A–B} \right]_n$	Statistical copolymer
$\left[(\text{A–A–A–A})_n \text{ –}(\text{B–B–B–B})_m \right]_n$	Block copolymer
$\left[\text{A–B–A–B–A–B–A–B–A–B} \right]_n$	Alternating copolymer
 $\left[\text{A–}\underset{[\text{B}]_m}{\text{A}}\text{–A–A–}\underset{[\text{B}]_m}{\text{A}}\text{–A–A–}\underset{[\text{B}]_m}{\text{A}}\text{–A–A} \right]_n$	Graft copolymer

Graft copolymers also consist of blocks of two different monomers, except that the main chain is composed of one type of monomer units and the branches of the other. The second monomer is polymerized onto the trunk of the chain or attached through reactive groups similar to grafting a branch of one tree onto the trunk of another.

9.4 POLYMER DEPOSITION TECHNOLOGIES

Whilst single crystals of polymers can sometimes be grown, the most common geometrical configuration of a polymer in an electronic device is a thin planar film. This results from the advanced planar microelectronics and integrated optics technologies that have grown up over the last 30 years. A number of techniques exist for the deposition of polymers in thin film form, the most well-known and well-used method being polymer *spin coating*.

Spin coating is used most often for depositing layers of photoresist as part of the patterning process in integrated circuit fabrication. This simple, yet effective, technique is illustrated in Figure 9.4. A relatively viscous polymer solution is placed onto the substrate to be coated which is then rotated at a fixed angular speed in the range 500-4000 rpm.

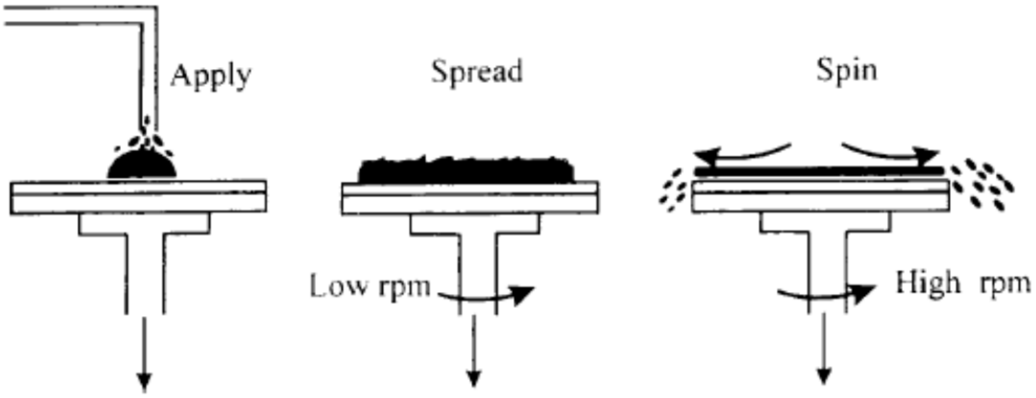


Figure 9.4 The polymer spin-coating process

The polymer solution flows radially outwards to form a thin solution layer that subsequently ‘sets’ as the solvent evaporates. The uniformity of the layer depends on a number of factors including the initial acceleration of the substrate and the rate of solvent evaporation, both of which can be easily controlled. The film thickness, *d*, depends on the solution viscosity, η, rotation speed ω, solution density ρ and spinning time *t* and is given by:

$$d = \{\eta / 4\pi \rho \omega^2 t\}^{1/2}$$

A favoured technique by those who require highly ordered or partially ordered polymer films is *Langmuir – Blodgett* (LB) deposition. An LB film is an ultrathin organic assembly formed by the sequential transfer of floating Langmuir layers from a water surface onto a solid substrate. In its conventional form, it comprises two-dimensional solid sheets of well-packed organic molecules deposited in a predetermined, controllable sequence. The first materials investigated by Pockels, Langmuir and Blodgett were generally rod-shaped amphiphilic molecules such as octadecanoic acid. There is today a huge variety in the shapes and types of molecules that can be deposited as LB films, many of them being polymers.

Before the LB film can be fabricated, a floating Langmuir film must first be created at an air-water interface. Octadecanoic acid (Figure 9.5a) is an amphiphilic molecule, which illustrates the LB process nicely. The polar carboxylic group is attracted to other polar substances such as water, whereas the long alkyl chain is hydrophobic and is repelled by water. This amphiphilicity is responsible for giving octadecanoic acid its unique orientational properties; the polar headgroup is effectively dissolved in water, yet the alkyl chain protrudes from the water surface in a near normal direction. The carboxyl functionality is so strong that many alkanoic acids will spontaneously spread from a bulk crystallite placed in contact with a water surface. Usually, however, the material is dissolved in a solvent such as chloroform, minute droplets (each containing ~2µl) of which are then dropped carefully on the water surface. The solution spreads over the available water surface rapidly, the solvent evaporating fully over a period of a few minutes, leaving behind the randomly distributed octadecanoic acid solute molecules. Indeed, even at this point in the procedure, a monolayer film exists, albeit an inhomogeneous one. Uniformity in the surface density and thickness of the monolayer is improved by now reducing the water surface area available to the floating solute molecules, thus compressing the initially expanded monolayer to form eventually a closed-packed, two-dimensional sheet of octadecanoic acid. Typically its molecular surface density in the compressed state is ~5 x 10^{14} cm^{-2} and its thickness is ~2.5 nm.

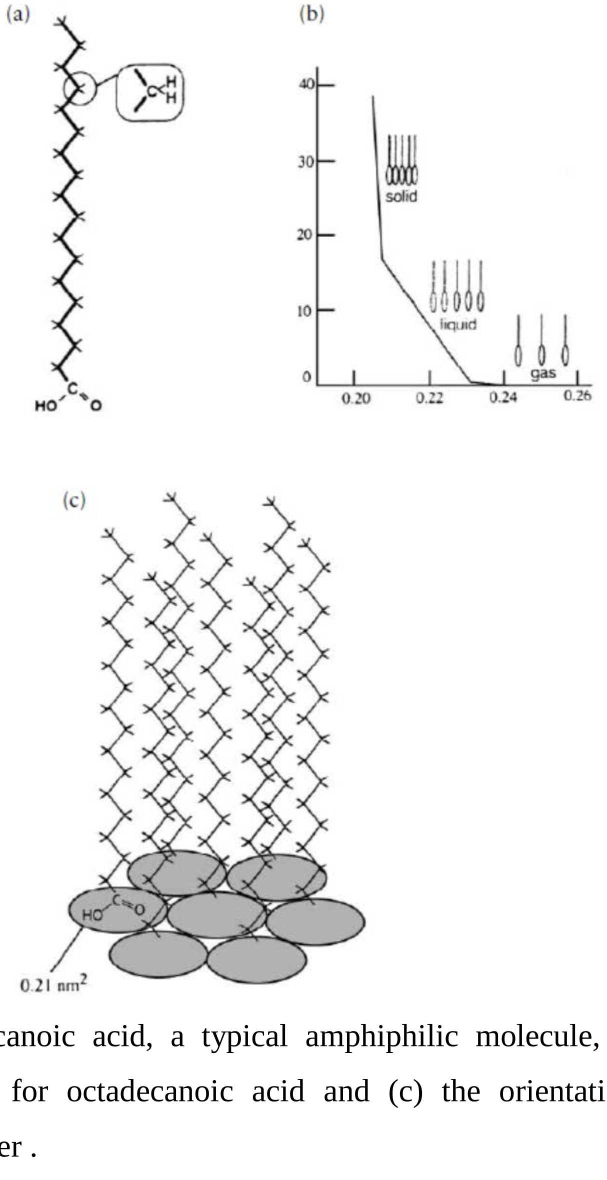


Figure 9.5 (a) Octadecanoic acid, a typical amphiphilic molecule, (b) a schematic surface pressure-area isotherm for octadecanoic acid and (c) the orientation of octadecanoic acid molecules in a monolayer .

A schematic diagram of a typical Langmuir trough is shown in Figure 9.5. The substrate insertion and withdrawal is normally achieved using a motor-driven micrometer screw to which is attached the clamp for the substrate. The sequential transfer process for a hydrophilic substrate withdrawn and inserted through an alkanoic acid monolayer is shown in Figure 9.6. The transfer of the first monolayer occurs due to the strong interaction between the carboxylic acid groups of the alkanoic acid molecules and polar sites (usually taking the form of free hydroxyl groups) on the hydrophilic substrate surface.

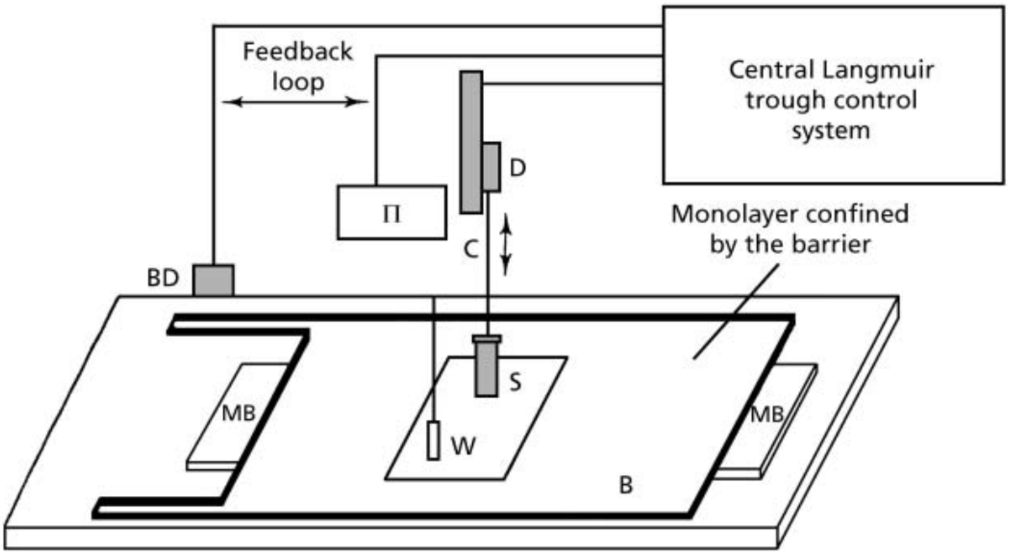


Figure 9.5. A typical Langmuir trough for the fabrication of LB films showing the surface pressure monitor (P), the wettable plate (W), the substrate (S) attached to the deposition mechanism (D) via the clamp (C), the moveable constant perimeter barriers (MB), the barrier drive motors (BD) and the water bath (B).

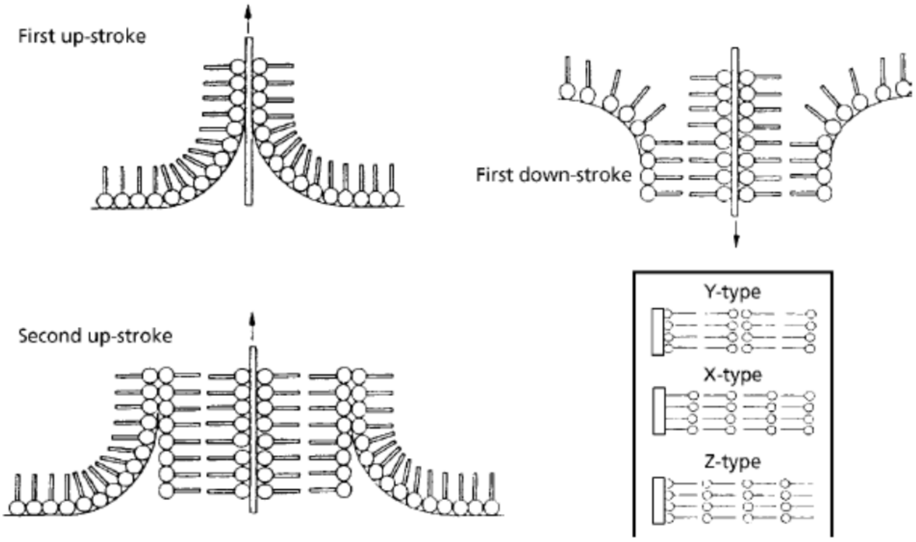


Figure 9.6. The sequential monolayer transfer process for the formation of LB films showing the formation of Y-type assemblies. Inset: Y, X and Z-type architectures .

Reinsertion of the monolayer-coated substrate results in the transfer of the second monolayer. There is a strong hydrophobic interaction between the alkyl chains within the first monolayer (the new substrate surface for the second layer) and those protruding from the water surface in the floating monolayer. A second layer is thus adsorbed and the surface of the coated substrate becomes hydrophilic again and is thus ready to accept the third monolayer upon its next withdrawal through the Langmuir film. This process can be repeated over many cycles and is referred to as Y-type deposition. Most rodshaped amphiphilic molecules follow this mode of deposition as do most pendant-chain polymers, like polysiloxanes. This deposition mode results essentially (except for the first monolayer) in a non centrosymmetric structure. Such assemblies are generally very stable over time owing to the strong stabilising interactions between hydrophobic chains from adjacent monolayers, and between hydrophilic carboxyl groups that can form hydrogen bonds across the interface between layers and in-plane sideways dimers that serve to strengthen the layers laterally.

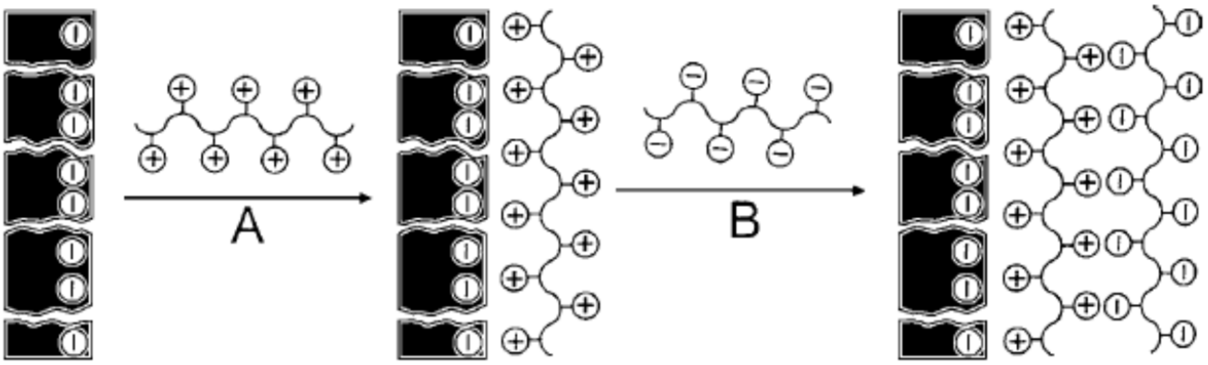


Figure 9.7 The layer-by-layer polyelectrolyte self-assembly process for charged polymers

Although the LB approach is applicable to an ever-increasing range of materials, water-soluble compounds are excluded. In a much lower technology, yet elegant, deposition process, water-soluble polymers can be deposited layer by layer to form similar molecular assemblies to LB films. This method, generally known as the polyelectrolyte layer-by-layer self-assembly process, makes use of the ionic attraction between opposite charges on anionic and cationic electrolytes. A solid substrate is positively charged by chemical cleaning methods and is placed in a solution of an anionic polyelectrolyte, such as poly (sodium 4-styrenesulfonate). The deposition process is depicted in Figure 9.7. During a period of around 10 minutes, a monolayer of the polymer is adsorbed onto the substrate surface via the electrostatic attraction.

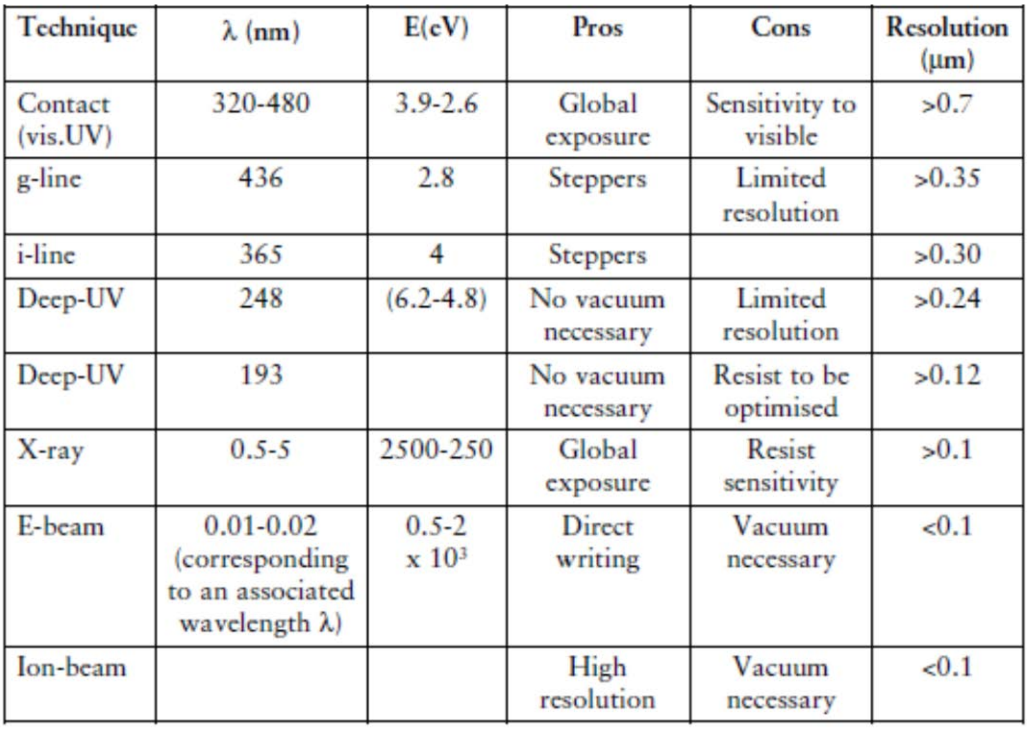
The process is self-regulating since a second monolayer of the anion cannot be adsorbed due to electrostatic repulsion since there are many unpaired anionic charges on the polymer which remain electrostatically unsatisfied. Therefore, after monolayer adsorption, the coated substrate is washed thoroughly in water and then placed in a solution of a cationic polyelectrolyte, such as poly(alylamine hydrochloride.

Again ionic attractions result in the adsorption of a cationic monolayer. This procedure can be repeated many times to build up alternate layer polymer films. The advantages of this method over LB deposition are that it is very easy to perform and applicable to water-soluble polymers; the drawback, however, is that the rate of deposition is limited to around 5 monolayers per hour compared to conventional LB deposition rates of 20 monolayers per hour. Recent developments in LB technology have led to much higher rates approaching 30 monolayers per minute.

9.5 PHOTOPOLYMERS AND PHOTORESISTS FOR MICROELECTRONICS

The formation of the desired patterns on the substrates requires the use of a resist coating in which the pattern is first generated. Depending on the wavelength of the irradiation, different minimum linewidth are obtained. The different techniques of lithography are summarised in Table 9.5. One can expect the development of deep-UV lithography techniques at 193 nm giving line dimensions as small as 0.12 µm, necessary for the latest VLSI.

Table 9.5 Microlithography techniques



Technique	λ (nm)	E(eV)	Pros	Cons	Resolution (µm)
Contact (vis.UV)	320-480	3.9-2.6	Global exposure	Sensitivity to visible	>0.7
g-line	436	2.8	Steppers	Limited resolution	>0.35
i-line	365	4	Steppers		>0.30
Deep-UV	248	(6.2-4.8)	No vacuum necessary	Limited resolution	>0.24
Deep-UV	193		No vacuum necessary	Resist to be optimised	>0.12
X-ray	0.5-5	2500-250	Global exposure	Resist sensitivity	>0.1
E-beam	0.01-0.02 (corresponding to an associated wavelength λ)	0.5-2 x 10^3	Direct writing	Vacuum necessary	<0.1
Ion-beam			High resolution	Vacuum necessary	<0.1

The resist layer, usually a polymeric film (0.5-1.0 µm thick) is spin-coated onto the substrates and then exposed to radiation. One of the key points for obtaining high resolution is to obtain polymers with sensitivities well adapted to the chosen type of radiation and with response speeds which allow high production rates. The main requirements of a resist are solubility, adhesion, etch resistance, sensitivity to radiation and contrast.

The source of radiation is generally a mercury lamp or a mercury-rare gas (xenon) discharge lamp. This gives a maximum of radiation in the 350-450 nm range. The steppers use classically a high-pressure mercury lamp with two radiations: i-line (436 nm) and g-line (365 nm). The recent development of excimer lasers emitting in the deep-UV region has introduced the use of deep-UV radiation (150-300 nm), and especially the 248 and 193 nm wavelengths.

Resists are generally deposited onto substrates by spin coating; therefore, solubility in organic solvents is necessary. The resist must possess good adhesion properties to various substrates, such as metal, silicon dioxide, silicon nitride or semiconductors, throughout the various steps of integrated circuit manufacture. Poor adhesion leads to loss of resolution. Wet chemical etching requires good adhesion and chemical stability of the resist towards acidic or basic etching solutions. Most dry etching techniques involve a high radiation flux and temperatures often higher than 80 °C. Polymers exhibiting high Tg values and containing radiation stable groups (e.g., aromatic structures) have higher dry etch resistance.

The sensitivity, σ, is related to the ability of a polymer to undergo structural modification on irradiation. Sensitivity is said to increase as the dose required to produce the lithographic image decreases.The sensitivity of a positive resist, $\sigma0$, is the dose required to achieve complete solubility of the exposed region under conditions where the unexposed region remains completely insoluble. The sensitivity of a negative resist is conventionally defined as the dose at which 70% of the original film thickness has been retained after development, $\sigma_{0,7}$. The required sensitivity varies with the type of irradiation and is expressed as energy/surface (e.g., J/m^2). The contrast, γ, is related to the ability of a polymer to give vertical sidewalls. Resolution (defined as the smallest linewidth which can be achieved) depends on the contrast. The parameters are determined from the sensitivity curve of the resist, which expresses the normalised film thickness, e_r/e_0 (er is the thickness after development and irradiation, e_0 is the initial thickness of the resist), as a function of $\log_{10}$ (Dose):

For a negative resist:

$$\gamma = (\log_{10} \sigma_1/\sigma_0)^{-1}$$

where σ_1 is the dose required to reach the required thickness after development

Positive Photoresists

All near-UV positive photoresists are two-component systems; the polymeric material is a low molecular weight novolac polymer (Figure 9.8) and the sensitiser is a derivative of a 1,2-diazonaphthoquinone (DNQ) (20%-50% by weight). DNQ forms a complex with the phenol groups of the novolac resin and prevents the dissolution of the latter in an aqueous base. Exposure of the resist to UV light results in photodecomposition of the sensitiser to an unstable ketocarbene. This reacts with water to produce the base-soluble indene carboxylic acid, which no longer inhibits dissolution of the novolac polymer in aqueous base (Figure 9.9).

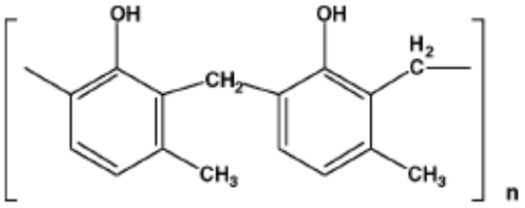


Figure 9.8 Novolac polymer

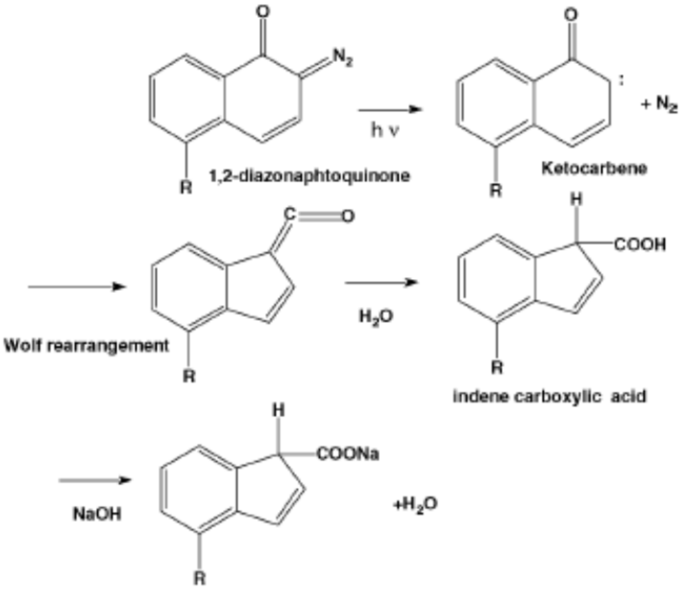


Figure 9.9 Photodecomposition of 1,2-diazonaphthoquinone

Negative Photoresists

In this type of photosensitive system, UV sensitive groups, such as chalcone, cinnamate, styrylacrylate, diphenyl-cyclopropene carboxylate, cinnamilidene malonate, *p*-carboxycinnamate, *p*-phenylene-bis acrylate and styrylpyridinium, etc., are included in the main chain or in the side chain of the polymer. UV irradiation gives rise to crosslinking. The crosslinking of poly(vinylcinnamate), which is the key process occurring in the KPR resist (Kodak), is shown in Figure 9.10.

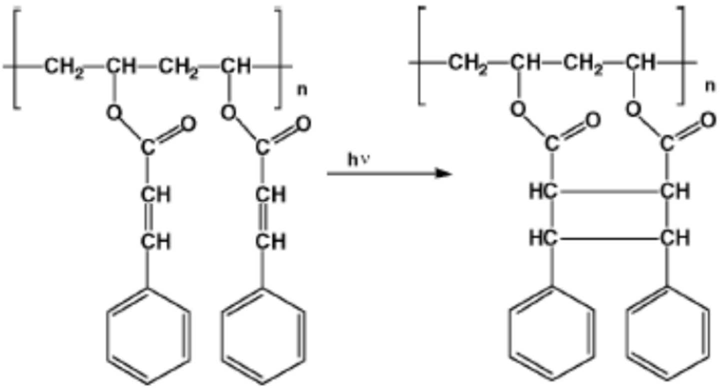
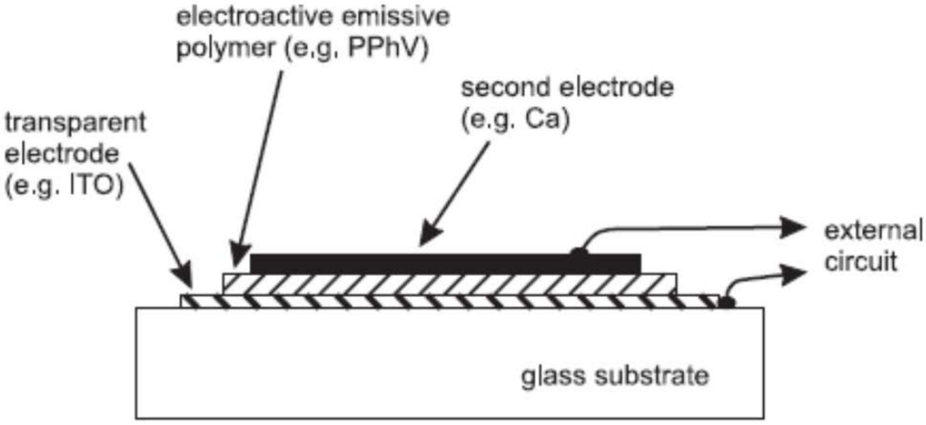


Figure 9.10 The KPR resist, a negative resist which crosslinks through the double bonds under irradiation.

9.6 CONJUGATED POLYMERS FOR OPTOELECTRONICS

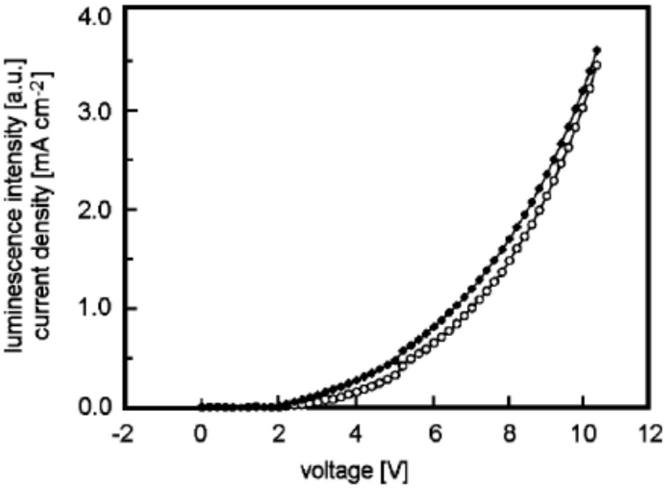
The large majority of polymers, first of all the broadly used commodity materials polyethylene, polypropylene, poly(ethylene terephthalate) or polystyrene, have similar electrical and optical properties. They are insulators and they are colorless, i.e., they possess no mobile charges and the lowest electronic excitations are in the UV region. There exists a peculiar class of polymers with quite different properties; these are polymers with conjugated double bonds in the main chain. They are semiconductors or conductors and interact with light. Figures 9.11 and 9.12 depict a characteristic observation.



Figures 9.11 Principal set-up of a polymer LED device

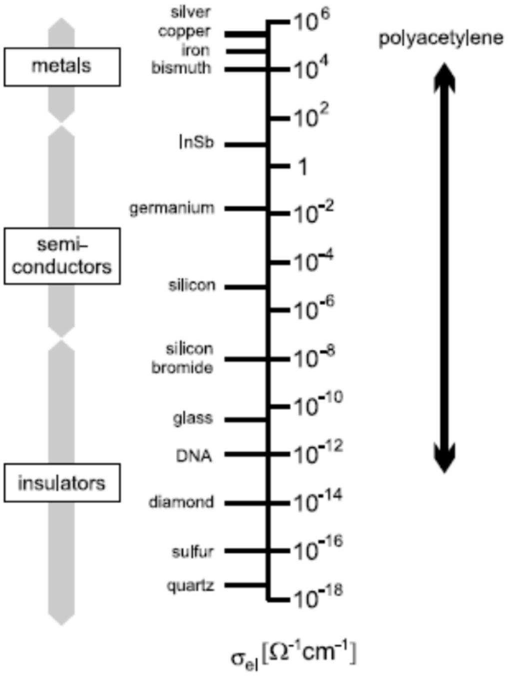
A thin film of poly-(phenylene vinylene) (PPhV) covered with two different electrodes made ofan indium-tin oxide alloy (ITO) and calcium, respectively, is placed on a glasssubstrate and then connected to an external source of current. Nothing happens up to a voltage of 2V, but then a flow of current sets in and grows with further rising voltage. Simultaneously with the onset of current the PPhV film begins to emit light, with an intensity proportional to the current. The effect is known as *electroluminescence*. A setup with such a performance is called a *light emitting diode*, abbreviated *LED*. The active element in this *organic light emitting diode*, shortly *OLED*, is the *conjugated polymer* poly(phenylene vinylene). All conjugated polymers possess conjugated π electrons, i.e., electrons that are delocalized rather than being part of one valence bond. Excitation energies of conjugated π electrons are usually in the visible range and they are, therefore, addressed as being optically active.

It was a second, really spectacular property that initiated the interest in conjugated polymers. In 1977 it was found that films of polyacetylene (PAc) increase their conductivity tremendously when they are exposed to iodine vapor, from a basic value at the lower end of the semiconducting range up to values comparable to metals. Figures 9.13 represents these iodine dopant induced changes in a comparison with other materials. They encompass the enormous range of seventeen orders of magnitude.

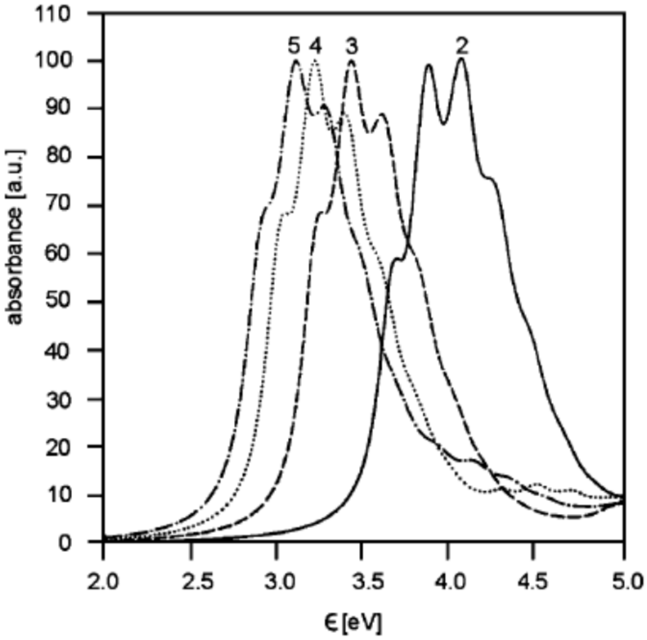


Figures 9.12**.** ITO/PPhV/Ca LED: Variation of the current density (*filled symbols*) and the luminescence intensity (*open symbols*) with the applied voltage

The figure 9.14 shows absorption spectra in the visible range, which were obtained for a series of oligomers of phenylene vinylene. The four presented absorption bands all have a similarly fine structure and are shifted against each other; the higher the number of monomers, the lower the frequency. A first conclusion can be drawn from the invariance of the profile of the bands. Since it is independent of the number of coupled monomers it must be a property of the monomer itself. We therefore first ask about the elementary optical excitations of a monomer.



Figures 9.13**.** Range of conductivities covered by doped PAc in a comparison with other materials



Figures 9.14**.** Optical absorption spectra of two-ring, three-ring, four-ring, and five-ring oligomers of phenylene vinylene. A shift to lower energies with increasing length of the molecule is observed.

Figures 9.15**.** presents itschemical structure on the left-hand side and on the right-hand side a sketch of the distribution of the *π electrons*. One bond of each double bonds is set up by two *σ* electrons, the other is formed from overlapping atomic *pz* orbitals oriented perpendicularly to the (*xy*-)plane of the planar molecular skeleton. The *conjugation*, i.e., the resonance interaction between the π bonds results in delocalized π electron states. Here, these states are occupied by the eight π electrons of a phenylene vinylene unit. There is a highest occupied molecular π orbital abbreviated *HOMO*. Above it there exists a *gap* and it extends up to next level, the lowest unoccupied molecular π orbital, shortly called *LUMO*.

Figures 9.15**.** (*left*) Structure unit of PPhV. (*right*) Clouds of π electrons in a unit, placed above and below the plane of the CC backbone. The *lines* connecting the carbon atoms represent the σ bonds

For conjugated polymers the gap energy is in the range of 1.5 eV to 3 eV, i.e., in the range of visible light and the near infrared, similar to anorganic semiconductors. As always in the ground state of a multielectron system with an even number of electrons, the spins compensate each other so that a singulet state, denoted S_0, results.

The absorption bands in Figures 9.14**.** are to be assigned to the lowest possible one-electron excitation. This is accomplished by a transfer of the electron at the HOMO level to the next higher state. The LUMO level would indeed be occupied if a ninth electron were added to the monomer. If, however, the electron in the HOMO level is excited, this level is left empty, and the then positively charged monomer exerts an attractive Coulomb force. The force reduces the excitation energy, which now falls below the gap energy. The thus formed state can be thought of as a coupled electron–hole pair and is named an *exciton*. In order to be optically active, the transition must have a non-vanishing transition dipole moment. One requirement for this is an unchanged total spin. Hence, this exciton state is necessarily a singulet state and is denoted S_1 (Figure 9.16).

In the introduction to this chapter a light-emitting diode was described as being set up of a polymer film (PPhV) and two different electrodes (ITO and Ca) on the surfaces (Fig. 9.11). Figure 9.12 then showed that for voltages above a critical value a current coupled to the emission of light flows. Figure 9.17 now displays the spectrum of the emitted light as well. The curve denoted EL represents this spectrum with a broad band with fine structure. The band is to be assigned to the vibronically broadened annihilation of an exciton. Also included in the figure is the *photoluminescence spectrum (PL)* observed after an excitation of poly(phenylene vinylene) by photons with energies above 2.6 eV. The two emission bands EL and PL are practically identical. Hence, both an electrical or optical excitation create the same excitons, which then annihilate under the emission of photons. The figure also shows the optical absorbance spectrum of poly(phenylene vinylene). It sets in at the frequency of the exciton and shows a strong rise at 2.4 eV. Here the photon energy reaches the value that is required to excite one electron from the HOMO to the LUMO level, i.e., to create a free electron and a free hole.

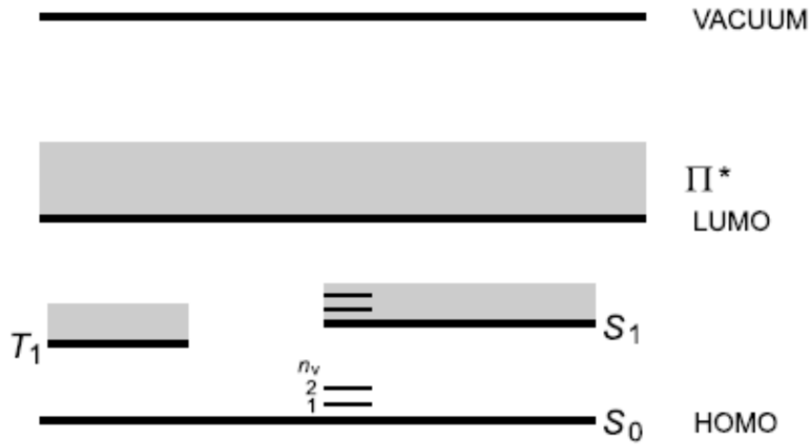


Figure 9.16 **.** One-electron levels and multielectron states created by excitations of one π electron in an ordered domain of a conjugated polymer. Prior to the excitation, in the ground state *S*0 the electron occupies the HOMO level. *S*1 denotes a first order singulet exciton with an energy at the lower bound of the associated vibronic band. The creation of an exciton is in general coupled to simultaneous vibrational excitations; *n*v are vibrational quantum numbers. *T*1 denotes a first order triplet exciton at the lower bound of an associated vibronic band. The LUMO level is at the lower edge of the band Π^* of states with one non-bonded electron. The VACUUM level is that of electrons with vanishing kinetic energy outside the sample.

A discussion of the functioning of a polymer LED can be based on the energy level scheme given in Fig. 9.18. It deals with the conditions in an LED onto which an external voltage is applied. Polymer LEDs are set up using two different electrodes. One of them, in the ITO example, has a high work function (as given by the energy that has to be supplied to transfer an electron from the Fermi level to the vacuum level). The second one, here calcium, has a low work function. When the electrodes are attached to the polymer film, *electrical double layers* spontaneously form at the interfaces; as always happens when two different conductive materials are brought in contact with each other. The layer on the metal side is very thin, the oppositely charged depletion layer near the surface of the polymer film has some extension. As a consequence, steps in the electrostatic potential *V* arise at both metal–polymer interfaces.

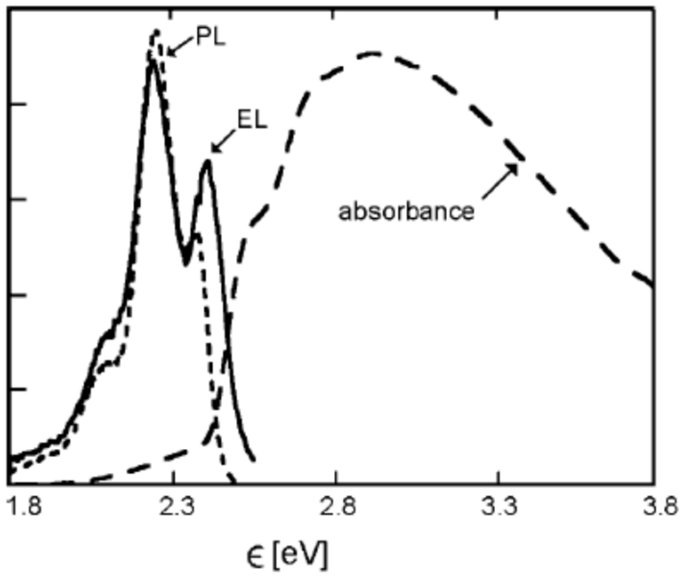


Figure 9.17**.** PPhV: Luminescence spectra after an electrical (EL) and a photonic (PL) excitation, shown together with the optical absorption spectrum.

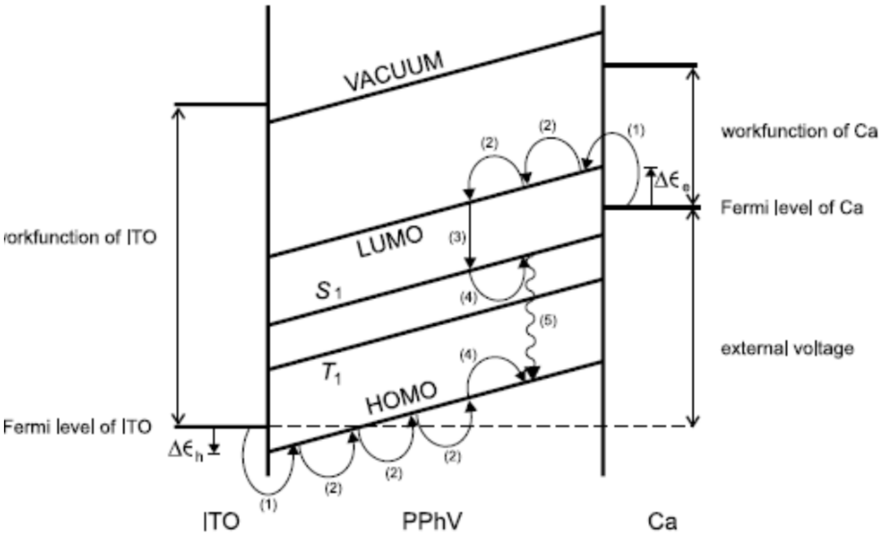


Figure 9.18**.** Energy level diagram for a ITO/PPhV/Ca LED under a voltage as given by the difference between the anode and cathode potential. Work functions of the two electrodes. Barriers to the injection of electrons, ΔE_e, and holes, ΔE_h.

The step heights have such values that in the initial equilibrium state, i.e., the state established before the application of an external voltage, the chemical potential of the electrons, which now includes also an electrostatic contribution (-eV) , becomes identical in all three materials. Applying an external voltage leads to deviations from the current-free equilibrium state, and a current starts to flow. If an LED is activated by an external voltage, from the cathode electrons change into the polymer film to the LUMO level of mobile free electrons and holes are simultaneously ejected by the anode and enter the HOMO level

So, the working mechanism of a LED consists of four main steps:

(1) By applying a bias voltage, charges of opposite signs are injected in the active material and form positive and negative polarons.

(2) The charges + (-) move in the material towards the negative (positive) electrode, driven by the applied electric field.

(3) Polarons of opposite sign couple to generate the excitons, *S*.

(4) S recombine radiatively by photon emission

The role of many factors are almost completely understood, but there is still a lot of work to do because improvements both in lifetime and in electroemission efficiencies are possible.

The structural control can be extended to several physicochemical parameters:

- The band gap of a conjugated polymer is responsible for the PL and EL peak position,
- The ionisation potential (IP), the electronic affinity (EA) of the polymer strongly affect the charge injections from the electrodes into the polymers, and this is related to the PL and EL efficiency,
- The solid state packing of the macromolecules influences the stability and the emission efficiencies,
- The surface polarity of the polymer is responsible for the adhesion between the active polymer and the electrodes, which is an important factor in charge injection, and resistance to oxidation and to temperature of the active polymeric layer are fundamental to the lifetime of the device.

Optimisation of all the above parameters is extremely difficult on a chemical synthesis basis, but the great efforts provided by many groups in this area have lead to exciting results for both improvements in macromolecular chemistry and in technological application.

*Polyphenylenes (*PPV) (Table 9.3) was one of the first polymers studied, for its good PL and EL efficiencies. Several synthetic routes have been reported for the preparation of this polymer. Its insolubility is a great problem in the preparation of high molecular weight materials. For example, step growth polymerisation, such as Wittig condensation between terephthaldicarboxaldehydes and arylene-bisphosphylidenes gives very low molecular weight polymers, because when the growing chain is formed by 6-10 repeating units, it becomes insoluble and chain growth stops. The low molecular weight polymers synthesised in this way cannot be used for thin film preparation because they are insoluble. To overcome these difficulties, a soluble form of a precursor was synthesised. A thin film of the precursor was formed by spin coating techniques; a subsequent thermal treatment of the film lead to the insoluble conjugated thin film of PPV. In Figure 9.19 the procedure followed to prepare PPV is reported. Several modifications of this general procedure have been introduced, consisting of changing the conditions of the transformation process and/or the nature of the chemical species to be eliminated. Another procedure, consists of the polymerisation of dichloro-*p*-xylene with potassium tert-butoxide in organic solvents.

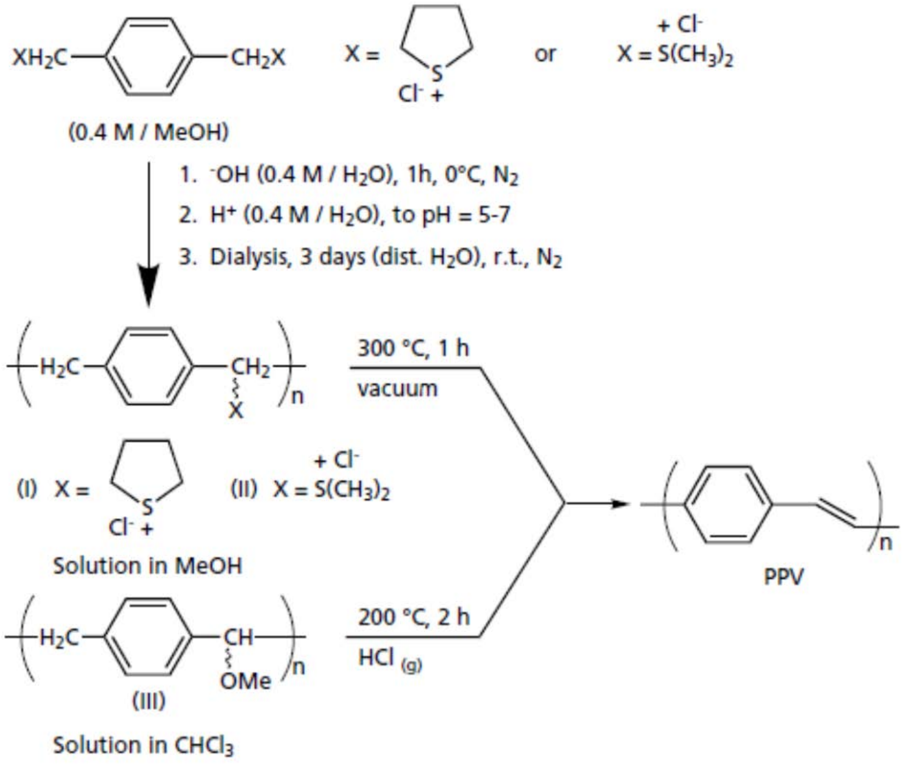


Figure 9.19 Synthetic procedure for PPV

*Polyfluorenes, Poly(phenylene ethynylenes), Polythiophenes (*see tabl.9.3) also used for LED**.** The using of a series of polymeric structures it is possible to prepare LEDs covering all the visible range (Figure 9.20).

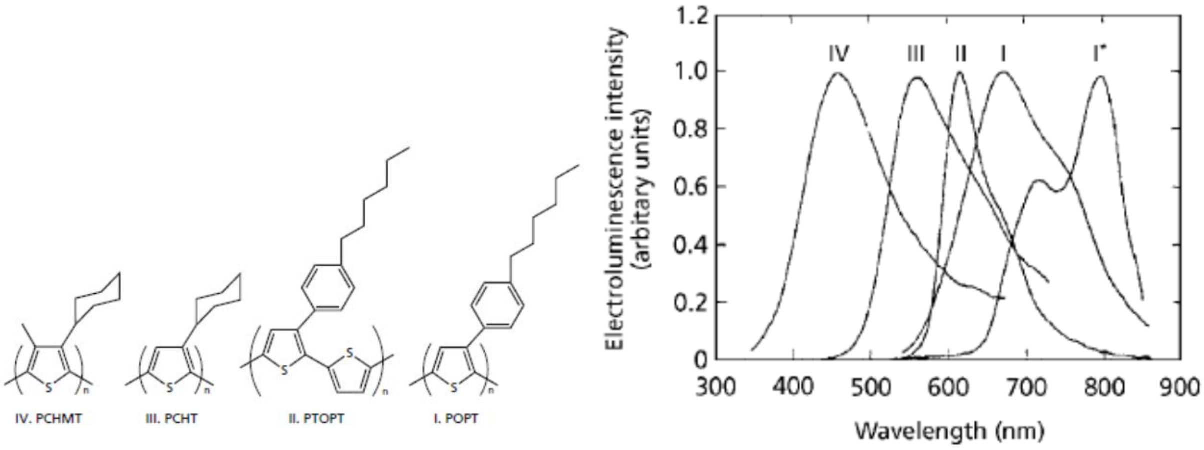


Figure 9.20 Polymeric and copolymeric structures emitting from 450 nm to 800 nm

9.7 CONDUCTING POLYMERS IN MOLECULAR ELECTRONICS

Molecular electronics (ME) is so named because it uses molecules to function as ‘switches‘ and ‘wires’. ME is a term that refers both to the use of molecular materials in electronics and to electronics at molecular level. It is as yet not very clear how molecular electronic devices will operate, but it is conjectured that active molecules are needed, either in isolation or becoming active by association with other molecules. It is thought that electronics is likely to imitate some of the basic functions of macroscopic devices such as memories, sensors and logic circuits.

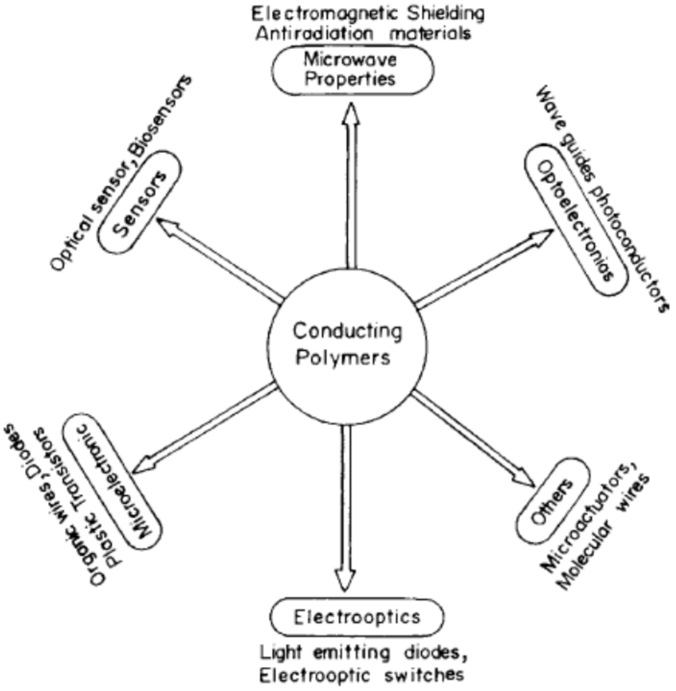


Figure 9.21 Applications of conducting polymers in molecular electronics

Organic molecules such as conducting polymers, proteins and pigments are being considered as alternatives for carrying out the same functions that are presently performed by semiconductors (e.g., silicon) and metals. Among them, conducting polymers (or conjugated polymers) have been considered as highly promising for molecular electronics. These conducting polymers offer a unique combination of properties that make them attractive materials for use in molecular devices (Figure 9.21). The conductivity of these polymers can be tuned by chemical manipulation of the polymer backbone, by the nature of the dopant, by the degree of doping and by blending with other polymers. In addition they offer lightweight, processibility and flexibility. Because of these advantages, the use of conducting polymers in molecular electronics is rapidly evolving from physics, chemistry, biology, electronics and information technology. These molecular electronic materials differ from conventional polymers by having a delocalised electronic structure that can accommodate charge carriers such as electrons and holes.

Besides this, these organic materials exhibit Peierl's instabilities due to the built-in high anisotropic interactions and undergo substantial geometric modifications due to electronic excitations. This results in various charge transfer processes and a substantial degree of disorder leading to various localised states in the forbidden gap due to localisation. Conducting polymers exhibit the behaviour of both metal and semiconductor. Polyacetylene, the first and simplest electrically conducting polymer, doped with iodine (I_2) has been shown to have an electrical conductivity of 1000 S cm^{-1}. Some of the conducting polymers that have recently generated much interest are shown in Figure 9.22. It can be seen that conducting polymers behave as anisotropic semiconductors with band gaps in the range 1.4-3.2 eV.

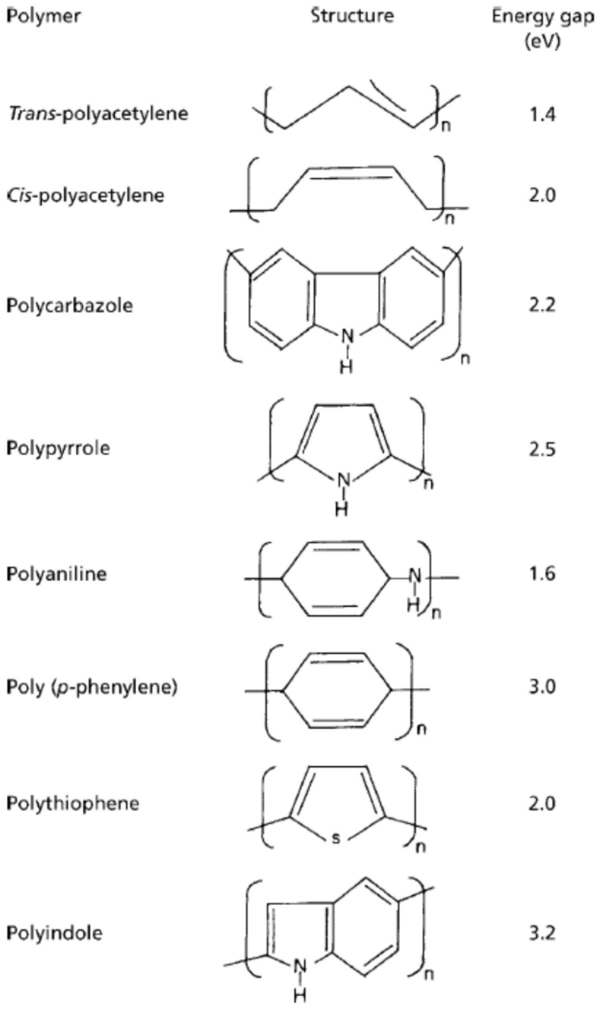


Figure 9.22 Structures and band gaps of some important conducting polymers

Characterisation of a conducting polymer is important before it can be used for any technological application. The morphology of conducting polymer films has been investigated using optical microscopy, SEM/TEM, scanning tunneling microscopy (STM) and AFM techniques Studies conducted on the morphology of conducting polypyrrole films containing different anions such as NO_3^-, F^-, ClO_4^-,BF_4^- and $CH_3C_6H_4SO_3^-$, respectively, have revealed that the topology of the growing conducting polypyrrole surface is influenced by the nature of the electrolyte. For instance, it has been shown that the fibrillar structure of polyacetylene is often advantageous since it can store up to about 7% of electrical charge. AFM and STM methods have recently provided valuable information on the presence of microdomains in poly(3-hexylthiophene)/stearic acid films (Figure 9.23).

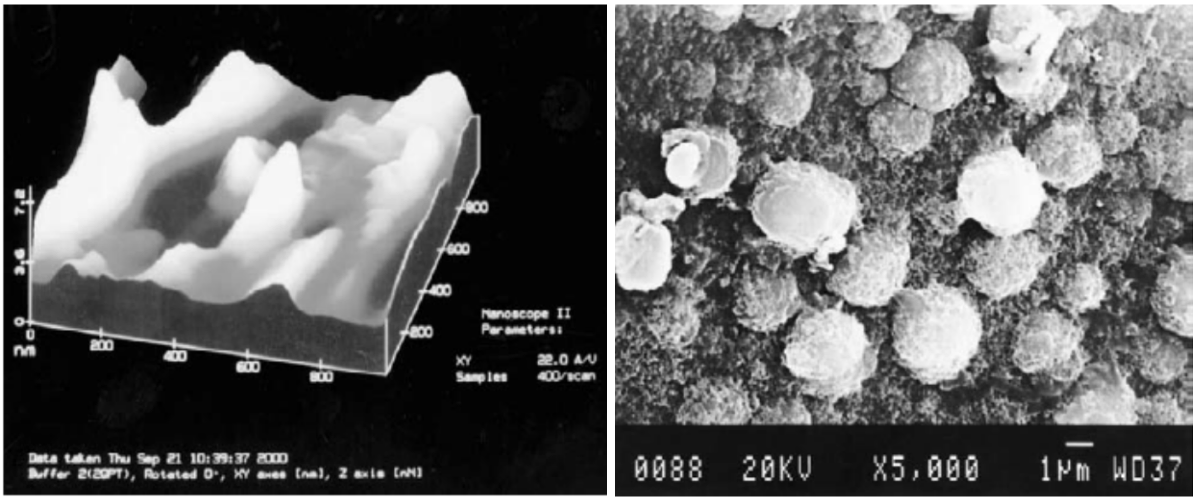


Figure 9.23 Atomic force micrograph of conducting poly(3-hexylthiophene) (left) and Scanning electron micrograph of conducting poly(3-hexylthiophene) (right)

There has been an increased interest towards the possible applications of conducting polymers as the active elements in electronics. The characteristics of conducting polymer/inorganic semiconductor interfaces have been considered as very important since it has been indicated that restrictions on Schottky barrier devices can be overcome by using conducting polymer contact layers. Besides this, the ability to manipulate the interface characteristics by changing the polymer dopant allows switchable devices to be fabricated. Semiconducting polymers such as polyacetylene, polypyrrole and polyaniline have recently been used for the fabrication of Schottky barrier diodes, like metal-insulatorsemiconductor (MIS) diodes and p-n junction diodes . A schematic diagram for a Schottky device is shown in Figure 9.24.

Schottky diodes formed between metallic AsF_5^- doped $(CH)_X$ and n-type GaAs indicate high electronegativity. The polyacetylene has been found to exhibit p-type behaviour when it is used for fabrication of Schottky diodes with low work function metals. Heterojunctions have been fabricated using electrochemically prepared polypyrrole and metal (indium, titanium, aluminium and tin). It has been revealed that carrier concentration, estimated as 1.5×10^{20} and 5×10^{17} cm^{-3} in doped and undoped polypyrrole samples, respectively, plays an important role in controlling the junction characteristics and hence the performance of these Schottky devices. The electrical characteristics of the junctions have been found to be dependent on the work function of polypyrrole, estimated as 4.42-4.49 eV.

It has been recently revealed that it is possible to fabricate all vacuum deposited metal (Pb, Al, In, Sn)/polyaniline/metal Schottky devices. It has been shown that the barrier height and the ideality factors determined are dependent on the work function of the metal used in the fabrication of these devices. The improved ideality factor obtained as 1.2 for an Al/polyaniline/Ag device has been attributed to more intimate contact of the metal with the vacuum deposited polyaniline electrode.

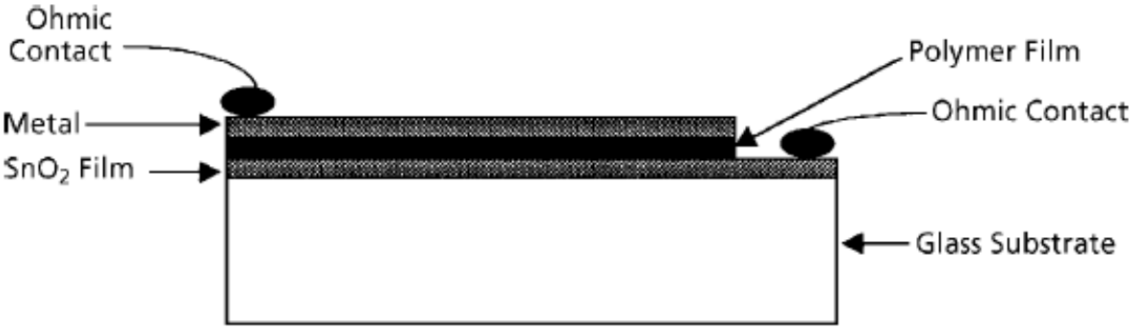


Figure 9.24 Schematic diagram of a conducting poly(3-alkylthiophene) based Schottky device

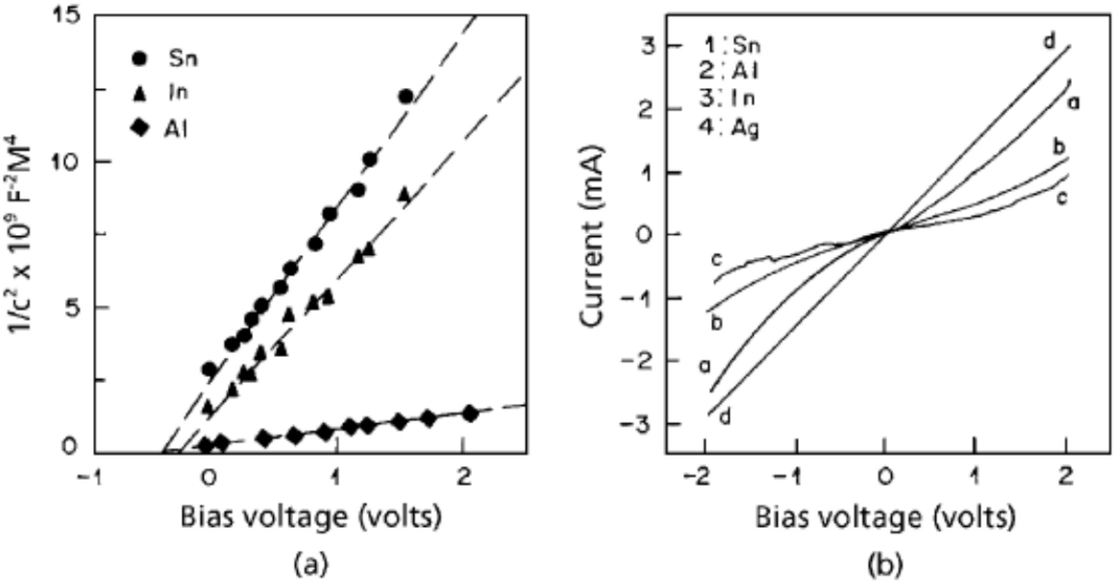


Figure 9.25 Current-voltage and capacitance-voltage characteristics of metal/conducting polymer junctions

Electrical properties of the poly(3-cyclohexylthiophene)/metal junctions were compared with those of the poly(3-*n*-hexylthiophene)/metal junctions in Figure 9.25.

9.8 POLYMER FOR MICROACTUATORS

Actuators that generate movements and forces, such as bending, expansion and contraction driven by stimulation of electrical, chemical, thermal and optical energies, are different from rotating machines such as electric motors and internal combustion engines. There are many sorts of soft actuators made of polymers, gels and nanotubes. Particularly, biomimetic actuators are interesting because of the application to artificial muscles that will be demanded for medical equipment, robotics and replacement of human muscle in the future.

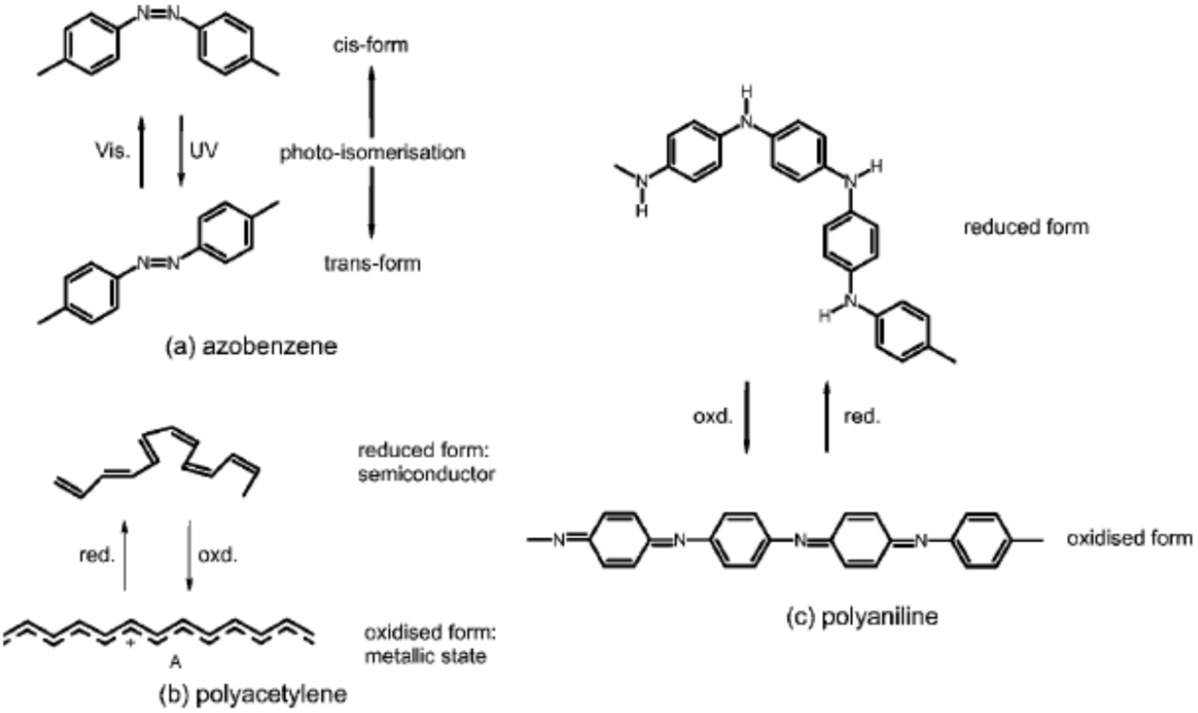


Figure 9.26 Conformational changes of molecular structure, (a) photoisomerisation of azobenzene, (b) and (c) extension and contraction of polyacetylene and polyaniline, respectively, upon oxidation and reduction

The driving force originates from the conformational change of peptide molecules by the chemical energy cycles. Similarly, various sorts of stimulating energies can change the conformation of molecules. For example, the *cis-trans* photoisomerisation in azobenzene, as shown in Figure 9.26 a, is a well-known phenomenon. The *cis* form converts to the longer *trans* form upon illumination by UV light, and the *trans* form reverts to the *cis* form upon illumination by visible light. The electrical conductivity in polyacetylene (Figure 9.26 b), the representative of conducting polymers, dramatically increases from insulator to conductor upon chemical or electrochemical oxidation. This results from the delocalisation of π electrons and in the change of polymer conformation. At the pristine stage (or in the reduced state), the polyacetylene is flexible because of the single bond in the bond alternation. The single bond has more freedom in the rotational and bending modes than that of the double bond. In the oxidised state, the bond alternation is reduced and the molecular structure becomes more planar than that of the reduced state. Similarly, polyaniline, shown in Figure 9.26 c, changes its bond alternation from the benzenoid form to quinoid form upon oxidation. This results in the deformation of polymer conformation.

In fact, films, fibres or blocks of conducting polymer expand and contract upon electrochemical oxidation and reduction, respectively. This process is tentatively named ‘electrolytic deformation’ or ‘electrolytic expansion’. The mechanisms of electrolytic deformation have been classified by three principal mechanisms: (1) the insertion and removal of bulky ions, (2) conformational change of polymer structure due to the delocalisation of π electrons, and (3) electrostatic repulsion between likely charged polycations (polarons and/or bipolarons). In mechanism (1), the insertion is induced by the neutralisation of polymers for oxidation and reduction.

The magnitude of expansion or contraction depends on the volume of ions and cannot be larger than the total volume of the inserted ions. For mechanism (2), the expansion of the polymer is associated with an expanding spring and depends on the morphology of polymer structure. The electroexpansion of hydrogels is explained by mechanism (3). To accomplish the larger expansion ratio in the electrolytic deformation of conducting polymers, it is desirable to utilise mechanisms (2) and (3).

Conducting polymers are prepared by either chemical or electrochemical oxidation of monomers, like pyrrole, thiophene and aniline. For the measurement of electrolytic deformation, it may be preferable to use soluble polymers, such as polyaniline (soluble in *N*-methyl-2-pyrrolidinone (NMP)). Others are hardly soluble in usual organic solvents. However, polymers with substituted long alkyl chains have been found to be soluble in organic solvents. Usually, the conducting polymers prepared by the electrochemical methods are obtained as thin films. Such conducting polymers are, however, difficult to process. The polyaniline prepared by the chemical oxidation is a powder and the base form is named emeraldine base and is soluble in NMP. The emeraldine film can be prepared by casting the NMP solution containing its concentration of 2-10 wt.% on a glass plate.

The cast film obtained by this method can be stretched mechanically to more than 3 times the original length. The films have been examined along the stretch direction, the perpendicular direction and also the thickness direction to investigate the anisotropic behaviour. For a measurement of electrolytic deformation, a bimorph actuator has been fabricated, and the ratio of expansion is estimated from the bent curvature of the actuators. This method is effective for qualitative measurement and demonstration, since the expansion is tremendously magnified. Even with an expansion rate at the level of 1%, the bending is clearly observed.

Electrochemistry and Expansion Behaviour in Polyaniline Film.

Polyaniline in an aqueous acid solution takes three typical redox stages depending on the degree of oxidation, as shown by the cyclic voltammogram curve in Figure 9.27. The half redox potential, *E*1/2 is defined as *E*1/2 = (*E*a + *E*c)/2, where *E*a and *E*c are anodic and cathodic potentials at the peak currents for oxidation and reduction, respectively.

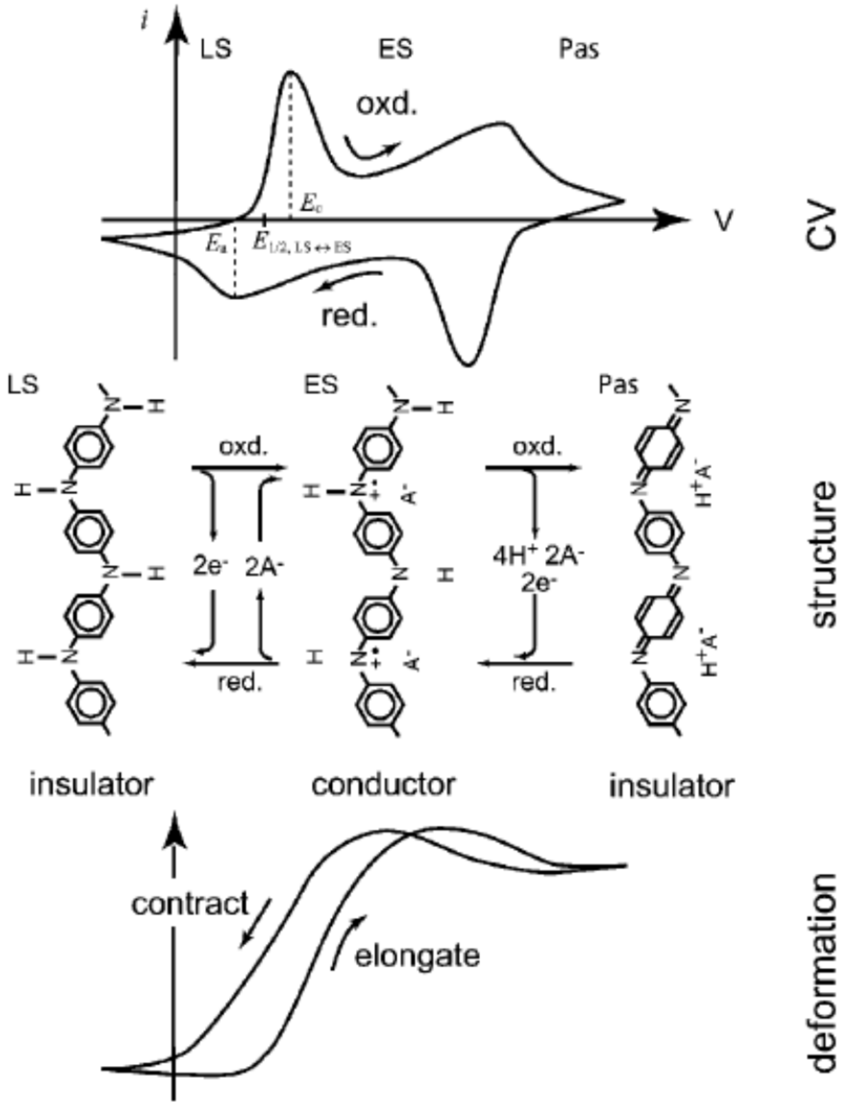


Figure 9.27 Typical cyclic voltammogram (upper), redox behavior in chemical structures (middle) and expansion and contraction of polyaniline film along the stretched direction (lower). *Ea* and *Ec* are the anodic peak and cathode peak, respectively, *E*1/2, LS-ES = 1/2 (*Ea* + *Ec*).

The most electrically conductive state is the emeraldine salt (ES) that is between pernigraniline salt (Pas) at the high potential side and the leuco-emeraldine salt (LS) at the low potential side. The LS and the Pas are the most reduced and oxidised states, respectively, and have been found to be insulating. In an oxidation process from the LS to the ES, two electrons are withdrawn and two chloride ions are doped for every four benzene units. For the oxidation from the ES to the Pas, two electrons are withdrawn and two protons are released. The LS/ ES reaction is reversible, however, the hydrolysis occurs at the higher oxidised Pas states. The bottom of Figure 9.27 shows a typical expansion behaviour in a polyaniline film along the stretched direction. By the potential sweep from the LS to higher potentials, the film starts to expand and shows the maximum expansion at the ES. Then the film contracts slightly. When the potential was returned from the Pas state, the film expands slightly, and then contracts below the potential of $E1/2$, LS↔ES and returns to the original length. This extension and contraction behaviour is very similar to the weight change measured by a quartz crystal microbalance . The result indicates that the electrolytic expansion closely relates to the insertion and exclusion of dopant ions in the film.

The unstretched film and the stretched film perpendicular to the stretched direction show monotonous expansion during the oxidation LS→ES→Pas and vice versa . The result possibly indicates that the dopant ions settle between the polymer chains. In the expansion of the film perpendicular to the stretched direction, irreversible expansion, namely, a creeping effect, was observed even under light load during the redox cycles. Actuators fabricated by conducting polymers are soft, flexible, and lightweight, with low voltage drive and strong contraction force. Tweezers, microvalves, and directors of optical fibre are some of the technological applications. Two types of bimorph actuators have been proposed . One is the backbone type, which consists of two conducting polymer films stuck together on a double-sided adhesive tape, which can also be replaced by a solid polymer electrolyte. The other is the shell type, in which conducting polymer films stuck on adhesive tapes are sandwiched onto a sheet of electrolyte media. The shell-type actuator is self-standing and works in air. There are some advantages in the use of these bimorph structures. One of these relates to the oxidation and the reduction of the polymer film process resulting in the bending force doubling. Since dopants transfer from one film to the other film through the electrolyte media, the electrolyte can be minimised. The electrolytic actuator can also be used both for positioning and rechargeable battery.

9.9 POLYMERS FOR ANOTHER APPLICATIONS

Conjugated polymers are used also for preparation of Polymer Field-Effect Transistors, Solar Cells; Liquid-Crystal Polymers are used for Liquid-Crystal Displays, Special Connectors, Fiber-Optic Cables; Ionomers are used for Fuel Cells. Polymers also are used in Lithium Ion Polymer Batteries, in different Sensors.

There are the solid polymeric materials in various forms, ‘commodity polymers’ for widespread applications, as well as speciality polymers for specific utilizations. In industry these are called thermoplasts, which expresses that they can be shaped and brought into forms of choice by thermomechanical treatments at elevated temperatures. Polymer fibers comprise a second large class of materials and are mostly used for the production of textiles and woven products. Fibers are generally obtained by spinning processes carried out on the melt or concentrated solutions at elevated temperatures, which is followed by a fixing accomplished by rapid cooling. Again, the temperature range for uses is limited. If a fiber is heated to a too high temperature it shrinks. Rubbers, technically addressed as *elastomers*, constitute the third class of polymeric materials. Both synthetic and natural products are utilized. The essential step in rubber production is the cross-linking process. ‘Natural rubber’, for example, is obtained by heating cis-polyisoprene in the presence of sulfur. This *vulcanization process* creates cross-links between the polyisoprene chains, composed of short sequences of sulfur atoms. For high cross-link densities the large deformability characteristic of a rubber is lost, and one obtains stiff solids. This is know as the class of *duromers or thermosets*, also known as *resins*. Various adhesives based on the mixing of two reacting components belong to this class of polymers. The shapes of these compounds are rather stable and remain unaffected by heating, up to the point of chemical decomposition.

Chapter 10. Piezoelectric Materials and Applications

10.1. Intoduction

Mechanical energy is one of the most ubiquitous energies that can be reused in our surroundings. The sources of mechanical energy can be a vibrating structure, a moving object, and vibration induced by flowing air or water. The energies related to induced vibrations or movement by flow of air and water at large-scale are wind energy and hydroelectric energy, respectively, which are not within the scope of this chapter. Instead, the mechanical energies here can be classified as so-called ‘‘low-level’’ vibrations and movements (Fig.10.1) [1]. Such potential ‘‘low-level’’ vibrations and movements are summarized in Table 10.1 [2] and Table 10.2 [3].

Mechanical waste energies usually can be harvested by using vibration-toelectricity conversion [3–5]. The most distinguished characteristic of this kind of waste energy harvesting is initially identified for low power generations. Therefore, one of the targeted applications is to power small electronic devices. However, recent development indicates that it can also be used for large-scale applications [6]. Vibration-to-electricity conversion can be realized through three basic mechanisms, including electromagnetic [7 – 9], electrostatic [10], and piezoelectric [11, 12] transductions.

Fig. 10.1. Power and energy scale for piezoelectrics, photovoltaics, and conventional power.

Among the three mechanisms, piezoelectric transduction has received the greatest attention. This is because piezoelectric materials have larger power densities and higher feasibility for practical applications than the materials used in the other two mechanisms [4]. For example, voltage outputs in electromagnetic energy harvesting are typically very low and thus must be amplified to a level sufficiently high to charge storage devices. In contrast, however, piezoelectric energy harvesters output voltages that can be used directly. In electrostatic energy harvesting, the materials should be subject to an external applied voltage to trigger the relative vibratory motion of the capacitor elements, which outputs alternative electrical currents [10]. Such external applied voltages are required in piezoelectric energy harvesting. Another advantage over electromagnetic devices is that piezoelectric harvesting devices can be fabricated at both macro-scale and micro-scale, due to the well-established deposition techniques for thick-films and thin-films piezoelectric materials [11, 13]. Comparatively, it is hard to fabricate electromagnetic energy harvesters at micro-scale, due to the poor properties of planar magnets and the limited number of turns that can be realized when using planar coils [4].

Table 10.1. Sources of mechanical energy around us each and every day that can be harvested for electricity [2].

Human body/motion	Transportation	Infrastructure	Industry	Environment
Breathing, blood flow/pressure, exhalation, walking, arm motion, finger motion, jogging, talking	Aircraft, automobile, train, tires, tracks, peddles, brakes, turbine engine, vibration, noises	Bridges, roads, tunnels, farm, house structure, control-switch, water/gas pipes, AC system	Motors, compressor, chillers, pumps, fans, vibrations, cutting and dicing, noise	Wing, ocean current/ wave, acoustic wave

Table 10.2. Acceleration (m s^{-2}) magnitude and frequency of vibration mode of potential vibration sources [3].

Vibration source	a (m s^{-2})	f_{peak} (Hz)
Car engine compartment	12.0	200
Base of 3-axis machine tool	10.0	70
Blender casing	6.4	121
Clothes dryer	3.5	121
Person nervously tapping their heel	3.0	1
Car instrument panel	3.0	13
Door frame just after door closes	3.0	125

Small microwave oven	2.5	121
HVAC vents in office building	0.2–1.5	60
Windows next to a busy road	0.7	100
CD on notebook computer	0.6	75
Second story floor of busy office	0.2	100

Research and development in dielectric, piezoelectric and ferroelectric materials has advanced at an unimaginable rate in the past decade, driven mainly by three factors:

(*i*) new forms of materials have been prepared in a multitude of sizes (e.g. buck single crystals and ceramics, nano-structured films, tubes, wires and particles) by a variety of techniques (e.g. melt growth, physical deposition and to soft chemical synthesis);

(*ii*) novel and intricate structural and physical properties have been discovered in these materials (e.g. morphotropic phase boundary and related phenomena, domain engineering under fields, effects of nanostructures and superlattices);

(*iii*) the extraordinary potential offered by these materials to be used in the fabrication of a wide range of high-performance devices (such as sensors, actuators, medical ultrasonic transducers, micro electromechanical systems (MEMS), microwave tuners, ferroelectric non-volatile random access memories (FeRAM), electro-optical modulators, etc.).

In this Chapter, harvesting mechanical waste energy with piezoelectric effect will be also considered. The main content will be started with a detailed description of piezoelectrics, including history of piezoelectric materials, principle of piezoelectric effect, and types and preparation of piezoelectric materials. After that, the progress in the development of different devices, including mechanical waste energy harvesters, based on piezoelectric effect will be summarized. Potential mechanical waste energy sources, including vehicle suspensions, civil structures, rail way tracks, roads, ocean waves, streams, and human motions, have been explored for respective potential applications.

10.2. Hystory of Piezoelectrics

The history of piezoelectricity – as a physical phenomenon to be used intentionally – goes back to the beginning of the 18th century. At that time the Dutch brought a precious stone called *tourmaline* from the East Indies to Europe. Tourmaline had a peculiar feature: while being heated, the material attracted other materials such as ashes. Almost half a century later the Swedish botanist and physician Carl Linneaus – also famous as developer of biological nomenclature – had a hunch that this phenomenon might have something to do with electricity. And indeed, within a decade, the German physicist Franz Aepinus confirmed that this ‘peculiar feature’ was electric. This phenomenon was later known as *pyroelectricity*: the ability of a material to generate a temporary voltage when it is being cooled or heated. Pyroelectricity lead to the discovery of piezoelectricity, which was to a large extent a French affair. When Charles-Augustin de Coulomb assumed that electric charge might be produced by pressure, René-Just Haüy and later also Antoine César Becquerel tried to apply their knowledge of pyroelectricity to perform experiments to investigate Coulomb's assumption. However, they were not very successful.

In 1880 the brothers Pierre and Jacques Curie finally discovered the direct piezoelectric effect, also with pyroelectricity as a basis. They observed that by pressing in a certain direction on crystals of tourmaline, quartz, cane sugar and Rochelle salt (also known as Seignette's salt), these crystals were able to generate charge on certain positions of their surfaces. The German physicist Wilhelm G. Hankel gave this phenomenon the name 'piezoelectricity' – named after the ancient Greek *piezein* that means to press or to squeeze, and *elektron* meaning amber, describing substances that (like amber) attract other substances when rubbed. One year later, the French-Luxembourgian physicist Gabriel Lippmann predicted the inverse piezoelectric effect, which was experimentally verified by the Curie brothers in that same year.

10.2.1. Brief History of Modern Piezoelectric Ceramics

The history of modern piezoelectricity and piezoelectric ceramics is close related to the history of ferroelectricity and ferroelectric ceramics. The history of ferroelectrics can be tracked back to Rochelle salt (sodium potassium tartrate tetrahydrate, $KNa(C_4H_4O_6)\cdot 4H_2O$), which was synthesized more than 400 years ago, initially for medicinal purposes [14 - 18]. It is in this same crystalline material that pyroelectric (thermal-polar), piezoelectric (stress-polar), and ferroelectric were discovered subsequently. Before this discovery, ferroelectricity was only a hypothetical property of solid materials at the turn of the twentieth century. However, the practical application of this material is largely limited due to its water solubility. It was after the discovery of ferroelectric ceramics (barium titanate, $BaTiO_3$), this class of materials became extremely useful for a variety applications.

The first ferroelectric ceramic material is barium titanate ($BaTiO_3$ or BT), which was discovered in the mid-1940s [14, 15, 17, 18]. Before the discovery of $BaTiO_3$, the most widely used materials for capacitors were steatite, mica, TiO_2, $MgTiO_3$, and $CaTiO_3$, with dielectric constant of not higher than 100. During the World War II, there was pressing needs for high dielectric constant materials to fabricate high capacitance capacitors. Before publication was available in the literature, $BaTiO_3$ had already been studied as a high dielectric constant material concurrently. In the later open publications, it was concluded that the source of the high dielectric constant in BaTiO3 is due to its ferroelectric properties [14].

The history of ferroelectric ceramics also includes the report of lead zirconatetitanate ($PbZr_{1-x}Ti_xO_3$, or PZT) piezoelectric ceramics, the development of transparent electro-optical lead lanthanum zirconate titanate ($Pb_{1-x}La_xZr_{1-y}Ti_yO_3$, or PLZT), the research on lead magnesium niobate ($PbMg_{1/3}Nb_{2/3}O_3$, or PMN) relaxor ferroelectric ceramics and the discovery of many other nonperovskite ferroelectric ceramics [14, 17]. Among these, PZT has been demonstrated to possess best performances as piezoelectric ceramics.

Recently, there has been a concern with PZT, due to the toxicity of Pb. Regulations and legislations have been established globally to restrict the use of leadcontaining materials. For example, according to the directive for the Restriction of the use of certain Hazardous Substances in electrical and electronic equipment (RoHS) adopted by the European Parliament in the year 2006, the maximum allowed concentration of lead is established to be 0.1 wt% in homogeneous materials for electrical and electronic equipment used in households as well as industry. Therefore, lead-containing piezoeletric materials will be prohibited eventually. Similar regulations have been established worldwide. As a consequence, there is significantly increasing interest in developing lead-free piezoelectric ceramics all around the world. Although major progress has been made in materials research [19, 20], there are still obstacles blocking the successful industrial implementation of lead-free piezoelectric ceramics. One key problem is their poor piezoelectric properties and fatigue degradation problem. Therefore, PZT will still be dominant materials for piezoelectric applications at least in the near future.

10.2.2. Practical Applications

So far the ‘founding fathers’ and the scientific history. This interesting phenomenon of materials that can convert mechanical energy into electric energy – and vice versa – cried out for practical applications. And they came. During World War I, in 1917 Paul Langevin developed the predecessor of *sonar*, a device to detect other objects under water. He managed to make a *quartz*-based transducer to send ultrasonic waves, and a receiver to detect the returning echo. By measuring the time span between the emitted wave and the wave that returned after bouncing off an object, a submarine should be able to determine the distance to that object. To date, sonar is still a major application of piezoelectric technology, where modern (ceramic) materials are being used.

Sonar as a successful application of a piezoelectricmaterial stimulated others to discover new piezoelectric materials and to develop new devices. An eye-catching example is the use of Rochelle salt as a single crystal needle in the pick-up part of early electronic phonographs, starting in 1935. Around 1950 Rochelle salt was replaced by *piezoelectric ceramics*, and in turn they were replaced by magnetic cartridges in the 1970’s. Fromthen, it took about ten years before compact disc players massively replaced phonographs.

The second World War had a large influence on the development of new piezoelectric materials. Independent from each other – due to World War II – Japanese, Soviet Union and American research groups discovered the so-called *ferroelectrics*, a new kind of man-made materials with much better piezoelectric properties than their natural counterparts. *Barium titanate* ceramics were the first materials in this series.

One significant example of the use of piezoelectric crystals was developed by Bell Telephone Laboratories. Following World War I, Frederick R. Lack, working in radio telephony in the engineering department, developed the "AT cut" crystal, a crystal that operated through a wide range of temperatures. Lack's crystal didn't need the heavy accessories previous crystal used, facilitating its use on aircraft. This development allowed Allied air forces to engage in coordinated mass attacks through the use of aviation radio.

Development of piezoelectric devices and materials in the United States was kept within the companies doing the development, mostly due to the wartime beginnings of the field, and in the interests of securing profitable patents. New materials were the first to be developed — quartz crystals were the first commercially exploited piezoelectric material, but scientists searched for higher-performance materials. Despite the advances in materials and the maturation of manufacturing processes, the United States market did not grow as quickly as Japan's did. Without many new applications, the growth of the United States' piezoelectric industry suffered.

In contrast, Japanese manufacturers shared their information, quickly overcoming technical and manufacturing challenges and creating new markets. Japanese efforts in materials research created piezoceramic materials competitive to the U.S. materials but free of expensive patent restrictions. Major Japanese piezoelectric developments included new designs of piezoceramic filters for radios and televisions, piezo buzzers and audio transducers that can connect directly to electronic circuits, and the piezoelectric igniter, which generates sparks for small engine ignition systems (and gas-grill lighter) by compressing a ceramic disc. Ultrasonic transducers that transmit sound waves through air had existed for quite some time but first saw major commercial use in early television remote controls. These transducers now are mounted on several car models as an echolocation device, helping the driver determine the distance from the rear of the car to any objects that may be in its path.

10.3. Piezoelectricity

The word ‘‘piezoelectricity’’ is derived from the Greek ‘‘piezein’’, which means to ‘‘squeeze’’ or ‘‘press’’ [14–17]. There are two piezoelectric effects: direct effect and converse effect. The direct effect (designated as a generator) is identified with the phenomenon whereby electrical charge (polarization) is generated from a mechanical stress, whereas the converse effect (designated as a motor) is associated with the mechanical movement generated by the application of an electrical field. Therefore, piezoelectric energy harvesting is to use the direct effect (generator). Properties of piezoelectric materials are generally characterized by k_p, k_{33}, d_{33}, d_{31}, and g_{33}.

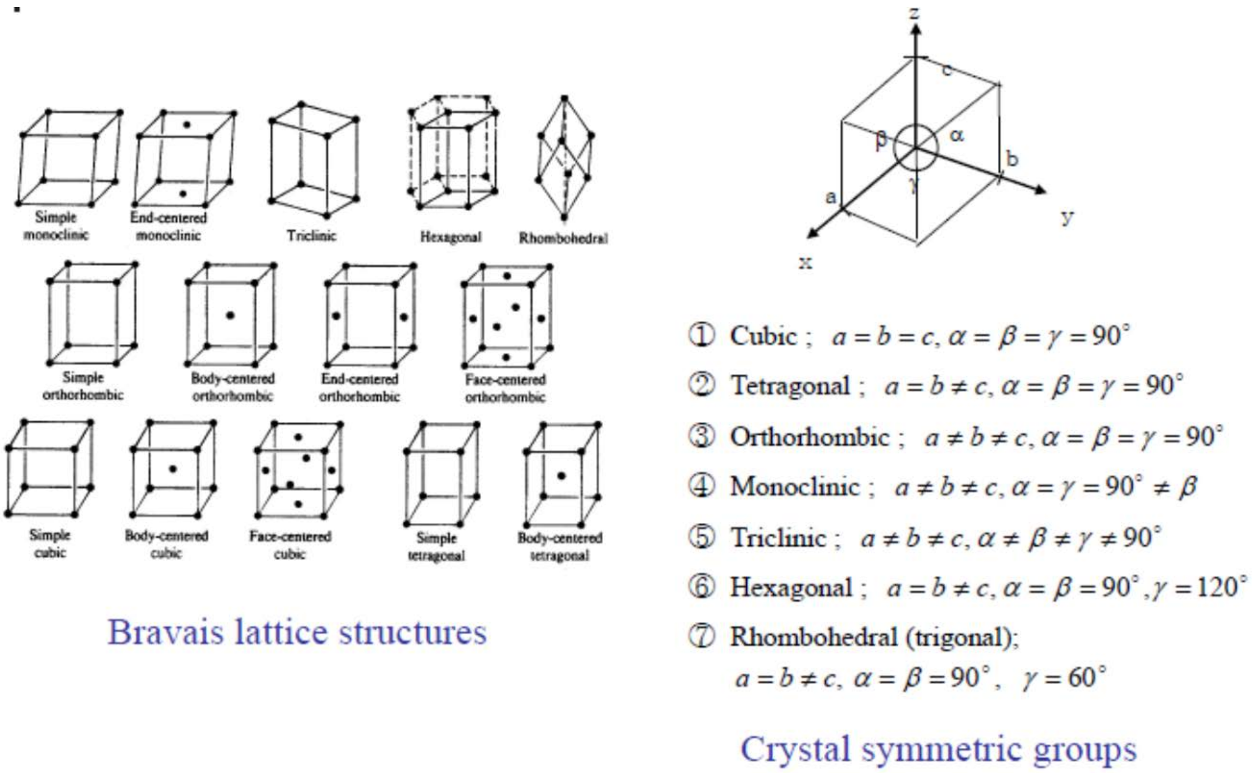
The k factors (e.g., k_{33}, k_{31}, and k_p), which are also called piezoelectric coupling factors, are convenient and direct measurements of the overall strength of the electromechanical effects, i.e., the ability of the ceramic transducer to convert one form of energy to another. They are defined as the square root of the ratio of energy output in electrical form to the total mechanical energy input (direct effect), or the square root of the ratio of the energy available in mechanical form to the total electrical energy input (converse effect). Because the conversion of electrical to mechanical energy (or vice versa) is always incomplete, k is always less than unity. Commonly used as a figure-of-merit for piezoelectrics, the higher k values are most desirable and constantly sought after in new materials. For ceramics, k_p is a typical measure used to compare piezoelectric properties of ferroelectric materials-values ranging from 0.35 for $BaTiO_3$ to as high as 0.72 for PLZT [14].

The d coefficients are called piezoelectric coefficients, having magnitudes of $\times 10^{-12}$ C N^{-1} (or pC N^{-1}) for the direct effect and $\times 10^{-12}$ m V^{-1} (or pm V^{-1}) for the converse effect, respectively. Subscript is used to describe the relative direction of inputs and outputs. For example, d_{31} means that this piezoelectric coefficient relates to the generation of polarization (direct effect) in the electrodes perpendicular to the vertical direction (3) and to the stress mechanically applied in the lateral direction (1), while d_{33} indicates the polarization generated in the vertical direction (3) when the stress is applied in the same direction. There are also other similar symbols [14].

g factors are called open-circuit coefficients, another parameters used to evaluate piezoelectric ceramics for their ability to generate large amounts of voltage per unit of input stress. The g constant is related to d constant: $g = d / K\varepsilon_0$ (K is relative dielectric constant and ε_0 is the dielectric constant of free space). High-g-constant piezoelectric ceramics are usually ferroelectrically hard materials whose polarizations are not readily switched and thus they possess lower K values.

Piezoelectricity is usually discussed together with ferroelectricity and pyroelectricity, because they have interesting inter-relationships in terms of crystal structures. All crystals can be categorized into 32 different classes. In the theory of point groups, these classes are determined by using several symmetry elements: (*i*) center of symmetry, (*ii*) axis of rotation, (*iii*) mirror planes, and (*iv*) several combinations of them. The 32 point groups are subdivisions of seven basic crystal systems that are, in order of ascending symmetry, triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral (trigonal), hexagonal, and cubic (Fig. 10.2). 21 classes out of the 32 point groups are noncentrosymmetric, which is a necessary condition for piezoelectricity to exist. 20 of them are piezoelectric. Of the 20 piezoelectric crystal classes, 10 crystals are of pyroelectric properties.

Within a given temperature range, this group of materials is permanently polarized. Compared to the general piezoelectric polarization produced under stress, the pyroelectric polarization is developed spontaneously and kept as permanent dipoles in the structure. Because this polarization varies with temperature, the response is termed as pyroelectricity. Within the pyroelectric group, there is a subgroup that has spontaneous polarization, which is called ferroelectric materials. On one hand, the polarization in a ferroelectric material is similar to the polarization in a pyroelectric one. On the other hand, there is difference between the two polarizations because the ferroelectric polarization is reversible by an external applied electric field, provided that the applied field is less than the dielectric breakdown of the materials. Therefore, materials that can be defined as ferroelectrics must have two characteristics: the presence of spontaneous polarization and reversibility of the polarization under electric field [14, 15]. Figure 10.3 shows their inter-relationship, together with general dielectrics. This inter-relationship is important because ferroelectric materials have best piezoelectric properties than nonferroelectric materials (Fig.10.4).



Bravais lattice structures

① Cubic ; $a = b = c, \alpha = \beta = \gamma = 90^\circ$

② Tetragonal ; $a = b \neq c, \alpha = \beta = \gamma = 90^\circ$

③ Orthorhombic ; $a \neq b \neq c, \alpha = \beta = \gamma = 90^\circ$

④ Monoclinic ; $a \neq b \neq c, \alpha = \gamma = 90^\circ \neq \beta$

⑤ Triclinic ; $a \neq b \neq c, \alpha \neq \beta \neq \gamma \neq 90^\circ$

⑥ Hexagonal ; $a = b \neq c, \alpha = \beta = 90^\circ, \gamma = 120^\circ$

⑦ Rhombohedral (trigonal); $a = b \neq c,\ \alpha = \beta = 90^\circ,\ \gamma = 60^\circ$

Crystal symmetric groups

Fig. 10.2. Crystal lattice structure and symmetric groups.

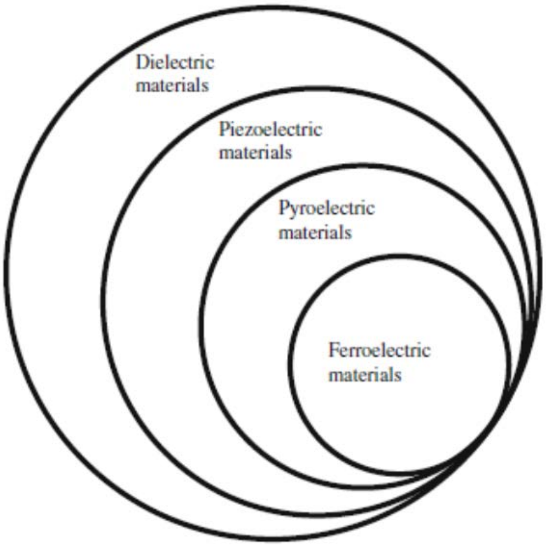


Fig. 10.3. Inter-relationship among piezoelectric, ferroelectric, pyroelectric and dielectric materials. Ferroelectric materials have superior pyroelectric, piezoelectric and dielectric properties than nonferroelectric materials.

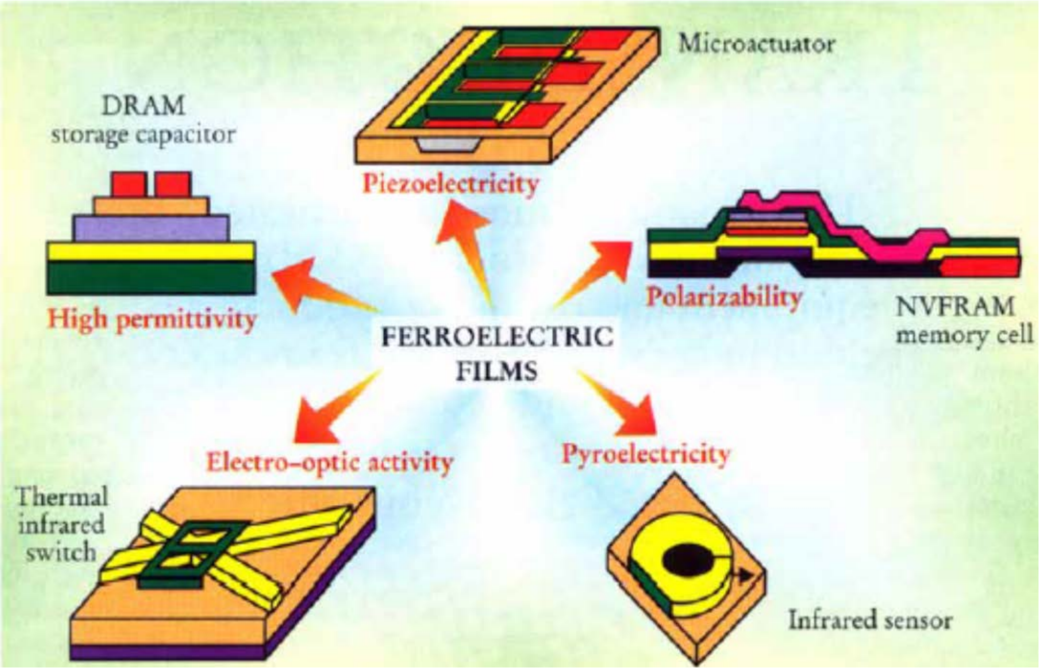


Fig. 10.4. Possible application of the ferroelectric films.

Ferroelectric materials can be grouped into four subcategories according to their crystal structures: perovskite group, pyrochlore group, tungsten-bronze group, and bismuth layer structure group, among which the perovskite group is the most important and thus the most widely studied. Perovskite is usually expressed as ABO_3. A typical ABO_3 unit-cell structure is shown in Fig. 10.5, taking $PbTiO_3$ as an example [14, 15, 17]. It consists of a corner-linked network of oxygen octahedra, creating an octahedral cage (B-site) and the interstices (A-sites). Ti^{4+} ions occupy the B-site while Pb^{2+} ions occupy the A-site. Figure 10.5 also shows the paraelectric and ferroelectric states of $PbTiO_3$. Most ferroelectric materials undergo a structural phase transition from a high temperature paraelectric phase into a low temperature ferroelectric phase. The paraelectric phase always has a higher symmetry than the ferroelectric phase. The temperature of the phase transition is called as the Curie temperature (T_C). Different ferroelectric materials have different values of T_C, which can be either lower than liquid nitrogen (LN) temperature or higher than 1,000 °C. For a given material (composition), the T_C is closely related to microstructure (grain size and distribution, density, porosity, and pore size and distribution, and impurity, and so on). Generally, the T_C of a given material decreases with decreasing grain size [21].

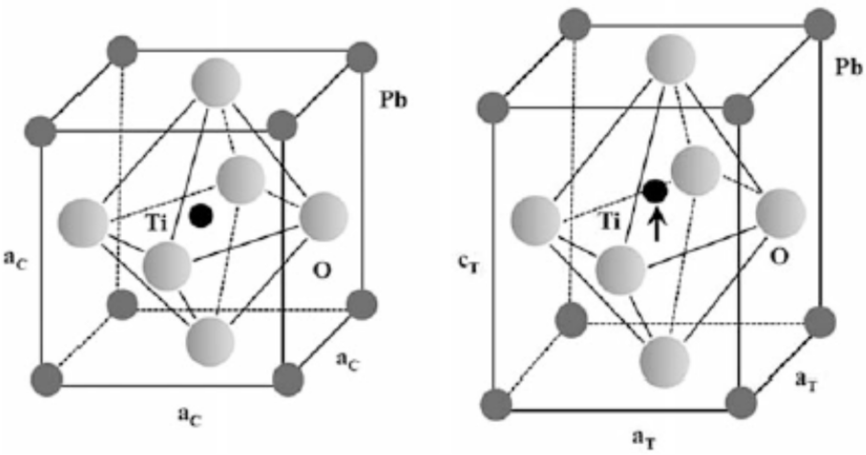


Fig. 10.5. Schematic perovskite structure of $PbTiO_3$, with cubic (C) structure in the paraelectric state ($P_s = 0$) and tetragonal (T) structure in the ferroelectric state ($P_s \neq 0$).

In the ferroelectric state, the ability of displacement of the central Ti^{4+} ion is the cause for the reversibility of polarization. The switch of many adjacent unit cells is referred to as domain reorientation or switching. The homogeneous areas of the material with the same polarization orientation are referred to as domains, with domain walls existing between areas of unlike polarization orientation [2, 3]. For as-prepared piezoelectric ceramics, the domains are randomly oriented and thus the net polarization of materials is zero because of their cancellation effect (Fig. 10.6). Therefore, the as-prepared piezoelectric ceramics have no piezoelectric effect. To show piezoelectric properties, the newly obtained polycrystalline piezoelectric ceramics must be poled at strong external DC electric fields (10–100 kV cm^{-1}). Poling is to electrically align the orientation of the randomly distributed domains within the grains and to make ferroelectric ceramics act like a single crystal possessing both ferroelectric and piezoelectric properties. Poling is usually conducted at elevated temperatures, because polarization would be more compliant at high temperatures [14, 15]. After poling during cooling, the applied voltages should be remained until the temperature is sufficiently low.

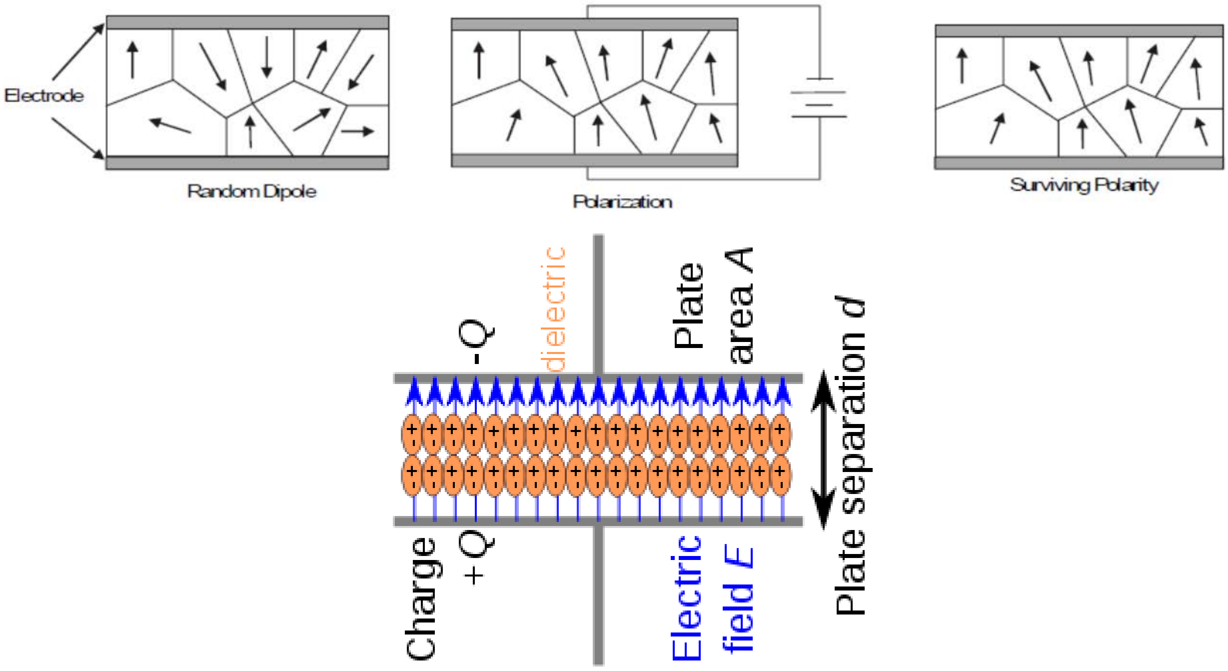


Fig. 10.6. Polarization of ceramic material to generate piezoelectric effect.

10.3.1. Piezoelectric Direct and Converse Effects

The domains of the piezoelectric ceramic element are aligned by the poling process. In the poling process the piezoelectric ceramic element is subjected to a strong DC electric field, usually at temperature slightly below the Curie temperature. When a poled piezoelectric ceramic is mechanically strained it becomes electrically polarized, producing an electrical charge on the surface of the materials (direct piezoelectric effect), piezoelectric sensors work on the basis of this particular property. The electrodes attached on the surface of the piezoelectric material helps to collect electric charge generated and to apply the electric field to the piezoelectric element.

When an electric field is applied to the poled piezoelectric ceramic through electrodes on its surfaces, the piezoelectric material gets strained (converse effect). The converse effect property is used for actuator purposes. Figure 10.7 shows the converse piezoelectric effect.

Based on the converse and direct effects, a piezoelectric material can act as a transducer to convert mechanical to electrical or electrical to mechanical energy. When piezoelectric transducer converts the electrical energy to mechanical energy it is called as piezo-motor/ actuator, and when it converts the mechanical energy to electrical energy it is called as piezo-generator/ sensor. The sensing and the actuation capabilities of the piezoelectric materials depend mostly on the coupling coefficient, the direction of the polarization, and on the charge coefficients (d_{31} and d_{33}). Figure 10.8 in the form of block diagrams shows the transducer characteristics of the piezoelectric materials.

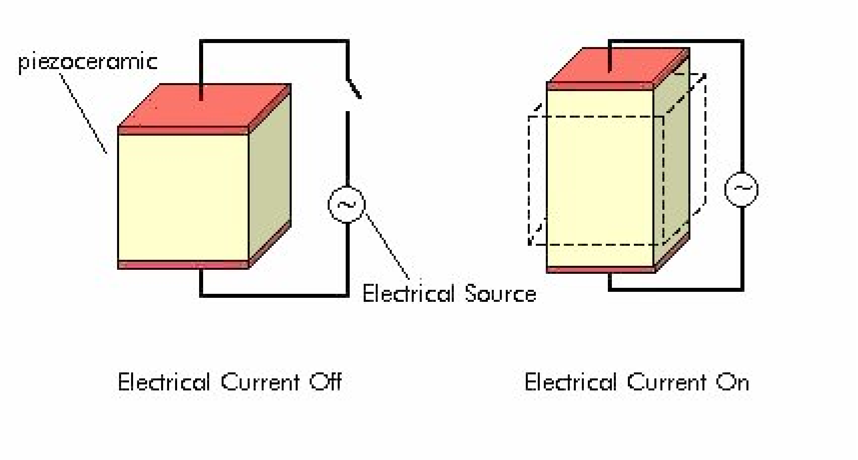


Fig. 10.7. Piezoelectric material.

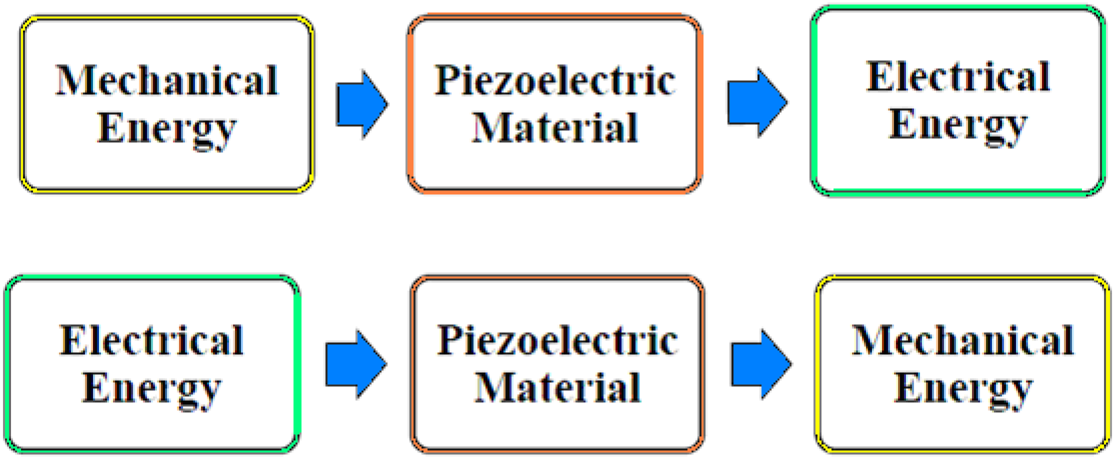


Fig. 10.8. Piezoelectric transducer.

10.3.2. Piesoelectric Effect Basics

A piezoelectric substance is one that produces an electric charge when a mechanical stress is applied (the substance is squeezed or stretched). Conversely, a mechanical deformation (the substance shrinks or expands) is produced when an electric field is applied. This effect is formed in crystals that have no center of symmetry. To explain this, we have to look at the individual molecules that make up the crystal. Each molecule has a polarization, one end is more negatively charged and the other end is positively charged, and is called a dipole. This is a result of the atoms that make up the molecule and the way the molecules are shaped. The polar axis is an imaginary line that runs through the center of both charges on the molecule. In a monocrystal the polar axes of all of the dipoles lie in one direction (Fig. 10.9). The crystal is said to be symmetrical because if you were to cut the crystal at any point, the resultant polar axes of the two pieces would lie in the same direction as the original. In a polycrystal, there are different regions within the material that have a different polar axis. It is asymmetrical because there is no point at which the crystal could be cut that would leave the two remaining pieces with the same resultant polar axis. Figure 10.9 illustrates this concept.

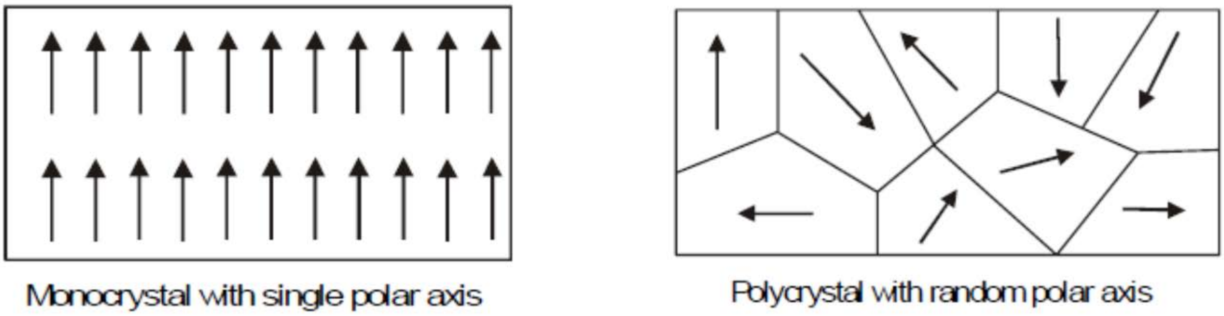


Fig. 10.9. Mono vs. poly crystals.

In order to produce the piezoelectric effect, the polycrystal is heated under the application of a strong electric field. The heat allows the molecules to move more freely and the electric field forces all of the dipoles in the crystal to line up and face in nearly the same direction (Fig. 10.6).

The piezoelectric effect can now be observed in the crystal. Figure 10.10 and Fig. 10.11 illustrate the piezoelectric effect. Figure 10.10(a) shows the piezoelectric material without a stress or charge. If the material is compressed, then a voltage of the same polarity as the poling voltage will appear between the electrodes (b). If stretched, a voltage of opposite polarity will appear (c). Conversely, if a voltage is applied the material will deform. A voltage with the opposite polarity as the poling voltage will cause the material to expand (d), and a voltage with the same polarity will cause the material to compress (e). If an AC signal is applied then the material will vibrate at the same frequency as the signal (f).

The nature of the piezoelectric effect is closely related to the occurrence of electric dipole moments in solids. The latter may either be induced for ions on crystal lattice sites with asymmetric charge surroundings (as in $BaTiO_3$ and PZTs) or may directly be carried by molecular groups (as in cane sugar). The dipole density or polarization (dimensionality [$C{\cdot}m/m^3$]) may easily be calculated for crystals by summing up the dipole moments per volume of the crystallographic unit cell [22]. As every dipole is a vector, the dipole density ***P*** is a vector field. Dipoles near each other tend to be aligned in regions called Weiss domains. The domains are usually randomly oriented, but can be aligned using the process of *poling* (not the same as magnetic poling), a process by which a strong electric field is applied across the material, usually at elevated temperatures. Not all piezoelectric materials can be poled [23].

Of decisive importance for the piezoelectric effect is the change of polarization ***P*** when applying a mechanical stress. This might either be caused by a re-configuration of the dipole-inducing surrounding or by re-orientation of molecular dipole moments under the influence of the external stress. Piezoelectricity may then manifest in a variation of the polarization strength, its direction or both, with the details depending on (*i*) the orientation of ***P*** within the crystal, (*ii*) crystal symmetry and (*iii*) the applied mechanical stress. The change in ***P*** appears as a variation of surface charge density upon the crystal faces, i.e. as a variation of the electric field extending between the faces caused by a change in dipole density in the bulk. For example, a 1 cm^3 cube of quartz with 2 kN (500 lbf) of correctly applied force can produce a voltage of 12500 V [24].

Piezoelectric materials also show the opposite effect, called converse piezoelectric effect, where the application of an electrical field creates mechanical deformation in the crystal.

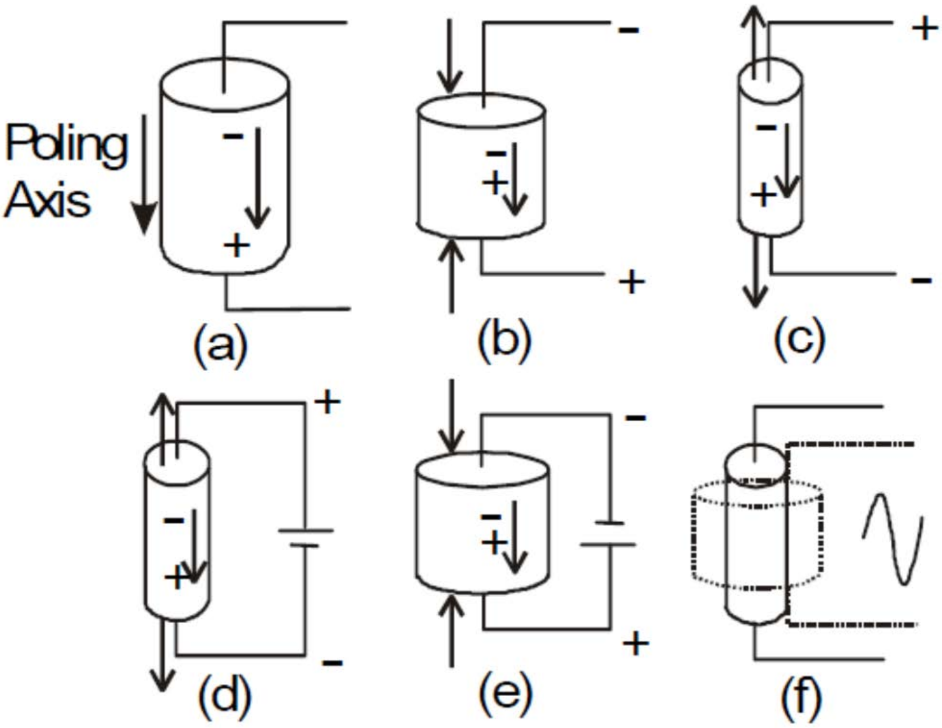


Fig. 10.10. Examples of piezoelectric effect.

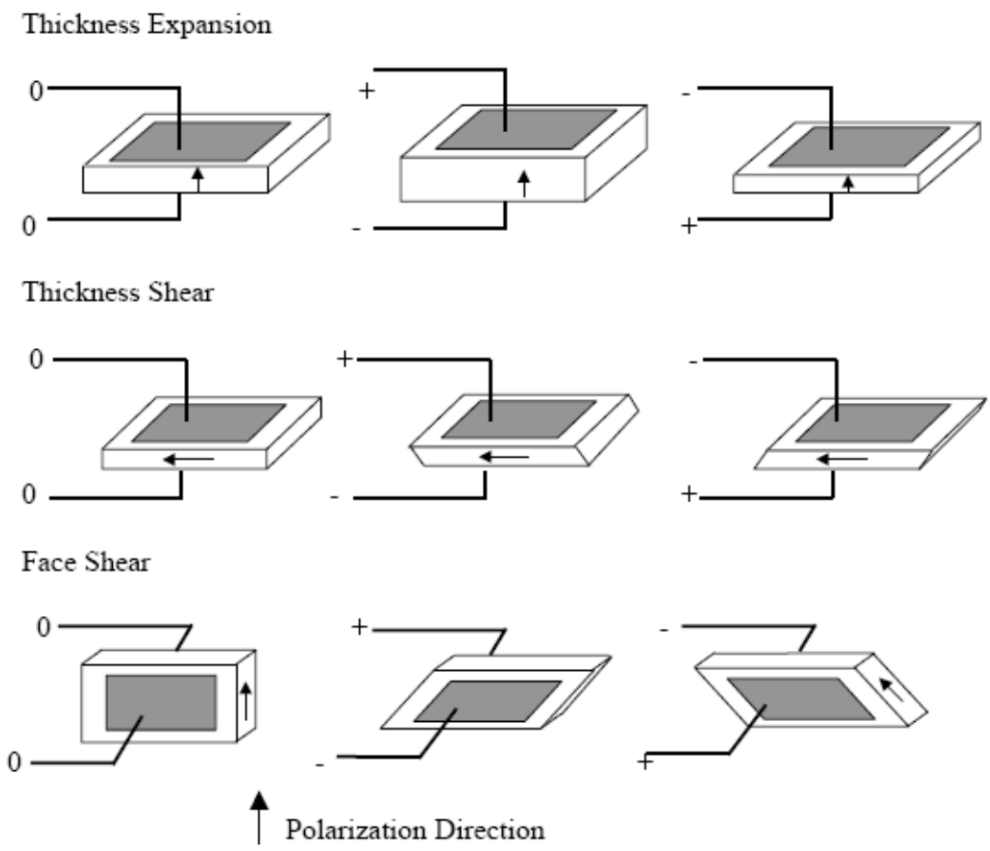


Fig. 10.11. Basic piezoelectric modes.

Piezoelectric Benders. Piezoelectric benders are often used to create actuators with large displacement capabilities (Fig. 10.12). The bender works in a mode which is very similar to the action of a bimetallic spring. Two separate bars or wafers of piezoelectric material are metallized and poled in the thickness expansion mode. They are then assembled in a + -+ - stack and mechanically bonded. In some cases, a thin membrane is placed between the two wafers. The outer electrodes are connected together and a field is applied between the inner and outer electrodes. The result is that for one wafer the field is in the same direction as the poling voltage while the other is opposite to the poling direction. This means that one wafer is increasing in thickness and decreasing in length while the other wafer is decreasing in thickness and increasing in length, resulting in a bending moment.

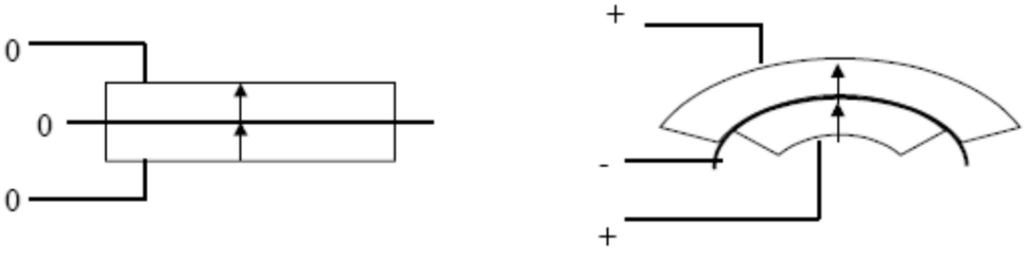


Fig. 10. 12. Schematic image of piezoelectric bender.

10.3.3. Poling and post poling

Piezoelectric ceramic materials, as stated ealier, are not piezoelectric until the random ferroelectric domains are aligned. This alignment is accomplished through a process known as "poling". Poling consists of inducing a DC voltage across the material. The ferroelectric domains align to the induced field resulting in a net piezoelectric effect (Fig. 10.6). It should be noted that not all the domains become exactly aligned. Some of the domains only partially align and some do not align at all. The number of domains that align depends upon the poling voltage, temperature, and the time the voltage is held on the material. During poling the material permanetly increases in dimension between the poling electrodes and decreases in dimensions parallel to the electrodes. The material can be depoled by reversing the poling voltage, increasing the temperature beyond the materials Currie point, or by inducing a large mechanical stress.

Applied Voltage. Voltage applied to the electrodes at the same polarity as the original poling voltage results in a further increase in dimension between the electrodes and decreases the dimensions parallel to the electrodes. Applying a voltage to the electrodes in an opposite direction decreases the dimension between the electrodes and increases the dimensions parallel to the electrodes.

Applied Force. Applying a compressive force in the direction of poling (perpendicular to the poling electrodes) or a tensile force parallel to the poling direction results in a voltage generated on the electrodes which has the same polarity as the original poling voltage. A tensile force applied perpendicular to the electrodes or a compressive force applied parallel to the electrodes results in a voltage of opposite polarity.

Shear. Removing the poling electrodes and applying a field perpendicular to the poling direction on a new set of electrodes will result in mechanical shear. Physically shearing the ceramic will produce a voltage on the new electrodes.

10.3.4. Mathematical Description

It is best to start with an understanding of common dielectric materials in order to understand the piezoelectric effect. The defining equations for high permittivity dielectrics are:

$$C = \frac{K\varepsilon_0 A}{t} = \frac{\varepsilon_0 \varepsilon_r A}{t} = \frac{\varepsilon A}{t} \qquad (1)$$

and

$$Q = CV = \frac{\varepsilon A V}{t} \qquad (2)$$

where C is the capacitance, A is the capacitor plate area, $\varepsilon_r=K$ is the relative dielectric constant, $\varepsilon_0=8.86\times10^{-12}$ F/m is the dielectric constant of air, ε is the dielectric constant, V is the voltage, t is the thickness or plate separation, Q is the charge.

In addition, we can define electric displacement, D, as charge density or the ratio of charge to the area of the capacitor:

$$D=\frac{Q}{A}=\frac{\varepsilon V}{t} \tag{3}$$

and further define the electric field as:

$$E=\frac{V}{t} \text{ or } D=\varepsilon E. \tag{4}$$

These equations are true for all isotropic dielectrics. Piezoelectric ceramic materials are isotropic in the unpolarized state, but they become anisotropic in the poled state. In anisotropic materials, both the electric field and electric displacement must be represented as vectors with three dimensions in a fashion similar to the mechanical force vector. This is a direct result of the dependency of the ratio of dielectric displacement, D, to electric field, E, upon the orientation of the capacitor plate to the crystal (or poled ceramic) axes. This means that the general equation for electric displacement can be written as a state variable equation:

$$D_i=\varepsilon_{ij}E_j. \tag{5}$$

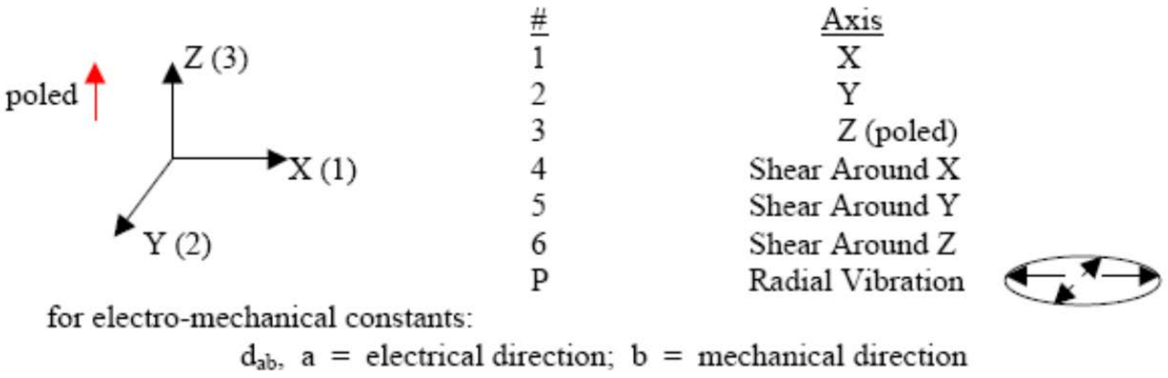
The electric displacement is always parallel to the electric field, thus each electric displacement vector, D_i, is equal to the sum of the field vector, E_j, multiplied by its corresponding dielectric constant, ε_{ij}:

$$\begin{aligned}D_1&=\varepsilon_{11}E_1+\varepsilon_{12}E_2+\varepsilon_{13}E_3\\ D_2&=\varepsilon_{21}E_1+\varepsilon_{22}E_2+\varepsilon_{23}E_3\\ D_3&=\varepsilon_{31}E_1+\varepsilon_{32}E_1+\varepsilon_{33}E_3\end{aligned} \tag{6}$$

Fortunately, the majority of the dielectric constants for piezoelectric ceramics (as opposed to single crystal piezoelectric materials) are zero. The only non-zero terms are:

$$\varepsilon_{11}=\varepsilon_{22},\varepsilon_{33}. \tag{7}$$

Axis nomenclature. The piezoelectric effect, as stated previously, relates mechanical effects to electrical effects. These effects, as shown above, are highly dependent upon their orientation to the poled axis. It is, therefore, essential to maintain a constant axis numbering scheme (Fig.10.13).



#	Axis
1	X
2	Y
3	Z (poled)
4	Shear Around X
5	Shear Around Y
6	Shear Around Z
P	Radial Vibration

for electro-mechanical constants:
d_{ab}, a = electrical direction; b = mechanical direction

Fig. 10.13. Axis numbering scheme.

For electro-mechanical constants:

d_{ab}, a is the electrical direction; b is the mechanical direction.

Electrical - mechanical analogies. Piezoelectric devices work as both electrical and mechanical elements. There are several electrical - mechanical analogies that are used in designing modeling the devices (Table 10.3).

Table 10.3. Electrical –mechanical analogies.

Electrical unit	Mechanical Unit
V Voltage (Volts)	f Force (Newtons)
I Current (Amps)	v Velocity (Meters / Second)
Q Charge (Coulombs)	s Displacement (Meters)
C Capacitance (farads)	C_M Compliance (Meters / Newton)
L Inductance (henrys)	M Mass (Kg)
Z Impedance	Z_M Mechanical Impedance
$I = dQ/dt$	$v = ds/dt$
$V = L\, dI/dt = L\, d^2Q/dt^2$	$f = M\, dv/dt = M\, d^2s/dt^{-2}$

Coupling. Coupling is a key constant used to evaluate the "quality" of an electro-mechanical material. This constant represents the efficiency of energy conversion from electrical to mechanical or mechanical to electrical.

$$k^2 = \frac{\text{Mechanical Energy Converted to Electrical Charge}}{\text{Mechanical Energy Input}} \tag{8}$$

or

$$k^2 = \frac{\text{Electrical Energy Converted to Mechanical Displacement}}{\text{Electrical Energy Input}} \tag{9}$$

Electrical, mechanical property changes with load. Piezoelectric materials exhibit the somewhat unique effect that the dielectric constant varies with mechanical load and the Young's modulus varies with electrical load.

Dielectric Constant

$$\varepsilon_{r(free)}(1-k^2) = \varepsilon_{r(clamped)} \tag{10}$$

This means that the dielectric "constant" of the material reduces with mechanical load. Here "free" stands for a state when the material is able to change dimensions with applied field. "Clamped" refers to either a condition where the material is physically clamped or is driven at a frequency high enough above mechanical resonance that the device can't respond to the changing *E* field.

Elastic Modulus (Young's Modules)

$$Y_{(open)}(1-k^2) = Y_{short} \tag{11}$$

This means that the mechanical "stiffness" of the material reduces when the output is electrically shorted. This is important in that both the mechanical loss (Q_M) and resonate frequency will change with load. This is also the property that is used in the variable dampening applications.

Elasticity. All materials, regardless of their relative hardness, follow the fundamental law of elasticity. The elastic properties of the piezoelectric material control how well it will work in a particular application. The first concepts, which need to be defined, are stress and strain (Fig. 10. 14). For a given bar of any material:

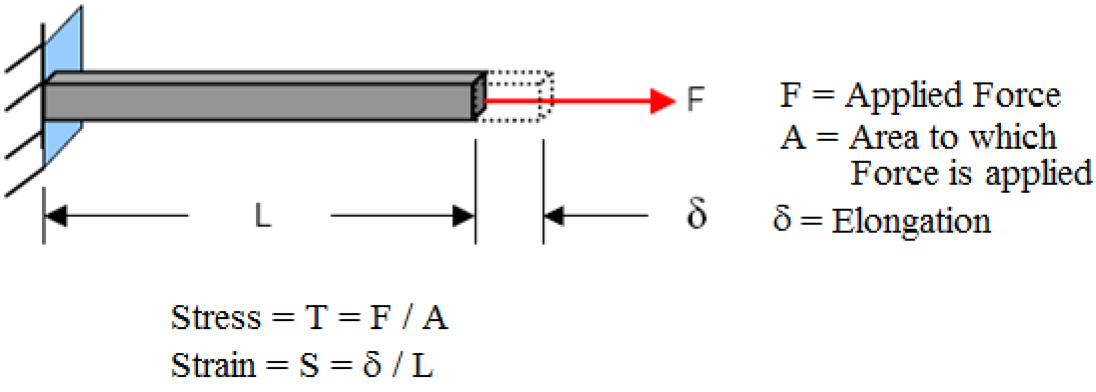


Fig. 10. 14. Elasticity of the materials.

The relationship between stress and strain is Hooke's Law which states that, within the elastic limits of the material, strain is proportional to stress.

$$S = sT \tag{12}$$

or, for an anisotropic material

$$S_i = s_{ij} T_j . \tag{13}$$

Note: The constant relating stress and strain is the modulus of elasticity or Young's modulus and is often represented by *Y*.

Piezoelectric Equation. It has been previously shown that when a voltage is applied across a capacitor made of normal dielectric material, a charge results on the plates or electrodes of the capacitor. Charge can also be produced on the electrodes of a capacitor made of a piezoelectric material by the application of stress. This is known as the Direct Piezoelectric Effect. Conversely, the application of a field to the material will result in strain. This is known as the Inverse Piezoelectric Effect. The equation, which defines this relationship, is the piezoelectric equation.

$$D_i = d_{ij} T_j \tag{14}$$

where D_i is the electric displacement (or charge density), d_{ij} is the piezoelectric modulus, the ratio of strain to applied field or charge density to applied mechanical stress.

Stated differently, d measures charge caused by a given force or deflection caused by a given voltage. We can, therefore, also use this to define the piezoelectric equation in terms of field and strain.

$$D_i = \frac{T_i S_i}{E_j} \tag{15}$$

Earlier, electric displacement was defined as

$$D_i = \varepsilon_{ij} E_j . \tag{16}$$

therefore,

$$\varepsilon_{ij} E_j = d_{ij} T_j \tag{17}$$

and

$$E_j = \frac{d_{ij}}{\varepsilon_{ij}} T_j \tag{18}$$

which results in a new constant

$$g_{ij} = \frac{d_{ij}}{\varepsilon_{ij}} . \tag{19}$$

This constant is known as the piezoelectric constant and is equal to the open circuit field developed per unit of applied stress or as the strain developed per unit of applied charge density or electric displacement. The constant can then be written as:

$$g = \frac{field}{stress} = \frac{volts/meter}{newtons/meter^2} = \frac{\Delta L/L}{\varepsilon V/t}. \qquad (20)$$

Fortunately, many of the constants in the formulas above are equal to zero for PZT piezoelectric ceramics. The non-zero constants are: $s_{11} = s_{22}$, s_{33}, s_{12}, $s_{13} = s_{23}$, s_{44}, $s_{66} = 2\,(s_{11} - s_{12})$; $d_{31} = d_{32}$, d_{33}, $d_{15} = d_{24}$.

Let’s consider more general presentation of above equations. Piezoelectricity is the combined effect of the electrical behavior of the material:

$$\vec{D} = \varepsilon\vec{E} \Rightarrow D_i = \varepsilon_{ij}E_j \qquad (21)$$

where ***D*** is the electric charge density displacement (electric displacement), ε is permittivity and ***E*** is electric field strength, and Hooke's Law:

$$\vec{S} = s\vec{T} \Rightarrow S_{ij} = s_{ijkl}T_{kl} \qquad (22)$$

where ***S*** is strain, *s* is compliance and ***T*** is stress.

These may be combined into so-called *coupled equations*, of which the strain-charge form is:

$$\vec{S} = s\vec{T} + d\vec{E} \Rightarrow S_{ij} = s_{ijkl}T_{kl} + d_{kij}E_k \qquad (23)$$

$$\vec{D} = d\vec{T} + \varepsilon\vec{E} \Rightarrow D_i = d_{ijk}T_{jk} + \varepsilon_{ij}E_j. \qquad (24)$$

In matrix form,

$$\{S\} = \left[s^E\right]\{T\} + \left[d^t\right]\{E\} \qquad (25)$$

$$\{D\} = \left[d\right]\{T\} + \left[\varepsilon^T\right]\{E\}, \qquad (26)$$

where [*d*] is the matrix for the direct piezoelectric effect and [d^t] is the matrix for the converse piezoelectric effect. The superscript *E* indicates a zero, or constant, electric field; the superscript *T* indicates a zero, or constant, stress field; and the superscript *t* stands for transposition of a matrix.

The strain-charge for a material of the 4mm (C_{4v}) crystal class (such as a poled piezoelectric ceramic such as tetragonal PZT or $BaTiO_3$) as well as the 6mm crystal class may also be written as:

$$\begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{bmatrix} = \begin{bmatrix} s_{11}^E \ldots s_{12}^E \ldots s_{13}^E \ldots .0 \ldots .0 \ldots ....0 \\ s_{21}^E \ldots s_{22}^E \ldots s_{23}^E \ldots 0 \ldots .0 \ldots ....0 \\ s_{31}^E \ldots s_{32}^E \ldots s_{33}^E \ldots 0 \ldots .0 \ldots ....0 \\ 0 \ldots ....0 \ldots ..0 \ldots s_{44}^E \ldots 0 \ldots ....0 \\ 0 \ldots ....0 \ldots ..0 \ldots .0 \ldots s_{55}^E .... \ldots 0 \\ 0 \ldots ....0 \ldots ....0 \ldots .0 \ldots ..0 \ldots .s_{66}^E = 2(s_{11}^E - s_{12}^E) \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} + \begin{bmatrix} 0 \ldots .0 \ldots .d_{31} \\ 0 \ldots .0 \ldots .d_{32} \\ 0 \ldots .0 \ldots .d_{33} \\ 0 \ldots d_{24} \ldots 0 \\ d_{15} \ldots 0 \ldots 0 \\ 0 \ldots ..0 \ldots .0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \qquad (27)$$

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 \ldots 0 \ldots 0 \ldots 0 \ldots d_{15}..0 \\ 0 \ldots 0 \ldots 0 \ldots d_{24}..0 \ldots 0 \\ d_{31}.d_{32}.d_{33}..0.....0....0 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} 0....0 \\ 0...\varepsilon_{22}.0 \\ 0...0...\varepsilon_{33} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \qquad (28)$$

where the first equation represents the relationship for the converse piezoelectric effect and the latter for the direct piezoelectric effect [25].

Although the above equations are the most used form in literature, some comments about the notation are necessary. Generally ***D*** and ***E*** are vectors, that is, Cartesian tensor of rank 1; and permittivity ε is Cartesian tensor of rank 2. Strain and stress are, in principle, also rank-2 tensors. But conventionally, because strain and stress are all symmetric tensors, the subscript of strain and stress can be re-labeled in the following fashion: 11 → 1; 22 → 2; 33 → 3; 23 → 4; 13 → 5; 12 → 6. (Different convention may be used by different authors in literature. Say, some use 12 → 4; 23 → 5; 31 → 6 instead.) That is why ***S*** and ***T*** appear to have the "vector form" of 6 components. Consequently, s appears to be a 6 by 6 matrix instead of rank-4 tensor. Such a re-labeled notation is often called Voigt notation. Whether the shear strain components S_4, S_5, S_6 are tensor components or engineering strains is another question. In the equation above, they must be engineering strains for the 6,6 coefficient of the compliance matrix to be written as shown, i.e., $2(s_{11}^E - s_{12}^E)$. Engineering shear strains are double the value of the corresponding tensor shear, such as $S_6=2S_{12}$ and so on. This also means that $s_{66}=1/G_{12}$, where G_{12} is the shear modulus.

In total, there are 4 piezoelectric coefficients, d_{ij}, ε_{ij}, g_{ij} and h_{ij} defined as follows:

$$d_{ij} = \left(\frac{\partial D_i}{\partial T_j}\right)^E = \left(\frac{\partial S_j}{\partial E_i}\right)^T \qquad (29)$$

$$\varepsilon_{ij} = \left(\frac{\partial D_i}{\partial S_j}\right)^E = -\left(\frac{\partial T_j}{\partial E_i}\right)^S \qquad (30)$$

$$g_{ij} = -(\frac{\partial E_i}{\partial T_j})^D = (\frac{\partial S_j}{\partial D_i})^T \quad (31)$$

$$h_{ij} = -(\frac{\partial E_i}{\partial S_j})^D = -(\frac{\partial T_j}{\partial D_i})^S \quad (32)$$

where the first set of 4 terms correspond to the direct piezoelectric effect and the second set of 4 terms correspond to the converse piezoelectric effect [26]. A formalism has been worked out for those piezoelectric crystals, for which the polarization is of the crystal-field induced type, that allows for the calculation of piezoelectrical coefficients d_{ij} from electrostatic lattice constants or higher-order Madelung constants [22].

Let's consider the physical meaning of various piezoelectric coefficients (d_{ij}, g_{ij}, s_{ij}, k_{ij}, and ε_{ij}). These coefficients play an important role in the performance of the piezoelectric materials [27].

Piezoelectric Constant (d_{ij}). It is defined as the ratio of the strain in *j*-axis to the electric field applied along the *i*-axis, when all external stresses are held constant. For example d_{31} is the ratio of strain along axis 1 to the electric field applied along the axis 3.

Piezoelectric Constant (g_{ij}). It is the ratio of strain developed along the *j*-axis to the charge (per unit area) deposited on electrodes perpendicular to the *i*-axis.

Elastic Compliance (s_{ij}). It is the ratio of the strain in the *i*-direction to the stress in the *j*-direction, given that there is no charge of stress along the other two directions.

Dielectric Coefficient (ε_{ij}). Determines the charge per unit area in the *i*-axis due to an electric field applied in the *j*-axis. The relative dielectric constant is defined as the ratio of the absolute permittivity of the material by permittivity of the free space.

Piezoelectric Coupling Coefficient (k_{ij}). This coefficient represents the ability of a piezoelectric material to transform electrical energy to mechanical energy and vice versa. This transformation of energy between mechanical and electrical domains is employed in both sensors and actuators made from piezoelectric materials. The *ij* index indicates that the stress, or strain is in the direction *j*, and the electrodes are perpendicular to the *i*-axis.

Loss. There are two sources for loss in a piezoelectric device. One is mechanical, the other is electrical.

Mechanical Loss:

$$Q_M = \frac{\text{Mechanical Stiffness Reactance or Mass Reactance}}{\text{Mechanical Resistance}} \quad (33)$$

Electrical Loss:

$$\tan\delta = \frac{\text{Effective Series Resistance}}{\text{Effective Series Reactance}} \quad (34)$$

10.4. Piezoelectric Materials

10.4.1. Single Crystals

Many materials, both natural and synthetic, exhibit piezoelectricity. The piezoelectric effect occurs only in non conductive materials. Piezoelectric materials can be divided in 2 main groups: crystals and cermaics.

Naturally occurring crystals: Quartz, Berlinite ($AlPO_4$), a rare phosphate mineral that is structurally identical to quartz, Sucrose (table sugar), Rochelle salt, Topaz, Tourmaline-group minerals, Lead titanate ($PbTiO_3$). Although $PbTiO_3$ occurs in nature as mineral macedonite [28, 29], it is synthesized for research and applications. The action of piezoelectricity in Topaz can probably be attributed to ordering of the (F,OH) in its lattice, which is otherwise centrosymmetric: Orthorhombic Bipyramidal (mmm). Topaz has anomalous optical properties which are attributed to such ordering [30].

Quartz, lithium nibonate ($LiNbO_3$), and lithium tantalite ($LiTaO_3$) are some of the most popular single crystals materials. The single crystals are anisotropic in general and have different properties depending on the cut of the materials and direction of bulk or surface wave propagation. These materials are essential used for frequency stabilized oscillators and surface acoustic devices applications [31].

Quartz. The most well-known piezoelectric material is quartz (SiO_2) (Fig. 10.15). Quartz is the trigonal crystallized silica (SiO_2) and is one of the most common crystals on earth's surface. Quartz belongs to the trigonal crystal system. The ideal crystal shape is a six-sided prism terminating with six-sided pyramids at each end. The building blocks of quartz are SiO_4 tetrahedra. In nature quartz crystals are often twinned, distorted, or so intergrown with adjacent crystals of quartz or other minerals as to only show part of this shape, or to lack obvious crystal faces altogether and appear massive. Quartz shows a strong piezoelectric effect perpendicularly to the prism axis (Fig. 10.16). Applying pressure on a quartz crystal generates an electrical polarization along the pressure direction. Alternatively, applying an electrical tension leads to a mechanical deformation of the crystal.

Fig. 10.15. Quartz crystals.

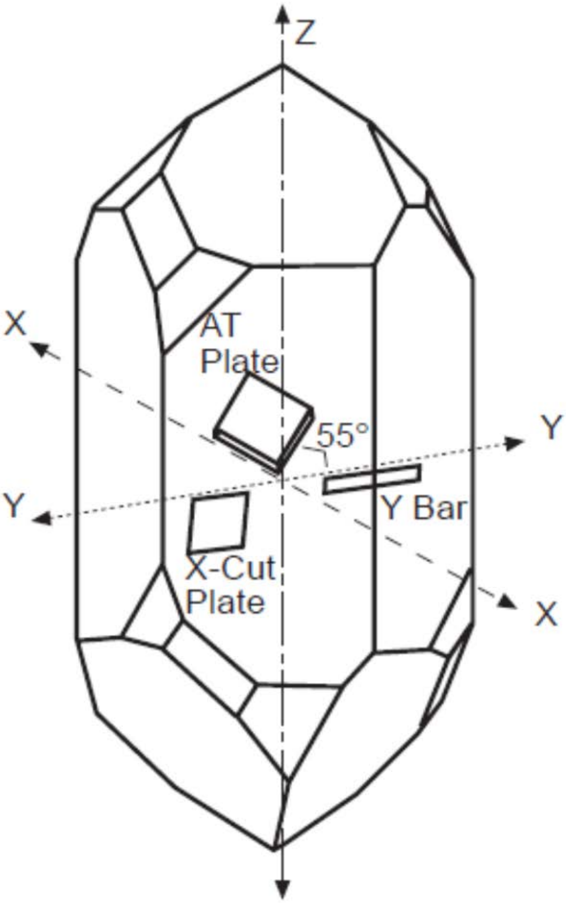


Fig. 10.16. Quartz crystal with 3 most important piezoelectric elements X-cut is used for expansion or contraction in thickness direction on application of electric field in X direction; Ybar expands and contracts in length when field applied in X direction. AT plate shows shear in side faces when field applied in thickness direction; it excels in low temperture variation on resonance frequency.

III-V and II-VI Semiconductors. A piezoelectric potential can be created in any bulk or nanostructured semiconductor crystal having non central symmetry, such as the Group III-V and II-

VI materials, due to polarization of ions under applied stress and strain. This property is common to both the zincblende and wurtzite crystal structures. To first order there is only one independent piezoelectric coefficient in zincblende, called e_{14}, coupled to shear components of the strain. In wurtzite instead there are 3 independent piezoelectric coefficients: ε_{31}, ε_{33} and ε_{15}. The semiconductors where the strongest piezoelectricity is observed are those commonly found in the wurtzite structure, i.e. GaN, InN, AlN and ZnO. ZnO is the most used material in the recent field of piezotronics. Since 2006 there have also been a number of reports of strong non linear piezoelectric effects in polar semiconductors [32]. Such effects are generally recognized to be at least important if not of the same order of magnitude as the first order approximation.

10.4.2. Piezoelectric Ceramics

Piezoelectric ceramics are widely used at present for a large number of applications. Most of the piezoelectric ceramics have perovskite structure. This ideal structure consists of a simple cubic cell that has a large cation “A” at the corner, a smaller cation “B” in the body center, and oxygen O in the centers of the faces. The structure is a network of corner-linked oxygen octahedral surroundings B cations.

For the case of Barium Titanate ceramic, the large cation A is Ba^{+2}, smaller cation B is Ti^{+4}. The unit cell of perovskite cubic structure of Barium Titanate is shown in Fig. 10.17. The piezoelectric properties of the perovskite-structured materials can be easily tailored for applications by incorporating various cations in the perovskite structure. Barium Titanate ($BaTiO_3$) and Lead Titanate ($PbTiO_3$) are the common examples of the perovskite piezoelectric ceramic materials [27,31].

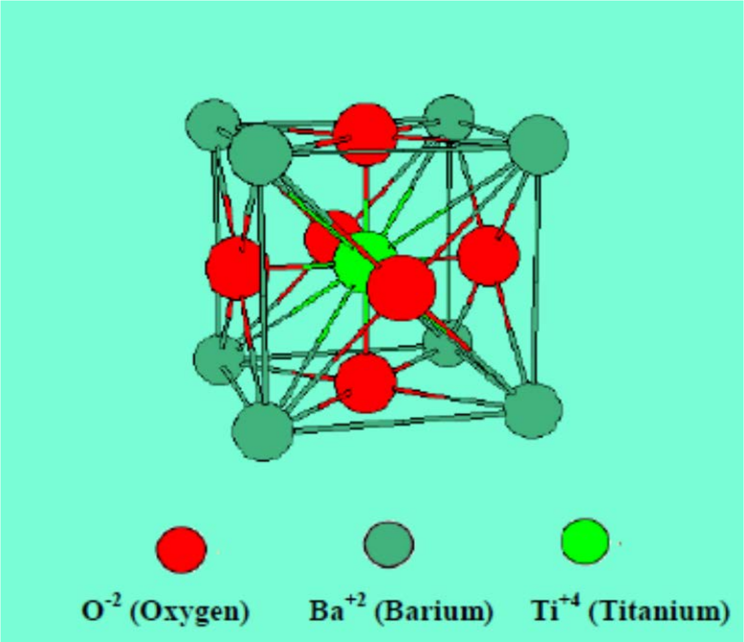


Fig. 10.17. Crystalline structure of a Barium Titanate (Perovskite structure).

Microstructures of piezoelectric ceramics. Fully dense piezoelectric ceramics (>95 % of theoretical density) are required by most applications for several reasons (Fig. 10.18). Firstly, a full densification ensures the ceramics to have their maximum performance. For example, the piezoelectric constant of piezoceramics usually decreases with decreasing density. This is because the pores or vacuum have no piezoelectric effect. Secondly, the presence of pores is generally a cause of high loss tangent, since porosity could provide a conduction path for electricity and thus could be a main contribution to dielectric loss tangent, as conduction loss. In this case, external electric fields cannot be applied to the materials effectively. Moreover, electrical conduction caused by porosity is also responsible for degraded electrical breakdown strength. Finally, mechanical strength, sometimes a critical requirement for some specific applications, especially as actuators, is directly related to density of the materials used.

Generally, the densities of piezoelectric ceramics increase with increasing sintering temperature. However, very high temperature is not suitable for lead- and bismuth-containing piezoelectric materials, which is due to the volatility characteristic of lead and bismuth. High temperature sintering also leads to abnormal grain growth (secondary grain growth). The presence of exaggeratedly grown grains is harmful to the performance for most piezoelectric ceramics. To effectively enhance the densification of piezoelectric ceramics at relatively low temperatures, various attempts have been made and have been widely reported in the literature. The main strategies that have been employed to reduce the sintering temperature of piezoelectric ceramics are the use of fine/ultrafine powders and the addition of sintering aids. If sintering aids are used, they must have relatively low melting points and no reaction with the piezoelectric phases during sintering process. Grain size is another important factor that determines piezoelectric properties of piezoceramics [33, 34]. The variation in grain size with sintering temperature is similar to that of density, i.e., grain size increases with increasing sintering temperature. There is a critical grain size for most piezoelectric ceramics. These properties can be optimized by adjusting materials processing parameters.

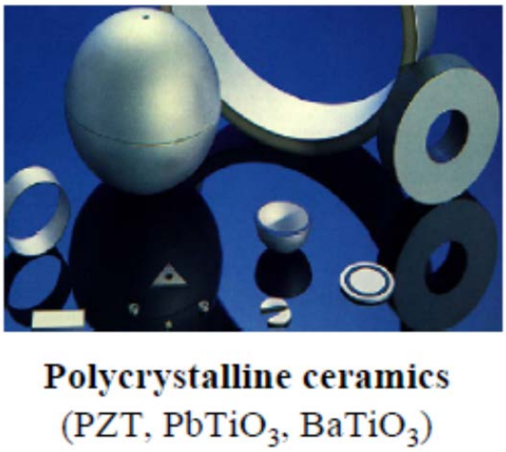


Fig. 10.18. Polycrystalline ceramics.

Typical piezoceramics. As mentioned above, lead zirconate titanate ($Pb(Zr_xTi_{1-x})O_3$ or PZT, $x = 0–1$) is the most widely used to fabricate piezoelectric ceramics. PZT is a solid solution of $PbTiO_3$ (PT) and $PbZrO_3$ (PZ). PT and PZ can form solid solution over composition range. Figure 10.19 shows a phase diagram of PZT. There exists an almost temperature-independent phase boundary at $x = 0.52–0.53$, which separates a rhombohedral Zr-rich phase from a tetragonal Ti-rich phase. Dielectric constant, piezoelectric constant, and electromechanical coupling coefficient all exhibit a pronounced maximum value for the composition corresponding to this phase boundary, which is generally referred to as the morphotropic phase boundary (MPB) [35, 36]. This is mainly attributed to the existence of a mixture of phases at the boundary and the presence of a larger number of reorientable polarization directions existing in the MPB mixed-phase region. This is reason why PZT is usually meant to be $PbZr_{0.52}Ti_{0.48}O_3$ or $PbZr_{0.53}Ti_{0.47}O_3$ used as piezoceramics in the open literature.

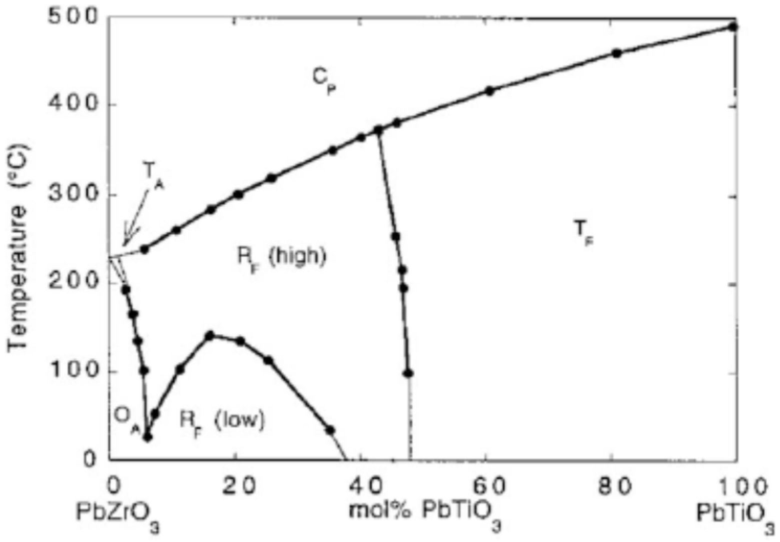


Fig. 10.19. Phase diagram of $PbZr_{1-x}Ti_xO_3$ [23]. There is a morphotropic phase boundary (MPB) at $x = 0.52–0.53$, with highest electrical properties

PZT ceramics. Conventionally, PZT powder is synthesized by using solid-state reaction method, with PbO (Pb_3O_4), ZrO_2, and TiO_2 oxide powders as starting materials. Mixtures of the three oxide powders with nominated compositions of PZT are ball milled and then calcined at a suitable temperature to form the desired perovskite phase. PZT piezoelectric ceramics are obtained by sintering the calcined powders at higher temperatures to achieve full densification.

Based on this basic composition, a large number of modifications have been adapted to cater for the requirements of different applications. According to international convention, piezoceramics are divided into two groups: soft and hard PZT ceramics. Soft and hard PZT ceramics refer to their mobility of the dipoles or domains and hence also to the polarization and depolarization behavior. Soft piezoceramics have high domain mobility and thus are easy to be poled. They usually have large piezoelectric charge coefficient, moderate values of permittivity, and high coupling factors. Hard PZT materials can be subjected to high electrical and mechanical stresses. Their properties change only little under these conditions, which makes them more suitable for high-power applications. They possess large piezoelectric coupling factors, high qualities, and very good stability at high mechanical loads and operating fields.

Commercially available oxide powders have relatively large grain/particle sizes. The synthesis of PZT with these powders requiring a high calcination temperature form the perovskite phase [37]. Figure 10.20 shows representative microstructural properties of a set of PZT ceramics synthesized by using the conventional solid-state reaction method [37]. PZT powders thus have rough particles that lead to high sintering temperature, which is not favorite owing to the volatility of Pb. Therefore, special carefulness should be taken when sintering these rough PZT powders to avoid the loss of Pb during the sintering process. One of the ways to address this problem is the use of fine PZT powders, which can be synthesized by using wet-chemical synthesis methods, such as chemical co-precipitation [38] and sol-gel [39]. Due to the high chemical homogeneity provided by these wet-chemical processing routes, PZT powders have ultrafine or nanosized grain/particles and thus can be sintered at significantly low temperatures. However, these chemical methods are relatively less cost-effective, because they require the use expensive starting chemicals and they also involve a multistep processing.

More recently, a novel technique, mechanochemical synthesis with high-energy ball milling, was employed to synthesize PZT powders [17]. In this method, the formation of designed compounds is due to the reactions of oxide precursors which are activated by mechanical energy, instead of the heat energy required in the conventional solid-state reaction process. The mechanical technique is superior to both the conventional solid-state reaction and the wet-chemistry-based processing routes for several reasons. Firstly, it uses cost-effective and widely available oxides as the starting materials. Secondly, it skips the intermediate temperature calcination step, leading to a simpler process. Thirdly, it takes place at room temperature in closely sealed containers, thus effectively alleviating the loss of Pb. Furthermore, due to their nanometer scale size and very high homogeneity, the mechanochemically derived PZT powders demonstrate much better sintering behavior than those synthesized by the conventional solid-state reaction and wetchemical processes.

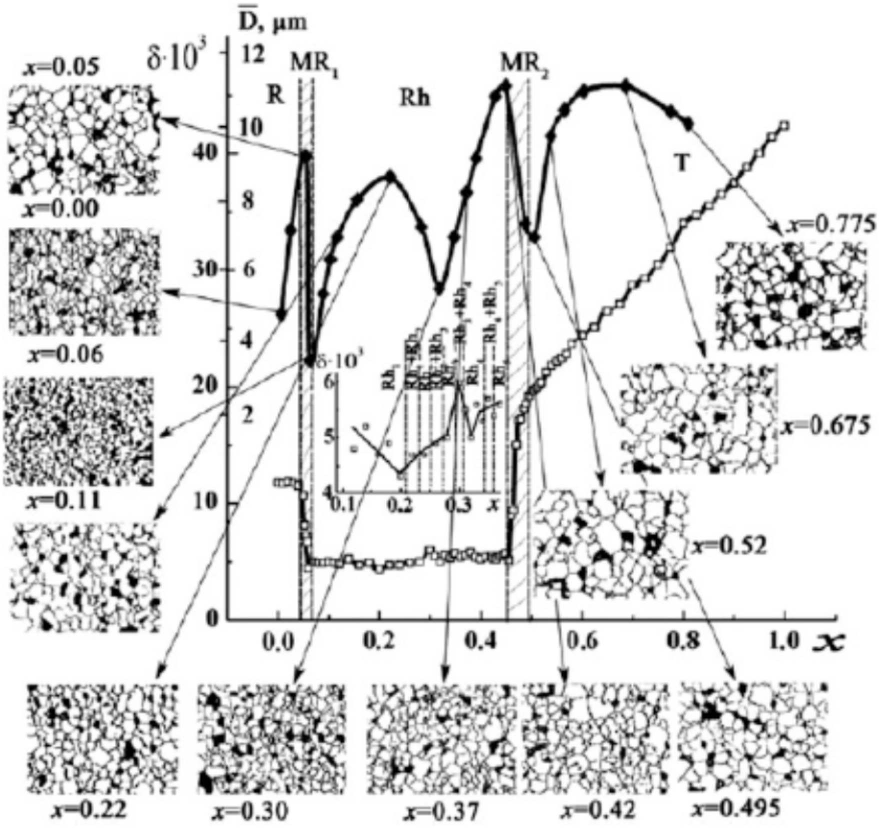


Fig. 10.20. Variation in mean grain size, D, and homogeneous deformation parameter, d, as a function of the content of PbTiO3 in the PZT system. The symmetries are designated as *R* rhombic, *Rh* rhombohedral, *T* tetragonal, and *MR1* and *MR2* denote morphotropic regions [37].

As an example, Fig. 10.21 shows XRD patterns of the mixtures (PbO, ZrO_2, and TiO_2) for the composition of $PbZr_{0.52}Ti_{0.48}O_3$, milled for different time durations, using a Fritsch Pulverisette five planetary high-energy ball milling system, with tungsten carbide (WC) vials and balls as the milling media [40]. After milling for 4 h, no PZT is formed, but the diffraction peaks from PbO are greatly broadened and weakened, indicating that the starting oxides have been significantly refined as a result of the high-energy ball milling. The diffraction peaks of PbO are further widened and reduced in the 8-h milled sample. At the same time, a trace of PZT can be observed. PZT with perovskite structure comes to the predominant phase in the samples milled for 15 and 24 h. It is worth mentioning that all the fours samples can be used to make PZT ceramics, which means that complete reaction of the precursor oxides is not necessary. This is because the reaction can be completed during the sintering processing. In this case, the milling time can be significantly reduced to save energy and time.

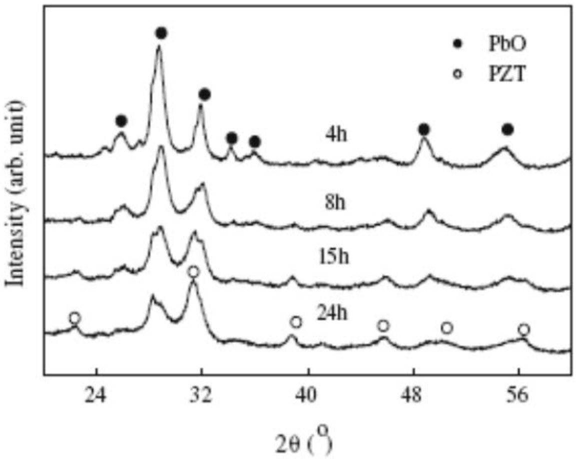


Fig. 10.21. XRD patterns of the mixtures for $PbZr_{0.52}Ti_{0.48}O_3$ milled for different times, using a planetary high-energy ball mill and tungsten carbide (WC) media [40].

10.4.3. Thin Film Piezoelectrics. PZT Films

Both zinc oxide (ZnO) and aluminum nitride (AlN) are simple binary compounds that have Wurtzite type structure, which can sputter-deposited in a c-axis oriented thin films on variety of substrates. ZnO has reasonable piezoelectric coupling and its thin films are widely used in bulk acoustic and SAW devices [31].

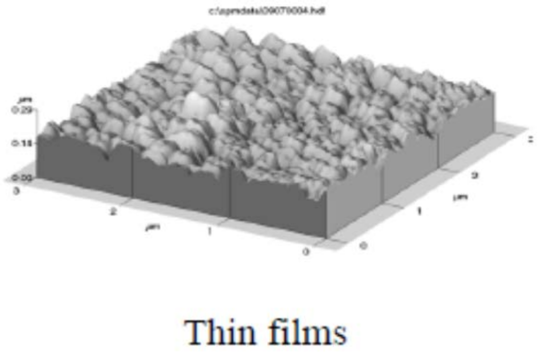


Fig. 10.22. AFM image of PZT thin film.

For applications in microelectromechanical system (MEMS), PZT thin films or thick films should be used [41]. Although there is no restricted definition, thin films usually have thicknesses of <1 μm and thick films are those with thickness of >1 μm. In practice, the thickness of a thick film can be up to hundreds of μm. PZT films can be deposited by using physical or chemical methods. Physical deposition includes sputtering, physical vapor deposition (PVD), and pulsed laser deposition (PLD), while chemical method includes solution, sol-gel, hydrothermal, and chemical vapor deposition (CVD). Among these physical and chemical methods, sol-gel has been most widely used to deposit PZT films, due to its effectiveness and efficiency.

Figure 10.23 shows a flow chart for deposition of PZT films by using a sol-gel technique [42]. $Pb(CH_3COO)_2{\cdot}5H_2O$ is dissolved in CH_3COOH at 120°C, while $Zr[CH_3(CH_2)_2CO]_4$ and $Ti[(CH_2)_2CHO]_4$ are stabilized with acetylaceton. They are then mixed to form 5 M solution with a composition of PZT, which is used to deposit PZT films by using spin-coating. After every coating, the samples are pyrolyzed at 350°C for 10 min, followed by annealing at 650°C for 30 min. A final annealing at 700°C for 30 min is conducted. PZT films with desired thickness can be obtained by repeating the pyrolysis-annealing step.

Figure 10.24 shows XRD patterns of the PZT films deposited for 1–6 times [42]. Pervoskite phase has been well formed in all films, which means that the annealing temperature of 700°C is sufficient. There is no pyrochlore phase in all samples. The sample coated for just one time has a random orientation. However, with increasing thickness, it becomes (100) orientated gradually. Cross-sectional SEM images of selected films are shown in Fig. 10.25 [42]. The thickness of the sample deposited for one time is ~0.5 μm. The total thickness of the films increases almost linearly with the number of deposition layers, which provides a very effective way to deposit thick PZT films. Comparatively, a single layer of PZT film deposited by using normal sol-gel process is much less than 0.1 μm. The use of acetic acid is the key to deposit thicker single layer of Pb containing films [43 - 45]. By the way, the films have a columnar microstructure, indicating their good crystallinity. Nevertheless, deposition of PZT thick films is still a challenge.

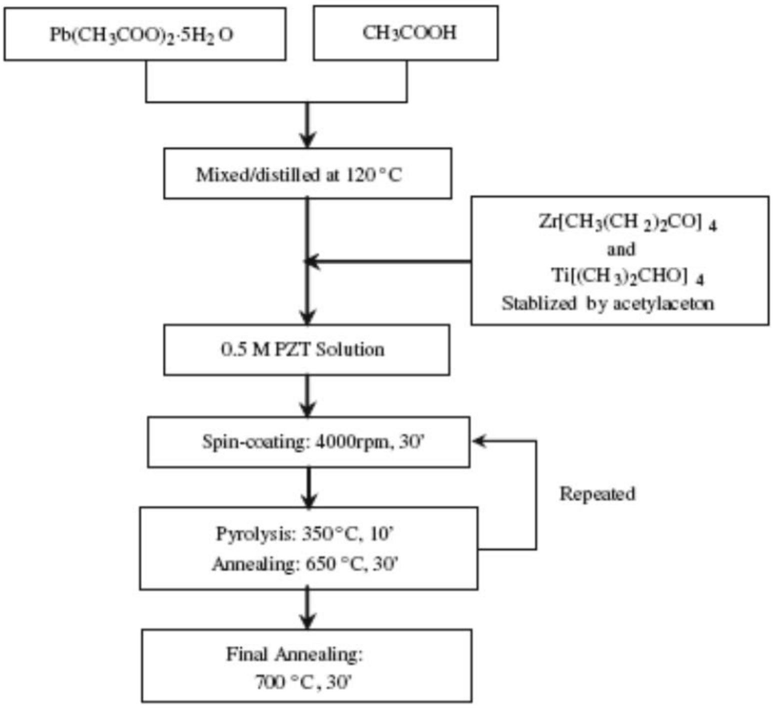


Fig. 10.23. Flow chart for deposition of PZT films with a modified sol-gel technique to produce single layer of PZT with sufficient thickness for thick film fabrication [42].

Figure 10.26 shows surface SEM images of selected PZT films [42]. All films have dense and crack-free microstructure. The average grain size slightly increases with increasing thickness, which can be readily attributed to increasing number of annealing times.

More recently, a new approach, called nanocomposite processing route, has been proposed to incorporated high-energy ball milling technique with sol-gel process to deposit thick films, which cannot be realized using the typical sol-gel with solution precursors [46 - 50]. This approach has combined the advantages of both high-energy milling and the sol-gel process. In this processing, commercial PZT powder is milled using a high-energy mill to produce nano-sized PZT powder. The nano-sized PZT powder is then milled for one more time with appropriate dispersants, which are used to modify its surficial characteristics. The modified PZT nano-sized powder can be readily mixed with sol-gel solutions without the presence of agglomerations or precipitations. This kind of mixture can be used as same as the normal sol-gel solutions to deposit thick films on various substrates. The slurries may also directly be used to other processing, such as tape casing, screen printing, and molding [47] (Fig. 10.27).

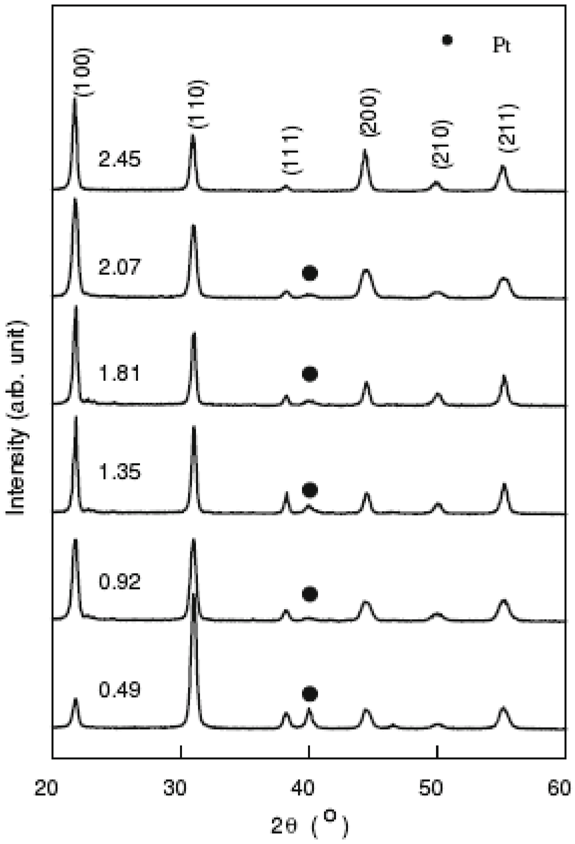


Fig. 10.24. XRD patterns of the PZT films deposited for 1–6 times (layers) with film thickness indicated [42]. There is a gradual increase in (100) orientation.

Thick films with thickness of up to 25 μm have been prepared through multilayer deposition with one layer being ~2 μm, which is nearly 50–100 times the normal sol-gel solution process. Figure 10.28 shows an example of such thick films [48]. The film is very uniform in thickness, with a dense microstructure and narrow distribution of grain sizes (~100 nm). Compared with the thick films prepared by using a precursor made of micro-sized PZT powder, the nanocomposite film has electrical properties, which has been attributed to the factor that the latter has a dense microstructure than the former. Such kind of composite thick films have been successfully combined with silicon technology to fabricate MEMS piezoelectric devices, so that they are suitable for applications in mechanical energy harvesting devices.

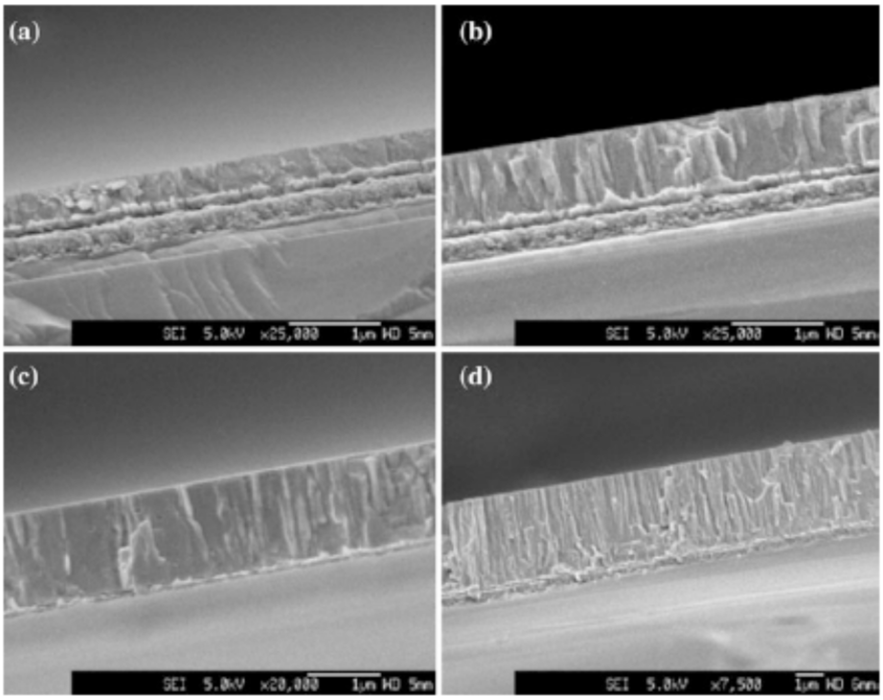


Fig. 10.25. Cross-sectional SEM images of the PZT films deposited for different times (layers): a-1, b- 2, c- 3, and d- 6 [42]. All the films have a columnar structure.

PZT ceramics have been prepared from the nanocomposite precursors at much lower temperature than that required by the conventional ceramic process [47]. Fully dense PZT ceramics can be achieved after sintering at a temperature as low as 800°C. Interestingly, no obvious grain growth is observed in the samples sintered at temperatures from 850 to 1,000°C, which means that the grain growth is stopped at ~850°C. The absence of grain growth at high temperatures has been attributed to the uniform size distribution and nonagglomeration of the nanocomposite precursors. This feature is particularly of importance when this kind of precursor is used to fabricate multilayer structured devices in which small thickness of single layer is required.

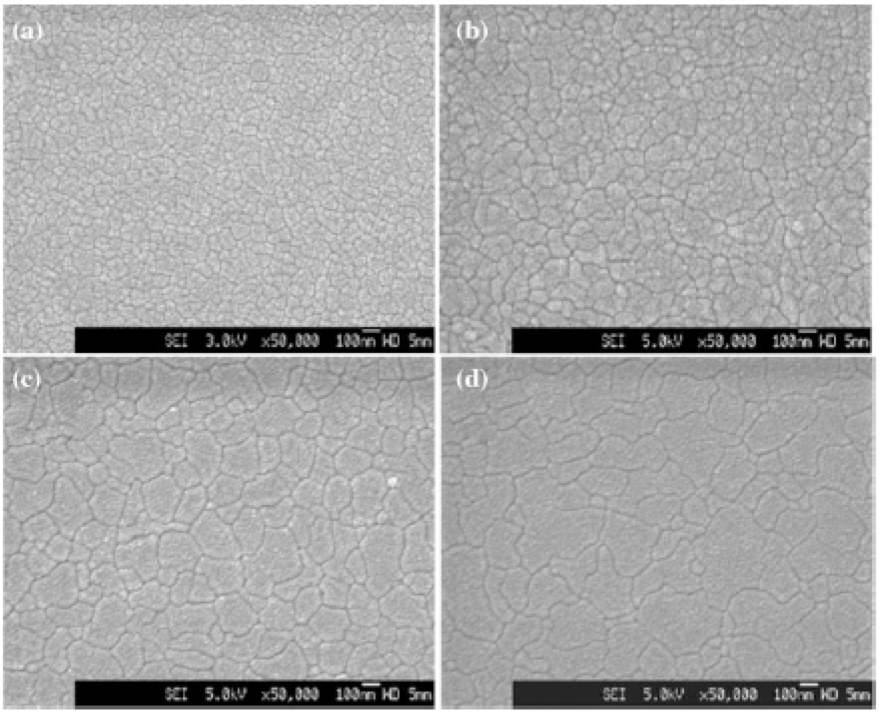


Fig. 10.26. Surface SEM images of the PZT films deposited for different times (layers): a- 1, b- 2, c- 3 and d- 5 [42]. Grain size of the films gradually increases with increasing thickness.

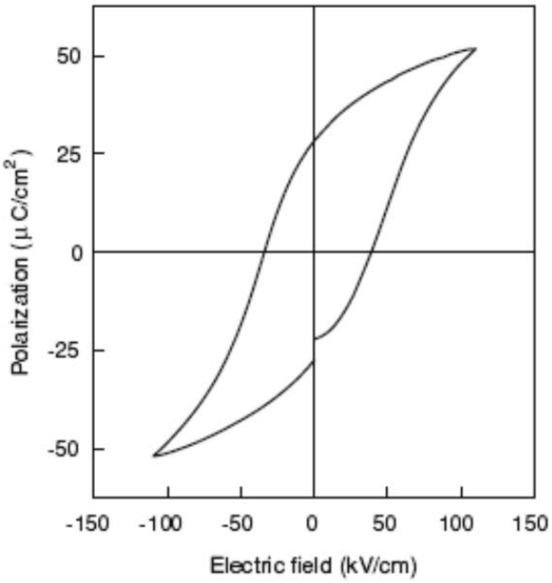


Fig. 10.27. Representative polarization-field (P-E) hysteresis curve of the PZT films showing their high quality [42].

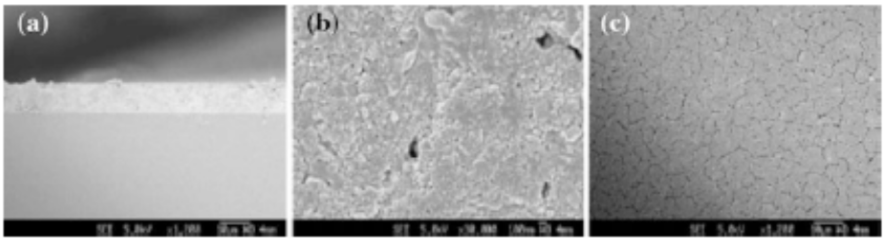


Fig. 10.28. SEM images of the thick film prepared with nanocomposite precursors: *a* and *b* crosssection and *c* surface [48].

10.4.4. Piezoelectric Polymers

The polymers like polypropylene, polystyrene, poly (methyl methacrylate), vinyl acetate, and odd number nylons are known to possess piezoelectric properties. However, strong piezoelectric effects have been observed only in polyvinylidene fluoride (PVDF or PVF2) and PVDF copolymers (Fig. 10.29). The molecular structure of PVDF consists of a repeated monomer unit $(\text{-CF}_2\text{-CH}_2\text{-})_n$. The permanent dipole polarization of PVDF is obtained through a technological process that involves stretching and poling of extruded thin sheets of polymer. These piezoelectric polymers are mostly used for directional microphones and ultrasonic hydrophones applications [31].

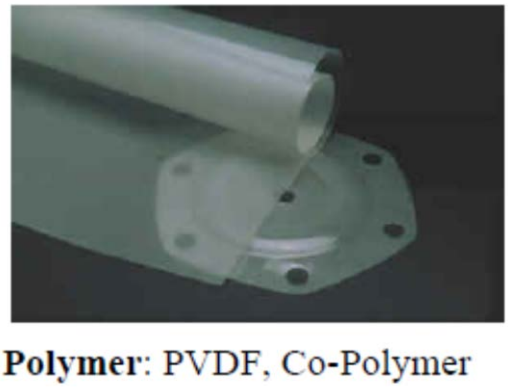


Fig. 10.29. Polymer: PVD, Co-Polymer.

Polyvinylidene fluoride or polyvinylidene difluoride (PVDF) has been found to show piezoelectric (ferroelectric) effect, due to the presence of a non-centrosymmetric unit cell and a net polarization in the material [51 - 53]. The common characteristics of piezoelectric polymeric materials are: a large dipole moment in the repeating unit, feasibility of crystallization in a noncentrosymmetric unit cell, and alignment of molecular dipoles. A dipole moment arises from a charge separation between adjacent atoms. In polymers, this can occur between covalently bonded atoms along the polymer chain or between adjacent chains.

PVDF has chains of CH_2CF_2 and is a semi crystalline polymer. Figure 10.30 shows the formation of PVDF polymer through the polymerization of monomer vinylidene fluoride. The hydrogen atoms having net positive charge and the fluorine atoms with net negative charge end up on opposite sides, as shown in Fig. 10.31. This develops a pole direction (indicated by the small ***P*** in Fig. 10.31). If the external applied electric field is in the opposite direction of the poled direction of the PVDF sheet, it will be stretched in the length direction. If the electric field is in the direction of the poled direction of the PVDF sheet, there will be a contraction in the length direction of the sheet. These two responses are shown in Figs. 10.32 and 10.33, respectively, which is the mechanism of piezoelectric effect of PVDF.

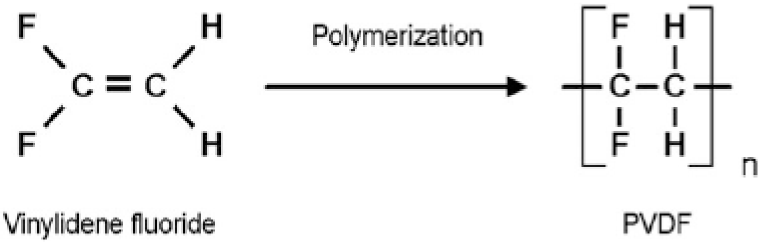


Fig. 10.30. Polymerization of vinylidene fluoride to form PVDF.

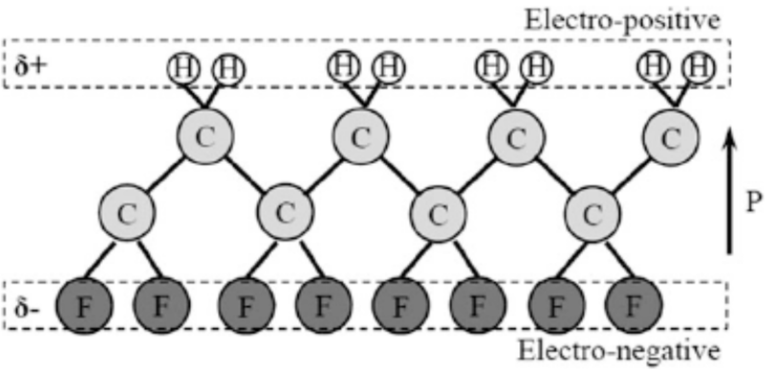


Fig. 10.31. Schematic diagram of atom arrangement in PVDF molecule.

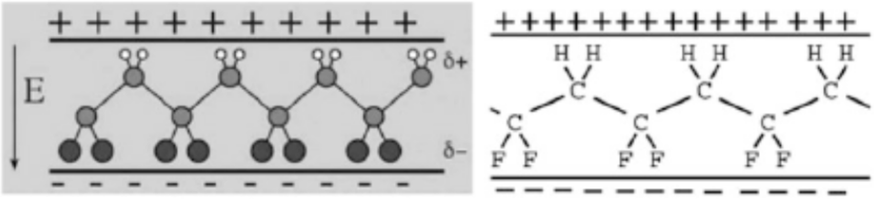


Fig. 10.32. Stretch in length direction occurred if the applied electric field is in the opposite direction of the poled direction of the PVDF sheet.

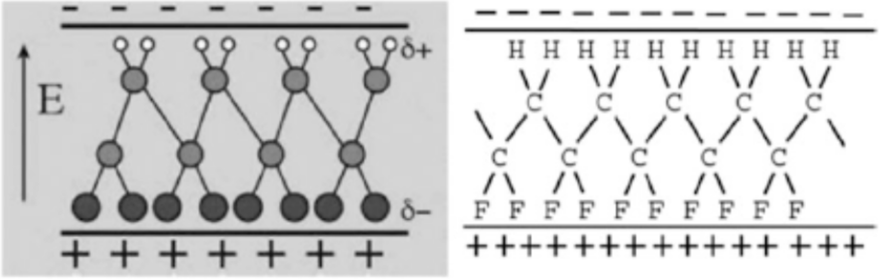


Fig. 10.33. Contraction in length direction occurred if the electric field is in the direction of the poled direction of the PVDF sheet.

It was later found that copolymers of PVDF with trifluoroethylene (TrFE) are new polymer materials with stronger piezoelectric effect. The copolymers have advantages over pure PVDF. The P(VDF-co-TrFE) copolymers can be modified by adjusting the compositions of the two components. The random copolymers P(VDF-co-TrFE)s are usually synthesized by copolymerization of VDF and TrFE monomers.

10.4.5. Piezoelectric Composites

Piezo-composites comprised piezoelectric ceramics and polymers are promising materials because of excellent tailored properties. These materials have many advantages including high coupling factors, low acoustic impedance, mechanical flexibility, a broad bandwidth in combination with low mechanical quality factor. They are especially useful for underwater sonar and medical diagnostic ultrasonic transducers [31].

Piezoelectric ceramics have high piezoelectric performances, but are hard and rigid for some applications that require flexibility and complicated shapes. To address this problem, composite materials are proposed, which consist of piezoelectric ceramic particles and polymer matrix. The presence of polymer matrix makes the composites to have mechanical flexibility. Composites are usually described by using a notation-connectivity. Figure 10.34 shows schematics of possible connectivity of composites with PZT and a polymer [54]. 0, 1, 2, and 3 are used to represent spherical particles, 1D wires, rods or fibers, 2D sheets and 3D matrix, respectively. To date, eight types of two-phase piezoelectric composites (piezocomposites) have been studied: 0–3, 1–3, 2–2, 2–3, 3–0, 3–1, 3–2, and 3–3. In these piezocomposites, the first number in the notation denotes the physical connectivity of the active phase (PZT) and the second number refers to the physical connectivity of the passive phase (polymer). There have been increasing reports on piezocomposites based on PZT and PVDF (Fig. 10.35) [55].

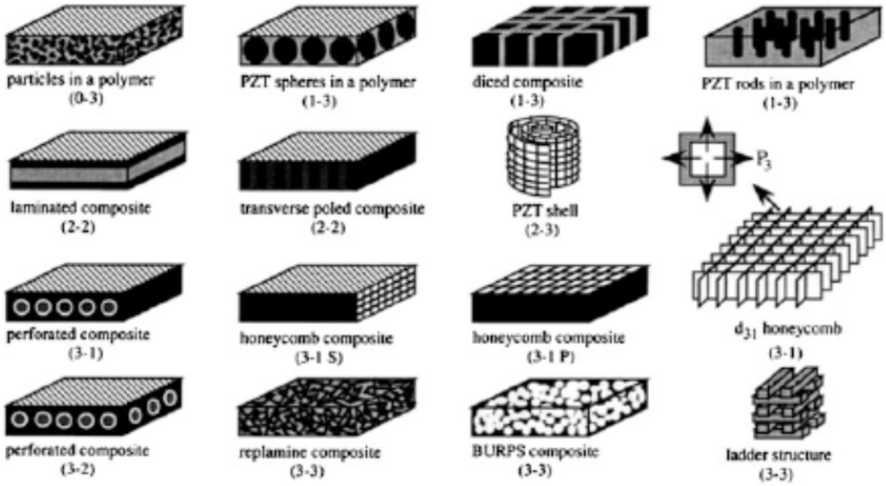


Fig. 10.34. Connectivity of constituent phases in piezoelectric ceramic-polymer composites [54].

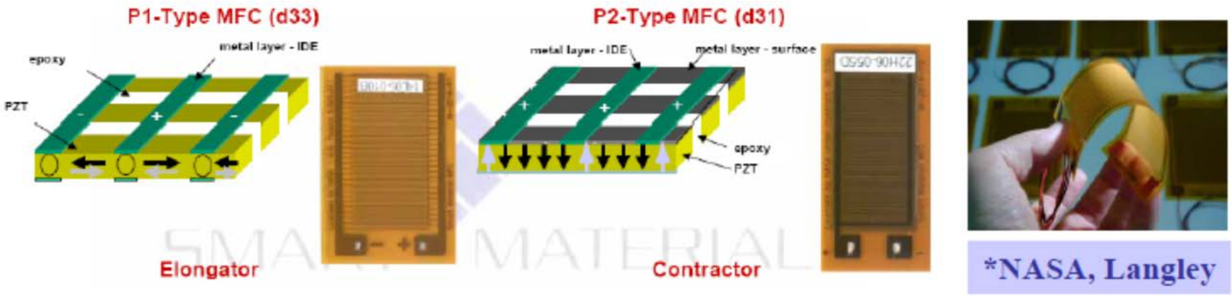


Fig. 10.35. Some types of piezoelectric ceramic – polymer composites.

10.5. Piezoelectric Material Applications

Currently, industrial and manufacturing is the largest application market for piezoelectric devices, followed by the automotive industry. Strong demand also comes from medical instruments as well as information and telecommunications. The global demand for piezoelectric devices was valued at approximately US$14.8 billion in 2010. The largest material group for piezoelectric devices is piezocrystal, and piezopolymer is experiencing the fastest growth due to its low weight and small size [56]. Piezoelectric crystals are now used in numerous ways.

The piezoelectric crystal bends in different ways at different frequencies. This bending is called the vibration mode. The crystal can be made into various shapes to achieve different vibration modes. To realize small, cost effective, and high performance products, several modes have been developed to operate over several frequency ranges. These modes allow us to make products working in the low kHz range up to the MHz range. Figure 10.36 shows the vibration modes and the frequencies over which they can work.

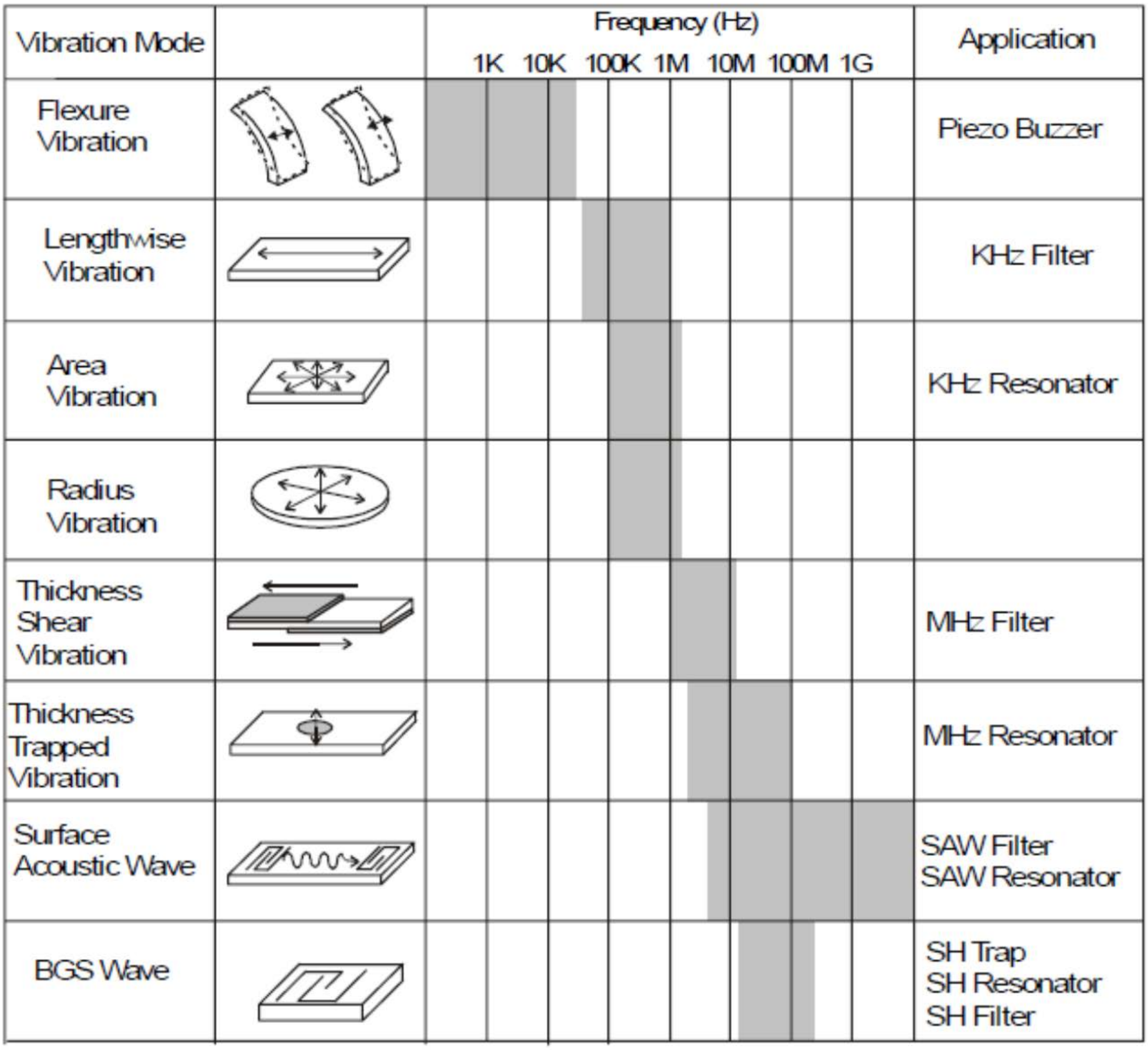


Fig. 10.36. Various vibration modes possible with piezoelectric ceramics.

10.5.1. Piezoelectric Sensors

Piezoelectric sensor are devices using the piezoelectric effect to measure acceleration, pressure, strain or force and converting them to an electrical signal. Piezoelements are suitable for the detection of dynamic processes. In static applications the piezoelectric charges are too small, in order to be detected. An amplifier is used to convert the piezoelectric charges into a measurable electrical tension. The principle of operation of a piezoelectric sensor is that a physical dimension, transformed into a force, acts on two opposing faces of the sensing element. The main sensor parameters are sensitivity, frequency response, stability or repeatability. Depending on the design of a sensor, different "modes" to load the piezoelectric element can be used: longitudinal, transversal and shear.

The varieties of piezoelectric microsensors have been developed. Among them are accelerometers, micro-resonators, surface acoustic wave (SAW) devices, underwater acoustic imaging sensors, etc. Detection of pressure variations in the form of sound is the most common sensor application, e.g. piezoelectric microphones (sound waves bend the piezoelectric material, creating a changing voltage) and piezoelectric pickups for acoustic-electric guitars. A piezo sensor attached to the body of an instrument is known as a contact microphone. Piezoelectric sensors especially are used with high frequency sound in ultrasonic transducers for medical imaging and also industrial nondestructive testing (NDT).

For many sensing techniques, the sensor can act as both a sensor and an actuator– often the term *transducer* is preferred when the device acts in this dual capacity, but most piezo devices have this property of reversibility whether it is used or not. Ultrasonic transducers, for example, can inject ultrasound waves into the body, receive the returned wave, and convert it to an electrical signal (a voltage). Most medical ultrasound transducers are piezoelectric.

When speaking of piezoelectric sensors, it is necessary to distinguish between the axial sensor (Fig. 10.37) and the bending sensor (Fig. 10.38). Axial sensors are subjected to a force in the polarization direction (33 mode) whereas the force on bending sensor is being exerted perpendicular to the polarization direction (31 mode).

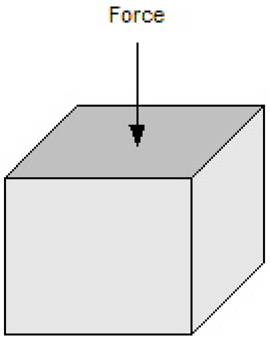


Fig. 10.37. Force action in axial sensors.

Axial sensors can either be made from bulk material (single layer) or frommultilayer (cofired component), where the latter introduces new options such as active feedback mechanisms serving the aim to constantly monitor the functionality of the sensor.

Bending sensors are commonly made as bulk or bimorph. Noliac bimorphs are co-fired structures where the middle electrode is screen-printed in the green state (unfired ceramic) and the external electrodes are applied after firing. Bimorphs can be polarized either in series or parallel, depending on customer requirements. Multilayer bending sensors (more than 2 active layers) with similar build-up structure as bending actuators are also available.

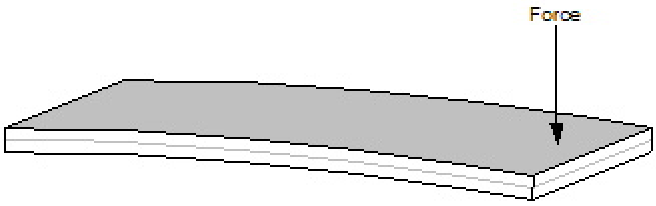


Fig. 10.38. Force action in bending sensors.

Piezoelectric Pressure Sensors. Piezoelectric Pressure Sensors measure dynamic pressures. They are generally not suited for static pressure measurements. Dynamic pressure measurements including turbulence, blast, ballistics and engine combustion under varying conditions may require sensors with special capabilities. Fast response, ruggedness, high stiffness, extended ranges, and the ability to also measure quasi-static pressures are standard features associated with quartz pressure sensors. There are two modes of operation for pressure sensors. Charge mode pressure sensors generate a high-impedance charge output. ICP (Integrated Circuit Piezoelectric) voltage mode-type sensors feature built-in microelectronic amplifiers, which convert the high-impedance charge into a low-impedance voltage output. Piezoelectric pressure sensors are available in various shapes and thread configurations to allow suitable mounting for various types of pressure measurements (Fig. 10.39). Quartz crystals are used in most sensors to ensure stable, repeatable operation. The quartz crystals are usually preloaded in the housings to ensure good linearity. Tourmaline, another stable naturally piezoelectric crystal, is used in some PCB sensors where volumetric sensitivity is required. Figure 10.40 illustrates the cross-section of a typical quartz pressure sensor. This particular sensor is a General Purpose Series with built-in electronics.

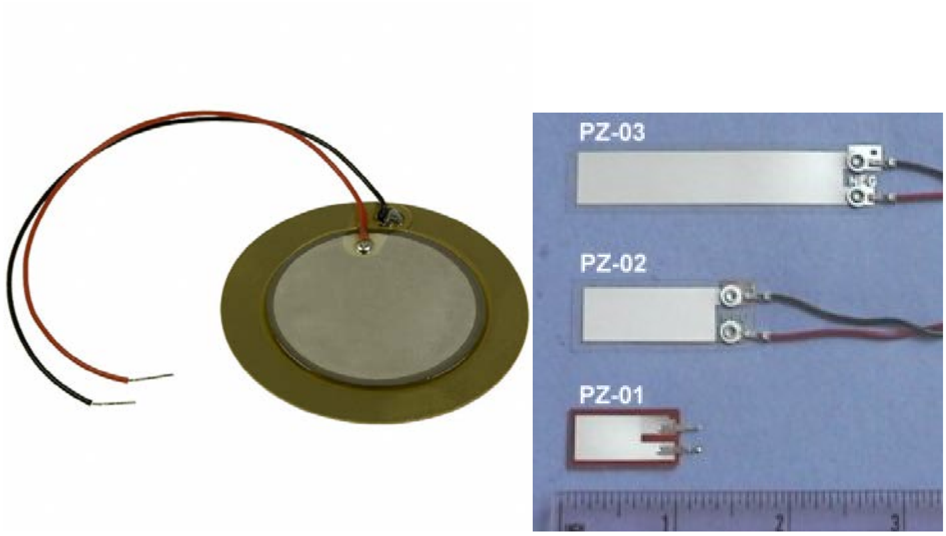


Fig. 10.39. Different designs of piezoelectric sensors.

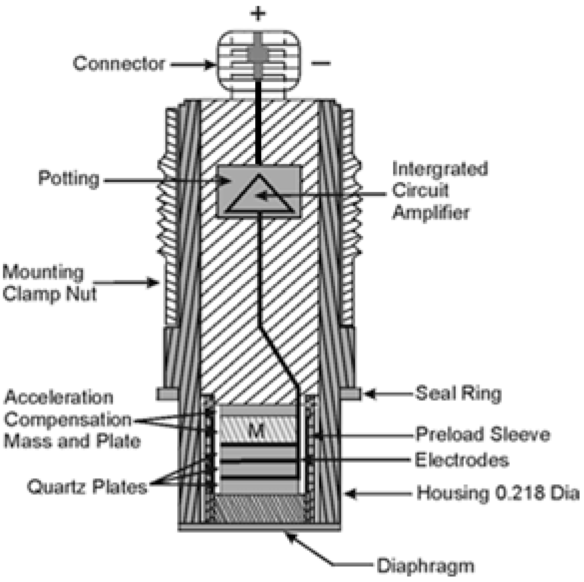


Fig. 10.40. Typical ICP Quartz pressure sensor.

In addition to those mentioned above, various piezoelectric sensor applications include:

- the detection and generation of sonar waves;
- the single-axis and dual-axes tilt sensing [57];
- power monitoring in high power applications (e.g. medical treatment, sonochemistry and industrial processing);
- piezoelectric microbalance are used as very sensitive chemical and biological sensors.
- the strain gauges;
- as a piezoelectric transducer in the penetrometer instrument on the Huygens Probe;

- piezoelectric transducers in electronic drum pads to detect the impact of the drummer's sticks, and to detect muscle movements in medical acceleromyography;

- automotive engine management systems with piezoelectric transducers to detect Engine knock (Knock Sensor, KS), also known as detonation, at certain hertz frequencies. A piezoelectric transducer is also used in fuel injection systems to measure manifold absolute pressure (MAP sensor) to determine engine load, and ultimately the fuel injectors milliseconds of on time;

- ultrasonic piezo sensors in the detection of acoustic emissions in acoustic emission testing;

- detection sound, e.g. piezoelectric microphones (sound waves bend the piezoelectric material, creating a changing voltage) and piezoelectric pickups for electrically amplified guitars;

- the generation of sonar waves.

10.5.2. Piezoelectric Generator and Energy Harvesters

Piezoelectric energy harvesters at the sub-micron-scale have infinite scope for research and application in upcoming technologies [58]. Devices for wireless technologies [59] and medical applications for replacing vital human organs need to be small and are required to be powered for their functioning as active devices [60]. Recent embedded electronic devices like RFID tags and remote sensors are giving stand alone and MEMS/NEMS devices a new dimension [61]. These devices, despite consuming very little power, require separate energy sources. The self-powering of these devices has remained an issue and is a new field of research in state-of-the-art technologies [62]. Piezoelectric materials can be used as the energy harvesting materials for devices for micro-/nanoelectromechanical systems, which can be made much smaller and lighter than conventional sources [63, 64].

Energy harvesting has been an active area of research in recent years due to growing awareness for environmental protection. Harvesting vibration energy based on piezoelectric materials utilizes the direct piezoelectric effect. In order to efficiently collect the low-quality electrical energy generated by a piezoelectric generator, the design of a high performance electronic interface has been an important issue in energy harvesting. In addition, many studies have also been conducted in the design of mechanical structures for energy harvesters as well as the development of piezoelectric materials with high electromechanical factors.

Power harvesting. The process of acquiring the energy surrounding a system and converting it into usable electrical energy is termed power harvesting (Fig. 10.41, Table 10.4). With piezo-materials, it is possible to harvest power from vibrating structures. It has been proven that micro-to mill watts of power can be generated from vibrating systems (Fig. 10.42).

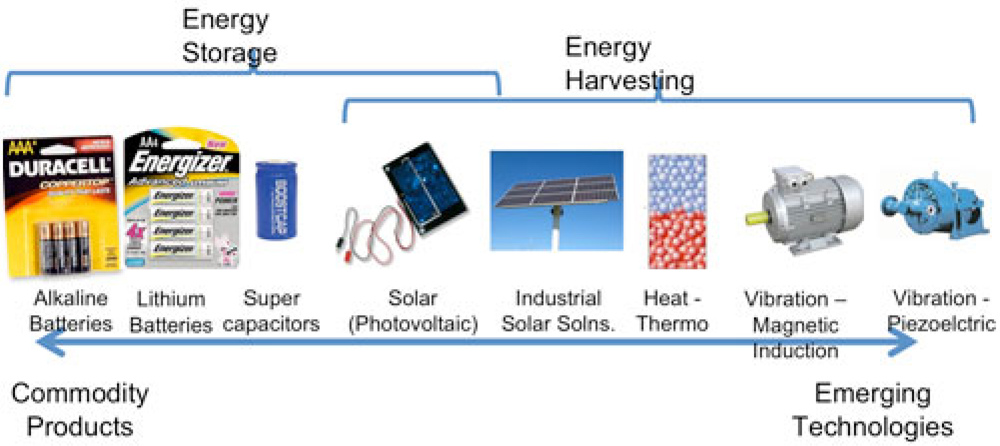


Fig. 10.41. Devices for energy storage and harvesting.

Table 10.4. Power available from energy sources.

Energy source	**Characteristics**	**Harvested power**
Light	Outdoor	100 mW/cm^2
	Indoor	100 μW/cm^2
Thermal	Human	60 μW/cm^2
	Industrial	~1-10 mW/cm^2
Vibration	~Hz-human	~4 μW/cm^2
	~KHz-machines	~800 μW/cm^2
RF	GSM 900 MHz	0.1 μW/cm^2
	WiFi	0.001 μW/cm^2

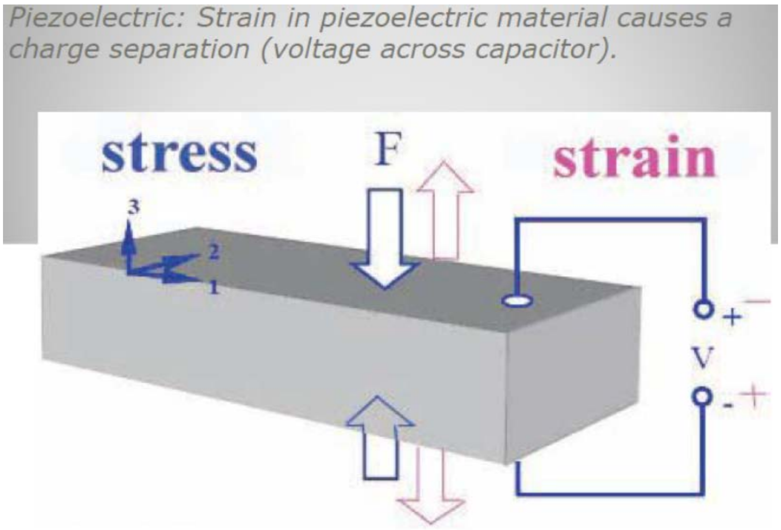


Fig. 10.42. Principle of electric energy generation based on piezoelectric.

Vibration energy harvesting based on piezoelectric materials can potentially be applied to power supplies for low-power electronic devices, such as structural health monitoring systems, wireless devices and semi-active vibration control systems. Desirable electromechanical coupling characteristics of piezoelectric materials facilitate the conversion of mechanical energy to electrical energy. Some non-linear techniques for piezoelectric voltage can be used to enhance the electromechanical conversion (Fig. 10.43).

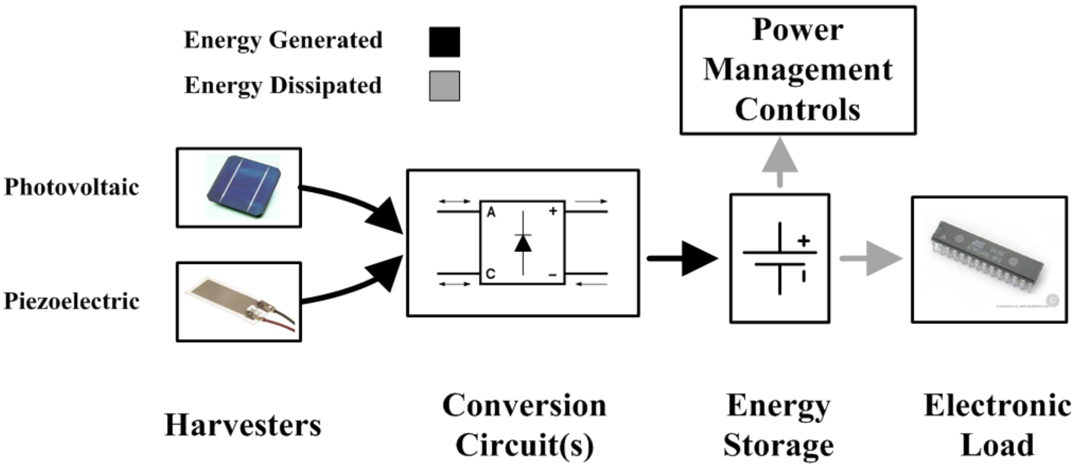


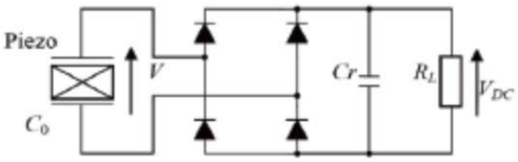
Fig. 10.43. Scheme of energy harvesting.

The principle of piezoelectric energy harvesting. The power generated by a piezoelectric element cannot to electrical energy. Some non-linear techniques for piezoelectric interface is necessary for the energy harvesting system to ensure the voltage is compatible with electric load or energy storage element. There are amny different schemes for an electric interface developed for the energy harvesting such as an AC-DC rectifier or a voltage doubler [65, 66]. Let's consider the principle of piezoelectric energy harvesting system based on the most commonly used interface, "standard interface", which simply rectifies the AC voltage to DC voltage. The circuit of a piezoelectric energy harvesting system based on the standard interface is shown in Fig, 10.44(a) [67].

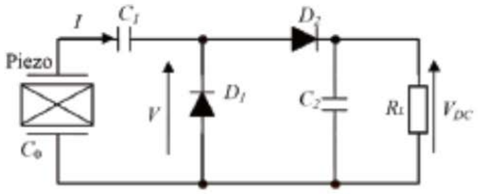
As shown in Fig, 10.44(a), the standard interface includes a diode rectifier and a filter capacitor. The terminal electric load is modeled by an equivalent resistor R_L. To calculate the power output of the interface under the condition of single mode vibration, some assumptions should be made. First, the mechanical displacement u must be purely sinusoidal and the open-circuit voltage V on the piezoelectric must also be sinusoidal. However, the rectifier is in the blocked state and the piezoelectric element is on the open-circuit state when the absolute value of V is lower than V_{DC}, the voltage across the capacitor. When the absolute value of V is above V_{DC}, a current I flow through the diodes. The current is divided into two parts, one to the capacitor and the other to the load. The energy E generated in the period T can be estimated by calculating the integration of product of V and I during a period, therefore the power can be obtained by dividing E by T and expressed as

$$P = \frac{4\alpha^2 u_m^2 \omega^2 R_L}{(2R_L C_0 \omega + \pi)^2}, \tag{35}$$

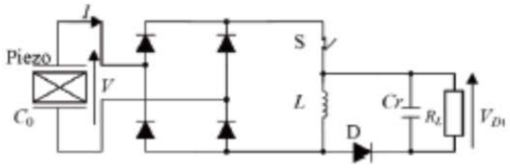
where u_m is the displacement amplitude of mechanical vibration. The other three popular interfaces are the voltage double interface [68], a synchronous electric charge extraction (SECE) [69], and synchronized switch harvesting on inductor (SSHI) [65, 69] are shown in Fig. 10.44(b-d).



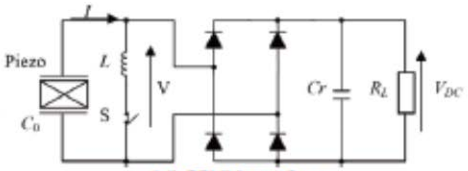
(a) Standard interface



(b) Voltage double interface



(c) SECE interface



(d) SSHI interface

Fig. 10.44. For electrical interfaces for energy harvesting. SECE: synchronous electric charge extraction, SSHI: synchronized switch harvesting on inductor.

A piezoelectric transformer is a type of AC voltage multiplier. Unlike a conventional transformer, which uses magnetic coupling between input and output, the piezoelectric transformer uses acoustic coupling. An input voltage is applied across a short length of a bar of piezoceramic material such as PZT, creating an alternating stress in the bar by the inverse piezoelectric effect and causing the whole bar to vibrate. The vibration frequency is chosen to be the resonant frequency of the block, typically in the 100 kilohertz to 1 megahertz range. A higher output voltage is then generated across another section of the bar by the piezoelectric effect. Step-up ratios of more than 1000:1 have been demonstrated. An extra feature of this transformer is that, by operating it above its resonant frequency, it can be made to appear as an inductive load, which is useful in circuits that require a controlled soft start [70]. These devices can be used in DC-AC inverters to drive cold cathode fluorescent lamps. Piezo transformers are some of the most compact high voltage sources.

The application of piezoelectric materials for energy harvesting has been realized at creation of high voltage and power sources (Fig. 10.45).

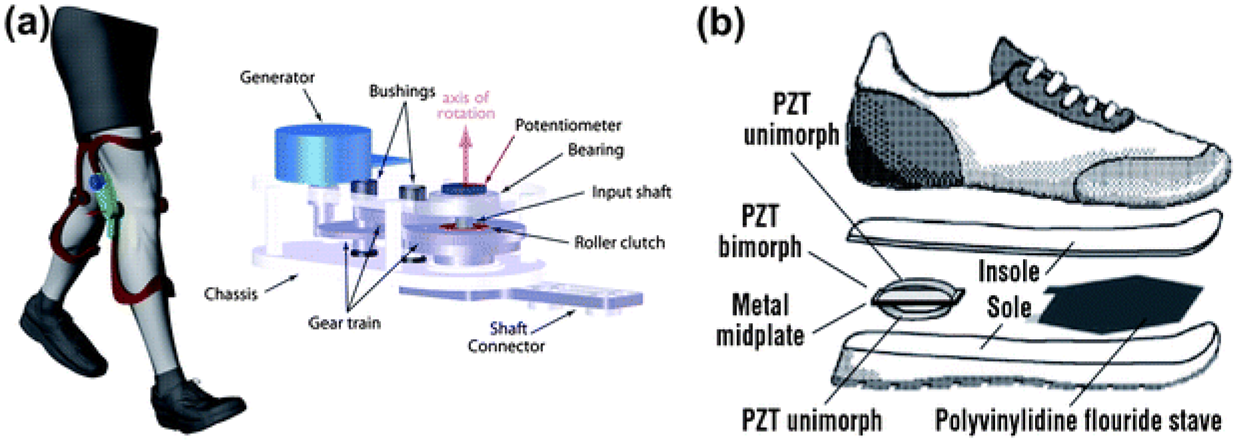


Fig. 10.45. Examples of energy harvesters based on piezoelectric: (a) piezoelectric knee-joint harvester, (b) piezoelectric generator embedded in boots.

Direct piezoelectricity of some materials can generate potential differences of thousands of volts. The best-known application is the electric cigarette lighter: pressing the button causes a spring-loaded hammer to hit a piezoelectric crystal, producing a sufficiently high voltage electric current that flows across a small spark gap, thus heating and igniting the gas. The portable sparkers used to ignite gas stoves work the same way, and many types of gas burners now have built-in piezo-based ignition systems.

A similar idea has been used at creation of battle field equipment with piezoelectric generators embedded in soldiers' boots. However, these energy harvesting sources by association have an impact on the body. The effort to harness 1–2 watts from continuous shoe impact while walking were abandoned due to the impracticality and the discomfort from the additional energy expended by a person wearing the shoes. Other energy harvesting ideas include harvesting the energy from human movements in train stations or other public places [71, 72] and converting a dance floor to generate electricity [73]. Vibrations from industrial machinery can also be harvested by piezoeletric materials to charge batteries for backup supplies or to power low-power microprocessors and wireless radios [74].

Small-scale mechanical energy harvesting has been studied intensively and extensively for more than a decade, with a number of potential applications in wireless sensors and electronic devices. Large-scale harvesting at 1 W–100 kW or even higher could lead to meaningful solution to the energy crisis. It can also find important applications in self-power active or semi-active vibration controls. Relatively, more challenges have been encountered in large-scale mechanical energy harvesting technology, with many questions to be answered. Although various sources of mechanical waste energy have been identified and studied for potential energy harvesting, it is important to further explore such sources. For practical applications, efficiency of vibrational piezoelectric energy harvesting has become a concern. Addressing this problem requires more and higher efficient transducers, understanding of motion mechanisms and sophisticated power electronic circuits. While waiting for the emergence of new piezoelectric materials with higher performances might be unrealistic at the moment, it is highly potential to develop novel physical or geometrical configurations with the materials that are currently available. For example, the optimization of electromagnetic motor should be paid more attention. If the back electromotive force coefficient ke is increased, the efficiency can be significantly enhanced. Motion mechanism is also a very important factor to improve the efficiency. It is understood that partial energy is lost when the kinetic energy is transformed into electric energy, such as eddy current and friction. In addition, parasitic voltage drops, the inherent resistances of the electronic components and the power consumption by the controller, are also responsible for power loss. Strategies to reduce such voltage drops and the parasitic power consumption should be further studied. Furthermore, large-scale vibrations are usually in irregular ways at time-varying frequencies and in most cases at low velocities, which all make it difficult to harvest them at acceptably high efficiencies. Therefore, it still cannot be an option for efficient power takeoff technology. In this respect, the development of novel mechanical motion rectifiers that can convert irregular oscillatory vibrations into regular unidirectional rotations is of specific significance, which should be a direction of the future research in this area.

Vibration control is another specific challenge closely associated with the efficiency of mechanical energy harvesting, especially for large-scale applications. For small-scale energy harvesting, the vibration of the host structure is not a concern. However, when a vibration energy harvesting is at a large-scale, it is necessary to consider various issues, for instance, the protection of the mechanical systems, the human occupants and passengers during the vibration. As a result, many techniques for small-scale energy harvesting may not be applicable to the large-scale harvesting. For example, mechanic resonance can be used to magnify the vibration and thus improve the output power of the harvester in small-scale harvesting. In contrast, it becomes much difficult in large-scale applications, e.g., regenerative vehicle suspensions. Therefore, the ability of simultaneous vibration control and energy harvesting should be considered for each component and advanced control algorithms are required.

In addition, the motion mechanism should be efficient in driving and driven modes, the power electronics should be able to withdraw energy from the mechanical systems and also to inject energy into them if necessary, and the control algorithms should be able to regulate the electrical output voltages and control the regenerated electrical currents or voltages simultaneously. Because large-scale mechanical energy harvesting is not just a problem of design, it is multiple disciplinary problem and thus requires system level approaches, which involves structure dynamics and vibration, mechanical structures and design, power electronics, materials development and system controlling. Nevertheless, waste mechanical energy harvesting will be one of the most important energy technologies in the near future.

10.5.3. Piezoelectric Actuators and Motors

Actuators. An actuator accepts energy and produces movement (action). The energy supplied to an actuator might be electrical or mechanical (pneumatic, hydraulic, etc.). An electric motor and a loudspeaker are both actuators, converting electrical energy into motion for different purposes.

Piezo actuators make use of the deformation of the piezoelectric material when an electric field is applied (Fig.10.46). As very high electric fields correspond to only tiny changes in the width of the crystal, this width can be changed with better-than-µm precision, making piezo crystals the most important tool for positioning objects with extreme accuracy — thus their use in actuators. Multilayer ceramics, using layers thinner than 100 µm, allow reaching high electric fields with voltage lower than 150 V. These ceramics are used within two kinds of actuators: direct piezo actuators and amplified piezoelectric actuators. While direct actuator's stroke is generally lower than 100 µm, amplified piezo actuators can reach millimeter strokes. In loudspeaker voltages are converted to mechanical movement of a piezoelectric polymer film. Piezoelectric elements can be used in laser mirror alignment, where their ability to move a large mass (the mirror mount) over microscopic distances is exploited to electronically align some laser mirrors. By precisely controlling the distance between mirrors, the laser electronics can accurately maintain optical conditions inside the laser cavity to optimize the beam output. A related application is the acousto-optic modulator, a device that vibrates a mirror to give the light reflected off it a Doppler shift. This is useful for fine-tuning a laser’s frequency. Atomic force microscope and scanning tunneling microscopes employ converse piezoelectricity to keep the sensing needle close to the probe.

Fig. 10.46. Metal disk with piezoelectric disk attached, used in a buzzer.

The high resolution of the change in length and large mechanical load capacity of piezoelectric ceramic are of particular interest in high-tech fields (semiconductors, optics and telecommunications, etc.), and also, to an increasing extent, in the automotive field (fuel injection systems), for micropumps, pneumatic valve technology and vibration damping.

Multilayer linear actuators. Ceramic multilayer linear actuator products harness the full potential of thepiezoelectric effect at very low operating voltages. Multilayer piezoelectric actuators are co-fired multilayer ceramics typically with a height up to 2-3 mm and with up to 100 ceramic layers. Multilayer piezoelectric stack actuators are made of two or several single actuators glued together. The purpose of the stacking is to obtain more displacement than can be achieved by a single actuator.

Multilayer bending actuators. Ceramic multilayer bender products harness the full potential of the piezoelectric effect at very low operating voltages. Multilayer piezoelectric benders are single co-fired ceramic components with ceramic layers and internal electrodes configured as to generate a bending mode (Fig. 10.47).

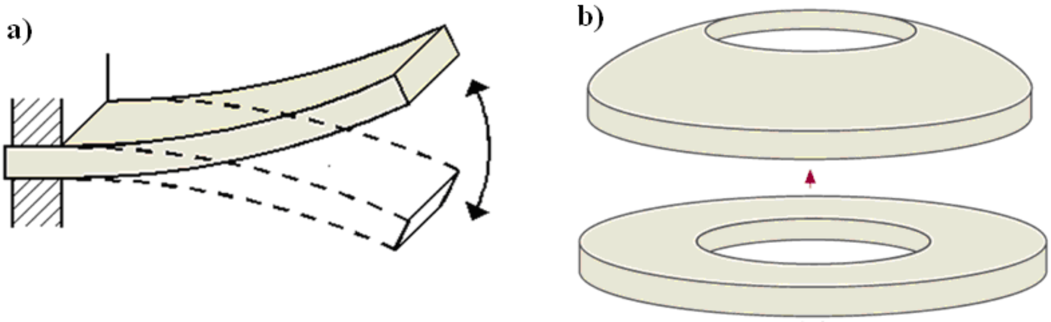


Fig. 10.47. Benders: (a) plate principle and (b) ring principle.

Shear actuators. Based on the piezoelectric shear mode, shear actuators can be made more compact and have higher resonant frequencies than conventional piezo actuators. In addition, their low capacitance results in reduced power requirements for a given generated displacement in dynamic operation.

The varieties of piezoelectric microactuators have been developed. Among them are ink droplet ejectors (printhead), piezoelectric transformers, piezoelectric scanning tunneling microscope tip, etc. The main performance criteria of actuators are generative force/momentum, displacement and frequency response.

Piezoelectric actuators started being commercially designed and manufactured about four decades ago, spreading our days into an ever increasing spectrum of applications related to positioning and motion control. An exhaustive list would be very long in various areas:

Precision Mechanics: *Stimulation or active cancellation of vibrations; Precision tool adjustment; Micro and nanopositioning stages; Rotary and linear piezo drives (Inch-Worm, Stick-Slip, Ultrasonic Motors); Microgrippers; Microvalves.*

Robotics: *Micro/nanomanipulators; Mobilemicrorobots.*

Microtechnology: *Microlithography; Wafer and mask alignment; Inspection system, e.g. profilometers.*

Optics, Nanometrology: *Image stabilization; Auto focus systems; Scanning mirrors; Adaptive optics; Interferometry; Scanning microscopy (AFM, STMetc.); Optic fiber alignment.*

Bio-Medical: *Living cell and genemanipulation; Microdispensers; Micro- to pico-liter pumps.*

Piezoelectric devices can perform actuation and micropositioning tasks with resolutions from micrometers down to sub-nanometer scale. Blocking force can range from several thousand Newtons (stacked actuators) to less than microNewtons (AFM, piezocantilever, etc.). In static operation the power consumption is extremely small while in dynamic operation, settling time can be far under a millisecond.

PiezoMEMS actuators started to emerge for several years, usually made of deposited films of AlN, $BaTiO_3$ or PZT ($Pb[Zr,Ti]O_3$). New single-cristal materials such as PMN-PT (lead magnesium niobate - lead titanate) and PZN-PT (lead zinc niobate - lead titanate) provide higher piezoelectric properties, energy density and electromechanical coupling factors. Several applications were reported such as in [75 - 77] but they are not yet silicon-integrated. In this trend for functional miniaturization, a new technique for molecular-level bonding and lapping of piezoelectric layers on silicon wafer has been recently developed. Adhesives are eliminated and interface mechanical strength increased. First attempts on composite bimorph cantilevers (Fig. 10.48) are promising. Dimensions were reduced from 15.0 x 2.0 x 0.4 mm (central-bottom in image) to only 6.0 x 0.04 x 0.2 mm.

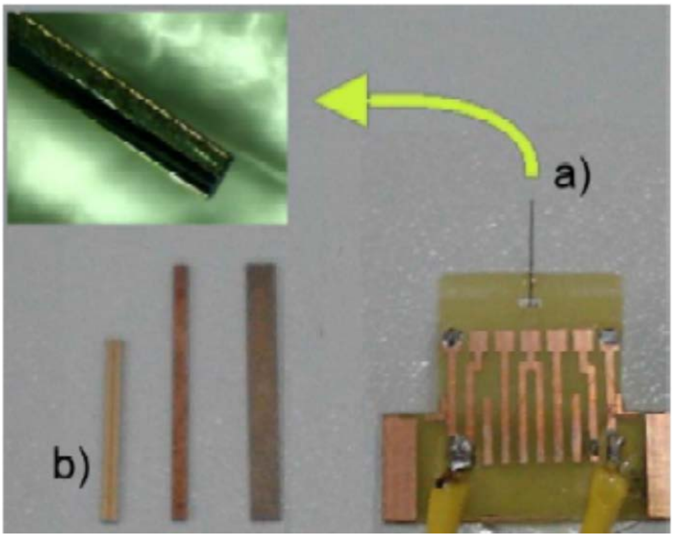


Fig. 10.48. Miniaturised PZT piezo cantilever (a) of 6.0x0.04x0.2mm. Comparison with regular-size actuators (b) [75].

Same examples of actuators and the action are listed below.

Loudspeakers: Voltage is converted to mechanical movement of a piezoelectric polymer film.

Piezoelectric motors: Piezoelectric elements apply a directional force to anaxle, causing it to rotate. Due to the extremely small distances involved, the piezo motor is viewed as a high-precision replacement for the stepper motor.

Piezoelectric elements can be used in laser mirror alignment, where their ability to move a large mass (the mirror mount) over microscopic distances is exploited to electronically align some laser mirrors. By precisely controlling the distance between mirrors, the laser electronics can accurately maintain optical conditions inside the laser cavity to optimize the beam output.

A related application is the *acousto-optic modulator*, a device that scatters light off soundwaves in a crystal, generated by piezoelectric elements. This is useful for fine-tuning a laser's frequency.

Atomic force microscopes and scanning tunneling microscopes employ converse piezoelectricity to keep the sensing needle close to the specimen [78].

Inkjet printers: On many inkjet printers, piezoelectric crystals are used to drive the ejection of ink from the inkjet print head towards the paper.

Diesel engines: High-performance common rail diesel engines use piezoelectric fuel injectors instead of the more common solenoid valve devices.

Active vibration control using amplified actuators.

X-ray shutters.

XY stages for micro scanning used in infrared cameras.

Moving the patient precisely inside active *CT* and *MRI* scanners where the strong radiation or magnetism precludes electric motors.

Crystal earpieces are sometimes used in old or low power radios.

High-intensity focused ultrasound for localized heating or creating a localized cavitation can be achieved, for example, in patient's body or in an industrial chemical process

Piezoelectric motors. A piezo motor is based on the change in mechanical shape of a piezoelectric material when an tension is applied (Fig.10.49). The material produces ultrasonic or acoustic vibrations and produces a linear or rotary motion. Types of piezoelectric motor include the well-known travelling-wave motor used for auto-focus in reflex cameras, inchworm motors for linear motion, and rectangular four-quadrant motors with high power density (2.5 watt/cm^3) and speed ranging from 10 nm/s to 800 mm/s. All these motors work on the same principle. Driven by dual orthogonal vibration modes with a phase shift of 90°, the contact point between two surfaces vibrates in an elliptical path, producing a frictional force between the surfaces. Usually, one surface is fixed causing the other to move. In most piezoelectric motors the piezoelectric crystal is excited by a sine wave signal at the resonant frequency of the motor. Using the resonance effect, a much lower voltage can be used to produce a high vibration amplitude. When we applied different driving voltage, frequency and phase angle to the composite type piezoelectric actuating element, so that the rotor can produce clockwise or counterclockwise rotation (Fig. 10.50) [79]. Stick-slip motor works using the inertia of a mass and the friction of a clamp (Fig. 10.51). Such motors can be very small. Some are used for camera sensor displacement, thus allowing an anti-shake function .

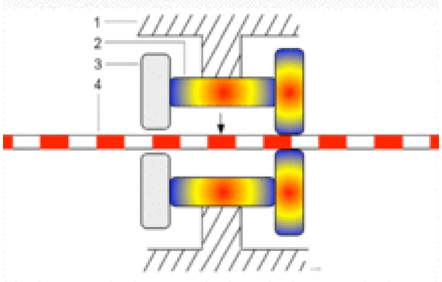


Fig.10.49. Piezoelectric motor.

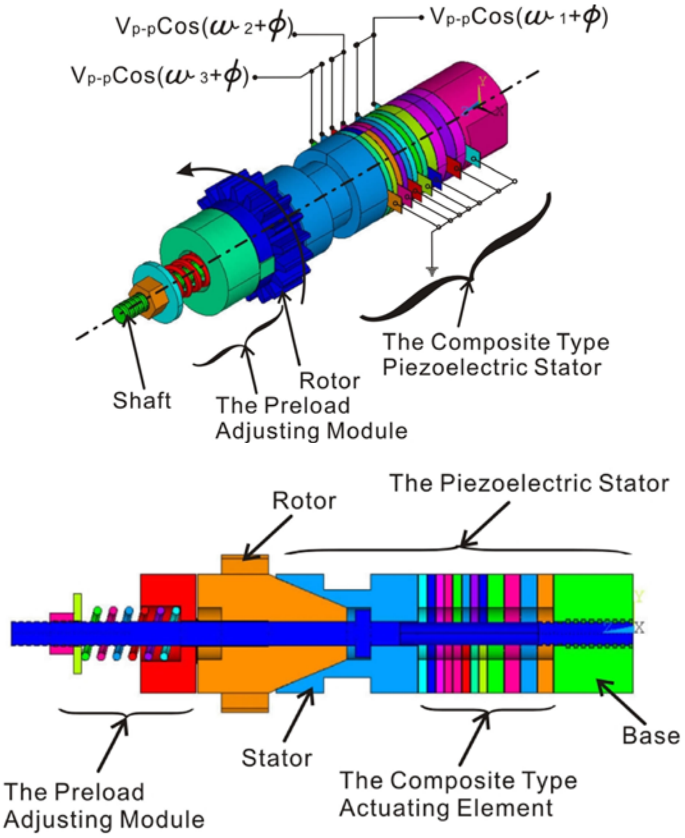


Fig. 10. 50. The composition and operation principle of the composite type piezoelectric motor [79].

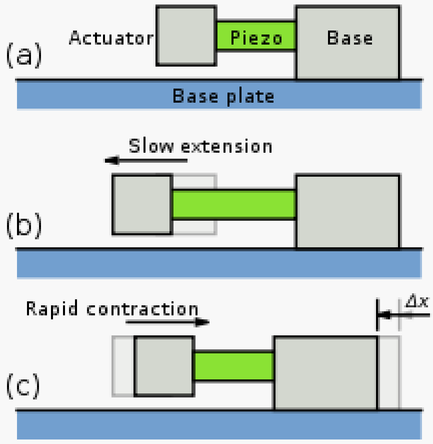


Fig. 10.51. A slip-stick actuator.

10.5.4. Piezoelectric Acustic Transducers

A transducer is a device that converts a signal in one form of energy to another form of energy [80]. Energy types include (but are not limited to) electrical, mechanical, electromagnetic (including light), chemical, acoustic and thermal energy. While the term *transducer* commonly implies the use of a sensor/detector, any device which converts energy can be considered a transducer. Transducers are widely used in measuring instruments.

Ultrasonic transducers (10kHz to 70kHz) are used in imaging and high power applications. The transducer can act as a sensor and an actuator in imaging applciations. In high power applications such as sonochemistry or medical treatment, the transducer sends ultrasound waves in the body and detect the returning echo.

Electromagnetic acoustic transducers (EMATs). The method of ultrasound generation is the use of electromagnetic acoustic transducers (EMATs) (Fig. 10.52) [81]. When a wire carrying current at the desired ultrasonic frequency is placed near the surface of a conductor, eddy currents are induced in the region near the surface of the conductor. In the presence of a static magnetic field, these eddy currents experience a Lorentz force given by:

$$\vec{F} = \vec{J} \times \vec{H} \qquad (36)$$

where ***F*** is the body force per unit volume, ***J*** is the induced current density, and ***H*** is the static magnetic field. This results in ultrasound generation at the frequency of the current flowing through the wire. EMATs do not require any couplant for ultrasound generation. This provides a non-contact method for ultrasound generation at elevated temperatures and in remote locations. However, it has very low efficiency.

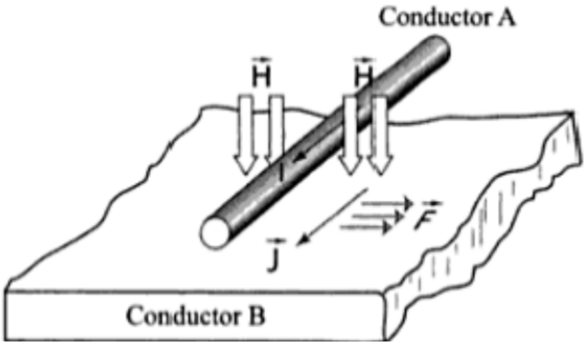


Fig. 10.52. Schematic of an elementary EMAT [82].

Piezoelectric transducers. The most common ultrasound generation technique uses piezoelectric transducers [83]. These transducers consist of a piezoelectric ceramic, which is cut into various shapes to produce different wave modes (Fig. 10.53). Before the development of piezoelectric ceramics, quartz crystals and magnetostrictive materials were used. Piezoelectric ceramics have become the dominant material for transducers because of good piezoelectric properties and ease of manufacturing various shapes and sizes. Piezoelectric ceramics also require low drive voltages and can be used up to 300°C. The operating frequency of the piezoelectric devices depends on the thickness of the piezoelectric element. This can be a limiting factor in applications where restrictions on the PZT strip thickness conflict with those necessary to achieve the required frequency. In addition, high frequency and three-dimensional (3D) imaging require closely spacedand extremely thin piezoelectric elements which are difficult to manufacture using the available macroscale fabrication processes [84, 85].

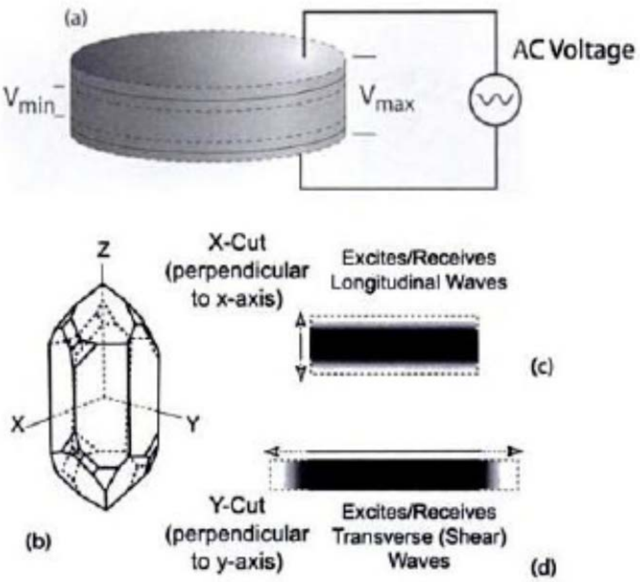
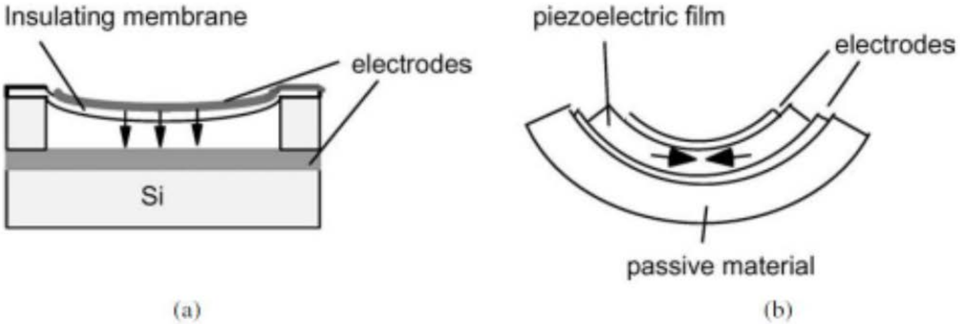


Fig. 10.53. Response of a piezoelectric disk to an alternating voltage. (a) Piezoelectric crystal, (b) axis definition and (c) X cut vibrations are longitudinal and (d) Y- cut vibrations are transverse [82].

Micromachined ultrasonic transducers (MUTs). With the development of microfabrication processes, micromachined ultrasonic transducers (MUTs) have emerged as an alternative approach for ultrasound generation and detection, especially for biomedical imaging applications [86]. Micromachined ultrasonic transducers consist of micromachined multilayered membrane resonators that are ten’s of micrometers in width or diameter, and a few micrometers in thickness. The kinetic energy stored in the membranes is low because of the negligible mass of these membranes. This results in a higher bandwidth and a lower acoustic impedance mismatch between the transducer and the human body [87]. There are two common types of MUTs based on their actuation principle: capacitive MUTs (CMUTs) and piezoelectric MUTs (PMUTs) (Fig. 10.54).



(a)
(b)

Fig. 10.54. Schematic of cross-section of (a) CMUT and (b) PMUT [88].

Capacitive MUTs utilize a vibrating membrane that is electrostatically actuated. The performance of CMUTs equals that of the conventional piezoelectric transducers in terms of the bandwidth, the dynamic range, the electromechanical coupling coefficient, and the generated sound pressure [89 - 91]. The coupling coefficient is defined as the ratio of the mechanical energy delivered by the transducer to the total energy stored in the transducer. Imaging results from both one-dimensional (1D) and two-dimensional (2D) CMUT arrays have been reported [92]. Although CMUTs have a high coupling coefficient [93], various operational and fabrication considerations limit the achievable coupling. The voltage bias necessary to achieve the maximum coupling is near the pull-in voltage of the CMUT elements, making it difficult to operate without causing any collapse of the elements [87].

Piezoelectric MUTs use the piezoelectric effect to vibrate the membranes. The PMUT consists of multilayered membranes one of which is a piezoelectric film. When a drive voltage is applied, the piezoelectric layer strains and generates the necessary deflection. This, in turn, generates ultrasound waves in the surrounding medium. Unlike conventional transducers, the operating frequency of the PMUT is not determined by the thickness of the piezoelectric layer; it is determined by the size and the material properties of the various layers in the membrane [94]. The acoustic impedance and the coupling coefficient of the PMUTs can also be controlled through the device design [86]. PMUTs also have high sensitivity, large displacement to voltage ratio, large pressure-to-voltage ratio, and large capacitance that cannot be realized with the conventional transducers [95]. PMUTS are less sensitive to parasitic capacitances than CMUTs because of the higher capacitance of PMUTs. Moreover, the lower electrical impedance of PMUTs allows a better matching to the readout circuit. Also, PMUTs are less sensitive to fabrication inconsistencies because the operation of PMUTs does not require a DC bias [96].

10.5.5. Other piezoelectric applications

Frequency standards. Quartz clocks employ a tuning a crystal oscillator made from quartz that uses a combination of both direct and converse piezoelectricity to generate a regularly timed series of electrical pulses that is used to mark time. The quartz crystal (like any elastic material) has a precisely defined natural frequency (caused by its shape and size) at which it prefers to oscillate, and this is used to stabilize the frequency of a periodic voltage applied to the crystal. □The same principle is critical in all radio transmitters and receivers, and in computers where it creates a clock pulse. Both of these usually use a frequency mutliplier to reach the megahertz and gigahertz ranges.

Reduction of vibrations and noise. Different teams of researchers have been investigating ways to reduce vibrations in materials by attaching piezo elements to the material. When the material is bent by a vibration in one direction, the vibration-reduction system responds to the bend and sends electric power to the piezo element to bend in the other direction. Future applications of this technology are expected in cars and houses to reduce noise. Further applications to flexible structures, such as shells and plates, have also been studied for nearly three decades. Piezoelectric ceramic fiber technology is being used as an electronic damping system on some tennis rackets.

Surgery. A recent application of piezoelectric ultrasound sources is piezoelectric surgery, also known as piezosurgery [97]. Piezosurgery is a minimally invasive technique that aims to cut a target tissue with little damage to neighboring tissues. For example, it has been reported the use of piezoelectric ultrasound sources in hand surgery for the cutting of bone, using frequencies in the range 25–29 kHz, causing microvibrations of 60–210 μm [98]. It has the ability to cut mineralized tissue without cutting neurovascular tissue and other soft tissue, thereby maintaining a blood-free operating area, better visibility and greater precision [99].

Photovoltaics. The efficiency of a hybrid photovoltaic cell that contains piezoelectric materials can be increased simply by placing it near a source of ambient noise or vibration. The effect was demonstrated with organic cells using zinc oxide nanotubes. The electricity generated by the piezoelectric effect itself is a negligible percentage of the overall output. Sound levels as low as 75 decibels improved efficiency by up to 50 percent. Efficiency peaked at 10 kHz, the resonant frequency of the nanotubes. The electrical field set up by the vibrating nanotubes interacts with electrons migrating from the organic polymer layer. This process decreases the likelihood of recombination, in which electrons are energized but settle back into a hole instead of migrating to the electron-accepting ZnO layer [100, 101].

Transportation Industry. The property of piezoelectric ceramics to withstand harsh environmental conditions present in automotives made it a candidate material for application in all branches of transportation industry. The following are example applications of piezoelectrics in passenger cars.

Knock sensors are placed near the engine in order to detect irregular combustions [102]. The measurement principle is the one also used in accelerometers. The piezoelectric material is placed between the vibrating structure and a seismic mass introducing the vibration forces into the piezo element. The piezo element itself converts the vibrations into an electric charge proportional to the applied force. Usually, piezoelectric ceramics (PZT) with specially tailored properties are used. The material has to withstand high temperatures (up to 200°C) as well as rapid temperature changes. Also, the piezoelectric coefficient of the material must be almost independent of the temperature and remain stable over the vehicle's lifetime. Only recently, first attempts were made to replace PZT by thin PVDF foil sensors [103].

Distance sensors are ultrasonic transducers used in vehicles as so called parking pilots. During backing, the transducer emits ultrasonic waves reflected by the obstacle and then, in turn, changed into an electrical system by the same transducer now acting as an ultrasonic sensor. From the travelling time the distance is calculated [102]. Fuel injection systems based on piezoelectric stack actuators set another milestone in the piezo – technology [104]. The advantage compared to the conventional magnetic systems was the much faster response of the piezoelectric actuators and the possibility to drive any desired profile of valve displacement. It was not only that this technology reduced the fuel consumption by more than 20%, it was also that piezoelectric stack actuator technology as a whole experienced a decisive push [105]. One key issue the early development phases was the question of the reliability of the actuators under harsh engine environments [106].

Potential applications. In recent years, several attempts at the macro-scale application of the piezoelectric technology have emerged [107 - 109] to harvest kinetic energy from walking pedestrians. The piezoelectric floors have been trialed since the beginning of 2007 in two Japanese train stations, Tokyo and Shibuya stations. The electricity generated from the foot traffic is used to provide all the electricity needed to run the automatic ticket gates and electronic display systems [110]. In London, a famous nightclub exploited the piezoelectric technology in its dance floor. Parts of the lighting and sound systems in the club can be powered by the energy harvesting tiles [111] However, the piezoelectric tile deployed on the ground usually harvests energy from low frequency strikes provided by the foot traffic. This working condition may eventually lead to low power generation efficiency [112].

In this case, locating high traffic areas is critical for optimization of the energy harvesting efficiency, as well as the orientation of the tile pavement significantly affects the total amount of the harvested energy. A Density Flow evaluation is recommended to qualitatively evaluate the piezoelectric power harvesting potential of the considered area based on the number of pedestrian crossings per unit time [112]. The potential application of a commercial piezoelectric energy harvester in a central hub building at Macquarie University in Sydney, Australia was examined. Optimization of the piezoelectric tile deployment is presented according to the frequency of pedestrian mobility and a model is developed where 3.1% of the total floor area with the highest pedestrian mobility is paved with piezoelectric tiles. The modelling results indicate that the total annual energy harvesting potential for the proposed optimized tile pavement model is estimated at 1.1 MW h/year, which would be sufficient to meet close to 0.5% of the annual energy needs of the building [113].

Examples of application. ☐

- Acoustic components: *Microphones, Ultrasonic detector, Hydrophones, Sonars.*
- ☐Electrical equipments:*Switches, Miniature electric fans.*

- ☐Medical instrumentation: *Catheter, Pedobarography, Osteogenesis, Medical echography, Blood pressure detector.*
- Robotics: *Artificial sensitive skin, Pressure sensors.*
- ☐Security devices:*Intruder alarm, IR alarms, Vibration sensors.*
- ☐Optical devices:*Laser diameter measurement, Variable mirrors.*
- ☐Transports:*Automotive detection on roads, Weighing vehicles.*
- ☐Pressure pick-ups: *Distribution of pressure on surfaces, Localization of impacts, Pick-ups for non destructive testing, Accelerometer, Keyboards.*
- ☐Sport and leisure: *Tennis court line, Display panel on sport grounds, Impact detector if shooting targets.*

10.6. Conclusions.

The background information on piezoelectric materials, including an introduction to the piezoelectric effect and application of piezoelectric materials has been presented. The wide range applications give us an idea about the growing popularity of these materials in various fields. Piezoelectric materials are more and more embedded in complex devices and high tech systems that make whole economies exist and function in an efficient way. Piezoelectric materials are among the materials that are widespread around us. Mobile phones, automotive electronics, medical technology, and industrial systems are only a few areas where piezoelectric components are indispensable.

Chapter 11. FERROELECTRIC MATERIALS

11. 1. INTRODUCTION

11.1.1. Basic Concepts

Ferroelectrics

Ferroelectrics are a class of polar dielectrics characterized by a reversible spontaneous polarization [1, 2]. Ferroelectrics possess a unique set of physical properties, that makes them attractive for application in a variety of electronic devices, such as nonvolatile memories and microelectromechanical systems. Their piezoelectric, electrooptic, and nonlinear optical properties are largely determined by arrangement of ferroelectric domains. Domains form during phase transition to relieve mechanical stress associated with transformation strains and to minimize the depolarization energy associated with the spontaneous polarization. The presence of the domain structure is one of the fundamental properties of ferroelectrics. Domain arrangements reflect the effects of sample conductivity, mechanical strains, and imperfections such as dislocations, vacancies, and impurities. Polarization reversal under an external electric field proceeds via domain nucleation and motion of domain walls.

A ferroelectric is a solid that exhibits a *spontaneous* dielectric polarization, P_S (when $T < T_C$, the Curie temperature), that can be *reversed* by the external application of an electric field, E. The spontaneous (remanent) polarization at $E = 0$ can have one of two values, $\pm|P_R|$, depending on the sign of E just before it was switched off. We define these quantities with respect to the dielectric hysteresis curve (Fig. 11.1). The spontaneous polarization is defined as the dipole moment per unit volume and its axis is usually a high symmetry direction of the crystal structure. Of the 20 noncentrosymmetric, piezoelectric crystal classes, only 10 (those with C_n, C_{nv}, and C_{1h} point symmetries) have a unique, “polar” axis because their primitive cell has a nonzero dipole moment. These are the 10 pyroelectric crystal classes. A crystal belonging to any of these classes *may* show ferroelectricity provided the polarization is reversible. The usual signatures of a ferroelectric system are: (a) a peak in the temperature dependence of the low frequency dielectric response function (ε) at $T \approx T_C$, and (b) a hysteresis in the P–E curve for $T < T_C$. In most ferroelectrics, the temperature dependence of ε_0 above the Curie point (i.e., in the paraelectric phase) can be described fairly accurately by the Curie–Weiss law,

$$\varepsilon = \varepsilon_0 + C/(T - T_0) \text{ for } T > T_0 \qquad (11.1)$$

where C is the Curie–Weiss constant and T_0 is the Curie– Weiss temperature. In general, a spatially coherent alignment of electric dipoles occurs only in limited regions (“domains”) of the crystal, such that a ferroelectric crystal normally has a polydomain structure reminiscent of a ferromagnet. However, such domains may be switched in the same direction (“poled”) by the application of a suitable external field. In an antiferroelectric, on the other hand, the ground state consists of an antiparallelly ordered array of local dipoles, which results in the net polarization being zero.

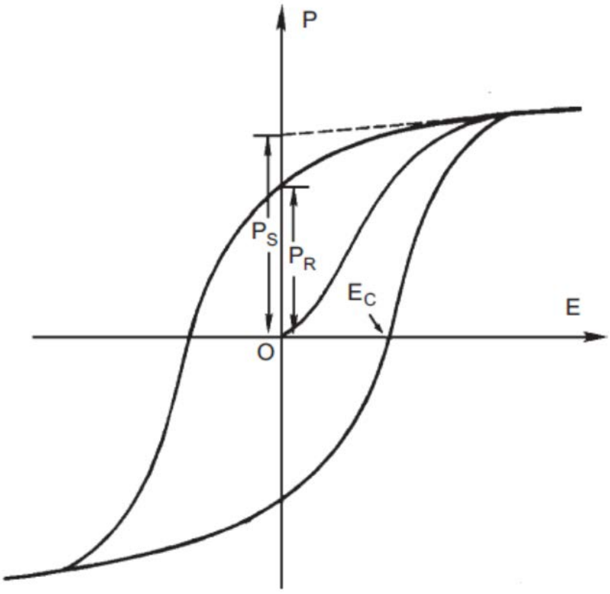


Fig. 11.1. The dielectric hysteresis curve (polarization vs applied electric field) for a typical ferroelectric.

The paraelectric to ferroelectric phase transition may be of the second order ($P_S \rightarrow 0$ continuously as $T \rightarrow T_C$, although dP_S/dT is discontinuous) or of the first order, in which case P_S has a discontinuity at T_C. Similar continuous or discontinuous behaviors are also exhibited by ε in the two cases. Conventionally, a ferroelectric is termed “displacive” when the elementary dipoles strictly vanish in the paraelectric phase, and “order–disorder” when these dipoles are nonvanishing but thermally average out to zero in the paraelectric phase. In a displacive transition, the displacements of the atoms in the low-temperature ferroelectric phase from their corresponding positions in the high-temperature paraelectric phase are small compared to the internuclear distances. Just below T_C such displacements can indeed be infinitesimally small in the case of a second order displacive phase transition. However, in an order–disorder transition, the displacements across T_C are comparable to the internuclear distances. Though it is often possible to rigorously distinguish between the two types of systems on the basisof the dynamics of their phase transition and the nature of the soft mode involved (whether propagating or diffusive), it is difficult to classify certain ferroelectrics and antiferroelectrics strictly as “displacive” or “order–disorder.”

The basic physics and phenomenological aspects of ferroelectricity are covered by a number of textbooks [3-7]. Some of the more common ferroelectrics and antiferroelectrics (out of more than a thousand known ones) are listed in Table 11.1.

Table 11.1. List of some common ferroelectrics (F) and antiferroelectrics (A).

Chemical formula	Structural family	*F/A*	T_C (℃)
$BaTiO_3$	perovskite	F	120, 5, -90
$PbTiO_3$	perovskite	F	490
$PbZrO_3$	perovskite	A	235
$NaNbO_3$	perovskite	A	354
$KNbO_3$	perovskite	F	435, 225, -10
$BiFeO_3$	perovskite	F	850
$LiNbO_3$	$LiNbO_3$ type	F	1210
$LiTaO_3$	$LiNbO_3$ type	F	665
$PbNb_2O_6$	tungsten bronze	F	570
$PbTa_2O_6$	tungsten bronze	F	260
$Sr_{0.6}Ba_{0.4}Nb_2O_6$	tungsten bronze	F	75, -213
$Ba_2NaNb_5O_{15}$	tungsten bronze	F	560, 300
KH_2PO_4	KDP	F	-150
$(NH_4)H_2PO_4$	KDP	A	-125
$(NH_2CH_3CO_2H)_3\ H_2SO_4$	TGS	F	49
$NaNO_2$	nitrite	F	163

Note: The third column lists the ferroelectric/antiferroelectric transition temperature as well as lower transitions

Size Effects in Ferroelectrics

The study of finite size effects in ferroelectric nanoparticles and thin films is an important contemporary problem. Though ferroelectric materials (which are also, by definition, piezoelectric) have been widely used as bulk capacitors and transducers, there are a number of interesting current and futuristic applications that involve these materials in microscopic form. These include sensors of various types and nanorobotic and microelectromechanical devices. In addition, interest in ferroelectrics is growing rapidly because its high dielectric constant can be utilized in dynamic random access memories (D-RAM), while its capacity for being polarized in opposite directions makes it an attractive candidate for nonvolatile random access memories (NV-RAM) and electrically erasable programmable read-only memories (EEPROM). An important motivation for the study of size effects in ferroelectrics is to determine the ultimate level to which a device based on such systems can be miniaturized, since certain properties may get degraded while others are enhanced with a decrease in the characteristic size of the ferroelectric element. From the point of view of basic physics, it is no less interesting to investigate model ferroelectric systems in which the particle size is smaller than the correlation length corresponding to the long range dipolar ordering. For example, the existence of the superparaelectric state (analogous to superparamagnetism in magnetic fine particles) in nanocrystalline ferroelectrics is yet to be established unambiguously.

Experimental Aspects

It is pertinent to make a few brief comments about the experimental techniques involved in these studies. Other than direct observations using electron microscopes and atomic force microscopes, the most popular method for characterizing the particle size is based on X-ray diffraction (XRD) line broadening. Assuming that the small crystallite size is the only cause of line broadening, the coherently diffracting domain size (d_{XRD}) can be related to the full width at half maximum (FWHM) of a particular diffraction peak by the Scherrer equation: $d_{XRD} = K\lambda/\beta(\theta)\cos\theta$, where λ is the X-ray wavelength, $\beta(\theta)$ is the instrument corrected FWHM of the diffraction line, θ is the diffraction angle, and $K \approx 1$ for equiaxed particles. Different aspects of ferroelectric phenomena are governed by different length scales, such as particle size, grain size, ferroelectric domain size, film thickness, and the lateral dimensions of the relevant devices. The important "intrinsic" length scales are the ferroelectric correlation length and the electron mean free path.

The important characteristics of a ferroelectric material are manifested mainly through its dielectric, piezoelectric, pyroelectric, acoustic, optical, electro-optic, and nonlinear optic properties, and a reduction in the crystallite size may affect many of these. Clearly, the ferroelectric T_C of a system may be measured by studying the temperature dependence of any one of a number of properties, such as the dielectric response, tetragonal distortion, mode softening, etc., though there are subtle differences in the information obtained from these distinct properties. It is, therefore, possible for the nature of the "size effect" to actually depend on the method used to observe it. Thus, structural measurements may show that a ferroelectric system has become pseudo-cubic at a particular critical size, while vibrational spectra may still show evidence for a low symmetry structure.

11.1.2. Ferroelectricity in Perovskites

Most of the technologically important ferroelectrics and antiferroelectrics are oxides that possess the well-known perovskite structure. These include $BaTiO_3$, $PbTiO_3$, $PbZrO_3$, $KNbO_3$, and mixed oxides such as $Pb(Zr_{1-x}Ti_x)O_3$ and $(Pb_{1-y}La_y)(Zr_{1-x}Ti_x)O_3$. Since much of the work on finite sized ferroelectrics also relates to this class of compounds, we briefly recapitulate the mechanism of ferroelectric ordering in them. Most perovskites are oxide ceramics with the general formula ABO_3. In their cubic unit cell, the cube corners are occupied by the "A" cations, which may be monovalent (e.g., K^+, Na^+, Ag^+) or bivalent (e.g., Ba^{2+}, Sr^{2+}, Pb^{2+}, Ca^{2+}, Cd^{2+}), while the body center is occupied by the smaller B cations, which are pentavalent (e.g., Nb^{5+} or Ta^{5+}) or tetravalent (e.g., Ti^{4+} or Zr^{4+}). The O^{2-} ions are situated at the face centers (Fig. 11.2a). The perovskite lattice is basically a three-dimensional network of BO_6 octahedra, as shown in Fig. 11.2b. The phenomenon of ferroelectricity is controlled by a delicate interplay between the long-range Coulomb forces that favor the ferroelectrically ordered state and short-range repulsive forces that favor the paraelectric state.

The cubic perovskite structure shown in Fig. 11.2a actually represents the high-temperature ($T > T_C$), high-symmetry (cubic) paraelectric phase. It is the off-centering of the B cation in the oxygen octahedron that leads—through a lowering of the symmetry—to the creation of a local dipole moment, and ultimately to ferroelectric or antiferroelectric ordering. The structural distortion from the high-symmetry (paraelectric) phase is therefore of fundamental importance to the stabilization of ferroelectric order in this class of systems. This concept is central to the understanding of finite size effects in displacive ferroelectrics. In the case of second order transitions of the displacive type, or of relatively "gentle" first order transitions, the static displacements involved in the transition from the high symmetry to the low symmetry structure are similar to the lattice vibrational amplitudes. The link between lattice dynamics and ferroelectricity is the Lyddane–Sachs–Teller equation:

$$\frac{\varepsilon(0)}{\varepsilon(\infty)} = \prod_i \frac{[\omega_{LO}(\vec{k}=0)]_i^2}{[\omega_{TO}(\vec{k}=0)]_i^2} \qquad (11.2)$$

Here $\varepsilon(0)$ and $\varepsilon(\infty)$ are the static and optical dielectric constants, respectively, while ω_{LO} and ω_{TO} are the longitudinal and transverse optical modes, the product being over all optic branches. If any one of the TO modes (the "soft mode") exhibits an anomalous temperature dependence of the type $\omega_{TO}^2=K(T-T_0)$, a behavior that is experimentally verified, then $\varepsilon(0)$ would diverge as $T \rightarrow T_0$, as expected across the ferroelectric transition.

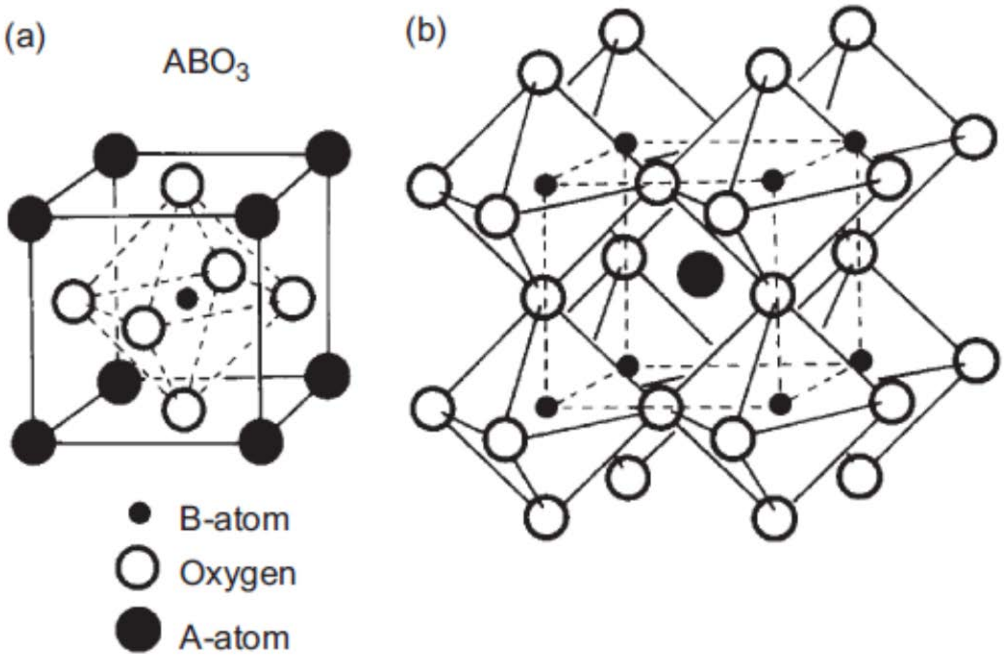


Fig. 11.2. (a) A unit cell of the cubic perovskite (ABO$_3$) structure, (b) The cubic perovskite lattice consisting of a three-dimensional network of BO$_6$ octahedra.

The nature of the soft mode can be visualized rather simply in the case of a typical perovskite (ABO$_3$) compound. The allowed normal modes of vibration (optical modes) are shown in Fig. 11.3. The first three modes are infrared active (i.e., they involve an oscillating dipole moment) while the fourth is "silent" (i.e., neither infrared nor Raman active). Generally, the lowest frequency TO mode is the soft mode (Fig. 11.3a). As $T \to T_0$, the frequency of only this mode decreases continuously and at T_0, the restoring force for this mode vanishes due to a cancellation of the long range (dipole) attractive forces and the short-range repulsive forces. The structure of the low-symmetry phase ($T < T_0$) is determined by the frozen-in soft mode displacements. Referring to Fig. 11.3a, it is clear (since the mode-softening takes place at $k \approx 0$) that a dipole moment would be produced along the z-direction in all the unit cells. Further, the lattice would undergo a *tetragonal distortion*, getting slightly elongated in the direction of the dipole moment and contracted perpendicular to it.

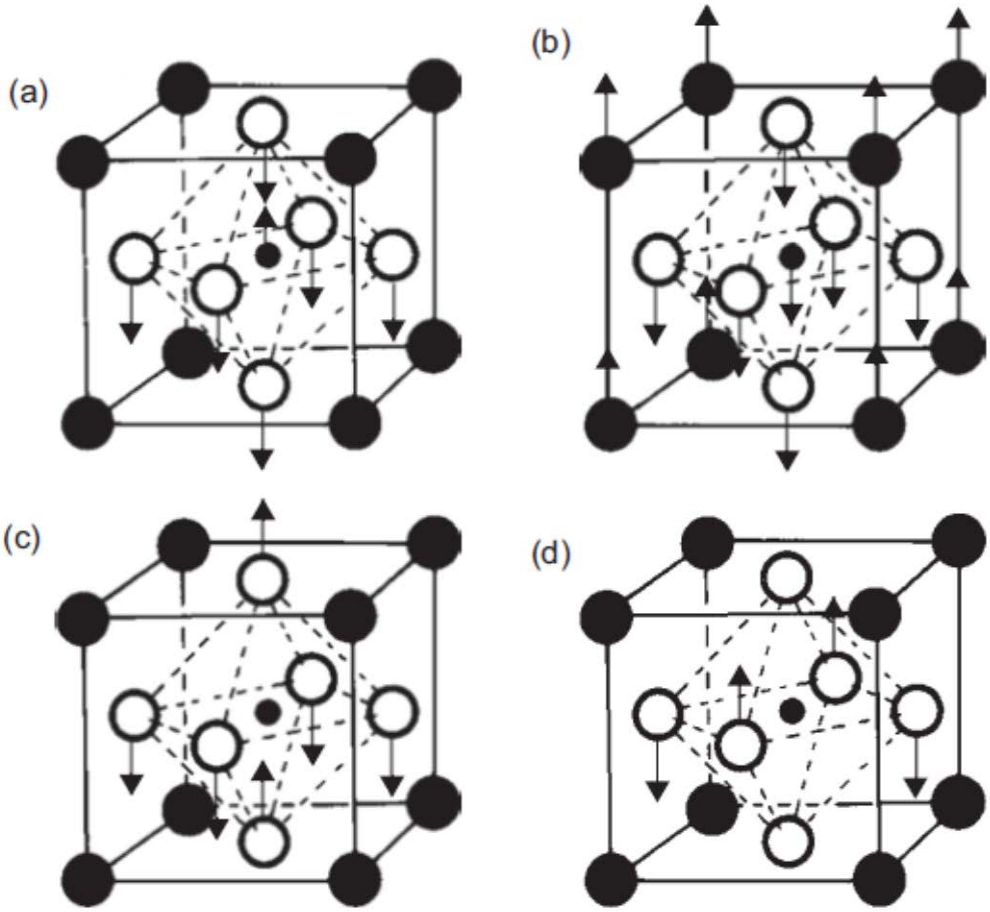


Fig. 11.3. The atomic displacements in the perovskite unit cell (see Fig. 11.2) for the four normal optical modes of vibration. Though only the vibrations along the *z*-direction are shown, each of these modes is actually triply degenerate.

11.1.3. Size Effects in Ferroelectrics. A Historical Perspective

Though nanoparticle physics, in general, picked up momentum only in the 1980s, research on finite size effects in ferroelectric materials started surprisingly early (in the early 1950s) [8]. The first systematic study [8] involved potassium dihydrogen phosphate, KH_2PO_4 (KDP), which is a rather complicated system in that it has characteristics of both displacive and order–disorder mechanisms. In this study of KDP fine particles embedded in an insulating medium with a low dielectric constant, it was found that the spontaneous polarization (which can be considered as the order parameter in ferroelectric systems) was hindered when the particle size was less than 150 nm. Below a critical particle size that depends upon the embedding medium, the depolarization field was thought to prevent polar, ferroelectric structures. Bulk ferroelectric behavior was exhibited only when the particle size was above 0.5 μm.

A contrasting result was obtained in $BaTiO_3$, in which the paraelectric phase (with a cubic perovskite structure) undergoes a tetragonal distortion of about 1% below $T_C \approx 125$ °C to a polar phase. Ball milled fine particles of $BaTiO_3$ in the size range of 0.03–2.3 μm showed a small *increase* in the T_C and a smearing of the transition with decreasing particle size [9, 10]. No critical size (for the destabilization of ferroelectricity) was observed in this size range. The authors concluded from their electron diffraction studies that the deviation from normal ferroelectric behavior was due to the existence of a 10–15 nm thick surface layer that was under a higher tetragonal strain than the bulk and had a *higher* T_C. This layer appeared to remain ferroelectric up to a few hundred degrees higher than the bulk T_C. They suggested that the misfit between the surface layer and the interior caused the broadening of the transition region and the lattice distortions.

The inconsistencies in the early results can be ascribed to the fact that some of the samples were prepared by mechanical grinding, in which case the observations often did not reflect intrinsic size effects but were influenced by residual surface stresses, inhomogeneities, and imperfect crystallization. In spite of such problems, it was able to make the key observation that the spontaneous tetragonal strain (c/a) in the ferroelectric phase decreases with decreasing particle size [8].

During the 1960s and 1970s there was considerable progress in the study of grain size effects in ferroelectric *ceramics* or sintered compacts. Such systems, and the results pertaining to them, differ in certain essential aspects from the quasi-free nanoparticles—which would appear to be more suitable for the study of “intrinsic” size effects. However, since ferroelectric ceramics are technologically important, one cannot ignore the possibility of enhancing some of their properties by controlling the size of the polycrystalline grains. The salient observations on ceramic samples of $(Pb_{1-y}La_y)(Zr_{1-x}Ti_x)O_3$ or PLZT [11], $Pb(Zr_{1-x}Ti_x)O_3$ or PZT [12], and other systems can be summarized as follows. With a decrease in the grain size—generally in the 1–5 μm range—(*i*) there is a decrease in the peak dielectric constant ε_{max}, (*ii*) the peak of the dielectric response near T_C becomes broader, and (*iii*) the ferroelectric T_C goes up. According to work [12], these observations can be explained in terms of an intergranular effect (involving the size dependence of the domain wall mobility) and an intragranular effect, where each grain is assumed to be a monodomain and one considers the effect of different-sized grains having different values of T_C. The dielectric response of the system was simulated using a distribution model [13], which assumes that the T_C for different grains obeys a Gaussian distribution. Below each individual T_C, the grains are assumed to have a constant dielectric constant, while the Curie–Weiss law is obeyed above T_C.

It was also made use of the concept of a diffuse phase transition (DPT) [12]. They considered the crystal or ceramic to consist of many noninteracting microregions that can independently switch from a ferroelectric to a paraelectric state. Several physical situations that could result in a DPT were proposed. In the case of solid solutions such as PZT and PLZT, microscopic variations in the local chemical composition could give rise to a T_C distribution and hence a DPT. The DPT effectively leads to a Curie *region* rather than the conventional Curie *peak*. A physical situation that can lead to a DPT even in *pure* compounds where there is no compositional variation will be introduced later.

11.2. SIZE EFFECTS IN PEROVSKITE FERROELECTRICS

During the period described, the study of ferroelectric fine particles was mainly an academic exercise, except in the technologically important case of piezoelectric ceramics. After a period of relative inactivity (1975–1987), there was a rather sudden renewal of interest in ferroelectric nanomaterials due to a revival of the idea of a ferroelectric memory device. The possible applications of ferroelectrics and piezoelectrics as microelectromechanical devices, sensors, etc. also provided a strong motivation for a spurt in both theoretical as well as experimental activities in this area. By this time a lot of progress had been made in the development and improvement of synthesis techniques for nanomaterials. These included wet chemical techniques such as sol–gel, co-precipitation, spray-pyrolysis, and microemulsion-mediated reactions (Section 11.7), and physical vapor deposition techniques such as sputtering, evaporation–condensation, and laser ablation (Section 11.8). This made it possible to study different compounds in the form of phasepure, ultrafine particles with a relatively narrow size distribution, as also high-quality ultrathin films on desired substrates.

11.2.1. BaTiO$_3$

Barium titanate has a cubic perovskite structure in its high symmetry, paraelectric phase (above $T_C \approx 125$ °C). With a decrease in the temperature, it transforms successively to three structurally distinct ferroelectric phases: a tetragonal phase ($T_C > T > 5$ °C), an orthorhombic phase (5 °C $> T > -90$ °C), and a rhombohedral phase ($T < -90$ °C). All three transitions are of the first order.

Early measurements involved sintered samples with grain size in the submicrometer to few micrometer range and relatively broad size distributions. In samples with grain size in the 1–50 μm range, the dielectric function (ε) was found to depend strongly on size in the ferroelectric phase but not so in the paraelectric phase [14]. The strong dependence of ε on the grain size was explained from the observation that the width of the 90° domains is proportional to the square root of the grain diameter and assuming that the value of ε is determined mainly by the 90° domain walls [15]. In the absence of ferroelectric domains above T_C, the value of ε in the high temperature phase is almost size independent. For sintered (but not hot-pressed) samples with 20 µm > d > 0.2 µm, the tetragonal to orthorhombic transition is relatively insensitive to the particle size, d [16]. There is, however, an increase in the thermal hysteresis with a decrease in d, indicating that the latent heat of transition is strongly size-dependent. The variation of the tetragonal distortion (c/a) in $BaTiO_3$ fine particles (0.1–1.0 µm) prepared by co-precipitation and hydrothermal routes has been studied in [17]. Identifying T_C as the temperature at which $c/a \rightarrow 1$, they determined the critical size for the existence of ferroelectricity (d_{crit}) in $BaTiO_3$ to be 120 nm. This implies that the $T_C < T_R$ (room temperature) when $d < d_{crit}$.

Detailed X-ray diffraction (XRD) measurements on smaller-sized (0.01–1.0 µm) $BaTiO_3$ nanocrystals indicate [18] that the structural phase change associated with the ferroelectric transition does not occur when d < 50 nm. This was supported by second harmonic generation (SHG) measurements and differential scanning calorimetry (DSC). Interestingly, *no change* was observed in the Raman spectra of the tetragonal structure, with a decrease in the particle size. Note that while XRD and SHG probe the average crystal structure over several unit cells, Raman spectral response is determined by the symmetry of individual unit cells. These results appear to imply that the polar distortion of the TiO_6 octahedra continues to be maintained irrespective of the particle size, though the *average* structure over many unit cells does become cubic. It was suggested that such local distortions (in apparently "cubic" nanoparticles) exist only at the time scales of light scattering ($\approx 10^{-13}$ s) [18]. These results are supported by EXAFS and XANES measurements of the local structure around Ti atoms [19], which show that the magnitude of the Ti atom off-center displacement *does not* depend on the particle size. A careful study of size effects in well-characterized sol–gel derived nanocrystalline $BaTiO_3$ showed that with decreasing size, the cubic–tetragonal transformation temperature shifts downward *only slightly* before getting substantially suppressed, while the tetragonal–orthorhombic transformation temperature shifts slightly upward before getting suppressed [20]. It was also pointed out that though hydroxyl defects are clearly observed by infrared (IR) spectroscopy in nanocrystalline samples synthesized at low temperature, the Raman activity observed in XRD-cubic material cannot be ascribed to their presence only.

A more precise understanding of the locally acentric structure in $BaTiO_3$ nanoparticles (15 – 155 nm) has been recently obtained from electron paramagnetic resonance measurements on Mn^{2+}-doped samples [21]. Comparatively larger crystallites consist of a regular ferroelectric core with a tetragonality gradient toward the outer surface, surrounded by a peripheral layer with strongly distorted symmetry. In particles smaller than ≈40 nm the regular core no longer exists at room temperature. Thus, the regular core undergoes a size-driven transition into the paraelectric phase, but the distorted surface layer does not participate in this transition. The size-driven transition is considerably smeared out by the grain-size distribution. The application of Landau theory to individual $BaTiO_3$ particles suggests a critical size of ≈50 nm at room temperature whereas in the size-distributed nanopowder a considerably lower mean size (≈25 nm) represents the effective threshold for the sizedriven transition into the paraelectric phase.

The value of d_{crit} obtained in these experiments (≈50 nm) contradicts an earlier study of a surfactant-stabilized *suspension* of $BaTiO_3$ particles [22]. Probing the spontaneous polarization through the electro-optic Kerr effect, it was claimed to have detected a permanent dipole moment even in 10 nm particles. However, such a claim may be contested because: (a) the nanoparticles were prepared by a relatively crude ball-milling technique and some relatively large particles may have remained, (b) such a mechanical attrition process may create surface damage and induce defects that show a “polar” response, and (c) the size determination process with arbitrary correction factors appears questionable. Studies on $BaTiO_3$-based glass ceramic composites also show that the T_C, ε_{max}, and dielectric peak broadening all scale systematically with the mean $BaTiO_3$ crystallite size, but the value of d_{crit} was found to be only 17 nm [23].

11.2.2. $PbTiO_3$

Lead titanate is a typical displacive ferroelectric with very good dielectric, pyroelectric, and piezoelectric properties. It has a cubic perovskite structure above T_C = 490 °C and transforms to a tetragonal ferroelectric phase below it. The room temperature tetragonal distortion, $(c-a)/a$, is as large as 6%, as compared to only 1% for $BaTiO_3$. Due to its obvious application potential, detailed size effect studies have been carried out in this system during the past decade.

$PbTiO_3$ fine particles ($22 < d_{XRD} < 52$ nm) by an alkoxide route have been prepared and a Raman study of samples with different average size as a function of temperature has been made [24]. They measured T_C indirectly using the principle that in displacive ferroelectrics, the frequency (ω_s) of the ferroelectric soft mode (which is usually the lowest TO mode) tends to vanish as $T \rightarrow T_C$ [25]. For each sample, ω_s was plotted as a function of T and the T_C was identified as the temperature at which ω_s extrapolates to zero. They found that T_C decreases monotonically when the average particle size, $d < 50$ nm, and its deviation from the bulk value, $T_C(\infty)$, can be fitted to the equation

$$T_C(d) = T_C(\infty) - C/(d - d_{crit}) \qquad (11.3)$$

with d_{crit} = 12.6 nm and C = 588.5 °C. Note that d_{crit} is the size at which $T_C \rightarrow 0$ K. Raman scattering measurements (at room temperature) on sol–gel derived $PbTiO_3$ nanoparticles (20–200 nm) has been also made and it has found that the ω_s decreases with decreasing particle size, implying a lowering of T_C with size [26]. Even when T_C is identified as the peak of the heat flow curve in DSC, it is found to vary with particle size following Eq. (11.3) but with d_{crit} = 9.1 nm.

However, results obtained from such indirect measurements can be influenced by surface depolarization fields that may affect the TO mode frequency.

The nanostructure of ultrafine ferroelectric $PbTiO_3$ particles (20–2000 nm) has been imaged by high-resolution transmission electron microscopy (HR-TEM) [27]. High resolution images and selected area diffraction patterns showed that all the particles had tetragonal structure, and that the c/a ratio and domain size both decreased with decreasing particle size. The particles became monodomain when $d < 20$ nm. A domain wall width of 1.4 nm was deduced from strain contrast shown by 90° domain walls, but there was no evidence of amorphous surface layers.

Because it has a large spontaneous polarization, a high T_C, only one structural transition (at T_C), and there is a satisfactory conceptual understanding of its transition mechanism, $PbTiO_3$ is certainly the candidate of choice for studying finite size effects in ferroelectric systems. However, the studies reported suffer from certain drawbacks: (*i*) the transition temperature is measured either indirectly, or using only a single method, (*ii*) there are no proper studies of the size dependence of the dielectric response, which is the single most important property of ferroelectrics, and (*iii*) no clear correlation is made between the dielectric properties and the structural parameters (particularly the ferroelectric distortion). The following study [28] fills up some of these lacunae and provides data that could lead to a clearer conceptual understanding.

Ultrafine $PbTiO_3$ particles ($20 < d_{XRD} < 200$ nm) were synthesized by co-precipitation. Though the average particle size was obtained from X-ray line broadening (d_{XRD}), specific surface area measurements (gas adsorption), and scanning electron micrographs (see Fig. 11.4), the X-ray domain size (d_{XRD}) is expected to be more relevant to the problem since ferroelectric phenomena in displacive systems are controlled by lattice vibrational mechanisms. The T_C was obtained from three complementary measurements: (*i*) dielectric response vs temperature, (*ii*) temperature dependence of the tetragonal distortion (c/a), which scales with the spontaneous polarization, and (*iii*) differential scanning calorimetry. XRD measurements were also made down to 15 K to confirm that the large deviation observed in c/a at room temperature is an inherent particle size effect.

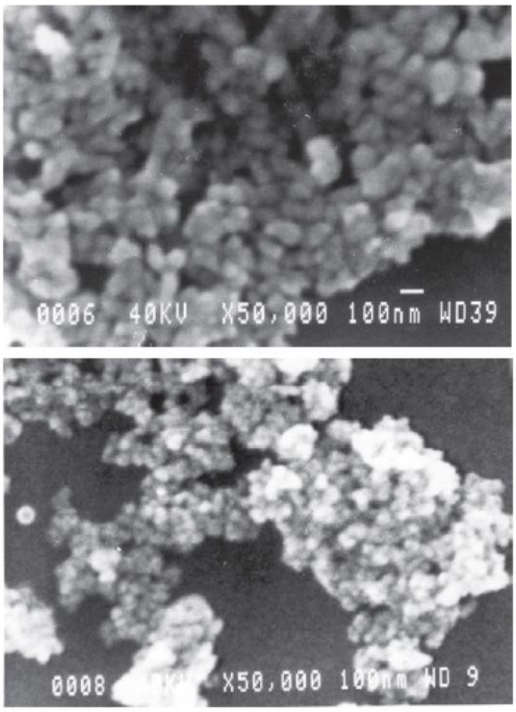


Fig. 11.4. Scanning electron micrographs (×50,000) of chemically synthesized samples of nanocrystalline $PbTiO_3$ with an average size of about 80 nm (top) and 40 nm (bottom).

Figure 11.5 shows the variation of the ferroelectric distortion (c/a) with particle size at room temperature. The unit cell dimensions do not change down to ≈100 nm, below which c decreases and a increases with decreasing size. The resulting reduction in c/a is especially rapid below ≈50 nm. To estimate the critical size at which $PbTiO_3$ transforms to the cubic (paraelectric) phase *at room temperature*, the c/a vs d_{XRD} data were fitted to an equation of the form

$$y = y_{\infty} - C\exp[C(d_{crit} - x)] \qquad (11.4)$$

with $x \equiv d_{XRD}$, $y \equiv c/a$, $y_{\infty} \equiv 1.065$, and $C \equiv y_{\infty} - 1$. Using the critical size (d_{crit}) as a fitting parameter, one finds — by extrapolation — that $c/a \rightarrow 1$ when $d_{crit} \rightarrow 7.0$ nm (ferroelectric order cannot be sustained below this size). The dielectric, thermal, and structural (variable temperature XRD) data also indicated that the T_C decreases monotonically with decreasing particle size.

A study of the temperature dependence of the low frequency dielectric response function, ε (Fig. 11.6, top), shows that with a decrease in the particle size, T_C decreases (Fig. 11.7, insert), ε_{max} decreases, and the peaks become increasingly broader. Samples with $d_{XRD} \leq 26$ nm exhibited no peaks in the ε vs T curves. In larger particles of $PbTiO_3$ ($d_{XRD} \geq 60$ nm), a decrease in frequency leads to an increase in ε_{max}, but the ferroelectric T_C remains constant at ≅500 °C. In smaller particles of $PbTiO_3$ (e.g., with $d_{XRD} = 31$ nm), however, the T_C moves up with an increase in frequency from 476 °C at 500 kHz to 488 °C at 1 MHz. Also, for a particular frequency, the temperature interval between the maxima in tan δ (dielectric loss tangent) and ε is larger in finer particles. These properties are typical of materials with a diffused phase transition and will be discussed later

(Section 11.4.4). With a reduction in the particle size, the nature of the changes in the "dielectric" and the "thermal" transitions (observed by DSC) were found to be qualitatively similar. Thus, the peak in the heat flow shifted toward lower temperatures and became broader and flatter with decreasing size in $PbTiO_3$ (Fig. 11.6, bottom).

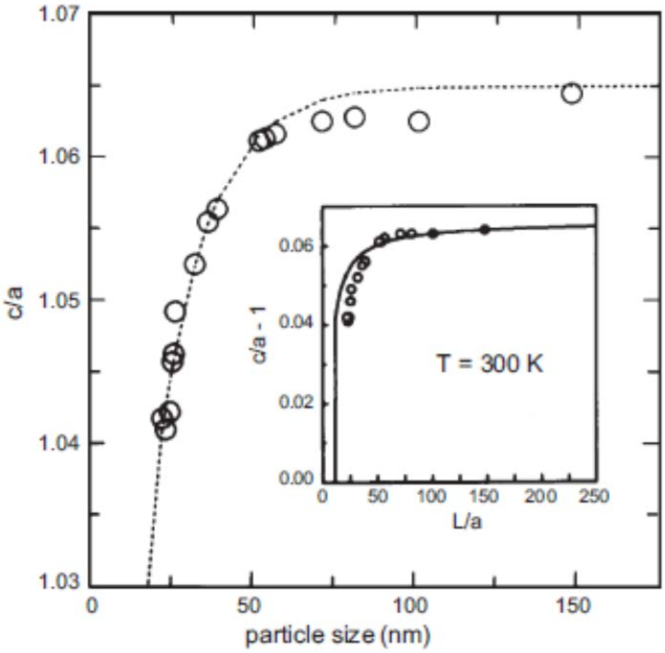


Fig. 11.5. Variation of the ferroelectric distortion (measured at room temperature) with the particle size (d_{XRD}) in nanocrystalline $PbTiO_3$. The dashed line represents a fit to Eq. (11.4). The inset shows ($c/a - 1$) as a function of the reduced particle size, where $L \equiv d_{XRD}$, and a is the lattice constant. The solid line represents a fit to the model described in Section 11.4.3.

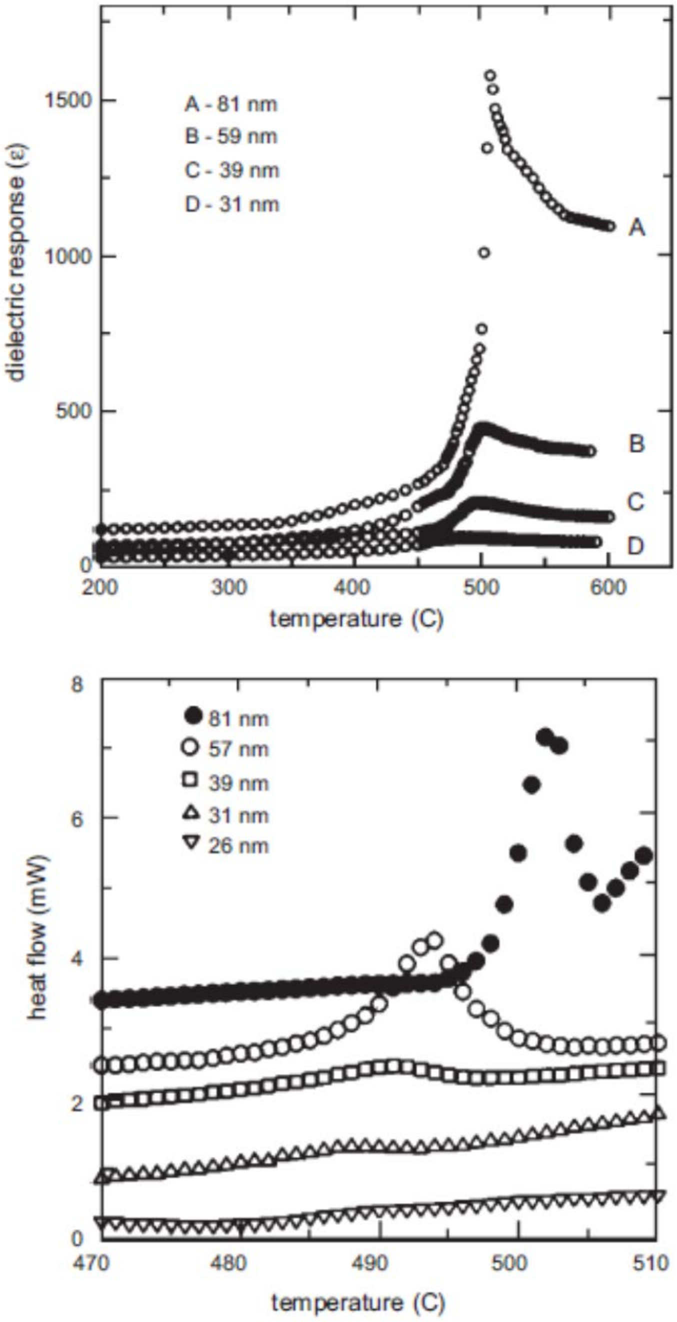


Fig. 11.6. Top: Temperature dependence of the dielectric response function (ε) for $PbTiO_3$ samples with different average particle size (d_{XRD}) measured at 1 MHz. Bottom: Differential scanning calorimetric study of the effect of particle size on the thermal transition at T_C in nanocrystalline $PbTiO_3$, scanned at 5 °C/min.

A quantitative estimate of the “diffuseness” of the ferroelectric phase transition in nanoparticles can be obtained by fitting the ε–T data ($T > T_{max}$, where T_{max} is corresponds to the peak in ε) to the equation

$$\frac{1}{\varepsilon} - \frac{1}{\varepsilon_{\max}} = \frac{1}{C}(T - T_{\max})^{\gamma} \qquad (11.5)$$

where C is a constant and the critical exponent $\gamma = 1$ for a perfect Curie–Weiss ferroelectric. According to the prediction of the local compositional fluctuation model, $\gamma = 2$ for a system with a completely diffuse transition. For systems exhibiting intermediate degrees of diffuseness we expect γ to have values between 1 and 2. Figure 11.7 shows that the ferroelectric phase transition becomes increasingly diffuse with a reduction in the particle size (see Section 11.4.4 for a discussion).

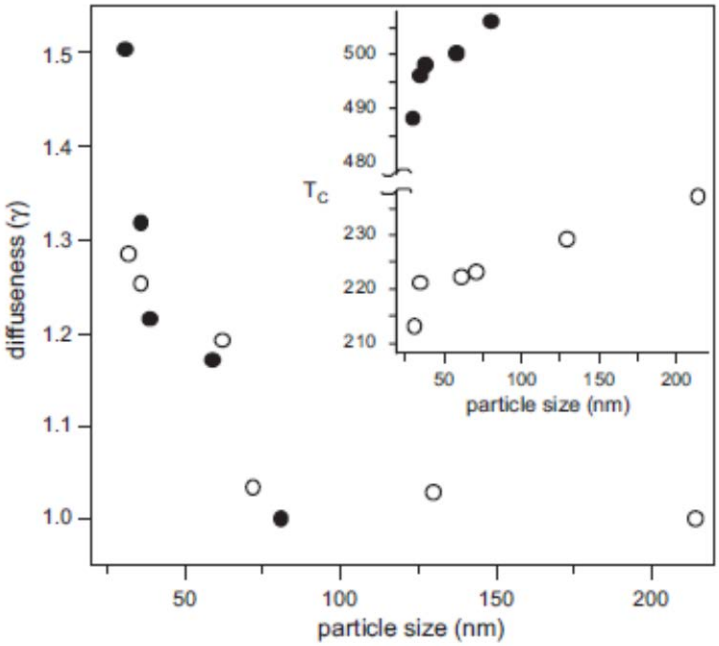


Fig. 11.7. Particle size dependence of the diffuseness coefficient (γ in $PbTiO_3$ (solid circles) and $PbZrO_3$ (open circles). The variation in the "dielectric" T_C with particle size for the same two systems is shown in the inset.

11.2.3. $PbZrO_3$

Though size effects have been studied for over 35 years in ferroelectrics, there were no systematic studies in nanocrystalline antiferroelectrics prior to the study of $PbZrO_3$ [29] in 1997. At 300 K, $PbZrO_3$ has an orthorhombic structure with eight formula units per crystallographic unit cell. This structure is derived from a cubic perovskite prototype (the high-temperature paraelectric phase) by *antiparallel* displacements of the Pb ions along one of the original [110] directions—which becomes the *a*-axis of the orthorhombic phase. Ignoring the displacements of the Pb ions (so as to compare the paraelectric and ferroelectric unit cells), the ordered phase has a pseudotetragonal unit cell whose lattice constants (a_T , c_T) are related to the real orthorhombic unit cell (a, b, c) through $a_T = a/\sqrt{2}$ and $c_T = c/2$. At $T_C \approx 500$ K, $PbZrO_3$ shows a strong dielectric anomaly.

Figure 11.8 shows the dependence of the pseudo-tetragonal distortion (c_T / a_T) on the particle size in sol–gel-derived ultrafine particles of $PbZrO_3$ ($d_{XRD} \geq 30$ nm). The crystallographic unit cell remains relatively undistorted (from the bulk) down to ≈100 nm but thereafter shows an increasing tendency to become cubic. From a downward extrapolation of these data, one may conclude that the crystal structure would become perfectly cubic below a size of 23 nm. With decreasing particle size, there is a monotonic reduction in the T_C as observed by both dielectric (Fig. 11.7, inset) and thermal (Fig. 11.8, inset) measurements. The ferroelectric phase transition also becomes increasingly diffuse with a reduction in the particle size (Fig. 11.7). The broadening of the transition is also reflected by the fact that the heat of transition (measured by DSC) decreases with decreasing size (Fig. 11.8, inset). In general, the nature of size effects in antiferroelectric $PbZrO_3$ is qualitatively similar to those in the ferroelectrics $PbTiO_3$ and $BaTiO_3$.

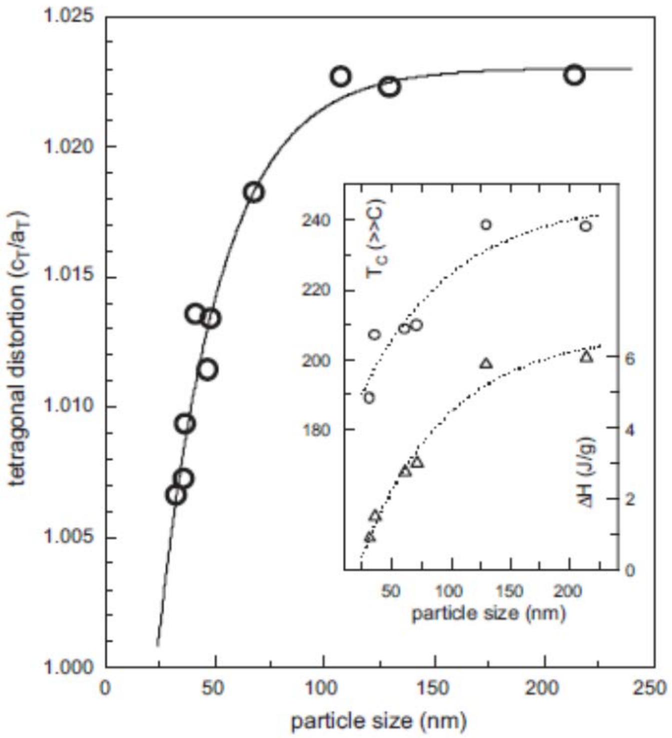


Fig. 11.8. The particle size dependence of the pseudo-tetragonal distortion of the $PbZrO_3$ unit cell, measured at room temperature. The solid line represents a least-squares fit to an empirical function of the form $1 - e^{-x}$). The inset shows the variation in the “thermal” T_C (obtained from DSC) and the heat of transition with particle size.

11.2.4. Pb(Zr,Ti)O$_3$

It has been shown that when only intrinsic effects are considered (i.e., contributions from domain walls, defects, and impurities are disregarded) a decrease in particle size should, in general, cause a reduction in the T_C, a reduction in ε_{max}, and a broadening of the dielectric response peak [30]. However, in solid solutions, a broadening in the dielectric response (and other critical parameters) may also take place because of the disordering of the two ions occupying equivalent crystallographic positions. A number of ferroelectric mixed oxides or solid solutions such as PZT, barium strontium titanate (BST), and strontium bismuth tantalate (SBT) are technologically important and studies of particle size effects have been carried out in a few of these compounds. Since it has a number of important applications, it is not surprising that the majority of such studies involve PZT.

PZT or $Pb(Zr_{1-x}Ti_x)O_3$ is obtained by a random replacement of the Ti^{4+} ions in $PbTiO_3$ by Zr^{4+} ions. The ferroelectric T_C increases monotonically from 235 °C ($x = 0$) to 490 °C ($x = 1$). With decreasing x, the structure transforms from tetragonal (as in $PbTiO_3$) to rhombohedral at $x \approx 0.5$ to orthorhombic (as in $PbZrO_3$) at $x \approx 0.05$. The tetragonal to rhombohedral phase boundary is almost temperature independent (morphotropic), though its exact position and width depends crucially on the purity and local homogeneity of the sample. The piezoelectric properties of PZT are particularly good in the neighborhood of the morphotropic phase boundary and may be further improved by suitable doping. Both the cubic–rhombohedral as well as the cubic–tetragonal transitions get progressively suppressed below ≈0.6 μm. Though the ε–T curves get flattened with decreasing size there no discernible change in T_{max} in the rather narrow size range (0.5–0.8 μm) [31]. It has been also pointed out that the values of ε_{max} are higher in samples sintered at higher temperature (and have higher density) even though the X-ray domain size does not differ much between them. This is ascribed to the higher porosity and the discontinuity of electric flux between grains in low-density samples.

A relation between the grain size (t) and the ferroelectric domain size (d) from a simple model involving the competing effects of the depolarization energy and the domain wall energy had been obtained [32]:

$$d = [(\frac{\sigma}{\varepsilon^* P_0^2})t]^{1/2} \tag{11.6}$$

In PZT ceramics, it has been found [33] that though this relation was valid in the grain size range of 3–10 μm, there was a deviation at lower grain sizes, probably due to the clamping of domain walls at the grain boundaries. It has been attempted [34] to separate out the role of extrinsic (i.e., due to domain wall, lattice defects, and impurities) and intrinsic contributions to particle size effects in Nb-doped PZT. They suggest that extrinsic contributions to the dielectric and piezoelectric properties get drastically suppressed when the grain size falls below about 2 μm while intrinsic properties get affected below ≈ 0.8 μm, mainly through a progressive reduction in the spontaneous strain.

The dielectric property of *amorphous* nanoparticulate PZT (synthesized by the sol–gel process) was studied in work [35]. The results were explained using a model that assumes that the glassy phase contains frozen in dipoles with random magnitudes and orientations formed by a distortion of the dielectrically soft prototype units (the BO_6 octahedra), which are expected to retain their identity even in the glassy phase.

11.2.5. Other Solid Solutions

Size effects have also been investigated in a number of other mixed ferroelectric or antiferroelectric oxides. The effect of grain size reduction on the piezoelectric properties in *sintered* ceramics of $(Pb_{1-y}La_y)(Zr_{1-x}Ti_x)O_3$ or PLZT was studied in work [36]. The authors measured the piezoelectric coefficients (d_{31} and g_{31}) and the planar coupling factor (k_p) in hot-pressed, low-porosity, electrically polarized PLZT samples ($y = 0.2$, $x = 0.35$), with grain sizes (G) in the 1–14 µm range. The size dependence of each property (X) could be expressed by a general equation of the form $X = A + B/G^{0.5}$. An increase in the coercive field with decreasing size has been also observed. This paper provided an early guide to the optimization of piezoelectric properties of fine-grained ceramics.

A detailed study of size effects for different values of x in BST has been carried out [37]. Deviations from “bulk” properties in solid solutions typically occur at much larger particle sizes than in pure compounds. Thus, in the case of BST, d_{crit} = 317 nm (for $x = 0.3$), 246 nm (for $x = 0.5$), and 176 nm (for $x = 0.7$). The diffuseness coefficient (γ) again increases with decreasing size. However (unlike in $PbTiO_3$ or $PbZrO_3$) the transition becomes *completely* diffuse ($\gamma = 2$) below 300 nm (for $x = 0.3$). It is emphasized that in solid solutions such as BST it is difficult to obtain a homogeneous cation distribution, which may be the main reason for the extremely diffuse transition and reduced values of ε_{max}.

In $(Pb_{1-x}La_x)TiO_3$, Raman studies of the soft mode indicate that both the phase transition temperature *and* pressure get reduced with a reduction in particle size [38]. A detailed study [39] of size effects in $Pb(Sc_{0.5}Ta_{0.5})O_3$ in the 10–160 nm range revealed that c/a reduces to unity (at 0 °C) when $d_{XRD} \leq 53$ nm. The T_C decreases gradually and the phase transition becomes more diffuse with decreasing size. Representative studies on size effects in ferroelectric solid solutions (mostly based on the perovskite structure) are summarized in Table 11.2.

Table 11.2. Stidies of size effect in ferroelectric solid solutions (other than PZT) based on the perovskite structure.

System	Sample specs	Nature of study	Ref.
$(Pb, La)(Zr, Ti)O_3$	(1) sintered ceramics	Piezoelectric properties	[36]
or PLZT	(2) sintered ceramics	ε-T, T_C	[40]
	(3) sintered ceramics	electro-optic properties	[41]
$(Pb_{1-x}La_x)TiO_3$	(1) 0.6 - 1.1 µm, sinter	ε-T, T_C	[42]
or PLT	(2) $0 \leq x \leq 0.2$, sol-gel	Raman at high press. & temp.	[38]
	(3) sol-gel	Raman, soft mode	[43]

$(Pb_{1-x}Sr_x)TiO_3$	Sintered ceramic	S/TEM, AFM, DPT	[44]
$Pb(Sc_{0.5}Ta_{0.5})O_3$	10-160 nm	ε-T, T_C, DPT, XRD-T, DSC, d_{crit}=10-20 nm	[39]
$Pb(Zn_{1/3}Nd_{2/3})O_3$-$BaTiO_3/PbTiO_3$	model	compositional fluctuation at nanoscale	[45]
$Pb(Mg_{1/3}Nd_{2/3})O_3$-$PbTiO_3$	granular thin film	(1) local piezoelectric properties using SFM	[49]
$(Pb, Ba)TiO_3$	$0 \leq x \leq 1$, sol-gel	micro-Raman, soft mode	[46]
$(Ba_{1-x}Sr_x)TiO_3$ or BST	x = 0.3, 0.5, 0.7, sol-gel	(1) ε-T, T_C, DPT (2) barrier height, relaxation frequency	[37] [47]
$(Sr_xBa_{1-x})Nb_2O_6$ or SBN	sol-gel	ε-T, T_C, DPT, Raman	[48]

DSC – differential scanning calorimetric; DPT – diffuse phase transition; SFM – scanning field microscopy.

11.3. SIZE EFFECTS IN OTHER TYPES OF FERROELECTRICS

The preceding part of this chapter shows that nanoparticles of *displacive* systems such as $BaTiO_3$, $PbTiO_3$, and $PbZrO_3$ exhibit certain basic similarities such as (*i*) a monotonic decrease in the T_C with particle size below ≈100 nm, (*ii*) existence of a critical particle size below which ferroelectricity cannot be observed, (*iii*) a broadening of the ferroelectric transition peak and behavior reminiscent of a diffuse phase transition at low sizes, and (*iv*) an interrelation between the T_C, the ferroelectric distortion, and the particle size.

The behavior of a typical *order–disorder* system such as $NaNO_2$, however, is somewhat different. Bulk $NaNO_2$ undergoes a first order ferroelectric–paraelectric transitionat 165 °C, which is associated with a relatively large latent heat. Nanocrystalline $NaNO_2$ (down to 5 nm) by three different methods has been prepared and the ferroelectric phase transition by differential thermal analysis has been studied [50]. Very little change in T_C was observed down to 5 nm, whereas displacive systems start showing deviations in the T_C at sizes as large as 100 nm. This clearly emphasizes the importance of the size-induced *structural distortions* in nanoparticles of displacive systems. The other obvious difference in the two types of system is that thermal as well as dielectric phase transitions are usually not observed at all in nanoparticles of displacive systems that are smaller than ≈25 nm, while even 5 nm $NaNO_2$ nanoparticles show a clear transition peak. Size effects, therefore, appear to be much weaker in order–disorder ferroelectrics.

Again, no change in T_C (≈440 K) was found in a study of 40–50 nm $NaNO_2$ particles impregnated in the free spaces between the 200 nm SiO_2 spheres that form the threedimensional (3D) lattice in synthetic opal [51]. However, a giant enhancement in the dielectric constant of the $NaNO_2$-impregnated opal (at 100 Hz) above 550 K by about 10^7 times has been reported! It has been ascribed to the melting of the dispersed $NaNO_2$ in the host lattice and the formation of electrolytic drops. This is similar to the large dielectric constants observed in electrolyte-saturated porous ceramics. More recently, the temperature evolution of the structure of nanodispersed $NaNO_2$ confined within a porous glass with an average pore size of ≈7 nm has been studied across T_C with the help of neutron diffraction [52]. This study suggests that there is a lattice softening above T_C, manifested by a pronounced growth in the thermal vibration parameters. The appearance of an ionic current due to oxygen diffusion in the lattice above T_C is presumed to lead to the giant increase in the dielectric response.

KNO_3 is also believed to be an order–disorder ferroelectric. On heating it transforms from an orthorhombic ferroelectric phase (II) to a trigonal paraelectric phase (I) at 130 °C. On cooling, phase I transforms to a metastable trigonal phase (III) at 125 °C, which goes to phase II at 115 °C. The effect of particle size on the II– I, I–III, and III–II phase transitions has been studied and it has been found that the III–II transition shifted to lower temperature, while the II–I transition temperature increased with decreasing size [53]. However, these measurements were made on powdered single crystals, in which surface stress effects and deformations could play a major role. It is also essential to extend these results to the submicrometer size range and use samples with a narrower size distribution to in order to obtain more reliable results.

11.4. CONCEPTUAL APPROACHES

11.4.1. Size Induced Structural Phase Transition

All the perovskite-type oxide ferroelectrics appear to exhibit a decrease in the tetragonal (ferroelectric) distortion with a reduction in particle size. It is important to note that this effect is a manifestation of a much more general phenomenon and is *not* restricted to ferroelectrics. Detailed studies show that in a large number of partially covalent oxides, the crystal lattice tends to transform into structures of higher symmetry when the particle size is reduced [54]. In systems such as Fe_2O_3 and Al_2O_3, the size-induced lattice distortion is large enough to actually produce a crystallographic transition to a high-symmetry structure [55]. In most other cases, there is a gradual reduction in an asymmetry parameter—such as (c/a - 1) - with decreasing size.

A decrease in the particle size may cause a change in the bulk lattice parameters via surface stresses [56, 57]. The *anisotropic* nature of the lattice expansion in the nanoparticles of most oxides can be understood in terms of a tendency of such systems to become increasingly *ionic* with decreasing size [58]. As a result, the interatomic bonds lose their directional character and the crystal tends to assume a structure with a comparatively higher symmetry. Interestingly, the size dependence of many important physical properties (such as superconductivity and magnetism, in addition to ferroelectricity) is often directly or indirectly connected to the anisotropic lattice distortion that accompanies a reduction in the particle size [54].

11.4.2. Overview of Theoretical Models

Experiments on perovskite-type ferroelectrics and antiferroelectrics indicate that a reduction in the average particle size leads to a decrease in the Curie temperature, the ferroelectric distortion (c/a), the heat of transition, and the soft mode frequency. Various attempts have been made to obtain a conceptual understanding of these observations. It is beyond the scope of this chapter to provide a complete review of the theoretical studies of size effects in ferroelectric nanoparticles. We will merely attempt to present a flavor of the area and briefly discuss some simple, general models. Attempts to explain the basic experimental features using the Landau–Devonshire phenomenological theory have been reasonably successful. For a finite-sized, inhomogeneous ferroelectric with a second order phase transition, the total free energy can be written in the form [59]

$$\int[\frac{1}{2}A(T-T_{C\infty})P^2+\frac{1}{4}BP^4+\frac{1}{2}D(\nabla P)^2]dv+\int\frac{1}{2}D\delta^{-1}P^2dS \qquad (11.7)$$

where P is the polarization, and T_C refers to the “bulk” value of T_C. D is connected to the correlation length (*1*): $D = \xi^2 \left|A(T - T_{C\infty})\right|$. The volume and surface integrals represent, respectively, the free energies of the interior and the surface. Compared to the analogous expression for an infinite, homogeneous ferroelectric, Eq. (11.7) contains additional contributions due to the gradient term and the surface term. The “extrapolation length,” ″, turns out be particle size dependent. The authors of work [59] carry out numerical calculations to obtain the size dependence of the polarization as well as the T_C and show that both decrease with decreasing size and vanish as $d \rightarrow d_{crit}$. A few years later, the phenomenological model for the second order phase transition in spherical ferroelectric particles has been analytically solved and the temperature and size dependence of both the static and dynamical dielectric susceptibilities has been obtained [60].

In these treatments, the surface charge is assumed to be fully compensated, such that the depolarization fields can be ignored. The effect of the depolarization field and a space charge layer—which tend to break up the particles into domains of different polarization has been considered in [61]. For simplicity, they assume cubic particles with alternating domains separated by 180° domain walls and incorporate the depolarization energy as well as the domain wall energy into the Landau free energy density. Since the multidomain ferroelectric state disappears in small enough particles, the presence of a depolarization field is found to substantially lower the ferroelectric T_C.

More recently, the macroscopic phenomenological approach by calculating the size dependence of the Landau–Ginzburg–Devonshire coefficients [A, B, and D in Eq. (11.7)] has been extended [62]. Other than predicting the size driven phase transition to a paraelectric state for $d < d_{crit}$, the model can explain the changes in the other physical parameters and appears to give a relatively better description of the changes in the latent heat than previous models. It also yields a simpler free energy expression, which allows some of the physical properties to be discussed analytically. Other than the macroscopic Landau theories, there have also been reasonably successful attempts to investigate size effects with the help of a microscopic pseudo-spin theory based on the Ising model in a transverse field [63, 64].

11.4.3. Microscopic Model for Nanoferroelectrics

The nature of size-induced changes in displacive ferroelectrics such as $PbTiO_3$ can be understood with the help of the simple model [65, 66]. Interestingly, the model also allows one to make certain predictions about order–disorder ferroelectrics which appear to be consistent with the available data. The model is based on two simplifying assumptions: (a) the finite sized ferroelectric system is quasi-free (i.e., it consists of loosely aggregated, unclamped nanoparticles) and (b) the ferroelectric nanoparticles are not electrically isolated. Under these conditions, we can ignore the effects of external strain and depolarization.

Experiments show that a reduction in the size of ABO_3- type displacive ferroelectrics and antiferroelectrics is accompanied by a monotonic decrease in the ferroelectric T_C, which is the temperature at which the high symmetry paraelectric structure transforms to a low symmetry phase in which the B ions no longer occupy the centrosymmetric positions in the BO_6 octahedra. When the particle size is made sufficiently small ($d < d_{crit}$), the ferroelectric system reverts to the paraelectric phase, *however low the temperature*. In other words, the ferroelectric to paraelectric transition may be made to occur as a function of increasing temperature as well as decreasing size. The following microscopic model naturally leads to a size-induced structural phase transition of the first order.

The displacement of the centrosymmetric *B*-ion to an offcentric position at the ferroelectric T_C is accompanied by a distortion in the cubic (paraelectric) unit cell to one of lower symmetry (e.g., in PbTiO3, a tetragonal one with sides $a \times a \times c$. The tetragonal distortion $(c/a - 1)$ is experimentally found to scale with the order parameter. The displacement (d) of each such ion clearly requires some energy, and the system would undergo a ferrodistortive transition only when it gains a *larger* amount of energy from the interaction of the resulting dipoles, which are aligned in some particular fashion.

The dipolar interaction energy of the stable ordered phase, ($\sim Jzd^2$), should be comparable to kT, where z is the coordination number. Thus, dipolar alignment may take place provided $d \geq \sqrt{(kT/Jz)}$. Note that this argument itself implies the possibility of driving the transition by varying either the temperature or the system size (which controls the average value of z). So the model describes the temperature driven transition in the bulk material, as well as the size driven transition in finite systems.

Phenomenologically, we can describe the interaction energy between the nearest neighbor dipoles by $Jd_i d_j \alpha_i \alpha_j$, where $\alpha = \pm 1$ is an Ising variable, which accounts for the fact that the central ion can be displaced in two opposite directions. J is positive for ferroelectrics and negative for antiferroelectrics. The effective Hamiltonian for the problem is thus the sum of an elastic part and an Ising part. From symmetry considerations, we obtain

$$H = \sum_i \left(\frac{1}{2}\lambda_2 d_i^2 + \frac{1}{4}\lambda_4 d_i^4\right) - J\sum_{i,j} d_i d_j \alpha_i \alpha_j \qquad (11.8)$$

This Hamiltonian is solved by treating the Ising variable via an inhomogeneous mean-field theory and the displacement variable within a saddle point approximation. The free energy density is obtained in the form $F(d) = E(d) + I(d)$, where the elastic part, $E(d)$, is given by

$$E(d) = \frac{1}{2}\lambda_2 d^2 + \frac{1}{4}\lambda_4 d^4 \qquad (11.9)$$

and the Ising part, $I(d)$, is given by

$$I(d) = \frac{1}{2}Jzm^2 d^2 - T\ln 2\cosh\left(\frac{Jzmd^2}{kT}\right) \qquad (11.10)$$

where $m \equiv \langle\alpha\rangle = \tanh(Jzmd^2/T)$. With increasing d, $E(d)$ increases monotonically and rapidly, while $I(d)$ decreases more slowly. The saddle point values of d, obtained by minimizing $F(d)$, are $d = 0$ and $d = d_0 = \sqrt{(Jzm^2 - \lambda_2)/\lambda_4}$. Clearly, it is this two-minima structure of the free energy that is responsible for the first order ferroelectric ($d = d_0$) to paraelectric ($d = 0$) transition.

Taking the model parameters (λ_2, λ_4, and J) for $PbTiO_3$ from experimental data [28], it is possible to plot the phase diagram in the temperature–particle size plane (Fig. 11.9). The phase boundary for displacive systems ($\lambda_2 > 0$, solid line) represents first order transitions for all $T > 0$. Thus, the theory predicts a depression in T_C with reducing size. In systems with $\lambda_2 < 0$ (which usually describes order–disorder ferroelectrics), size effects are greatly suppressed, as shown by the dotted curve in Fig. 11.9. One would need to study very small particles at temperatures very close to T_C to observe such effects. This is also in qualitative agreement with experimental observations [50]. Thus, in general and to the first order, particle size effects in ferroelectrics appear to arise mainly from a lower average coordination number for the atoms close to the surface of a nanoparticle.

This model has the important advantage of simplicity—having only three fitting parameters—while the phenomenological Landau theories [59, 61] have about twice as many. Also, by making J negative, the model can be easily extended to describe displacive-type *anti*ferroelectrics.

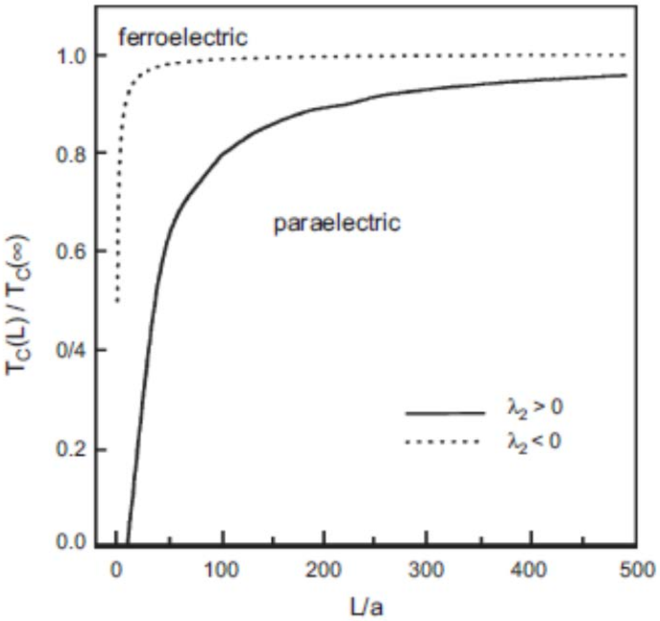


Fig. 11.9. The phase diagram of nanocrystalline ferroelectrics (reduced temperature against reduced particle size) obtained from the model described in Section 11.4.3. The solid line separates the ferroelectric phase from the paraelectric phase in typical displacive systems, while the dotted line does the same for typical order–disorder systems ($L \equiv d_{XRD}$, a = lattice constant).

11.4.4. Diffused Phase Transition

Let's now discuss another important experimentally observed feature in ferroelectric and antiferroelectric nanoparticles. The dielectric behavior (ε vs T) of nanoparticles of ferroelectric $PbTiO_3$ [28] and antiferroelectric $PbZrO_3$ [29] can be summarized as follows. With a decrease in the particle size: (*i*) the peaks in the ε–T curve become broader, (*ii*) there is an increase in the separation between the peaks in the tan δ (dielectric loss) vs T and ε vs T curves, (*iii*) there is an increasing deviation from the Curie–Weiss law (increase in γ), (*iv*) both the ε vs T and tan δ vs T curves show marked frequency dispersion, (*v*) the decrease in the tetragonal distortion (c/a) as $T \rightarrow T_C$ becomes more gradual in samples with smaller particle size, (*vi*) values of T_C obtained from different measurements (dielectric, thermal, and structural) do not necessarily coincide, and (*vii*) with a decrease in frequency, T_C decreases while ε_{max} increases.

The broadening of the dielectric and thermal peaks at the ferroelectric transition and most of the other observed features are known characteristics of the so-called "relaxor" ferroelectrics which exhibit a diffuse phase transition (DPT) [67–70]. A quantitative measure of the degree of diffuseness can be obtained by fitting the dielectric data to Eq. (11.5). The fitted value of the critical exponent (γ) lies between 1 (for an ideal Curie–Weiss ferroelectric) and 2 (for a ferroelectric with a totally diffused transition). Other than nanocrystalline ferroelectrics, a DPT is also exhibited by many complex perovskite ferroelectrics, random dipole ferroelectrics, and ferroelectrics with a graded composition. The near-T_C dielectric response of such relaxor ferroelectrics is often substantially higher than what might be expected from the LST relation [Eq. (11.2)], making them technologically important. The feature common to all relaxor ferroelectrics is that they possess structural inhomogeneities or compositional variations at the submicrometer scale.

Applying the Landau–Ginzburg free energy formulation to an inhomogeneous medium, a thermodynamic approach to DPT in mesoscopically inhomogeneous ferroelectrics has been provided [71]. They find that a coherent coupling can take place across the interface between two chemically different dipole-ordered regions. They assume each localized cluster to have a mean-field character, such that the DPT can be understood as an inhomogeneous condensation of localized soft modes.

DPTs were discovered in perovskite solid solutions of the type $A(B_{1-x}C_x)O_3$, in which, generally, there is no long range crystallographic ordering of the B and C cations. Local compositional fluctuations are therefore present throughout the crystal, so that it could consist of microscopic domains of ABO_3 and ACO_3 as well as intermediate compositions (Fig. 11.10, top). Since ABO_3 and ACO_3 would—in general—have different values of T_C, this would lead to a microlevel T_C distribution about a central value corresponding to the T_C for $A(B_{1-x}C_x)O_3$. There could also be an intermixture of ferroelectric and paraelectric regions if only one of the constituents of the solid solution is ordered at the observation temperature. The scale of heterogeneity is believed to be 10–20 nm [69].

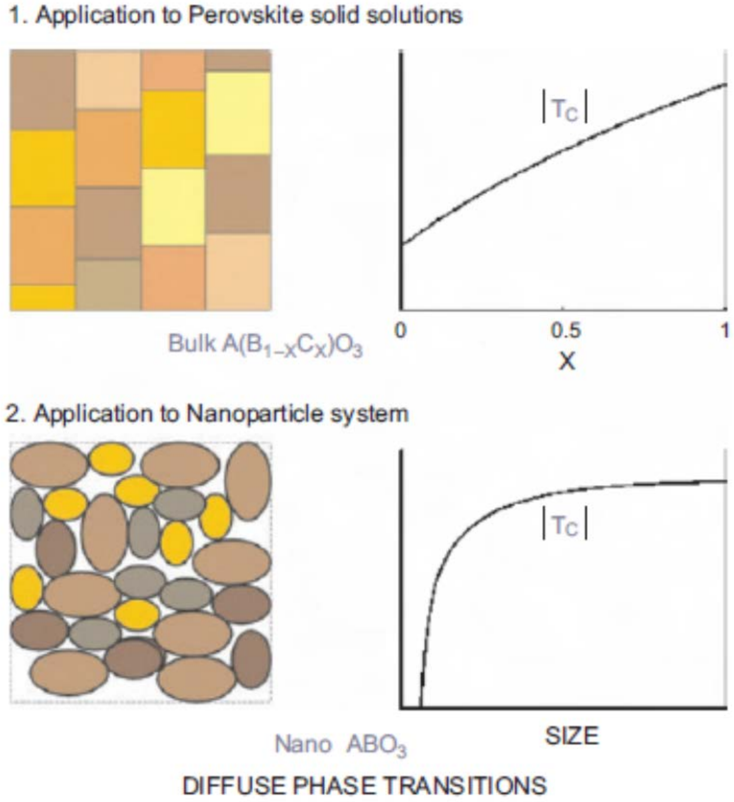


Fig. 11.10. A schematic equivalence of the phenomenon of diffuse phase transition in a typical ferroelectric solid solution (top) with a variation in the local stoichiometry and a nanocrystalline ferroelectric system (bottom) with a particle size distribution. In both situations, different shades represent local regions with different values of T_C.

One can now understand the conceptual similarity between a solid solution and a nanocrystalline system in a simple manner [28]. A nanocrystalline ferroelectric with an average size in the range of 10–50 nm would consist of a collection of crystallites with varying sizes (due to the inherent size distribution) centered around the mean size (Fig. 11.10, bottom). Recall that in the size range being discussed, particles of different sizes also have different values of T_C. Since the variation of T_C with size is more conspicuous in smaller particles, the latter should exhibit a more diffused phase transition. Many workers have invoked the T_C-distribution model [12] to explain the dielectric behavior of sintered ceramics, in which a complicated stress distribution occurs due to the high compaction. In the comparatively loosely aggregated nanocrystalline material studied by us, the observed strain and the T_C appear to depend directly on the particle size. However, in this case, the strain dependence is *inherently* a size effect and is not an artifact of sintering or pressure compaction.

11.5. FERROELECTRIC THIN FILMS

11.5.1. Size Effects in Thin Films

During the past decade, a very large amount of developmental work has been carried out on the synthesis and characterization of ferroelectric thin films. This has been fueled mainly by the intense interest in the development of ferroelectric memory devices. A detailed discussion of the technological aspects and the related synthesis problems are not within the scope of this chapter. Readers interested in applications aspects of ferroelectric thin films may refer to the excellent reviews [72, 73]. Here, we attempt to cover some of the important aspects of inherent size effects in ferroelectric thin films. We will also present an interesting effect in antiferroelectric thin films which appear to show ferroelectric behavior below a critical thickness.

The important length scales in thin films are the film thickness, the grain size (in granular films), and the lateral device size (in thin film devices), and it is interesting to observe the consequences of any of these quantities approaching the ferroelectric critical size (d_{crit}). We also point out that in thin films, it is often difficult to distinguish genuine size effects from extraneous effects related to crystalline quality, microstructural heterogeneities, and mechanical stresses induced by the substrate.

11.5.2. Intrinsic Effects

A large body of experimental data indicates that size/thickness effects in ferroelectric thin films are manifested mainly in the following five ways [74, 75]. A decrease in the thickness (t) usually leads to (*i*) a larger coercive field, (*ii*) a lower T_C, (*iii*) a smaller remanent polarization, (*iv*) a lower dielectric constant, and (*v*) a broadening of the dielectric peak at the ferroelectric transition.

In bulk ferroelectrics, the measured value of the coercive field E_C (the external field required to reverse the macroscopic polarization, see Fig. 11.1) is almost always much smaller than the theoretical or intrinsic value predicted by the Landau–Ginzburg mean field theory. This is due to domain nucleation and domain wall motion. The energy required to switch the polarization of such small nucleation volumes is much smaller than that required for the collective polarization reversal of the entire crystal. Thus, the obvious way to increase E_C would be to use particles that are so small or films that are so thin that domain nucleation is inhibited. Recently, the intrinsic value of E_C in Langmuir–Blodgett films of a polyvinylidene fluoride (PVDF)-based polymeric ferroelectric with $t \approx 1$ nm has been able achieved [76]. The measured E_C is in good agreement with theoretical predictions and is *independent* of t in the range 1–15 nm.

Considering only intrinsic effects, the shift in the T_C and the jump in the spontaneous polarization (P_S) at the T_C as a function of the film thickness (t) are given, respectively, by:

$$\Delta T_C = At^{-1} \quad \text{and} \quad \Delta P_S(T_C) = Bt^n \qquad (11.11)$$

where A and B are constants independent of temperature and thickness, while the exponent n is of the order unity. In addition, the film thickness may affect other types of phase transitions. For example, the morphotropic phase boundary (see Section 11.2.4) that separates the tetragonal phase from the rhombohedral phase in $Pb(Zr_{1-x}Ti_x)O_3$ (PZT) shifts from ≈0.5 in the bulk to about 0.2 in thin films [77]. There have been numerous attempts to calculate the intrinsic, thickness dependent electric susceptibility, χ_f. Starting from the continuum, phenomenological free energy, the expression has been obtained [78]:

$$\chi_f = \chi_b[1 + \frac{C_0}{(L - L_0)}] \qquad (11.12)$$

where C_0 is independent of L, the film thickness.

The study of ultrathin ferroelectric films also gives us an opportunity to investigate the possible existence of the superparaelectric (SPE) state. This is analogous to the phenomenon of superparamagnetism observed in ultrafine magnetic particles. The latter occurs when the particle size is small enough and the temperature is low enough such that the magnetic anisotropy energy of individual particles is smaller than the thermal energy. Under this condition the magnetization vector of individual particles continues to flip spontaneously by a coherent reorientation of atomic spins. Similarly, the SPE state would occur when the energy required to coherently flip the polarization vector of individual ferroelectric particles is smaller than kT. A simple calculation based on the *intrinsic* coercive field and remanent polarization of typical ferroelectrics, tells us that for the switching energy to approach kT, one requires a particle having a volume ≈1 nm^3 [73].

11.5.3. Extrinsic Effects

Changes in ferroelectric properties usually occur when the film thickness is below ~200 nm, but the actual value of this “critical thickness” often depends sensitively on the nature of the microstructure (and hence on the synthesis conditions), interfacial stresses, crystallographic orientation, and epitaxy. As mentioned earlier, it may be difficult to distinguish between particle size effects and thickness effects since the microstructure often changes continuously with t—thinner films being usually composed of smaller particles. An example of this in the study of free standing polycrystalline $PbTiO_3$ films has been provided [79]. It has been found that the domain structure and dynamic response change drastically when the thickness is reduced below 200 nm, because the thinner films are composed of nanosized, singledomain grains. From a study of the effect of granularity on the dielectric response of apparently dense radio-frequency (rf)-sputtered thin films of $BaTiO_3$, it has been found [80] that ε decreases considerably and the ferroelectric state becomes unstable when the coherently scattering X-ray domain size (d_{XRD}) ≤ 30 nm. Since the thinner films are also usually composed of smaller particles, this phenomenon is apt to be misinterpreted as a “thickness” effect.

The issue of thickness induced phase transitions in ferroelectrics is interesting and important. It had been suggested [81] that a ferroelectric exhibiting a second order transition in the bulk should exhibit a new type of first order transition. This was ascribed to the presence of thickness-dependent depolarization fields in a ferroelectric thin film deposited on a semiconductor surface.

Another type of thickness-dependent phase transition has recently been reported [82]. The study involves the two antiferroelectric compounds $BiNbO_4$ (which is antiferroelectric up to 360 °C, ferroelectric between 360 and 570 °C, and paraelectric above 570 °C) and $PbZrO_3$ (which has a $T_C \approx 230°C$. In high-purity $PbZrO_3$ samples, it is possible to stabilize a rhombohedral *ferroelectric* phase in a temperature range of 220–230 °C. Polycrystalline $BiNbO_4$ and *c*-axis oriented $PbZrO_3$ films ($100\text{ nm} \leq t \leq 900\text{ nm}$) were pulsed laser deposited on Si (100) wafers. Atomic force micrographs showed that the thinner films were (expectedly) composed of smaller grains in both systems. Dielectric hysteresis (P vs E) measurements showed that in both $PbZrO_3$ and $BiNbO_4$ thin films, the thicker films are antiferroelectric, while—below a characteristic thickness—they show ferroelectric behavior. Note that the critical thickness (t_C) below which ferroelectricity is found to be stabilized is quite large—about 500 nm. The observed stabilization of ferroelectricity in thin films of normally antiferroelectric materials may be ascribed to the effect of the electric field produced at the semiconductor–dielectric interface. For a typical semiconductor–insulator interface, the interfacial field $E_{if} \approx 100\text{ kV cm}^{-1}$, which is high enough to induce ferroelectricity in many antiferroelectric crystals. The nucleation of $PbZrO_3$ and $BiNbO_4$ grains on the substrate takes place at 700 °C (which is above their respective T_C’s) in the presence of a strong electric field (E_{if}) that tends to orient the local dipoles along the field, leading to ferroelectric ordering. In thicker films ($t > t_C$), the grains are larger and have multiple domains, which stops this sequential polarization process and the system shows its original antiferroelectric property.

In general, the piezoelectric coefficients of thin films are appreciably lower than those of the corresponding bulk crystals or ceramics. This can be mainly ascribed to (a) the clamping of the film by the substrate, which retards field induced motion and (b) a reduction in the extrinsic contribution to the piezoelectric response—such as a reduction in the mobility of certain types of domain walls [73].

11.6. COMPOSITE STRUCTURES

The formation of micro- or nanodispersed ferroelectric phases in ceramics, polymers or glasses with different welldefined geometric structures has opened up a new dimension in the engineering and optimization of desired physical properties. This affords, for example, the interesting possibility of combining the flexibility of a polymer with the piezoelectric or pyroelectric property of an oxide. In addition to the specific properties of its two phases, a biphasic composite is characterized by the scale of the dispersion and the nature of connectivity. In the Newnham convention, the connectivity of a biphasic composite is specified by n–m ($0 \leq n, m \leq 3$) since each phase may be self-connected in zero, one, two, or three dimensions in an orthogonal coordinate system. Thus, a “1–2” connectivity pattern implies that the first phase is self-connected in the form of 1D chains or fibers, while the second phase is self-connected in the form of 2D layers. The convention does not specify whether the fibers of the first phase are parallel or perpendicular to the layers of the second. By convention, the first index (n) usually denotes the ferroelectric phase while the second (m) denotes the matrix (polymer, glass, other ceramic, etc.). Three of the more common and practically realizable connectivity patterns relevant to ferroelectric-polymer systems are 0–3, 1–3 and 3–3 (see Fig. 11.11). A 0–3 composite consists of dispersed ferroelectric particles in a matrix, a 1–3 composite represents isolated ferroelectric fibers in a matrix, while a 3–3 composite consists of two interpenetrating networks. Similarly, a 3–1 composite may consist of a ferroelectric matrix with parallel cavities filled with a polymer [83].

11.6.1. Piezoelectric Properties

Early activity in this area mostly involved piezoelectric–polymer composites [84] and, in view of the strong application potential of such materials, this trend has persisted. PZT, the most studied piezoelectric, has a high piezoelectric strain coefficient (d_{33}) but low voltage coefficient (g_i),[1] which is a drawback in many transducer applications.

On the other hand, piezoelectric polymers such as PVDF have comparatively low d values, but their g values are high due to a low dielectric constant. Thus, one expects a PZT/PVDF composite to have a combination of desirable piezoelectric properties in addition to having lower density and

---

[1] The third rank tensors d (piezoelectric strain coefficient) and g (piezoelectric voltage coefficient) are conventionally defined as $d_{ij} = dP_i/d\sigma_j$ and $g_{ij} = dE_i/ d\sigma_j$ ($i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5, 6$) where P_i is the polarization, E_i is the electric field, and σ is the stress tensor (of rank 2). The two coefficients are related by $g_{ij} = d_{ij}/\varepsilon_0\varepsilon$ where ε is the dielectric constant and ε_0 is the permittivity of free space.

higher flexibility than sintered PZT. PZT–polymer composites with 0–3, 1–3, and 3–3 connectivities have been intensely studied and are widely used in ultrasonic imaging and hydrophones. The superior properties of such composites over their monolithic counterparts arise from the macroconnectivity of the stiff, high-permittivity PZT elements in the compliant, low-permittivity polymer matrix. The polymeric matrices used include PVDF, PVA, PVC, and epoxy resins. The presence of the polymer reduces the dielectric constant and the density, leading respectively to a larger voltage output and lower acoustic impedance. Recently, hot-pressed 0–3 PZT–PVDF (50/50 volume fraction) has been found to show reasonably high d_{33} and good mechanical properties [85]. There are several studies of ceramic–ceramic nanocomposites as well. For example a dispersion of $BaTiO_3$ nanoparticles in a sol–gel matrix, also of $BaTiO_3$, has been prepared [86]. Films having this composite structure were made by spin coating. Similar 0–3 ceramic–ceramic nanocomposites were also formed using PZT [87] and some of these may find application as ultrasonic transducers. Explicit formulas are available for calculating the effective piezoelectric coefficients of 0–3 composites of ferroelectric particles in a ferroelectric matrix in terms of the composition and the constituents' piezoelectric, dielectric, and elastic properties [88].

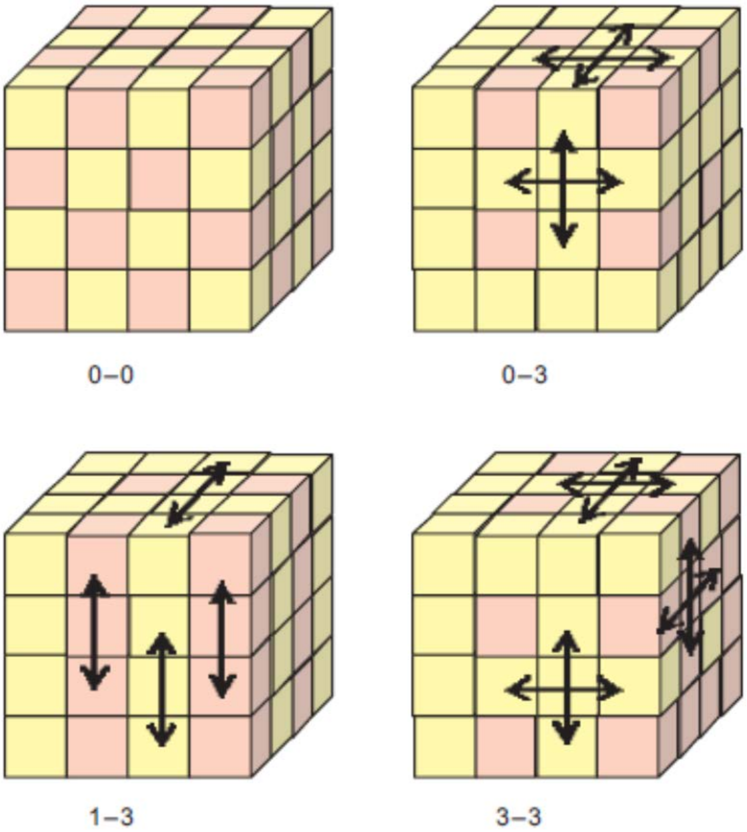


Fig. 11.11. Four of the ten possible connectivity patterns in a biphasic composite system. The pink cells represent the piezoelectric phase, while the yellow cells represent the polymer.

Even though the largest hydrostatic figures of merit for biphasic PZT–polymer composites have actually been obtained from 3–*n* types of structure, such as 3–3 [89], 3–2, and 3–1 [90], most of the older techniques used for the synthesis of these architectures are laborious and result in poorly controlled structures. The use of a robotic direct-write deposition technique ("robocasting") to fabricate a range of such 3–*n* structures has been made and a systematic study of the piezoelectric and dielectric properties as a function of connectivity and PZT/polymer has been presented [91].

11.6.2. Dielectric Properties

A proper choice of composite architecture can lead to a substantial enhancement in the dielectric constant. Recently, a dielectric constant of over 130 has been observed [92] in a 0–3 $BaTiO_3$–cyanoresin composite (51/49 volume fraction), while the dielectric constant of the pure polymer is only 21 (at 1 kHz). This is one of the highest dielectric constants observed so far in $BaTiO_3$-based polymer composites. The large enhancements in the dielectric constant of a 0–3 composite where the matrix was an *irradiated* copolymer has been reported [93]. High-energy irradiation converted the polymer from a normal ferroelectric to a relaxor ferroelectric, leading to a reduction in the *P*–*E* hysteresis at high fields. The dispersed ferroelectric phase was $Pb(Mg,Nb)O_3$–$PbTiO_3$, which is itself a relaxor ferroelectric with a high room temperature dielectric constant. Such a composite showed a weakly temperature dependent dielectric constant of over 250. At optimal composition, the breakdown strength too was very high (120 MV/m), leading to a very large energy storage density of 15 J/cm^3. However, the value of the ceramic content in this case is so high that the material probably loses its flexibility.

A three-phase composite with a dielectric constant above 500 has been reported in [94]. The composite consists of an insulating polymer matrix and a homogeneous distribution of small particles of $BaTiO_3$ and Ni. Such a three-phase polymer–matrix composite is flexible and can be easily fabricated into various shapes at low processing temperatures. The electrical properties of such composites can be analyzed by percolation theory when the metallic filler concentration is close to the percolation threshold. Similarly, the incorporation of metallic Ag particles into SBN [95] and Pt particles in PZT [96] leads to significant improvements in the dielectric properties and also to lower processing temperatures. A drastic increase in the dielectric constant of a PZT/glass composite on introducing 4–15 nm PbS particles has been observed [97]. In samples with PbS volume fraction varying between 0.07 and 0.14, the dielectric constant ranges between 900 and 3500 (at 1 kHz).

The glass ceramic composite has also been used as a vehicle for basic studies of size effects in ferroelectrics. Thus, dispersed nanometer-sized $BaTiO_3$ particles in a residual glass matrix has been grown [98]. They find that the T_C, ε_{max}, and the width of the ε–T curve all scale systematically with the mean size of the $BaTiO_3$ crystallites. The extrapolated critical size for the destabilization of ferroelectric order was found to be 17 nm.

11.6.3. Pyroelectric Properties

Ferroelectric–polymer composites also show excellent pyroelectric properties such as the figure of merit p/ε, where p is the pyroelectric coefficient and ε is the dielectric constant. For example, the value of p/ε for a PZT/polymer composite has been reported to be up to six times higher than in monolithic PZT [99]. Similar results have also been reported for polymer composites involving $PbTiO_3$, $BaTiO_3$, and TGS. Taking the depolarization of the ferroelectric particles into consideration, a general formula for the pyroelectric coefficient of a composite in terms of the dielectric constant of the two components, the volume fraction and the depolarization coefficient have been derived [100]. The pyroelectric coefficient of a 0–3 $PbTiO_3$/PVDF–TRFE composite as a function of the relative polarizations of the ceramic and the copolymer has been studied [101]. They find that the composite with the ceramic and the copolymer polarized in the same direction exhibits the strongest pyroelectricity but relatively weak piezoelectricity.

11.6.4. Optical Properties

Several ferroelectric oxides such as $BaTiO_3$, (Pb, La)TiO_3, $LiNbO_3$, and $KNbO_3$ have useful nonlinear optical (NLO) properties because of their polar crystallographic structure. Transparent TeO_2-based glasses containing dispersed ferroelectric crystallites are promising NLO glass candidates. The advantage of using a TeO_2-based glass as matrix is that its refractive index is close to that of several ferroelectric crystals, which leads to low scattering losses at the glass-ceramic interface. Other types of glasses have also proved useful. Transparent glass–ceramic composites in which ≈40 nm $SrBi_2Ta_2O_9$ particles were dispersed in lithium borate glass appears to be a promising candidate for NLO applications [102]. At room temperature, the system shows a pronounced second harmonic generation (at 532 nm) and a ferroelectric hysteresis with a low dielectric loss.

From a calculation of the electro-optic Kerr constant of a composite containing nonlinear, centrosymmetric dielectric microcrystals, it has been shown that the nonsphericity of the microcrystals may lead to an enhancement of the Kerr constant provided the permittivity of the glass ceramic phase is much smaller than that of the inclusions [103].

11.6.5. Magnetoelectric Properties

The magnetoelectric (ME) effect is characterized by the appearance of an electrical polarization (magnetization) on applying a magnetic (electric) field. A material with a high ME coefficient should find application as magnetoelectric sensors in radioelectronics, microwave electronics, and optoelectronics. Other than the “intrinsic” ME effect shown by many single phase materials (Cr_2O_3, yttrium iron garnets, rare earth ferrites, etc.), an “extrinsic” ME effect may be exhibited, for example, by a ferrite/piezoelectric (ceramic/ceramic) composite, so-called since neither of the two components is individually a ME system. Thus, when a magnetic field is applied to such a composite, the ferrite particles change shape due to magnetostriction, and the strain is passed on to the piezoelectric phase, which results in an electric polarization. The ME effect has, in fact, been observed in composites such as $CoFe_2O_4/BaTiO_3$, (Ni–Co–Mn) ferrite/PZT, etc. Theoretical calculations of the ME effect in composites have been made using Green’s function techniques and considering linear and nonlinear coupling between piezoelectric and piezomagnetic particles. Interestingly, a giant linear ME effect is predicted in, for example, a composite of ferromagnetic rare earth iron alloys filled with a ferroelectric polymer such as PVDF [104]. A three-phase multiferroic composite consisting of particles of Tb–Fe–Dy alloy (Terfenol-d), dispersed (with a volume fraction f) in a mixture of PZT and PVDF has been studied [105]. The ME response of such a composite is due to the large magnetostriction of Terfenol-d and the large piezoelectricity of PZT, and the coupling between the two through a flexible matrix. The percolation threshold for such a system (as measured from dielectric, piezoelectric, and ME properties) occurs at $f_T \approx 0.12$. For $f < f_T$, the system exhibits increasing piezoelectric and ME response. But these quantities drop sharply as f_T is approached and vanish at f_T , above which volume fraction the composite only shows magnetostriction. The expected giant ME effect remains unrealized as yet.

11.7. SYNTHESIS OF NANOCRYSTALLINE FERROELECTRICS

A wide variety of chemical techniques have been successfully used for synthesizing nanocrystalline ferroelectrics with controlled mean size and size distribution (see [106] for a recent review). These include sol–gel, co-precipitation, spraypyrolysis, freeze-drying, microemulsion-mediated reactions, hybrid dry–wet processes, etc. Conventional “dry” powdermixing processes are usually unsuitable because the temperatures involved are often too high—resulting in grain growth and a broad size distribution. Some of the more useful wet techniques are briefly described.

11.7.1. Sol–Gel Process

A *sol* is a dispersion of solid particles (less than ≈100 nm in size) in a liquid, while a *gel* is a rigid (or semisolid), porous interconnected network. The sol–gel process [107, 108] consists of the following generic steps (also described schematically in Fig. 11.12):

(*i*) preparation of a homogeneous solution of the precursors (e.g., metal alkoxides) in an organic solvent miscible with water [or with the reagent used in step (ii)];

(*ii*) conversion of this solution into a sol by treatment with suitable reagents;

(*iii*) gelation of the sol via polycondensation ;

(*iv*) shaping the gel (or viscous sol) into the desired form (granular, spherical, thin film, fiber, etc.);

(*v*) low temperature sintering (usually <500°C) of the shaped gel into the final nanocrystalline product.

The precursor metal alkoxides, $M(OR)_n$ (where M is the metal and R is an alkyl group) can be synthesized by reacting the corresponding metal chloride with an alcohol in the presence of a base such as ammonia. The general reaction is

$$MCl_n + nROH + nNH_3 = M(OR)_n + nNH_4Cl$$

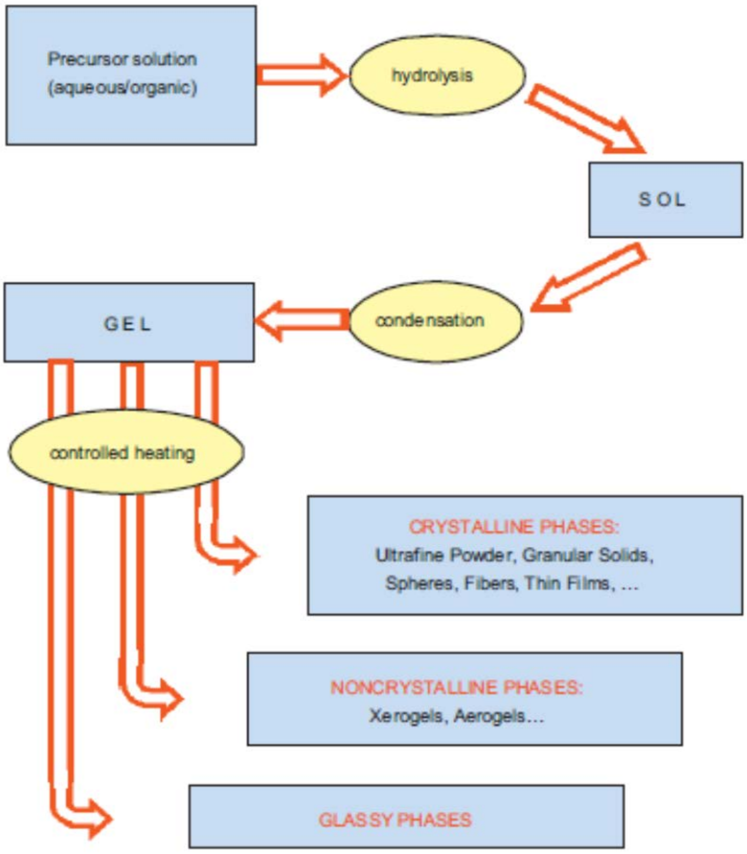


Fig. 11.12. A schematic flow chart of the sol–gel process for the synthesis of low-dimensional oxide/ceramic materials.

The metal alkoxides can then be hydrolyzed to yield oxides, hydrated oxides, and hydroxides in the amorphous, crystalline, or nanocrystalline form. A number of ferroelectric oxides have been successfully synthesized using the basic sol–gel process as well as its many variations. A detailed discussion of the preparation of a variety of ultrafine Ba and Pb based perovskite compounds and their solid solutions has been given [109]. A comparative study of the advantages of the sol–gel technique over the conventional oxide–mixing route in the case of ferroelectric $(Ba_{0.2}Pb_{0.8})TiO_3$ has been provided [110]. Complex alkoxides to prepare submicrometer $Pb(Zr_{1-x}Ti_x)O_3$ has been used [111]. $PbTiO_3$ fine particles by the basic sol–gel technique have been prepared [26]. They dissolved lead acetate in 2-ethylhexanol at 80°C and added a stoichiometric amount of titanium *n*-butoxide to the solution, which was then hydrolyzed by adding water. The resulting precipitate was washed, filtered, and dried. $PbTiO_3$ fine particles of different average size in the 20–200 nm range were obtained by calcining the dried precipitate at different temperatures between 550 and 700 °C. Earlier, similar results starting from a stoichiometric mixture of titanium tetra-isopropoxide and lead isopropoxide, both dissolved in isopropyl alcohol had been achieved [24]. Several hybrid techniques based on the basic sol–gel have also been successfully used. Nanocrystalline $PbZrO_3$ has been synthesized by using NH_4OH to co-precipitate Pb^{2+} and Zr^{4+} ions from an aqueous solution of $Pb(NO_3)_3$ and $ZrOCl_2{\cdot}8H_2O$ [26]. The mixed precipitate was converted to a stable sol by ultrasonication and gelated using 2-ethyl hexanol in presence of a surfactant. The gel was dried and calcined to obtain nano-$PbZrO_3$.

11.7.2. Co-Precipitation and Complexation

In this process, it is essential to achieve conditions that allow simultaneous and not sequential or selective precipitation of the relevant cations. This would ideally allow a molecular level mixing of the components and may be expected to lead to a totally homogeneous product. However, even in this case the resulting precipitate would be a mixture of different phases, probably with different particle sizes. The formation of the final product would then proceed via solid-state atomic diffusion among these particles, resulting in particle growth and a relatively broad size distribution. Wherever possible, it is preferable to form a *complex* of all the cations that would *directly* lead to the desired phase on thermal decomposition. The decomposition temperature should obviously be optimized to minimize particle growth.

As a simple example, let's consider the synthesis of nanocrystalline $PbTiO_3$ by coprecipitation from $Pb(NO_3)_2$ and $TiCl_4$ [28]. $TiCl_4$ (dissolved in HCl) was precipitated as $Ti(OH)_4$ by adding NH_4OH before being converted to soluble $Ti(NO_3)_4$ by adding con. HNO_3. The presence of NH_4NO_3 in the solution prevents the formation of insoluble $PbCl_2$ during the subsequent step. An aqueous solution of $Pb(NO_3)_2$ was added to the $Ti(NO_3)_4$ solution in stoichiometric proportion. The cations were co-precipitated as hydroxides by adding the solution to NH_4OH while keeping the pH controlled at 9–10. The mixed hydroxide precipitate was heated to various temperatures between 460 and 850 °C to produce $PbTiO_3$ nanoparticles in the size range 20–200 nm.

Co-precipitation from organometallic precursors has been used to synthesize ultrafine $BaTiO_3$ [112]. Potassium titanyl oxalate solution was slowly added to a barium acetate solution, being continuously stirred. The precipitate was dried and calcined at 550 °C to produce $BaTiO_3$. A similar technique (using strontium nitrate instead of barium acetate) can be employed to produce $SrTiO_3$.

11.7.3. Microemulsion-Mediated Reactions

A microemulsion is a thermodynamically stable, optically isotropic dispersion of two immiscible liquids (usually water and oil) consisting of nanosized domains of one liquid in the other, stabilized by an interfacial layer of surface active molecules (surfactant). A water-in-oil microemulsion consists of essentially monodispersed droplets (10–100 nm in size) of an aqueous phase in a continuous hydrocarbon phase. The dispersed aqueous phase can be used as nanoreactors for the precipitation of uniform-sized nanoparticles of a variety of metals, oxides, halides, etc. [113, 114]. Typically, the two reactants are confined within the aqueous cores of two different suitably chosen microemulsion systems. On adding the two, the reactants come into contact in a controlled manner since the aqueous droplets continually collide, coalesce, and exchange contents. The resulting precipitated species remains confined within the core of the microemulsion and is subsequently separated, dried, and heat-treated—if necessary—to obtain the nanoparticulate solid (Fig. 11.13). The droplet size, and hence the particle size of the final product, is controlled essentially by the water/surfactant ratio, in addition to other microemulsion parameters.

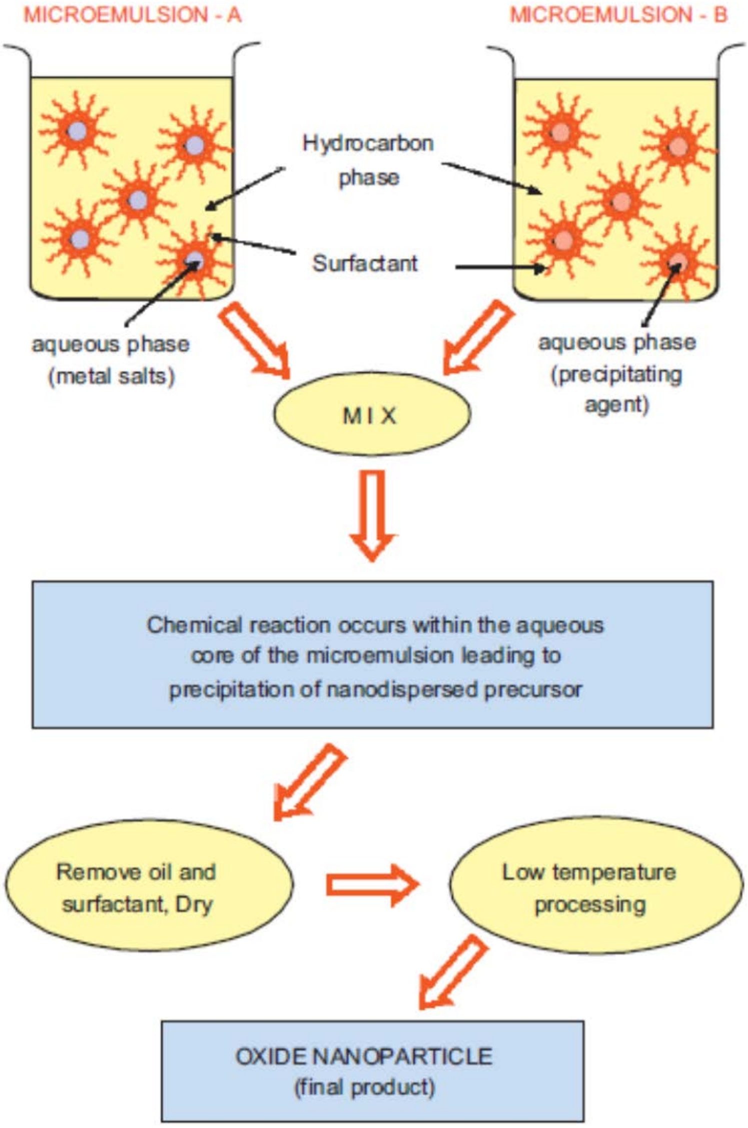


Fig. 11.13. Schematic diagram of the conventional scheme used for microemulsion-mediated precipitation of oxide nanoparticles.

A microemulsion-mediated sol–gel process for the synthesis of perovskite oxides such as $BaTiO_3$, $BaZrO_3$, $SrTiO_3$, and $SrZrO_3$ in the form of ultrafine particles stabilized as colloids in an organic solvent has been reported [115]. The metal alkoxide solution was prepared by refluxing Ba or Sr in isopropanol at 125 °C under dry N_2. The Ba/Sr isopropylate was cooled to room temperature and equimolar quantities of $Ti(nprop)_4$ or $Zr(nprop)_4$ added to it. Immediately thereafter, a water-incyclohexane microemulsion with Brij-30 (polyoxyethylene-4- laurylether) as surfactant was poured into the solution. This results in a transparent solution consisting of nanodispersed metal precipitates. The alcohol, cyclohexane, and water were removed from the solution by vacuum evaporation and the residue was washed in acetone using a centrifuge. The dried precipitates were sintered at various temperatures to obtain 9 nm $BaTiO_3$ (at room temp.), 18 nm $SrTiO_3$ (700 °C, 1 h), 11 nm $BaZrO_3$ (800 °C, 1 h), or 19 nm $SrZrO_3$ (900 °C, 1 h). A flow chart of the process is shown in Fig. 11.14.

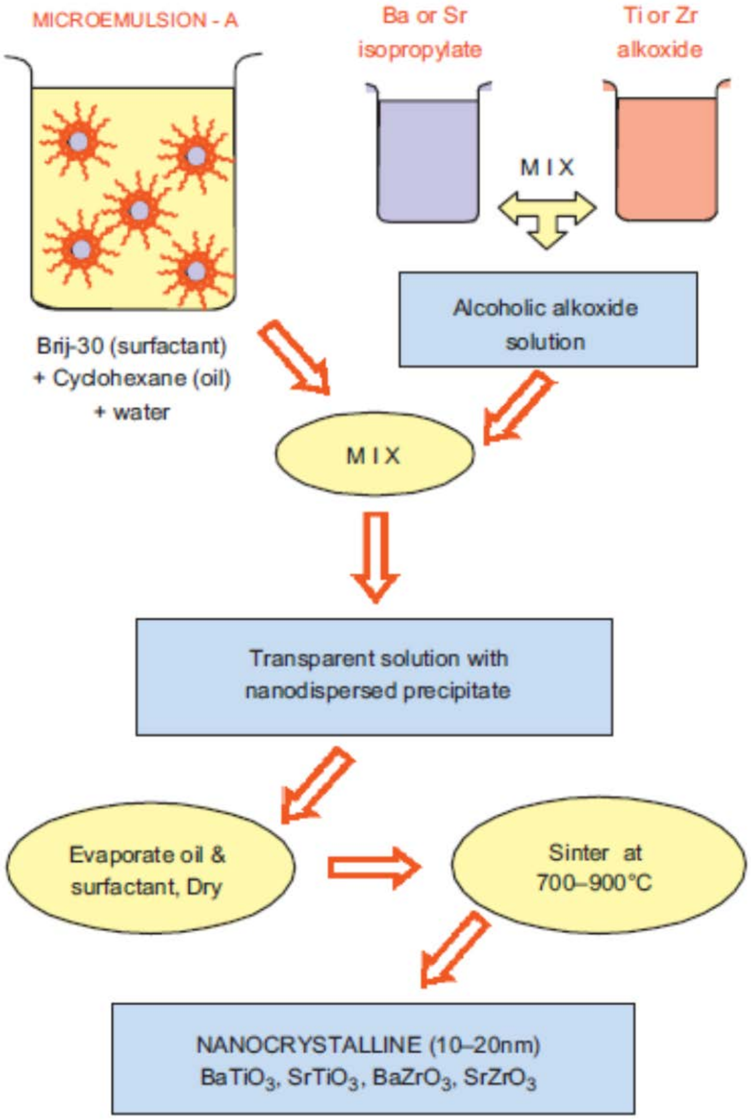


Fig. 11.14. Schematic diagram of the microemulsion-mediated sol–gel reaction for producing oxide nanoparticles.

A water–oil emulsion has been used to synthesize 100 nm (and larger) $PbTiO_3$ microparticles [116]. The aqueous phase contained a solution of equimolar lead nitrate and titanium oxynitrate (produced by hydroxide precipitation from $TiCl_4$ using NH_4OH, and dissolving the precipitate in HNO_3) dispersed in *n*-octane with 1 wt% SPAN-80 as surfactant. Due to the nature of the surfactant as well as the low surfactant/water ratio used, the resulting mixture is an *emulsion* (which phase-separates spontaneously with time) rather than a microemulsion. The advantage of this technique is that an additional precipitation step is unnecessary and the $PbTiO_3$ fine particles are obtained directly by removing the solvents and subsequent thermal treatment. The disadvantage of not using a microemulsion is that the resulting particle size was rather large: 100–200 nm for samples calcined at 700 °C.

11.7.4. Hydrothermal Reactions

Hydrothermal synthesis generally involves water as a catalyst as well as a component and is used to describe reactions at temperatures above 100 °C and pressures above 1 atm. This route offers certain important advantages such as ease of use, high product purity, good control over particle size, and low process cost. For example, $BaTiO_3$ powder has been synthesized by a low temperature, low pressure hydrothermal route [117] based on the reaction of TiO_2 (anatase) powder in an alkaline aqueous solution of $Ba(OH)_2$ at 90 °C. Hydrothermal routes generally involve the processing of an aqueous solution or suspension of the precursor gel powder at somewhat elevated temperatures and pressures. Such precursor gel powders are usually nmsized gels containing precipitated metal salts (such as chlorides, sulfates, and nitrates) or metal alkoxides. However, spherical gel powders have also been used in hydrothermal synthesis. Thus spherical nanoparticles of $PbTiO_3$, $BaTiO_3$, and $SrTiO_3$ from spherical TiO_2 gel powder have been synthesized [118], while spherical nanoparticles of $PbZrO_3$ and $Pb(Zr,Ti)O_3$ where prepared from spherical gels of ZrO_2 and $ZrTiO_4$, respectively. The TiO_2 and ZrO_2 spherical gel powders were first synthesized by the thermal hydrolysis of $TiCl_4$ and $ZrOCl_2$, respectively, while the spherical $ZrTiO_4$ gel powder was obtained from thermal hydrolysis of a mixture of $TiCl_4$ and $ZrOCl_2$ in an alcohol–water solution. The sources for the *A*-site cations of the perovskite compounds were $Pb(CH_3COO)_2{\cdot}3H_2O$, $Ba(OH)_2{\cdot}8H_2O$, and $Sr(OH)_2{\cdot}8H_2O$. The most interesting point about this process is that the spherical precursor gel powders retain their shape during the hydrothermal treatment to produce rather large (400–800 nm) but perfectly spherical particles of the perovskite oxide product.

Since the hydrothermal technique leads to considerable energy savings due to relatively low processing temperatures, several modifications of the basic process have been investigated. A combination of microwave and hydrothermal techniques to synthesize a number of ultrafine perovskite such as $BaTiO_3$, $SrTiO_3$, $(Ba,Sr)TiO_3$, $PbTiO_3$, $Pb(Zr,Ti)O_3$, $Pb(Mg,Nb)O_3$, etc. have been used [119]. The use of microwaves enhances the reaction kinetics by one or two orders of magnitude over the conventional hydrothermal process. Since the reaction vessel is sealed during the process, the synthesis of Pb- and Bi-based systems, for example, is environmentally safer. The microwave-hydrothermal process typically uses a microwave digestion system operating at 2.45 GHz and 600 W and at a maximum pressure of 200 psi. As a typical example, $BaTiO_3$ from a solution of $Ba(NO_3)_2$, $TiCl_4$, and KOH at 20 psi has been prepared. The product consisted of spherical, unagglomerated particles having a size in the range of 100–200 nm.

11.8. DEPOSITION OF FERROELECTRIC THIN FILMS

During the past couple of decades, major advances have taken place in the techniques for the deposition of uniform, homogeneous, device-quality thin films. Several different techniques have been successfully employed for the synthesis of high quality ferroelectric thin films suitable for basic studies as well as device applications. A process for depositing ferroelectric thin film based devices should generally possess the following characteristics [120]:

(1) possibility of integration with metallic or conductive oxide electrode layers with physicochemical properties different from the ferroelectric material,

(2) compatibility with integrated device processing with specific microstructure on specific substrates,

(3) production of highly oriented as well as polycrystalline films and heterostructures with specific properties,

(4) ability to produce patterned structures, superlattices, and layered heterostructures,

(5) perfect reproducibility of the deposition process,

(6) economical with high deposition rates.

The three most successful techniques that satisfy most of these criteria are pulsed laser ablation deposition (PLAD), metal-organic chemical vapor deposition (MOCVD), and plasma/ion-beam sputter deposition (PSD/IBSD). Here we provide brief descriptions of these processes. It is relevant to note that these techniques can also be suitably modified to produce nanocrystalline thin films of various materials (metals, oxides, nitrides, sulfides, etc.) with controlled grain size in the 5–100 nm range [121]. Some of the other techniques that have been used with varying degrees of success for producing ferroelectric films are chemical vapor deposition, metal-organic deposition, liquid source mist deposition, and sol–gel.

11.8.1. Pulsed Laser Ablation Deposition

The main advantage of PLAD is that it allows rapid and *stoichiometric* deposition of multication as well as multicomponent oxide films. A typical PLAD system requires an excimer laser (such as a KrF laser operated at 248 nm, with an output power density of ~1 Jcm^{-2}, and a repetition rate of 1–10 Hz), a rotating target holder, and a substrate holder that can be heated to about 800 °C. The process allows considerable control over the composition, microstructure, and properties. The process parameters that mainly determine these quantities are the ambient gas (usually O_2) pressure, the substrate temperature, the target-to-substrate distance, the laser energy, and wavelength.

Recent Applications

SBT films were grown by PLAD as a part of a metal–ferroelectric–insulator–semiconductor FET structure (SBT/SiON/Si) [122]. Ferroelectric $Sr_2(Ta_{1-x}Nb_x)_2O_7$ films on Pt/Ti/SiO_2/Si and SiO_2/Si(100) substrates have been deposited [123]. Excellent ferroelectric properties were observed [124] in (Pb,La)(Zr,Ti)O_3 deposited directly on metal surfaces. PZT thin films deposited on different substrates (Au coated Si, Al_2O_3, Corning glass, etc.) by reactive pulsed laser (Nd-YAG) deposition show good piezoelectric and ferroelectric properties [125]. It has been shown that the coercive field in PLAD-deposited thin-film ferroelectric capacitors consisting of PZT sandwiched between $La_{0.5}Sr_{0.5}CoO_3$ electrodes is comparable to bulk ceramic values [126]. Textured thin-film capacitors with a columnar microstructure were found to show lower switching voltages than epitaxial films. High quality $BaTiO_3$ thin films were epitaxially grown on $SrTiO_3$ (100) substrates from a single crystal target [127]. $BaTiO_3$/$YBa_2Cu_3O_7$ bilayer structures on $LaAlO_3$ (100) substrates have been also grew. $Ba_{0.5}Sr_{0.5}TiO_3$ thin films on conducting RuO_2 [grown epitaxially on (100) $LaAlO_3$] have been deposited [128], which showed pure (111) orientation normal to the substrate surface and exhibited dielectric constant above 600 and dielectric loss of 0.03–0.05 at 100 kHz under zero dc bias. It was able to deposit perfectly *c*-axis oriented, single phase, ferroelectric $PbTiO_3$ films directly on (100) Si [129].

11.8.2. Metal-organic Chemical Vapor Deposition

MOCVD outstrips all other techniques as far as compositional control, large-area film uniformity, and compatibility to ULSI fabrication technology are concerned. Since it is extensively used in commercial IC fabrication, the process parameters and instrumentation are understood extremely well. However, the major disadvantage still faced by this technique in the processing of ferroelectric oxides is the nonavailability of some of the metal-organic precursors with acceptable purity, vapor pressure, and chemical stability. A typical experimental setup (used for the synthesis of $PbTiO_3$ and PZT thin films) [130] consists of a cold-wall reactor with horizontal flow design using a resistively heated inconel susceptor. The system is equipped with three temperature and pressure controlled liquid-source stainless steel bubblers containing the organometallic precursors (for Pb, Ti, and Zr). In the case of low vapor pressure solid precursors, one directly sublimates the solid or evaporates the melt into the carrier gas, which is usually N_2, He, or Ar. For most materials, better thickness and compositional uniformity can be achieved when the processing occurs at pressures below about 100 Torr.

Recent Applications

The application of MOCVD to the synthesis of ferroelectric thin films has been reviewed [131]. Many common ferroelectrics such as SBT, BST, PZT, etc. have been successfully grown by

MOCVD and there is substantial volume of literature on the subject. In MOCVD-grown $BaTiO_3$ films on (100)Pt/(100)MgO substrates, the grain size was found to be strongly affected by $p(O_2)$ and showed a maximum when $p(O_2)$ was between 66 and 93 Pa. The dielectric constant increased from 93 to 640 with increasing grain size from 20 to 130 nm, showing a broad peak at 350–380 K [132]. The important issue of new, single source MOCVD precursors for SBT, SBN, and SBTN has been addressed [133]. It is clear that MOCVD remains the most promising technique for ferroelectric thin film based device fabrication and integration.

11.8.3. Plasma Sputter Deposition

There is extensive work on the PSD of ferroelectric thin films using multiple element targets as well as multicomponent oxide targets. The target (cathode) is connected to a dc or rf power supply (100 V to few kV) while the substrate (anode) may be grounded, electrically floating, or at small dc bias and is typically heated to a few hundred °C. After evacuating the chamber, a gas (typically a few millitorr of Ar) is introduced and serves as the medium in which the plasma discharge is initiated and sustained. It is advantageous to confine the plasma magnetically using a magnetron system. When a visible discharge is maintained between the electrodes, there is a flow of current accompanied by a deposition on the substrate. The process parameters that control the composition, microstructure, and film quality include the gas pressure, the substrate temperature, the target-substrate geometry, and the applied dc/rf voltage. Thus, oriented and epitaxial films can usually be grown when conditions are such that the mean kinetic energy of the sputtered atoms is rather low when they arrive at the substrate. The use of elemental metallic targets in the deposition of PZT films with controlled stoichiometry and properties has been demonstrated [134]. They used a disc-shaped target with several unequal sectors of Pb, Zr, and Ti. They suggest that such a process is governed by three factors: (a) the formation of a reproducible oxide layer on the target surface, (b) the stability of the oxide species during transport through the plasma, and (c) the nucleation and growth of the film on the substrate.

Recent Applications

PZT (and Nb substituted PZT) films on Si by rf-magnetron sputtering have been deposited and the effect of process parameters on dielectric, ferroelectric, and piezoelectric properties have been studied [135]. Ferroelectric PLZT thin films on 200 mm diameter $Pt/Ti/SiO_2/Si$ substrates in a multichamber production system have been deposited [136]. Using rf-magnetron sputtering from elemental targets, highly oriented (111) $PbTiO_3$ films were deposited on amorphous sapphire for optical studies [137]. SBT films have been deposited by rf-magnetron sputtering on a $Pt/Ti/SiO_2/Si(100)$ structure between 500 and 600 °C. The maximum remnant polarization value

($2P_r$) of the SBT film was 15 μC/cm^2 [138]. BST films have been grown on $SrTiO_3$, $LaAlO_3$, MgO, α-Al_2O_3, and $Y_3Fe_5O_{12}$ by rf sputtering [139]. In recent times, many other ferroelectric thin films have also been grown using this technique.

The IBSD technique typically uses an Ar^+ beam directed at 45° to the target normal. In a more successful modification, multiple ion beams have been used, where each beam is directed at different elemental targets such as Pb, Ba, Zr, and Ti [140].

11.8.4. Thick Film Printing

Thick film printing techniques for PZT have been developed to produce thick films in excess of 100 μm. Such thicknesses are desired for applications that require actuation forces that cannot be achieved with the much thinner sol-gel films. A thick film printing process whereby a PZT paste is made from a mixture of 95% PZT powder, 5% lead borosilicate powder, and an organic carrier was described in work [141]. The paste was then printed through a stainless steel screen using a thick film printer. Printing was performed on an oxidized Si substrate that is capped with a Pt electrode. After printing, the paste was dried and then fired at 850 °C to 950 °C. Printing could be repeated to achieve the desired thickness. The top electrode consisted of an evaporated Al film. It has been found that it was possible to perform plasma-based processing on the printed substrates, but that the porous nature of the printed PZT films made them unsuitable for wet chemical processing.

11.9. LOCAL PROBE MEASUREMENTS

Direct observation of ferroelectric domains provides a wealth of information on static and dynamic properties of ferroelectrics. A number of techniques, such as polarized optical microscopy, electron microscopy, surface decoration, pyroelectric emission, and etching, have been employed for visualization of domain structures [1, 2]. The main limitations of these methods have been either low spatial resolution or the destructive nature of the visualization techniques.

Rapid development of electronic devices based on ferroelectric thin films [142] generated a strong need for studies of ferroelectric properties at the nanoscale. Fortunately, this need appeared at the same time as new techniques for nanoscale characterization of materials became available. Specifically, scanning force microscopy (SFM) has emerged as a powerful tool for high-resolution characterization of virtually all types of materials, such as metals, semiconductors, dielectrics, polymers, and biomolecules [143-145]. In the field of ferroelectricity, the application of the SFM technique resulted in a real breakthrough providing an opportunity for nondestructive nanoscale visualization of domain structures in ferroelectric thin films. The employment of SFM makes possible nanoscale mapping of the surface potential, evaluation of local electromechanical properties, and dielectric constant measurements. In other words, characterization by means of SFM provides crucial information on the dielectric properties of ferroelectrics with unprecedented spatial resolution. Another promising development related to SFM is modification of the ferroelectric properties at the nanoscale via local polarization reversal induced by a conductive SFM tip.

Clearly, it is essential to understand the basic phenomena governing size effects in ferroelectric nanomaterials if we want to apply them to the next generation of high density electronic devices and microelectromechanical systems. In addition to some of the important studies described in the preceding sections, a number of new and interesting results have started arriving from various types of high resolution, local probe measurements such as SFM. These techniques allow us to directly visualize the ferroelectric domain structure and dynamics. SFM-based methods for ferroelectric domain imaging make use of either the piezoelectric property of ferroelectrics or the presence of surface charge associated with the presence of the spontaneous polarization [146].

Scanning force microscopy can be considered as a combination of a surface force apparatus and a surface profilometer, as it is based on local monitoring of the interaction forces between a probing tip and a sample [144]. The forces acting on the tip after it has approached the sample surface cause a deflection of the cantilever according to Hooke's law. This deflection can be detected optically or electrically with subangstrom accuracy and is controlled by a feedback device, which regulates the vertical position of the tip over the surface. By keeping the deflection constant while scanning the sample, a three-dimensional map of the surface topography can be obtained.

Besides this constant force mode, many other modes have been developed. The response of the cantilever to the externally modulated force (for example, due to an applied ac bias) can be used to map such physical properties as mechanical stiffness, friction, electric fields, density of electronic states, and so forth.

Depending on the type of tip-sample force interaction (attracting or repelling), the SFM can operate in two different regimes: noncontact or contact, respectively.

(a) Noncontact Mode: In the noncontact regime, the tip is scanned over the surface at a distance of 10–100 nm, which is controlled, for example, by monitoring the resonant frequency of the cantilever [147]. The tip-sample interaction in this regime is dominated by the long-range polarization and electrostatic forces. Because of this feature, noncontact SFM can be used for ferroelectric domain imaging by detecting the electrostatic field of the surface polarization charges. This mode of SFM is called electrostatic force microscopy (EFM) [148]. Quantitative information on local surface potential related to spontaneous polarization can be obtained by means of scanning surface potential microscopy (SSPM), or Kelvin probe force microscopy (KPFM) [149, 150], a technique complementary to EFM. Scanning capacitance microscopy (SCM) can be used to generate an image of trapped charges by measuring local tip-sample capacitance [149, 151]. SCM is not necessarily a noncontact mode of operation, as the probing tip can be placed in direct contact with the sample surface to measure the sample capacitance. General disadvantages of noncontact methods include susceptibility to screening effects, sensitivity to sample surface conditions, and low resolution in ambient air.

Electrostatic Force Microscopy

Imaging of ferroelectric domains in the noncontact mode is based on the detection of the modulated electrostatic interaction force between the probing tip and polarization charges. There are two modulation methods in the noncontact imaging.

In the conventional approach, the cantilever is made to oscillate near its resonant frequency by use of a piezoelectric bimorph. When the tip is brought close to the surface, the attractive force gradient acting on the tip alters the force constant k_0 of the cantilever as $k' = k_0 - \partial F/\partial z$. This in turn leads to a change in the resonant frequency and in the vibration amplitude. A feedback loop adjusts the tipsample distance to keep the amplitude of oscillation constant [147]. Obviously, in the case of the ferroelectric sample, there is an electrostatic contribution to the attractive force due to the Coulomb interaction between a surface polarization charge and an image charge Q_t in the probing tip. As the tip crosses the wall, it experiences a change in the force gradient, and the feedback loop alters the tip-sample distance to keep the gradient and the vibration amplitude constant. This produces a variation of contrast in the feedback signal image, which can be interpreted as an image of the domain wall (Fig. 11.15) [152]. Since the Coulomb force is proportional to the product of the polarization and image charges, the force gradient signal provides information only on the polarization magnitude and not the sign. This implies that the contrast of opposite 180° domains will be the same and that only domain walls will be visible because of the spatial variation of the charge density in the vicinity of a 180° domain boundary.

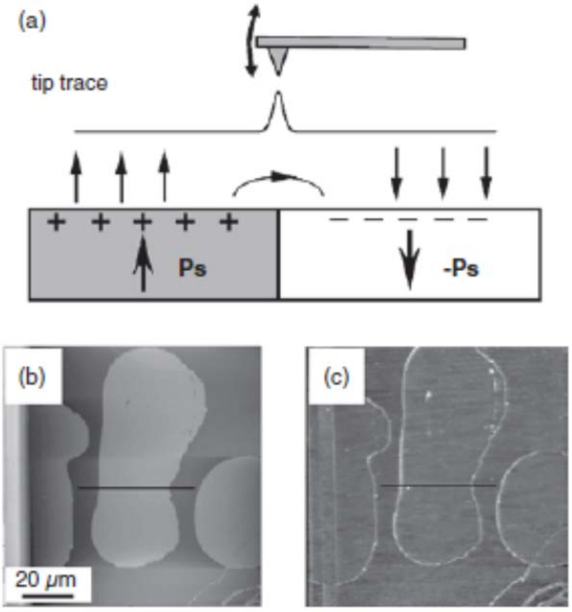


Fig. 11.15. (a) Sketch illustrating a mechanism of domain wall image formation in EFM. (b) Topographic and (c) EFM images of the etched lead titanate crystal. The domain structure consists of antiparallel *c*-domains, which appear with the same contrast in the EFM image, and domain walls appear as bright lines due to the surface charge gradient at the 180° domain walls. Topographic contrast is due to the differently etched surfaces of positive and negative domains [152].

(b) *Contact* (*Piezoresponse*) *Mode*: In the contact regime, the probing tip is in mechanical contact with the sample surface and senses repulsive short-range forces. The difference in mechanical, structural, electrochemical, dielectric, and piezoelectric properties of opposite ferroelectric domains can provide domain contrast in the SFM contact regime. Contact SFM methods of domain imaging include a topographic mode of atomic force microscopy (AFM), lateral (friction) force microscopy (LFM), piezoresponse force microscopy (PFM), and scanning nonlinear dielectric microscopy (SNDM). SFM-based methods of ferroelectric domain imaging are summarized in Table 11.3.

When the probe tip is in contact with the ferroelectric sample, the domain structure can be observed in the SFM piezoresponse mode by monitoring the local electromechanical oscillation of the sample caused by an ac voltage applied through the probe tip. The sample vibration due to an ac signal of frequency ω also has a second harmonic component at 2ω due to the electrostrictive effect and the dielectric response.

Contact domain imaging can be divided into static and dynamic, or voltage-modulated, methods. Static imaging methods, such as a topographic mode of AFM and lateral force microscopy, make use of the surface domain dependent properties of ferroelectrics, such as surface corrugations associated with the presence of different domains, difference in structure of polar faces of opposite domains, variations in friction forces, and so on. Dynamic methods, which include PFM and SNDM, are based on voltage modulation and detection of the electrical and mechanical responses of opposite ferroelectric domains to the applied ac voltage. The contact SFM imaging methods provide significant advantages, such as high lateral resolution (well below 10 nm), a possibility of the three-dimensional reconstruction of domain structure, and effective control of nanodomains. However, interpretation of the domain images could be complicated by cross-talk between different mechanisms involved in the domain contrast formation.

Table 11.3. Ferroelectric domain imaging by SFM

Ferroelectric domain imaging by SFM	
Non-contact imaging	
EFM	Electrostatic Force Microscopy
SSPM	Scanning Surface Potential Microscopy
KPFM	Kelvin Probe Force Microscopy
SCM	Scanning Capacitance Microscopy
Contact imaging	
LFM	Lateral (friction) Force Microscopy
PFM	Piezoresponse Force Microscopy
AFM	Atomic Force Microscopy
SNDM	Scanning Nonlinear Dielectric Microscopy

Studying the piezoelectric properties of individual grains in relaxor ferroelectric films [solid solution of $PbTiO_3$ and $Pb(Mg,Nb)O_3$] by SFM, a clear correlation between the values of the effective piezoelectric coefficients and the grain size has been observed [49]. Larger grains show stronger piezoelectric behavior, spontaneous polarization, and coercivity. A contrasting result was reported in [153], who deposited PZT films by reactive sputtering on conductive Nb:$SrTiO_3$ and patterned features down to 50 nm by e-beam lithography. Piezoelectric sensitive SFM in the contact mode revealed a strong *increase* of the piezoelectric response for feature sizes with lateral dimensions below 300 nm (in 200 nm films). This behavior is believed to be due to vanishing *a*-domains.

Scanning piezoresponse microscopy has been used [154] to show that in a ferroelectric thin film that has been depoled using the shrinking hysteresis loop method (similar to degaussing a ferromagnet) the zero-field, zeropolarization state actually consists of an array of up and down polarized domains, and each domain is roughly the size of the grains in the sample. The microscopic correlation between the crystallinity, domain arrangements, and switching behavior of SBT thin films deposited by PLAD and flash MOCVD have been studied [155]. They show that the domain contrast of an individual grain and its switchability are strongly dependent on the grain orientation.

Scanning probe techniques have also been used to demonstrate that single, 10 nm wide nanowires of $BaTiO_3$ are ferroelectric [156]. These nanowires, synthesized by a solution based method, are essentially single crystalline, defect-free ferroelectrics. Due to their one-dimensional structure, they should serve as the building blocks for fabricating a variety of nanoscale functional devices. Electrical polarization can be reproducibly induced on such nanowires but is retained only for a few days.

It appears safe to anticipate that in the next few years, many of the fundamental problems in nanoferroelectrics would be addressed and resolved using scanning probe techniques.

11.9.1. SFM of Ferroelectrics

Application of SFM to ferroelectric materials opened new possibilities for their characterization, control of ferroelectric properties at the nanoscale, and fabrication of regular nanodomain patterns for use in functional devices.

Characterization

Ferroelectric thin films are currently attracting significant interest because of the recent achievements in their integration into nonvolatile random access memory devices characterized by high-speed access, low power consumption, almost unlimited endurance, and extreme radiation hardness [157]. However, for ferroelectric memories to become high volume, mainstream devices competitive with other types of nonvolatile memories, two major requirements must be fulfilled. *The first* is the ability to manufacture at high device density in the gigabit range, where the constituent ferroelectric capacitors are reduced to submicrometer dimensions. *The second* is the necessity to ensure that ferroelectric memory devices of these dimensions can achieve the necessary performance, reliability, and lifetime requirements. These two requirements demand intensive studies of ferroelectric films to achieve a major breakthrough in understanding their nanoscale properties. For high-density memories, the ferroelectric capacitor must be scaled to area and thickness dimensions, which may approach the stability limits for the ferroelectric phase. In addition, the intrinsic capacitor-to-capacitor variability becomes an issue as the number of grains per capacitor becomes small. Despite significant progress in the development of ferroelectric films and devices over the last several years, there remains a great deal that is not known regarding the domain structure of thin films and its role in determining device switching speed and lifetime. The major limitation in acquiring this crucial information has been the lack of experimental methods to characterize ferroelectric switching behavior at the microscopic level. Most of such experiments have been performed by destructive methods, such as transmission electron microscopy.

In this respect, SFM has a significant advantage over other high-resolution techniques, as it provides the possibility of direct nondestructive high-resolution characterization of ferroelectric materials at the nanoscale level. Over the last 10 years SFM has proved to be an indispensable tool capable of dielectric, surface potentiometric, and electromechanical characterization of ferroelectrics at the micro- and nanoscale levels. Application of SFM has rendered direct information on local surface potential and charge distribution, nanoscale domain arrangement, and domain kinetics during phase transition [158-164]. SFM has been successfully applied to high-resolution studies of static and dynamic properties of domains in ferroelectric thin films and capacitors [165-196]. Recent achievements facilitated by the SFM approach include manipulation of domains as small as 20 nm in diameter [170, 197], direct study of domain nucleation and growth during polarization reversal [189], investigation of the mechanisms of polarization decay and ferroelectric fatigue at the subgrain level [167, 173, 179, 182], evaluation of the switching behavior of individual nanocapacitors via hysteresis loop measurements [190-195], and investigation of the dielectric breakdown mechanism [198]. Figure 11.16 illustrates several examples of SFM capabilities in the nanoscale characterization of ferroelectrics. These results demonstrate that SFM is well positioned to become a major characterization and testing tool for ferroelectric devices. Nanoscale SFM characterization in combination with phenomenological theory for ferroelectric films will lead to better understanding of their basic properties and significant improvement in the performance of nanoscale ferroelectric devices.

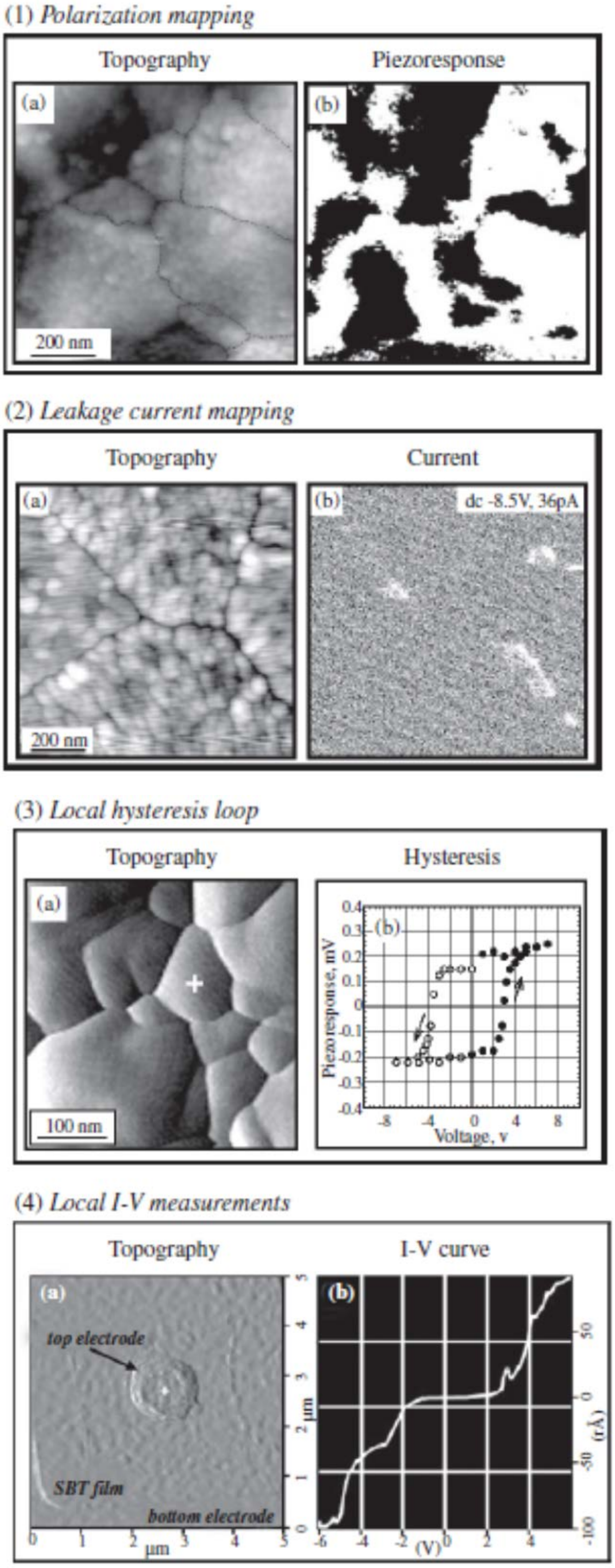


Fig. 11.16. Examples of SFM capabilities in high-resolution characterization of ferroelectrics. (1) Simultaneous acquisition of surface topography and polarization distribution [170]. (2) Mapping of leakage current sites [198]. (3) Measurement of local (within 20-nm size region) coercive voltage and piezoelectric coefficient [170]. (4) Measurement of current-voltage characteristics of a micrometer-size ferroelectric capacitor [198].

Control

SFM provides a unique opportunity for controlling ferroelectric properties at the nanoscale and direct studies of the domain structure evolution under an external electric field, which cannot be matched by previously available techniques. A conductive probing tip can be used not only for domain visualization but also for modification of the initial domain structure. Application of a small dc voltage between the tip and bottom electrode generates an electric field of several hundred kilovolts per centimeter, which is higher than the coercive voltage of most ferroelectrics, thus inducing local polarization reversal.

When an electric field is applied opposite to the polarization direction of a single-domain ferroelectric, the switching mechanism involves several steps [199]. *First*, new domains with the reverse polarization direction nucleate, mainly at the surface, and then grow through the sample thickness (forward growth). *Second*, the grown domains expand sideways as new domains continue to form. *Finally*, the growing domains coalesce to complete the polarization reversal of a ferroelectric. The contribution of the forward and sideways growth mechanisms is a function of the applied field and the electrodes and to a large extent determines the switching time. Until recently, observation of domain dynamics during switching in thin films has not been performed because of the absence of an appropriate experimental technique. It has become possible with the help of the PFM method. This approach allows direct studies of the mechanism of polarization reversal and measurements of the key parameters of domain dynamics, such as nucleation rate, domain wall velocity, spatial distribution of nucleation sites, etc.

However, a poor time resolution, which is determined by the time required for image acquisition (at a scan rate of 2 Hz and an image resolution of 256 × 256 pixels, it takes about 2 min to acquire an image), makes the *insitu* measurements of domain dynamics during fast switching processes difficult. Although SFM can readily be used to investigate slow polarization relaxation processes with characteristic times on the order of minutes and above, it is a challenge to deduce the mechanism of domain transformation when polarization reversal occurs in a matter of microseconds and faster.

This problem is usually circumvented by studying of domain structure dynamics in a quasi-static regime using *step-by-step switching*. This method has previously been used at the macroscopic level in classical switching experiments on correlating domain structure evolution and transient current in ferroelectric crystals [199] and later was applied to thin films [200, 201]. In this approach, partial reversal of polarization is generated by applying a voltage pulse shorter than the total switching time with subsequent piezoresponse imaging of the resulting domain pattern. By applying a sequence of voltage pulses of successively increasing duration and acquiring piezoresponse images after each pulse, a consistent picture of time-dependent behavior of domain structure can be obtained, providing information on the domain wall velocity, its spatial anisotropy, and its field dependence. To avoid data misinterpretation due to spontaneous back switching between the pulses, stability of the produced intermediate patterns should be checked by acquiring domain images at different time intervals after single pulse application.

An example of this approach is illustrated in Fig. 11.17, which shows a sequence of piezoresponse images of a growing domain in the PZT film. Based on the contrast of the domain, it is assumed that it fully extended through the film thickness after the first pulse and further growth occurs via the sideways expansion only. The diameter of the smallest stable domain produced by the first voltage pulse of 5 V varied from 20 nm to 40 nm. Since the electric field generated by a probing tip quickly decreases with distance, the rate of domain growth, after fast initial expansion, declines until the domain stops growing. To describe the sideways expansion of the domain it is necessary to take into account the field dependence of the domain wall velocity and the spatial distribution of the electric field generated by the probing tip [202].

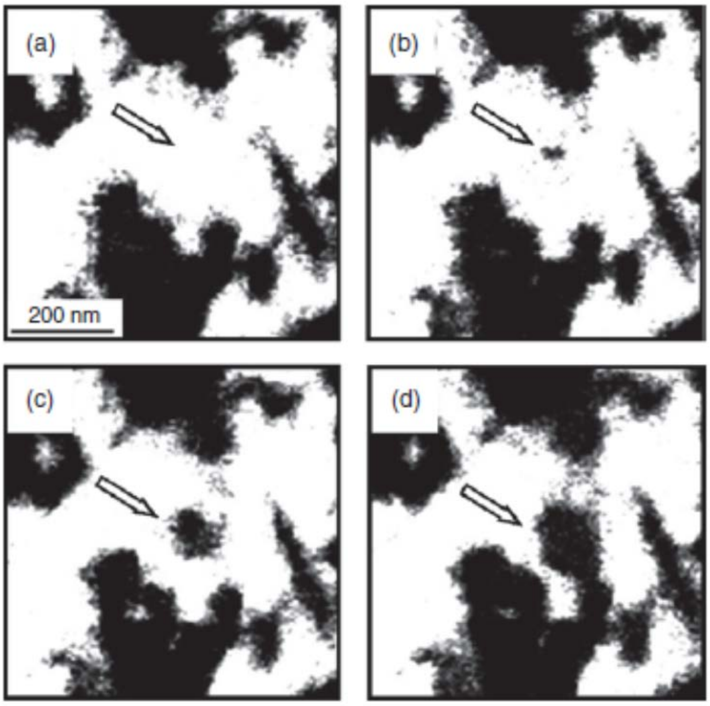


Fig. 11.17. Sequence of PFM images of PZT film showing a sideways growth of a domain in the center of the image. The growth is induced by application of 5-V voltage pulses of increasing duration to the same spot by a fixed probing tip. (a) Original domain structure. (b–d) Domain images after application of voltage pulses. Pulse duration: (b) 0.25 s, (c) 0.5 s, (d) 1 s.

Fabrication

SFM capability of domain control creates a possibility of developing ferroelectrics with specifically designed nanoscale domain patterns, which can find application in novel electronic devices. For example, it has been proposed to use SFM as a basis for high-density data storage with a ferroelectric layer as a recording medium and nanoscale domains as data bits [166, 197]. Recently, the feasibility of practical application of ordered domain structures for the fabrication of nanostructures has been demonstrated [203]. Manipulation of polarization of ferroelectric substrates in SFM opens a new way to assemble nanostructures consisting of oxide substrates, metal nanoparticles, and organic/biological molecules. The idea of using ferroelectric templates for nanofabrication is based on utilizing the local surface electronic properties and surface chemical reactivity, which can be controlled by switching the direction of spontaneous polarization. On the surface of a ferroelectric, an abrupt change in the normal component of the spontaneous polarization results in bound polarization charges and in the appearance of a depolarizing field [204]. Energy consideration requires that this field must vanish outside the ferroelectric, that is, it must be compensated for. In a ferroelectric capacitor, this field can be completely compensated (screened) by the charges on the electrodes. In addition, screening by charge absorption at the surface can be accompanied by accumulation of the free carriers in the ferroelectric sample just near its surface. Surface potential, determined by the relative contribution of the external (surface charges) and internal (nonequilibrium carriers) screening mechanisms, is a crucial parameter for assembling molecular structures. Therefore, investigation of the interface properties and the deconvolution of the external and internal screening in ferroelectrics is critical for developing new nanofabrication methods using domain patterned ferroelectric templates.

Several groups reported generation of regular domain patterns with nanoscale periodicity by SFM [166, 197, 201, 205-207]. Nanodomain patterns can be successfully fabricated not only in thin films but also in bulk crystals, in spite of the fact that the tip-generated electric field quickly decreases with distance to the bulk of the sample [206, 208]. Examples of SFM-fabricated nanodomain patterns in $BaTiO_3$, $Sr_{0.61}Ba_{0.39}Nb_2O_6$, and $RbTiOAsO_4$ single crystals are shown in Fig. 11.18.

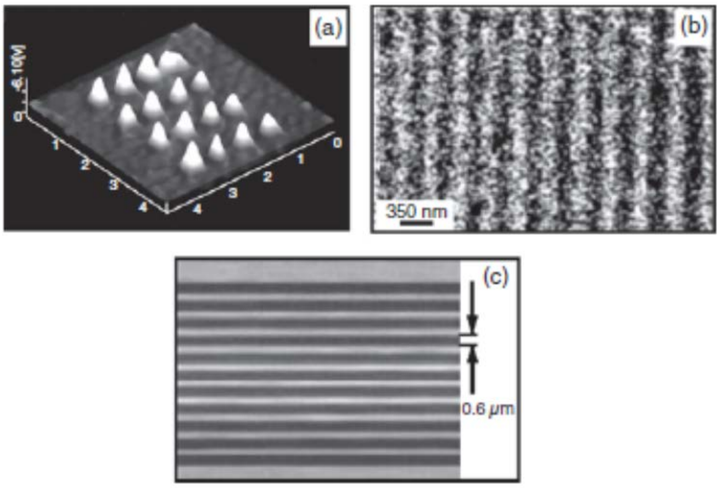


Fig. 11.18. SFM-generated nanoscale domain patterns in (a) $BaTiO_3$, (b) $Sr_{0.61}Ba_{0.39}Nb_2O_6$ [207], and (c) $RbTiOAsO_4$ single crystal [206].

For successful application of this domain-based approach to fabrication of novel electronic devices, several key issues related to the scale of domain engineering should be addressed.

First, reproducibility of high-resolution domain writing is necessary for the fabrication of reliable, large-scale ferroelectric templates. This requirement critically depends on the uniformity of the properties of the ferroelectrics at the nanoscale. This problem can be addressed based on statistical analysis of the PFM data of written domains in thin films [183, 209]. With this approach the effect of grain size and orientation on spatial homogeneity and reproducibility of the domain writing in thin films has been determined. The same approach can be used for bulk crystals and thin films to investigate the statistical variations in domain sizes as a function of the microstructure of the samples, their dielectric properties, and domain writing conditions.

Second to be considered is the AFM domain writing resolution, that is, the ability to scale the size of the written domains below 50 nm. There are a number of factors affecting a process of polarization reversal in SFM: microstructure, interface electrical and mechanical conditions, spatial distribution of an SFM-generated electric field, etc. The role of these factors can be different for different types of materials and samples. For example, the domain writing conditions in single crystals and thin films will be quite different.

Third, long-term stability of the written domain patterns is a key requirement necessary for the fabrication of large scale nanodomain patterns. This problem is closely related to the problem of thermodynamic stability of the ferroelectric phase at the nanoscale. One of the most important factors affecting domain stability is the screening of spontaneous polarization. EFM and SSPM studies can provide an insight into the process of polarization screening at the nanoscale and will open a way for local stabilization (and destabilization when necessary) of the domain patterns. This combined approach allows quantitative examination of screening processes as a function of time, type of interface, template microstructure, and ambient environment. Preliminary studies showed that domains as small as 20 nm in diameter could be written and detected by SFM [197]. In addition, it has been found recently that ferroelectric films as thin as 1.6 nm (two unit cells) still exhibit ferroelectric behavior [210]. SFM studies of domains of these dimensions can contribute to better understanding of the size effects and stability issues in ferroelectrics.

Fourth, and probably the most serious problem for the use of SFM in nanodomain fabrication, is throughput [211], which is limited by the tip scanning rate: whereas the write speed in ferroelectrics can be in the nanosecond range, the typical scanning speed of most scanning probe microscopes is in the range of tens of micrometers per second. The likely solution of this problem lies in the use of massive parallelism, as has recently been demonstrated by the IBM Millipede concept [212]. High data rates have been achieved through the use of large (32 × 32) two-dimensional arrays of SFM probes, which operate simultaneously.

11.10. DEVICE APPLICATIONS

The high level of interest in the synthesis, structure, and properties of nanoferroelectrics is obviously related to their current and potential use in devices of various types. Applications are envisaged in memory devices [213] such as D-RAM, NV-RAM, and EEPROM. Important nonmemory applications are in nanorobotic and microelectromechanical (MEMS) devices [214], and in sensors and detectors of various types.

Size effects have a direct bearing on applications. One of the main problems involving ferroelectric thin film capacitors of finite area has been pointed out in [72]. This arises from a reduction in the total quantity of switched charge due to the effect of fringing fields. Thus, in a circular disc of diameter 2 µm and thickness 200 nm (aspect ratio = 10:1) the fringe fields produce a 7% reduction in the switched charge. But this effect increases nonlinearly with decreasing aspect ratio and is about 23% for an aspect ratio of 5:1.

The ferroelectric NV-RAM works on the principle that at zero applied field, a ferroelectric system has two thermodynamically stable states described by $\pm P_R$ (see Fig. 11.1), each direction of polarization encoding either “0” or “1.” Since no external field is required to maintain these states, the memory device is nonvolatile. The polarization can be switched between the two states by the application of an external electric field, $E > \pm E_C$. The criteria for an ideal NV-RAM are fast read/write speed, high recording density, high endurance and retention, radiation hardness, compatibility with current integrated circuit (IC) technology, nondestructive readout capability, low power operation, and commercial viability. Property for property, ferroelectric-based memories are expected to beat most competing technologies. The recent discovery of a series of titanate-based layered perovskite compounds with very high values of spontaneous polarization might turn out to be very useful for NV-RAM applications. For example, a *c*-axis oriented $Bi_{4-x}Nd_xTi_3O_{12}$-based capacitor was found to have a switchable remanent polarization ($2P_r$) of over 100 µC/cm^2 and fatigue-free behavior [215].

Each memory cell consists of a ferroelectric capacitor sandwiched between metal electrodes and integrated with an access transistor. Thus the chip layout is composed of a matrix of capacitors and transistors. In order to achieve 1 Gbit NV memories, one needs to fabricate ferroelectric capacitors with lateral dimensions of 100 nm or less. The fundamental question to be answered is: what is the smallest size of the capacitor that will still behave like a ferroelectric? One is particularly concerned with the possibilities of depolarizing fields that might become large enough to inhibit switching. According to a review [216], the smallest switchable ferroelectric memory cells up to now have been fabricated by Mitsubishi–Symetrix (1 µm×1 µm) [217], NEC (0.7 µm × 0.7 µm) [218], U. Maryland (100 nm × 100 nm) [219], and MPI Halle (75 nm × 75 nm) [220]. The fabrication process involved is either e-beam lithography, focused ion-beam milling, or self-assembly. Materials and process parameters for ferroelectric D-RAM applications have been reviewed in [221]. As an example, Fig. 11.19 shows a sample of PZT nanodots patterned by e-beam lithography on a $SrTiO_3$ substrate by the group at MPI. See also Table 11.4.

Table 11.4. Approximate size of smallest switching single ferroelectric cells obtained by various research groups.

Research	Element dimension	Material used	Process	Ref.
Mitsubishi-Symetrix	1000 nm × 1000 nm	$(Ba, Sr)TiO_3$	e-beam lithography	[83]
NEC	700 nm × 700 nm	$SrBi_2Ta_2O_9$	e-beam lithography	[84]
University of Maryland	100 nm × 100 nm	$Pb(Nb, Zr, Ti)O_3$	focused ion beam lithography	[85]
MPI, Halle	100 nm × 100 nm 75 nm × 75 nm	$SrBi_2Ta_2O_9$ $Pb(Zr, Ti)O_3$	e-beam lithography	[86]

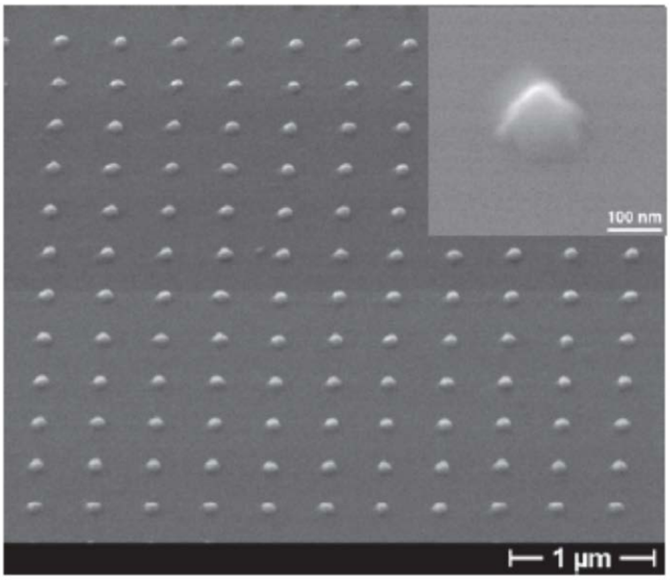


Fig. 11.19. Scanning electron micrograph of large-scale-ordering of PZT (30/70) nanoparticles on $SrTiO_3$ (111) substrate, patterned by electron beam lithography and crystallized at 650 °C. Inset shows magnified view of single dot.

To be useful as a memory element, a ferroelectric should ideally have a low dielectric constant, moderate spontaneous polarization, *TC* higher than the operating temperature range, switching speed in the nanosecond range, high chemical and environmental stability, and low fatigue properties. The need for low voltage operation (<5 V) demands that film thickness be in the submicrometer range and the coercive field be low. Two classes of materials, the perovskites (mostly PZT, doped PZT, and BST) and the layered perovskites SBT and $SrBi_2Nb_2O_9$ (SBN) have turned out to be particularly useful for this purpose. For many integrated ferroelectric thin film applications, SBT has now become the material of choice because: (a) it exhibits very little fatigue, (b) it has low leakage currents, and most importantly, (c) bulk properties are retained even in extremely thin films. In recent years, some success has also been achieved in depositing $PbTiO_3$ films on Si [222].

The simpler form of the ferroelectric memory is the destructive readout (DRO) consisting basically of a transistor and a capacitor. Each read operation consists of switching the capacitor, comparing with the linear response of an unswitched reference capacitor, and finally resetting the capacitor. The signal is measured and the state of the bit identified by means of a sense amplifier at the end of each bit line. The reset operation of the DRO device obviously enhances the fatigue. A more sophisticated version is the nondestructive readout (NDRO) where the information can be read over and over again, without the need to switch until the next write operation. The NDRO involves a field effect transistor (FET) with a ferroelectric thin film replacing the metal gate in a conventional FET. Here the polarization of the ferroelectric element (0 or 1) can be sensed without reversal by measuring the source–drain current. The magnitude of the current would depend on whether the gate is biased positively or negatively. Though such a device has many attractive features and some progress has been made in its development [223], a true ferroelectric memory FET has not yet been realized. The main problem is short memory retention and this is ascribed to the depolarizing field and the gate leakage current [224].

Ferroelectric thin films and nanoscale systems are also very useful in several nonmemory applications. These include IR detectors, gas sensors, surface acoustic wave devices, electro-optic switches and modulators, optical FETs, and waveguides. A lot of interest has been generated due to the futuristic applications of MEMS (which include transducers, actuators, accelerometers, etc.) that incorporate piezoelectric thin films such as PZT.

Piezoelectric materials play an important role in MEMS technology for sensing and mechanical actuation applications. In a piezoelectric material, mechanical stress produces a polarization, and, conversely, a voltage induced polarization produces a mechanical stress. Many asymmetric materials, such as quartz, GaAs, and zinc oxide (ZnO), exhibit some piezoelectric behavior. Recent work in MEMS has focused on the development of ferroelectric compounds, such as lead zirconate titanate, $Pb(Zr_xTi_{1-x})O_3$, or PZT for short, because such compounds have high piezoelectric constants that result in high mechanical transduction. It is relatively straightforward to fabricate a PZT structure on top of a thin freestanding structural layer (i. e., cantilever, diaphragm). Such a capability enables the piezoelectric material to be used in sensor applications or actuator applications for which piezoelectric materials are particularly well suited. Like Si, PZT films can be patterned using dry etch techniques based on chlorine chemistries, such as Cl_2/CCl_4, as well as ion beam milling using inert gases like Ar.

Recent developments in this field are now covered by a new journal (*Journal of MEMS*). Other applications of ferroelectric thin films include multilayer capacitors, boundary layer capacitors, varistors, etc.

GLOSSARY

Antiferroelectric A solid with an antiparallelly ordered array of local dipoles, resulting in a net zero polarization.

Coercive field (E_C) The external electric field required to reverse the macroscopic polarization of a ferroelectric.

Displacive ferroelectric A type of ferroelectric in which the elementary dipoles strictly vanish in the paraelectric phase. The atomic displacements in the low-temperature (ferroelectric) phase with respect to their corresponding positions in the high-temperature (paraelectric) phase are small compared to the inter-nuclear distances.

Ferroelectric A solid that exhibits a spontaneous, field reversible dielectric polarization, for $T < T_C$, the Curie temperature. Crystalline dielectric possessing spontaneous polarization the direction of which can be reversed by application of an electric field.

Ferroelectric domains Regions of uniform polarization in a ferroelectric.

Order–disorder ferroelectric A type of ferroelectric in which the elementary dipoles are non-vanishing but thermally average out to zero in the paraelectric phase. The atomic displacements on both sides of T_C are comparable to the inter-nuclear distances.

Piezoelectric effect Development of electrical polarity on the application of stress. This effect and its converse (the development of strain on applying an electric field) are both linear in nature, a reversal in the stimulus resulting a reversal of the response.

Piezoelectricity Property of a crystal to exhibit electric polarity when subject to mechanical stress (direct piezoelectric effect) or mechanical deformation upon application of an electrical field (converse piezoelectric effect).

Pyroelectric effect Change in the spontaneous polarization with temperature.

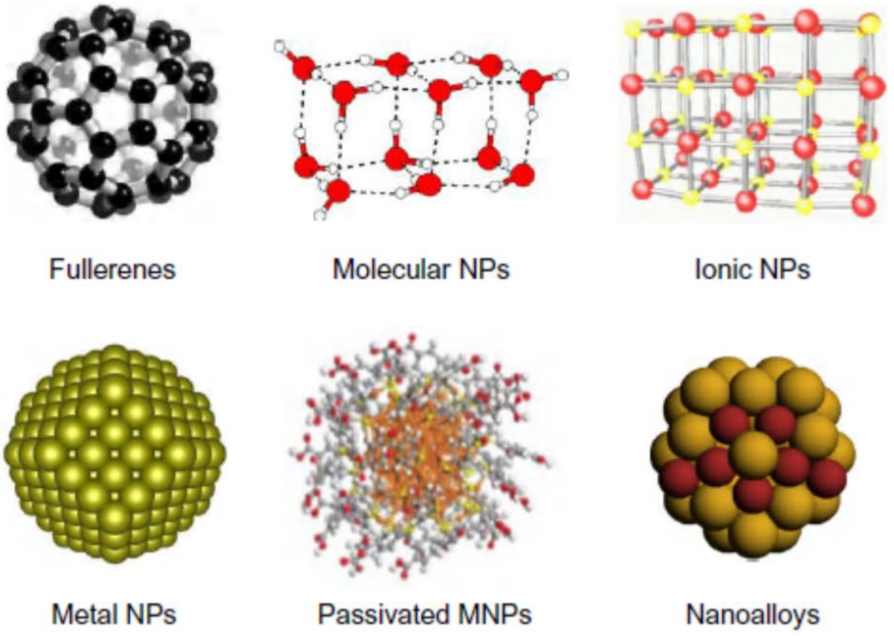
Relaxor ferroelectric In such systems, the temperature dependence of the dielectric constant (ε) displays a broadened peak. Unlike a normal ferroelectric, it obeys a modified Curie-Weiss law in which the inverse dielectric constant varies as some power (between 1 and 2) of the temperature.

CHAPTER 12. NANOPARTICLES OF METALS, SEMICONDUCTORS AND DIELECTRICS

12.1 INTRODUCTION

Nanoparticles (NPs) are clusters of atoms, ions or molecules, typically with dimensions of the order 1–100nm. This is a particularly interesting size range, which bridges the gap between small molecules (with discrete energy states) and bulk materials (with continuous energy states). The study of the evolution of the geometric and electronic structures of clusters and their chemical and physical properties is therefore of great fundamental interest. NPs also constitute a new type of material, since they often have properties, which are fundamentally different from those of discrete molecules or bulk solids. The high ratio of surface to interior atoms in clusters means that there are many similarities between NPs and bulk surfaces, with smaller NPs essentially being all surface.

NPs are formed by most of the elements in the periodic table and they can be classified according to the types of atoms of which they are composed and the nature of the bonding. Examples of NPs (see Figure 12.1) include: weakly bound rare gas and molecular NPs (e.g. Ar and H_2O); covalently bonded semiconductor NPs (e.g. carbon fullerenes and CdS); electrostatically bound ionic NPs (e.g. NaCl and MgO) and NPs of metallic elements (MNPs) .



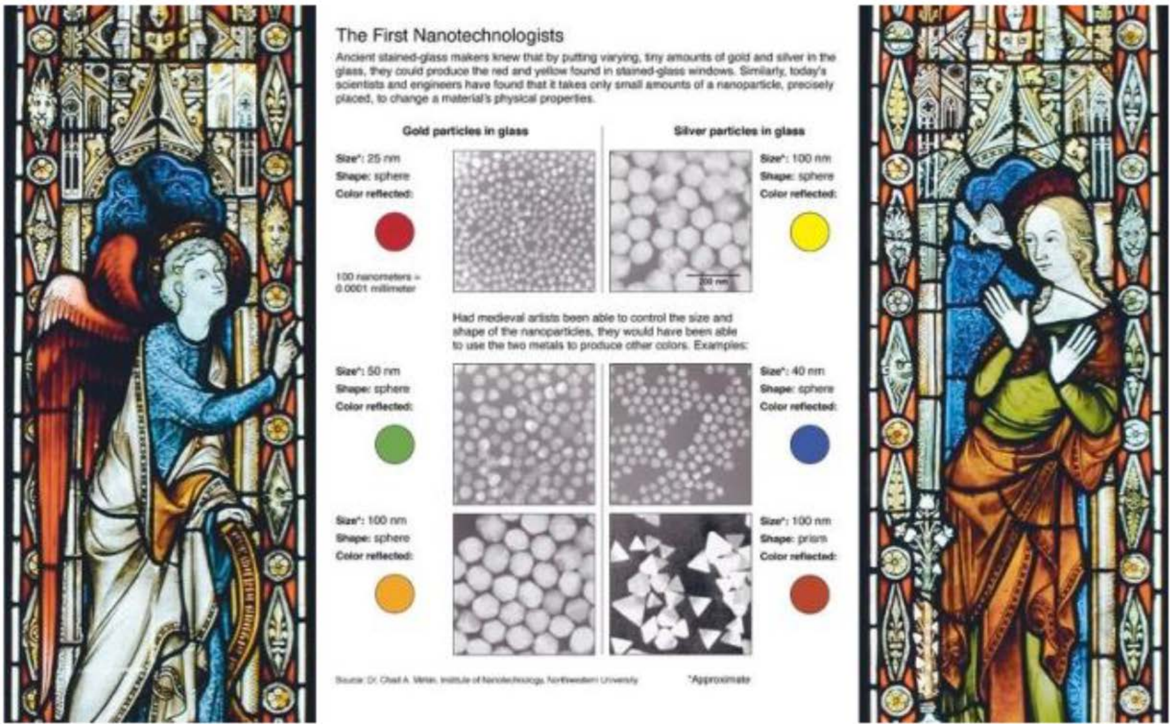
Fullerenes
Molecular NPs
Ionic NPs
Metal NPs
Passivated MNPs
Nanoalloys

FIGURE 12.1 Examples of types of nanoparticles

12.2 WHY IS INTEREST TO THE METAL NANOPARTICLES AND NANOALLOYS

Metallic elements form a wide variety of NPs, ranging from the s-block metals (such as the alkali and alkaline earth metals), p-block metals (such as aluminium—where the bonding involves both the s and the p orbitals) and the transition metals (where the bonding also involves the valence d orbitals). MNPs may be composed of a single metallic element or more than one metal—the subclass of “nanoalloys” (NAs). Several studies have shown the ability to control nanoparticles composition, starting with cores made of pure metal, *(e.g.*, Au, Ag, Ni, Co, Pt, Pd, Cu, Al); metal alloys (*e.g.*, Au/Ag, Au/Cu, Au/Ag/Cu, Au/Pt, Au/Pd, and Au/Ag/Cu/Pd, PtRh, Ni/Co, Pt/Ni/Fe); and more.

MNPs and NAs have been used (albeit unknowingly) for many centuries, primarily for their optical properties. For example, copper, silver and gold NPs are responsible for many of the colours of stained glass windows dating from the middle ages. As seen in Figure 12.2, variations in the size, shape or type of the nanoparticles have a dramatic effect on their color. For example, spherical gold nanoparticles with diameters of 25, 50, and 100 nm, look red, green, and orange, respectively. Composition also affects color - spherical silver nanoparticles with a 100 nm diameter exhibit a yellow color, compared to the orange color exhibited by gold nanoparticles of a similar shape and size. Shape is also significant - spherical silver nanoparticles that have 100 nm in diameter exhibit yellow color, while nanoparticles of a similar size and composition but a prism shape show exhibit red color.


The First Nanotechnologists

Ancient stained-glass makers knew that by putting varying, tiny amounts of gold and silver in the glass, they could produce the red and yellow found in stained-glass windows. Similarly, today's scientists and engineers have found that it takes only small amounts of a nanoparticle, precisely placed, to change a material's physical properties.

Had medieval artists been able to control the size and shape of the nanoparticles, they would have been able to use the two metals to produce other colors. Examples:

Source: Dr. Chad A. Mirkin, Institute of Nanotechnology, Northwestern University

*Approximate

Figure 12.2 Ancient stained-glass makers knew that by putting varying, tiny amounts of gold and silver in the glass, they could produce the red and yellow found in stained-glass windows

In fact, for large MNPs, many other properties, such as ionization energy (IE), electron affinity, cohesive energy, absorption frequency, electrical conductivity and chemical reactivity, show a regular variation with cluster size—that is, they obey a simple scaling law. By adopting a spherical cluster model, in which the N-atom cluster is approximated by a sphere of radius R, and noting that many properties depend on the ratio of surface to bulk (volume) atoms in a cluster, it can be shown that these properties (as for the melting temperatures mentioned above) typically vary as 1/R (or $N^{-1/3}$).

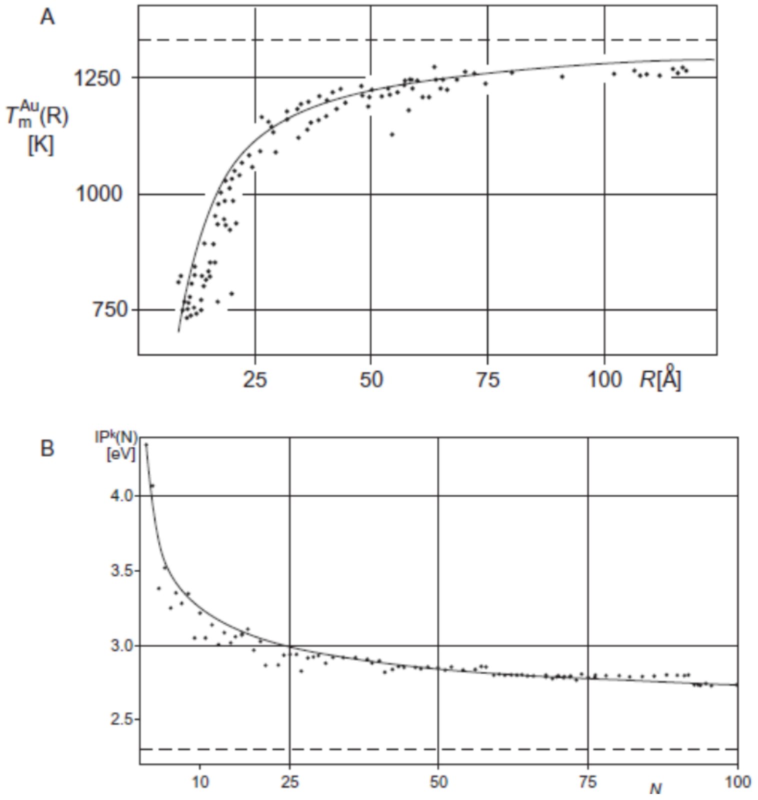


FIGURE 12.3 Examples of scaling laws for metal nanoparticles. (A) Melting temperatures of gold NPs; (B) Ionization energies of potassium NPs. The solid lines show fits to Equations (12.1), respectively, and the dashed lines indicated the bulk values

As another example of a simple MNP scaling law, the IEs (ionization potentials, IP) of potassium clusters, with up to 100 atoms, can be fitted, to a high degree of accuracy, to the following interpolation formula

$$\mathrm{IP}^{\mathrm{K}}(N)/\mathrm{eV} = 2.3 + 2.04\,N^{-1/3} \qquad (12.1)$$

One problem with studying bare MNPs (e.g. those created in cluster beams is that it is difficult to handle them on a preparative scale like conventional molecules. To enable the investigation of approximately uniformly sized MNPs, and to exploit their properties in various applications, it is necessary to protect or “passivate” them with a ligand surfactant shell, as this avoids coalescence at high cluster densities. Two types of stabilisation can be distinguished: electrostatic stabilisation due to Coulombic repulsion between MNPs, caused by ions adsorbed at the particle surface (e.g. sodium citrate gold sols); and steric stabilisation, due to the coordination of polymers and other bulky organic molecules, especially those with P-, N- and S-donor groups (e.g. thiols and thioethers).

MNPs can also be deposited (from solution or the gas phase, where size selection may be possible) onto a substrate (such as graphite, silicon or an inorganic oxide) or in an inert gas matrix. The desire to fabricate materials with well-defined, controllable properties and structures, on the nanometre scale, coupled with the flexibility afforded by intermetallic materials, has generated interest in bimetallic and multimetallic NAs. There are a large number of combinations of metallic elements and a wide range of elemental compositions which are possible for NAs. Bimetallic NAs can be generated with reasonably well-controlled size and composition. However, NA structures and the degree of elemental segregation or mixing can depend on the method and conditions of their generation.

One of the major reasons for interest in NAs is because their chemical and physical properties can be tuned by varying the composition and atomic ordering, as well as the size and shape/morphology of the particles. NAs may display not only magic sizes, but also magic compositions—compositions at which they possess special stability. Surface structures, compositions and segregation properties of NAs are also critically important in determining their chemical reactivity and, in particular, catalytic activity. Four main types of chemical ordering (mixing pattern) can be identified for NAs, these are shown in Figure 12.4.

-Core–shell segregated NAs (Figure 12.4.A) consist of a shell of one type of atom (B) surrounding a core of another (A), though there may be some mixing between the shells. This A core B shell configuration is common to a large variety of systems.

- Layered-segregated NAs (Figure 12.4. B) consist of A and B subclusters, with a mixed A–B interface. These are examples of so-called “Janus particles” (named after the two-faced Roman god) as they have two distinct faces, which may possess different chemical and physical properties. The map of TEM, of the elements in Cu–Ag NAs, indicating a potential transition from core–shell to Janus-type particles as a function of increasing size is shown in fig.12.5.

- Mixed NAs (Figure 12.4.C) may either be ordered (pseudocrystalline, left) or random (solid solution, right). The intermixed pattern is found in many systems, with the ordered tetragonal structure of CoPt NAs being important for their magnetic properties.

-Multishell “onion-like” NAs (Figure 12.4.D) present onion-like alternating concentric –A–B–A– shells. Metastable structures of this type were predicted from simulations of the growth of Cu–Ag, Ni–Ag and Pd–Ag NAs and for Co–Rh and Pd–Pt NAs.

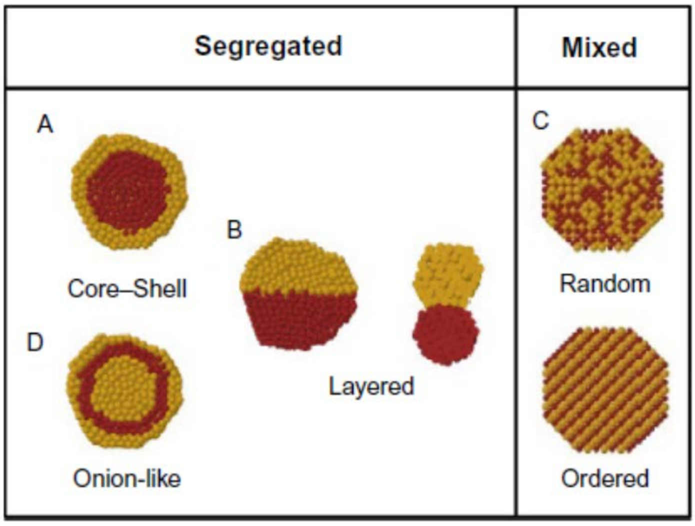


FIGURE 12.4 Types of chemical ordering observed for nanoalloys

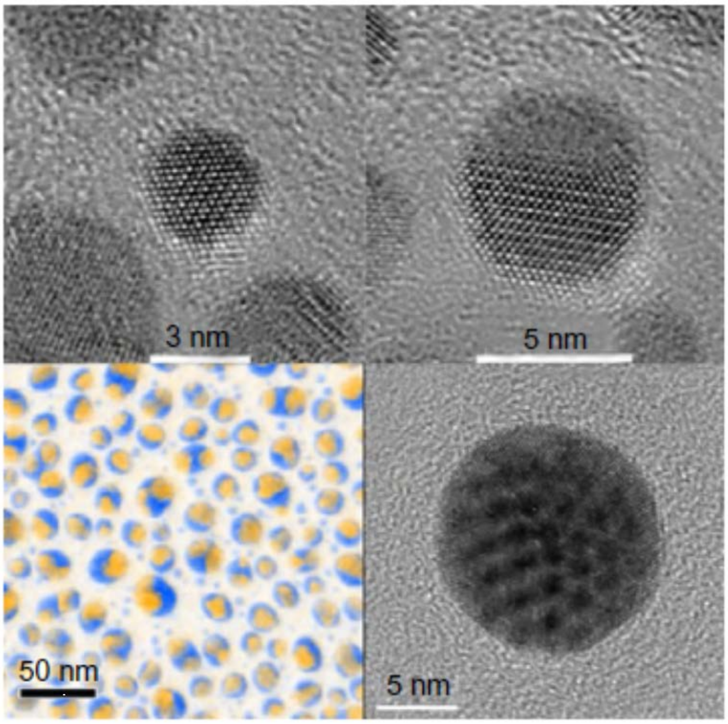


FIGURE 12.5 HRTEM and TEM (bottom left) images of Cu–Ag NAs

12.3 SYNTHESIS OF METAL NANOPARTICLES

MNPs and NAs can be generated in a variety of media: in the gas phase, in solution, supported on a substrate, embedded in a matrix or even in a bacterial cell. These methods are summarised briefly here.

Cluster Molecular Beams

Cluster molecular beams allow the study of “free” MNPs in the absence of interactions with other particles, ligands or substrates. In a typical cluster beam experiment, metal atoms are generated in the vapour phase from a metal target by one of the following methods: resistive heating (for volatile metals such as the alkali metals); laser vaporisation; pulsed electric vaporisation or ion/magnetron sputtering. NAs can be formed by using a bimetallic target or using two metal targets. After vaporisation, cluster nucleation and condensation are generally assisted by collisions with a cold inert carrier gas and further cluster–collisions are prevented by supersonic expansion through a nozzle into a region of high vacuum.

Ion Implantation

Implantation of metal ions (which are subsequently neutralised by electron capture) into bulk insulator using ion beams with energies of approximately 100keV has been used to generate MNPs with interesting magnetic and optical properties. NAs may be generated by sequential implantation of the different metal species.

Laser Ablation

One promising method for metallic nanoparticle formation is laser ablation. This method is good for production of bare nanoparticles. In this method, a high energy laser beam is focused on a bulky metal target that is located in a solvent. A large amount of energy is absorbed by the bulk metal target, locally increasing the temperature and leading to outer atom vaporization on the metal surface. The vapors condense upon meeting the solvent and nanoparticles are formed. This process is energetically favored as it minimizes the surface energy of the metal. Because the laser pulses are short enough, there is no evaporation of the solvent.

The laser ablation method is good for any combination of target and solvent. Metallic and ceramic nanoparticles can be formed both in organic and aqueous solvents. No protecting ligand is required when using laser ablation as the laser creates positive charge on the surface of the nanoparticles, protecting the nanoparticles from agglomeration.

One of the main advantages of the nanoparticles produced by “laser ablation” is their high purity. The purity is important for applications that require biocompatibility such as nickel titanium implants that are used in biomedical surgeries inside the human body.

Production of Nanoparticles by Chemical Route

A common approach for the formation of nanoparticles relies on the chemical route. The principle behind this process is that molecules that can interact with the surface of a nanoparticle influence nucleation and growth.

The synthesis process is shown in Figure 12.6. The main ingredients are a metal compound, (generally a positively charged metal salt) and metal salt that is dissolved in an appropriate solvent and introduced as a reducing agent. Surface-controlling agents are used to obtain very uniform particle sizes in the nanometer range. The final ingredient is a stabilization agent. The addition of a stabilizing agent to the solution prevents uncontrollable growth of the nanoparticles, grants control over the growth rate and particle size, prevents agglomeration of nanoparticles and increases nanoparticle solubility in the desired solvent.

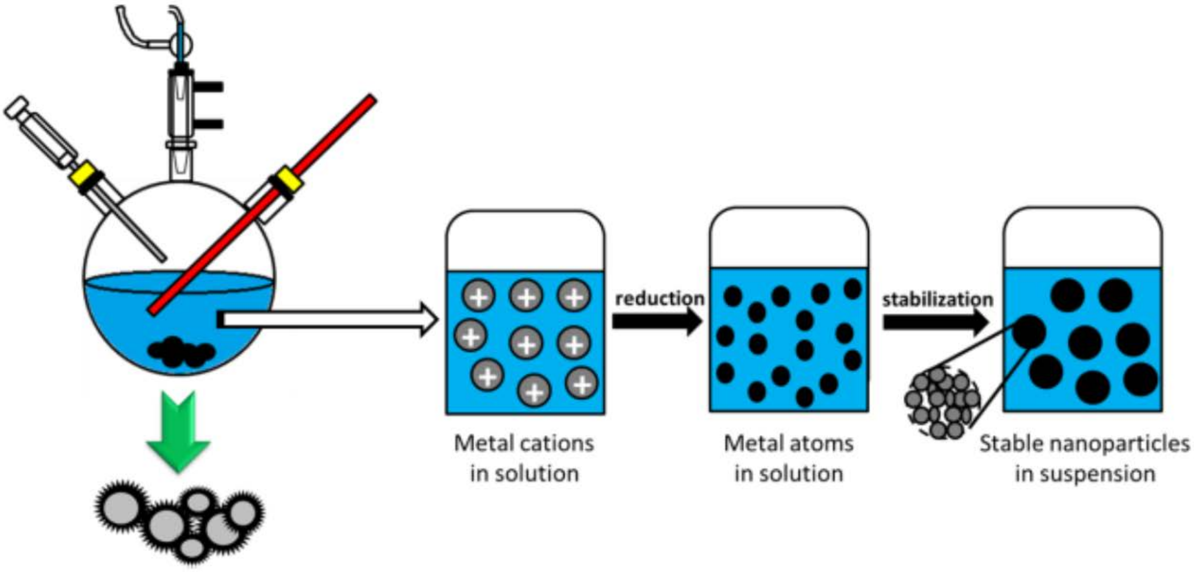


Figure 12.6 Illustration of nanoparticles production via a chemical route.

Production of Encapsulated Metal Nanoparticles

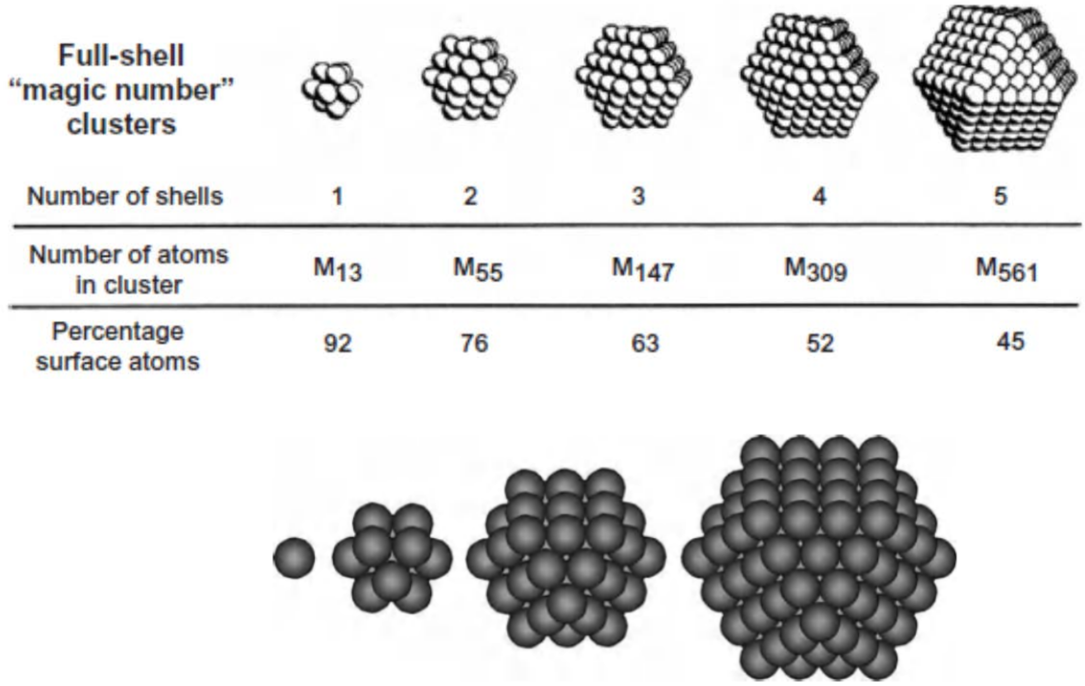
For the formation of metal nanoparticles encapsulated with organic ligands, it is common to use the precipitation method, or the so-called “Brust” method. The strategy exploited by this method is the chemical bonding of a thiol ligand monolayer on the surface of the gold nanoparticles during formation. This reaction is referred to as an attachment of self-assembly thiolated monolayer – it occurs spontaneously with no external assistance, as the bond between the thiolated molecules and the gold surface is an extremely strong covalent bond.

The Brust method is used to produce gold nanoparticles in organic liquids, such as toluene, that are generally not miscible with water. It involves the reaction of a chloroauric acid ($HAuCl_4$) solution with a tetraoctyl ammonium bromide (TOAB) solution in toluene, which serves as the gold phase transfer agent from water to the organic solvent, and sodium borohydride, which serves as a reducing agent. More specifically, when adding the TOAB dissolved in an organic phase to gold ions in water, the interaction with the gold ions transfer the gold to the organic solvent. In the organic solvent, the thiol molecule self-assembly can form on the nanoparticle surface, and the sodium borohydride serves as a source for the electrons necessary for the reaction synthesis, some of the phase transfer agent may remain bound to the nanoparticles, and this may affect physical properties such as solubility. In order to remove as much of this agent as possible, the nanoparticles must be further purified by extraction methods. Alkanethiol-protected gold nanoparticles can be easily precipitated and then redissolved, facilitating washing or transferring to fresh or alternative solvents.

12.4 PROPERTIES OF METAL NANOPARTICLES

“Magic numbers”

Using a mass spectrometer, it was discovered that certain masses of particles were produced in relatively large abundance. These "magic" sizes corresponded to the closing of atomic shells, analogous to stable nuclear shells from nuclear chemistry or the electronic shells which form the basis of chemical bonding and the periodic table. Schematic examples of a series of closed shell particles are shown in Figure 12.7.



Number of shells	1	2	3	4	5
Number of atoms in cluster	M_{13}	M_{55}	M_{147}	M_{309}	M_{561}
Percentage surface atoms	92	76	63	52	45

Figure 12.7. a) Full-shell magic number formation showing how the number of shells relates to the number of atoms in a particle and the percent of atoms present on the surface , b). A series of different clusters formed by addition of closed shells in fcc structure. Each of the presented clusters represents the core of the subsequent cluster. Starting from a single atom, cubo octahedra of 13, 55 and 147 Au atoms can be build. Note that the percentage of surface atoms is far above 50% for all structures

It was found a number of intense peaks in the mass spectra of sodium and other alkali metal clusters (see Figure 12.8a). The nuclearities corresponding to these peaks were termed “Magic Numbers” and were attributed to the enhanced stability of a cluster as compared with its immediate neighbours. The quantum mechanical “jellium” model was adapted from nuclear structure theory to account for these magic numbers. Agreement with experimental IPs was subsequently improved by allowing for ellipsoidal distortions of the clusters, for electron counts corresponding to incomplete jellium subshells. In the jellium model, the cluster is modelled by a uniformly positively charged sphere filled with an electron gas. The Schrodinger equation is solved for an electron moving within the sphere under the influence of an attractive central mean field potential. Magic numbers arise due to the complete filling of the resulting jellium orbitals (1s, 1p, 1d, 2s, 1f, etc.), corresponding to total jellium electron counts of 2, 8, 18, 20, 34, etc. It should be noted that the principal quantum number, n in the jellium model, because of its origin in nuclear physics, is distinct to that used in the definition of atomic orbitals (natom). They are related by: $n=n_{\mathrm{atom}}-l$, where *l* is the angular momentum quantum number. The jellium model ignores the positions of the atomic nuclei, which is a valid assumption if the valence electrons are quite weakly bound as in the alkali metals and (to a lesser extent) the noble metals (Cu, Ag and Au), though the model has also been extended to cover alkaline earth and other main group clusters (notably Al). Jellium shell closings have also been observed experimentally and predicted theoretically for small bimetallic NAs such as Cu–Ag NPs with tens of atoms.

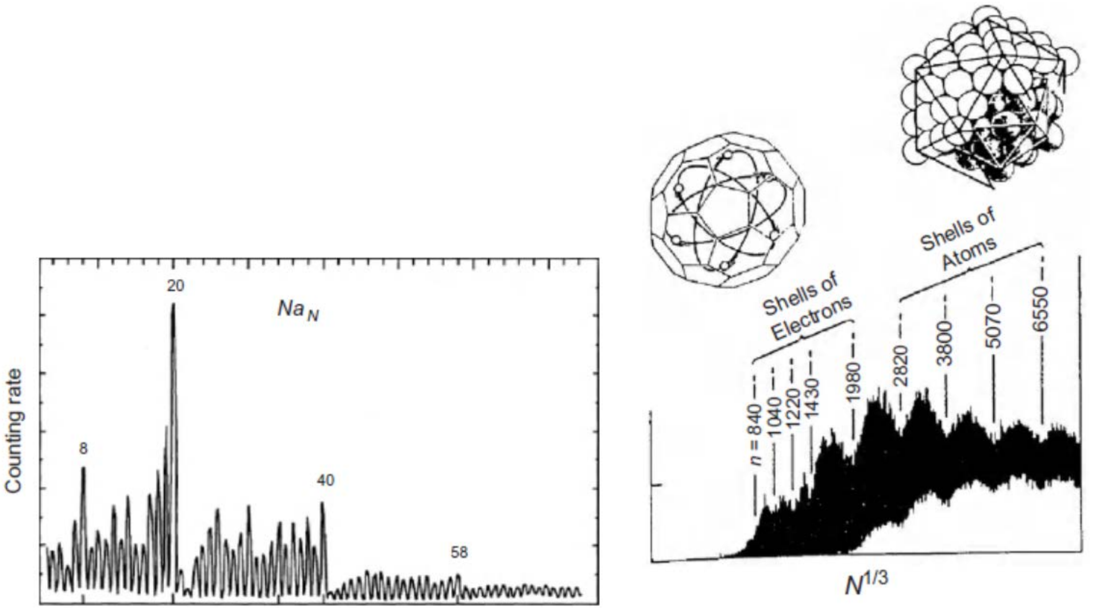


FIGURE 12.8 a)Mass spectrum of small sodium NPs, showing jellium magic numbers at 8, 20, 40 and 58 atoms (electrons), b) Long period oscillations observed in the mass spectra of large sodium NPs arising from electronic shells (left) and geometric (atomic) shells (right).

It was measured the mass spectra of sodium NPs with up to 25,000 atoms and observed two series of periodic oscillations in intensity which are approximately evenly spaced when plotted against N1/3, as shown in Figure 12.8b. For N<2000, the oscillations have a small period. The magic numbers correspond to dips in mass spectral (MS) intensity because near-threshold ionization is used to produce the cations and the more stable NPs tend to have higher IPs and smaller ionization cross sections. The magic numbers in this size regime have been attributed to the filling of electronic shells which occur due to the bunching together of jellium electronic energy levels. Longer period oscillations in this region have been attributed to the merging of electronic shells into dense blocks or supershells. Examples of polyicosahedral structures, which are again stabilised as core–shell particles when there is a large size mismatch between the (smaller) core and (larger) shell atoms, are shown in Figure 12.9.

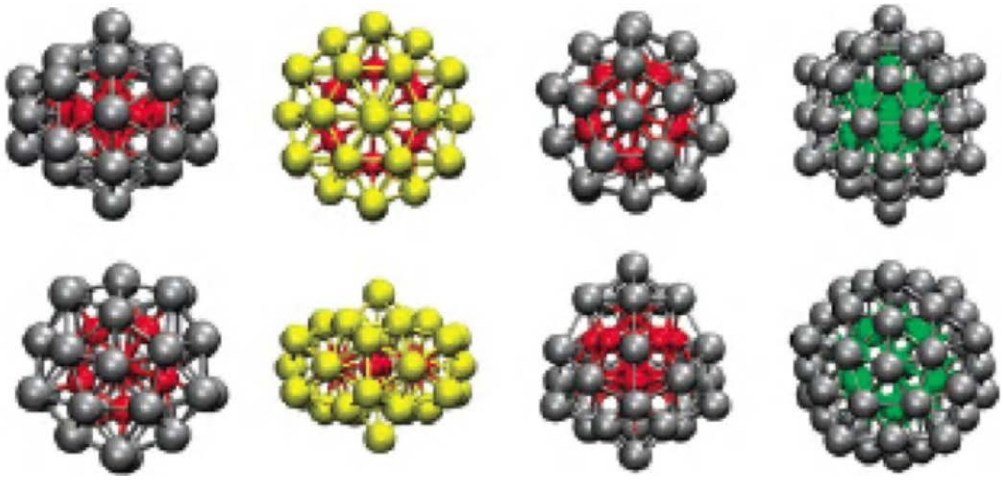


FIGURE 12.9 Examples of the four most stable core–shell polyicosahedra. Each cluster is shown from two different perspectives. (From left to right) Fivefold pancake of size 34, capped sixfold pancake of size 40, capped fivefold pancake of size 40 and anti-Mackay icosahedron of size 45.

Magnetic Properties

Ni, Fe and Co NPs are ferromagnetic (with higher magnetic moments per atom than in the bulk metals), small Rh NPs are also ferromagnetic (unlike bulk Rh which is paramagnetic). There is considerable interest in granular materials formed by embedding MNPs in a solid host. When embedded in non-magnetic metals, or even insulators, magnetic 3d MNPs (e.g. Cr, Fe, Co, Ni or mixtures of these metals) are known to exhibit Giant Magneto-Resistance (GMR), with a magnetic field induced change in resistance of as much as 50% for Fe clusters embedded in Ag. These GMR materials show considerable promise for applications as magnetic sensors and for magnetic recording. The combination of 3d metals (e.g. Co and Ni), with large local magnetic moments, and 4d metals (e.g. Rh), with strong spin–orbit coupling, may be an effective way of obtaining a high magnetic moment and anisotropy which is required in high-density magnetic recording. 3d–5d Fe–Pt and Co–Pt NAs are also candidates for ultra-high-density magnetic recording media due to their high magnetic anisotropy, with associated high magnetic susceptibility and coercivity.

Shapes of Metal Nanoparticles

In addition to their size and type, the shape of the nanoparticles can be engineered and controlled. The size of the nanoparticles can be determined by the ratio between the ingredients during the "cooking" step. Generally, a higher ratio of gold in comparison to the other components will result in larger nanoparticles, as seen in Figure 12.10(a). Different shapes, such as the nanorods illustrated in Figure 12.10 (c), can be obtained by changing other synthesis conditions (for example, by varying the temperature). Other modifications can result in more complex gold nanostructures (as seen in Figures 12.10 (h-o))

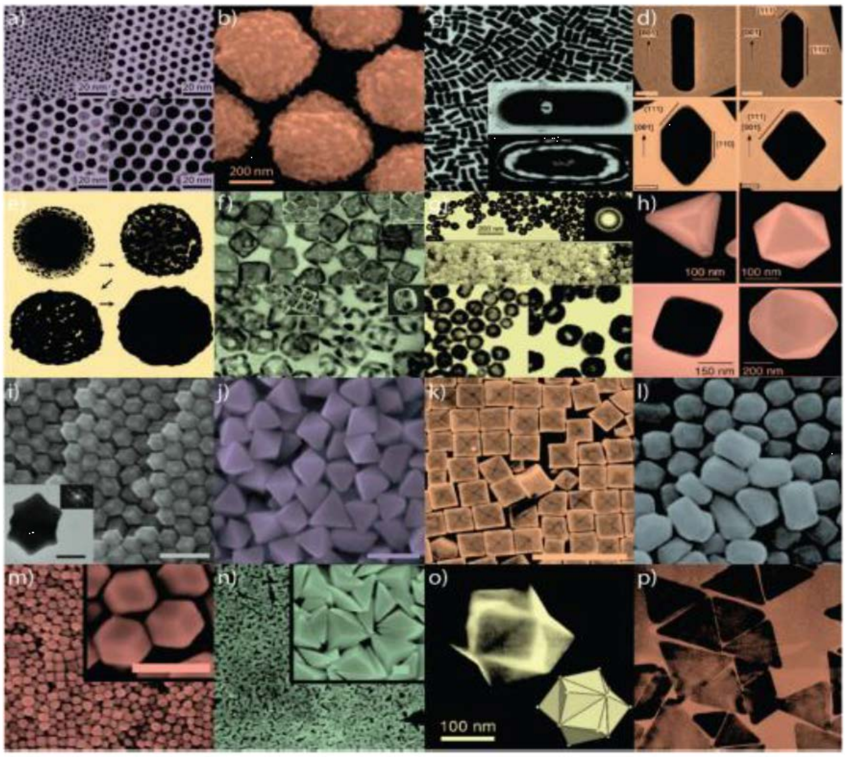


Figure 12.10 Gold nanoparticles of various sizes and shapes. (a) Nanospheres; (b) Nanorods; (c) Sharpened nanorods; (d) Nanoshells; (e) Nanocages; (f) Hollow nanospheres; (g) Cubes; (h) Rhombic dodecahedra; (i) Octahedra; (j) Concave nanocubes; (k) Tetrahexahedra; (l) Rhombic dodecahedra; (m) Obtuse triangular bipyramids; (n) Trisoctahedra; and (o) Nanoprisms

Nanoparticles are produced by chemical reduction of metal salts dissolved in a solvent with surfactant or polymeric ligands which passivate the cluster surface. Recently, using this method, not only the size but also the shape of particles is controlled to change the properties through controlling the growth rate of different crystallographic facets (Fig. 12.11).

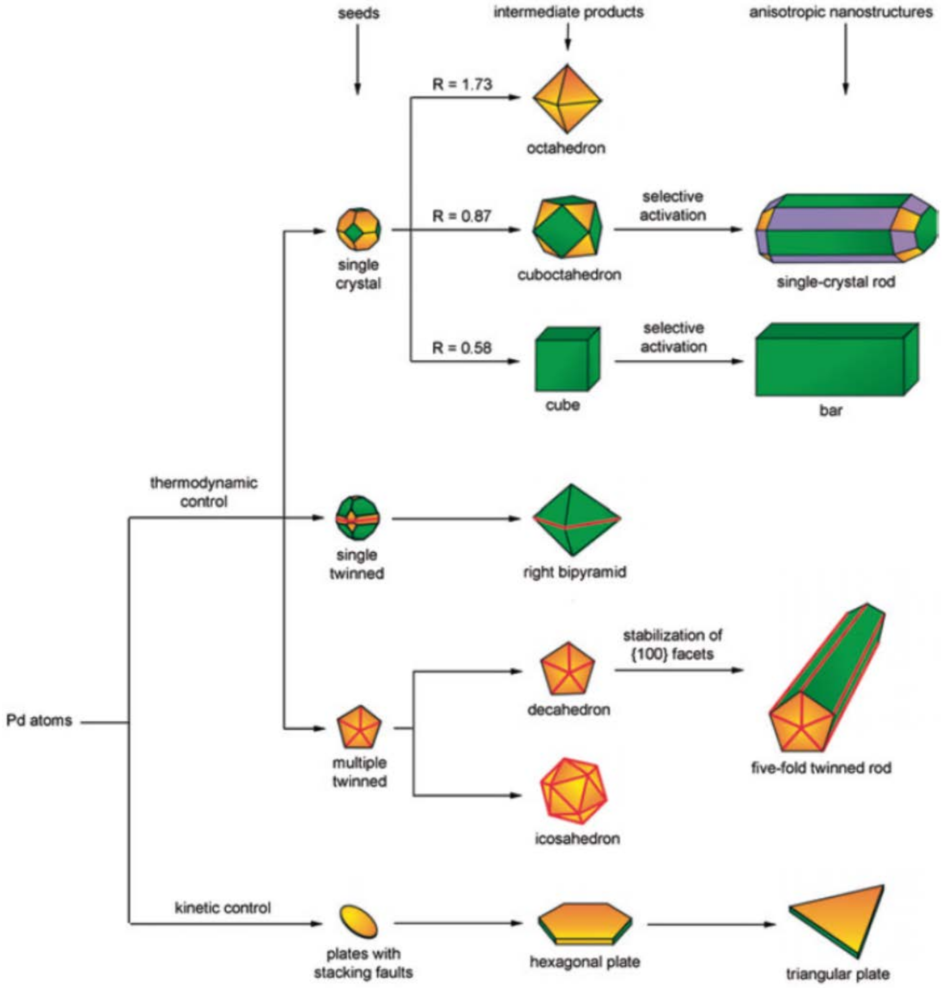


Fig. 12.11 A schematic of the reaction pathways that lead to Pd nanostructures with different shapes. The *green*, *orange* and *purple* colors represent the {100}, {111} and {110} facets, respectively

Plasmonics

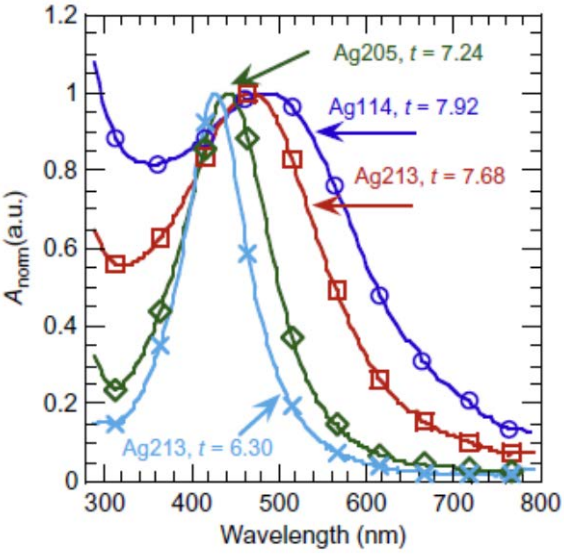
While for small metal clusters, the electronic spectrum consists of a number of well-defined lines corresponding to transitions between well-separated, quantized energy levels, in large MNPs, a single "surface plasmon" mode carries 100% of the oscillator strength so one observes a single peak in the electronic spectrum. The surface plasmon is a collective excitation of the conduction electrons, corresponding to a light induced correlated motion of the cluster's itinerent electrons in the attractive field of the positively charged ionic cores. The frequency of the dipolar surface plasmon of an MNP from Mie theory is red-shifted relative to the plasmon frequency of the bulk metal by an amount proportional to 1/R. The surface plasmon line width is also proportional to 1/R, becoming sharper for larger MNPs. Although other higher multipolar resonances are possible, the dipolar surface plasmon resonance is dominant for MNPs with radii significantly smaller than the wavelength of the exciting radiation (UV to near IR light): so that the atoms in the particle experience a constant electric field due to the light. For non-spherical particles (e.g. nanorods), better agreement with experiment is achieved using numerical treatments, such as the Discrete Dipole Approximation.

The optical absorption spectra of colloidal metal particles can be very complex, depending on the electronic structure of the metal, the size of the cluster, the dielectric constant of the medium (solvent or matrix) and the nature of the cluster surface-matrix interaction. At higher concentrations of colloidal particles (e.g. in particle arrays), there may also be interference effects due to coupling of the plasmons of the individual NPs. Other light scattering effects, such as opalescence, can also be observed. Considering NAs, the Ag–Au system has been widely studied, where the shape and frequency of the plasmon resonance have been shown to vary considerably with composition and the type of chemical ordering.

The plasmonic properties of MNPs has led to the development of the field of “nanoplasmonics”, which involves tailoring MNP plasmons for applications, for example, as sensors (e.g. for chemical and biological agents) and in optical computing.

Ag nanoparticles exhibit very clear differences in size-dependent optical properties. This is mainly due to the different energies of the interband (filled d shell to conduction sp band) transition onsets in the two metals. The close proximity of the interband transition to the classical plasmon energy in Au (it lies to the red of the plasmon energy) is the reason the energy damping (peak broadening) is so much greater for Au particles than for Ag where the interband transition energy onset lies significantly to the blue of the conduction band plasmon. This makes a determination of the size dependence of the linewidth for small Au particles difficult (i.e. the peak simply disappears). It also gives rise to a pronounced asymmetry to the peak shape. It has been a matter of significant controversy regarding the direction of the plasmon absorbance energy shift with particle size in the quantum size region. The details regarding relative weights assigned to the different theoretical factors permits a prediction of either a blue or a red plasmon shift with decreasing particle size. A blue shift occurred in the case of monodisperse, spherical nanoparticles of gold which seems to be the general consensus based upon dozens of papers in this field. However, the case of Ag is not as well established since both red and blue shifts have been reported. SEC and online optical characterization of Ag nanoparticles prepared using inverse micelle synthesis showed a clear red shift with decreasing size. The magnitude of this shift is relatively larger compared to gold.

Figure 12.12 shows the effect of size decreases on the position of the optical absorbance of Ag nanoparticles. The Ag samples whose spectra are exhibited in Figure 12.12 are all coated with an identical ligand, dodecanethiol, $C_{12}SH$, which keeps the particles from aggregating even when deposited onto a holey carbon grid and inserted into the high vacuum of an electron microscope. Figure 12.13 shows a TEM of 4nm Ag particles illustrating the ready formation of arrays which have been termed quantum dot or nanoparticle arrays.



Fuig.12.12 Normalized absorbance, A norm, versus wavelength obtained during SEC separation and online characterization of Ag nanoparticle solutions in toluene. A strong red shift and broadening of the absorbance with decreasing size (corresponding to longer elution times) is observed.

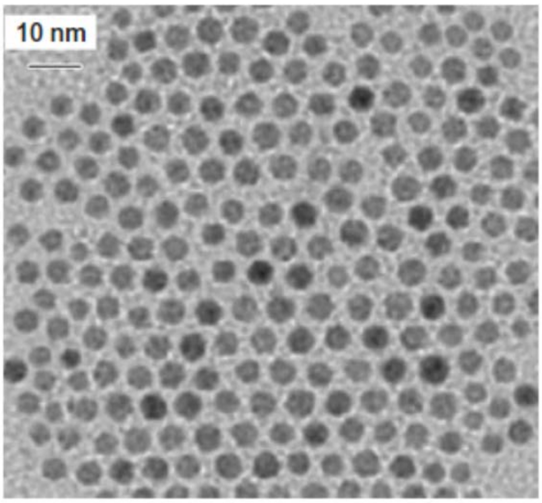


FIGURE 12.13 TEM of Ag nanoparticles with a core size of Dc =4.0 ±0.5nm stabilized with decanethiol. A solution of these particles has an absorbance peak at 440nm in toluene.

Figure 12.14 shows the effect of Ag deposition on the absorbance properties of a Au nanoparticle with a D=5.4nm core. The red-coloured parent solution with an absorbance peak at 518nm blue shifts in a systematic manner as thicker shells of Ag are formed around the core. This results in a yellowcoloured solution with the narrower symmetrical absorbance peak typically associated with a Ag nanoparticle. Au can also be deposited onto Ag particles. In this case, the absorbance peak shifts in the opposite manner as shown in the example of Figure 12.15. In this figure, a seed population of D=4.0nm Ag was used. It is interesting to note that even when the particle has a 31:1 Au:Ag ratio, corresponding to a 4-nm Ag core with a total size of ~12nm after Au deposition, the damping of the Au plasmon is much stronger than is observed in a pure Au particle of the same size.

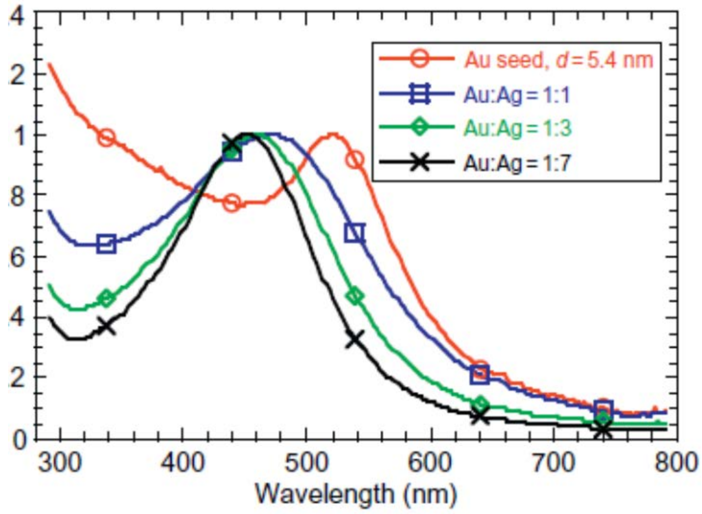


FIGURE 12.14Normalized absorbance versus wavelength obtained during chromatography for Au/Ag, core/shell nanoparticles. The effect of deposition of progressively larger amounts of Ag on Au nanoparticles with a core size of D=5.4nm is to blue shift the absorbance. The atomic ratios of Au:Ag are indicated for each case. The ratios correspond to Ag shell thicknesses (total thicknesses) of ~1.4, 3.2 and 5.3nm, respectively

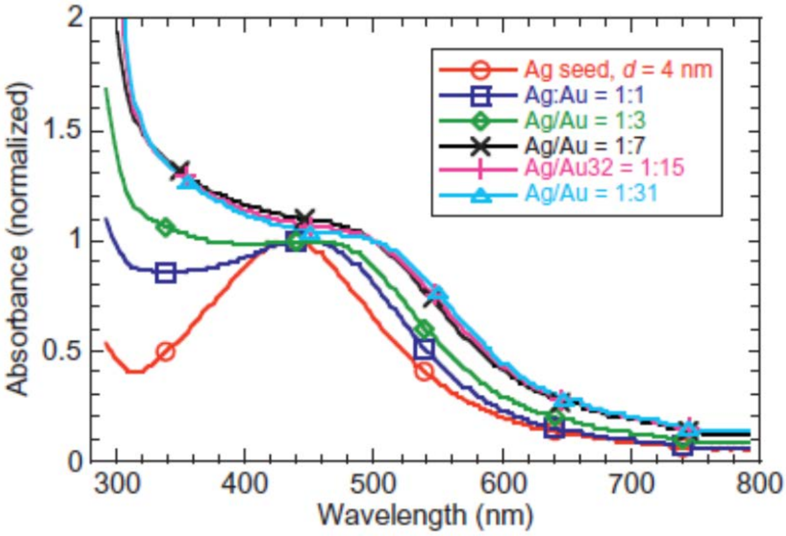


FIGURE 2.18 Normalized absorbance versus wavelength obtained using during chromatography of Ag/Au, core/shell nanoparticles. The effect of deposition of progressively larger amounts of Au on Ag nanoparticles with a core size of D=4.0nm on the optical absorbance is shown. The atomic ratios of Au:Ag are indicated for each case. The ratios correspond to Au shell thicknesses (total thicknesses) of ~1, 2.5, 4, 6 and 8.2nm

Molecular Modifications of Metal Nanoparticles

One important aspect of nanoparticles is the ability to encapsulate them or to modify their surfaces with a wide variety of molecular ligands such as alkylthiols, alkanethiolates, arenethiolate, alkyl-tri-methyloxysilane, dialkyl disulfides, xanthates, oligonucleotides, DNA, proteins, sugars, phospholipids, enzymes, and more.

The encapsulation (or the surface protection) of a nanoparticle is extremely important for a number of reasons. The capping ligands can protect the nanoparticles from agglomeration – with no protection nanoparticles will bind together and form larger particles, losing the unique nanostructure properties. In addition, the capping ligands allow surface functionalization that can help obtain specific properties. For example, by using a longer protecting ligand, the distance between the nanoparticles will be larger, thus affecting the analyte volume that can be adsorbed between adjacent nanoparticles. The third reason is a natural continuation of the first two reasons. Capping ligands allow the achievement of a hybrid combination of chemical and physical functions which can have a significant effect on the sensitivity and selectivity of the sensors.

Programmed Assembly of Functionalized Nanoparticles

One way to monitor changes in a solution of nanoparticles is by monitoring the changes in the solution optical properties. In the following example (Figure 12.19), the changes in solution absorption are examined. This system is designed to detect the presence of a specific DNA strand. For this purpose, the nanoparticles are modified via a thiol bond with single stranded DNA complementary to the target DNA. As seen in Figure 12.19 (A), the modification of the nanoparticles affects the amount of light absorbed by the solution, compared to the un-modified (or bare) nanoparticles, which have a characteristic peak at about 560nm. When the target DNA analyte is added to the solution, aggregation occurs - the nanoparticles move closer together and form clusters. This phenomenon changes the absorbance of the solution, as can be seen in the graph (Figure 12.19 (B)). The location of the absorbance peak is an indication of average nanoparticle size. When the peak is shifted to higher wavelength (for example, as in Figure 12.19, from 560 nm to 600 nm), it means that the average size of the particles in the solution is larger due to the aggregation. In addition, the width of the peak is much broader after aggregation – an observation that indicates higher distribution of nanoparticle size compared to the bare nanoparticles.

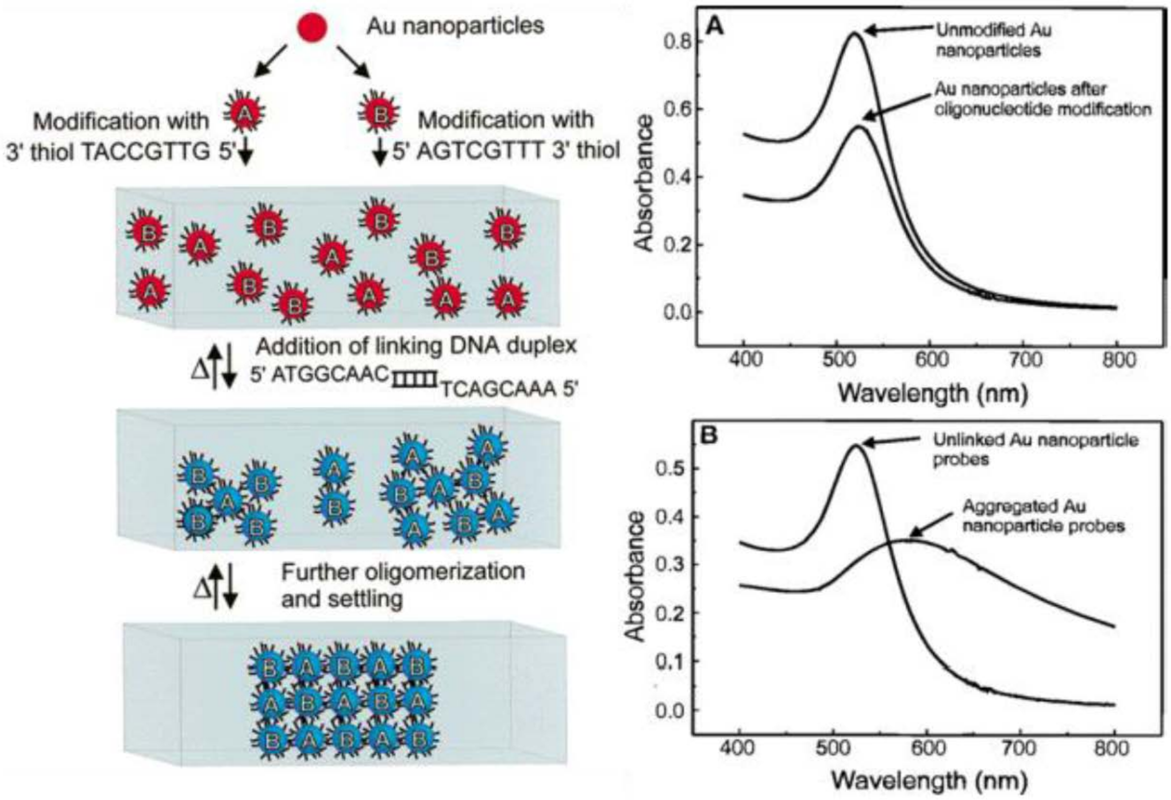


Figure 12.19 Nanoparticles for the detection of a specific DNA. (A) UV-vis spectra of Au nanoparticles in aqueous solution and Au nanoparticles functionalized with 5'-hexanethiol 12-base oligo-nucleotides. (B) Comparison of Au nanoparticles functionalized with 5'-hexanethiol 12-base oligonucleotides before and after treatment with a complementary 24-base oligonucleotide

12.5 APPLICATIONS OF METAL NANOPARTICLES

Biomedical Applications

An increasingly important area of application of MNPs is in biomedical applications, including: biodiagnostics (e.g. taking advantage of the sensitivity of plasmon resonances and other physical properties to the coordination of biomolecules), imaging (e.g. in fluorescence microscopy and Magnetic Resonance Imaging), drug delivery (typically employing relatively inert hollow gold nanospheres and nanorods) and other therapeutic applications (e.g. using the MNP as an agent for localised heating, radiation, etc.). The combination of the size of MNPs and the possibility of modifying their surfaces by coordinating surfactants to increase their lipophilicity or hydrophiliciy—or to target specific cells—makes them particularly attractive in medical applications.

The antibacterial properties of silver NPs are well known and they are now commonly used in wound dressings and in other disinfectant and antiseptic applications due to the increase in antibiotic-resistant bacteria—in fact, the use of silver in medicine predates the discovery of antibiotics. Silver NPs are also used in clothes and footwear to kill odour-causing bacteria.

There has been tremendous recent growth in the application of MNPs as biosensors for biodiagnostic applications—for examlple, for bioconjugation, as cellular labels and in assays for gases, metal ions and DNA/protein markers for disease. MNPs offer the possibility of enhanced sensitivity and selectivity over other materials. Because of their tunability (of composition and chemical ordering as well as size and shape), bimetallic NAs show particular promise as biodiagnostic agents. 50–100nm diameter Ag core Au shell NAs as tunable colorimetric probes for DNA detection, making use of the variation of the SPR frequency with composition and ordering and the sharp melting transitions of NP-labelled DNA.

Because of the widespread use of nanoparticles, including MNPs, in technological applications and even in consumer goods—such as silver NPs in clothes and proposed uses of NPs in food packaging—there are concerns over possible adverse environmental and toxicological effects of these NPs. Consumption of colloidal silver NPs, for example, in certain alternative medicines may cause argyria, a condition arising from accumulation of silver in the body, one of the manifestations of which is blue coloration of the skin.

Several studies have shown after nanoparticle cell binding, the binding and activation of membrane receptors and subsequent protein expression strongly depend on nanoparticle size. Most examined nanoparticles were in the range of 2 to 100 nm. While most nanoparticles were found to alter signaling processes essential for basic cell functions, nanoparticles with 40 nm and 50 nm diameters demonstrated the greatest effect.

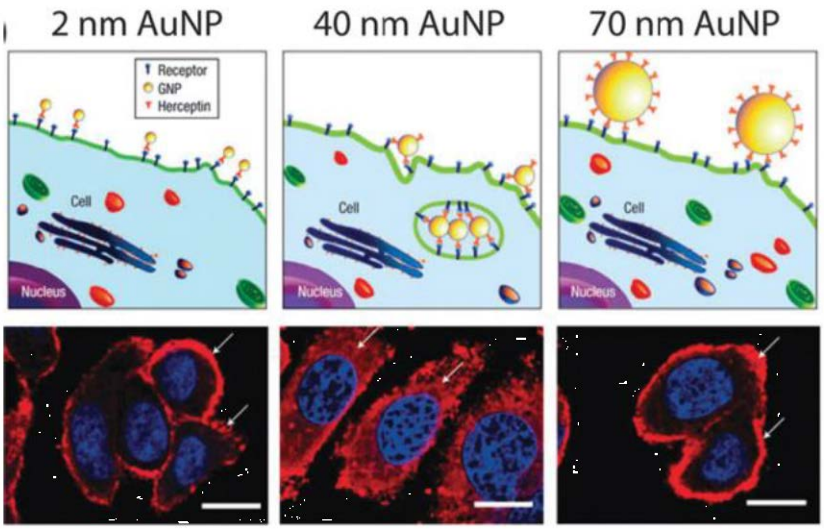


Figure 12.20 Illustration demonstrating binding of gold nanoparticles functionalized with herceptin antibodies, which recognize receptors on the cell surface. Arrows indicate ErbB2 receptors, and the nucleus is counterstained in blue

The top row in the Figure 12.20 is an illustration demonstrating binding of gold nanoparticles with diameters of 2 nm, 40 nm and 70 nm that were functionalized with herceptin antibodies, which recognize receptors on the cell surface (HER2/neu, ErbB2). As can be seen in the figure, 40 nm nanoparticles interact with the receptors more efficiently due to better matching between nanoparticle size and receptors on the cell surface, as well as improved penetration ability to the cell interior. This matching is quite clear in the second row of the figure, which presents fluorescence images of the cellular distribution of ErbB2 (red) after treatment with fluorescently-labeled nano-particles. As seen in these images, only the 40 nm nanoparticles successfully penetrated the cell membrane to the cytoplasm, while the other nanoparticle sizes only cover the cell membrane. These results show that the specific role of the nanoparticles in biological processes can be engineered via control over nanoparticle size.

DRUG AND DNA DELIVERY

A successful gene therapy requires the transport of nucleic acids into cells via delivery vehicles, as DNA is not able to efficiently traverse the cell membrane. The comparable size of DNA and small functionalized nanoparticles facilitates their interactions. Coated gold nanoparticles can produce effective transfection vectors; DNA delivery efficiency strongly depends on the structure of coating molecules and their concomitant ability to condense DNA.

The general process of the gene or DNA delivery can be schematically described as follows: the functionalized nanoparticles and the bare DNA are mixed together to form a complex of several nanoparticles with condensed DNA. This complex is not toxic to the cells due to nanoparticle functionalization and has the ability to cross the cellular matrix (Figure 12.21).

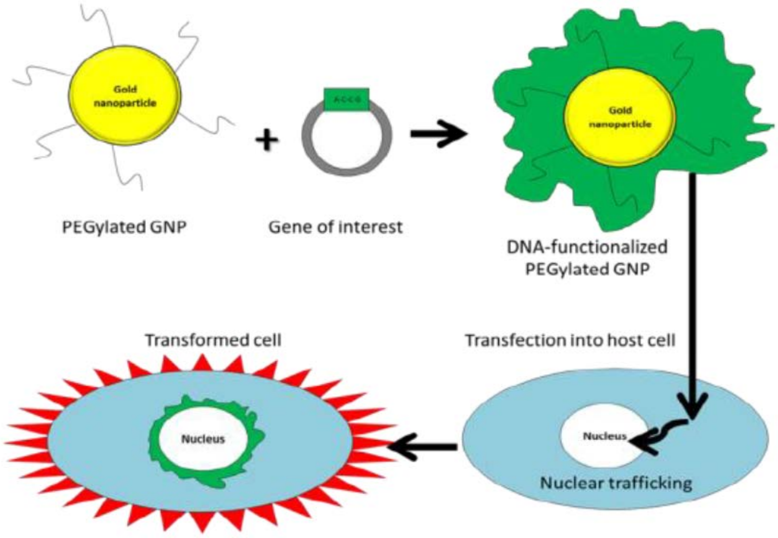


Figure 12.21 Scheme of gene or DNA delivery process using functionalized nanoparticles

The advantages of the described approach is that delivery does not cause cytotoxicity, the gene expression receives high amplification, and the carrier allows traversal of cellular membranes.

NANOPARTICLE BASED SENSORS FOR DETECTION OF DNA

DNA-capped gold nanoparticles are therefore a promising approach for the detection of DNA strands by controlling solution temperature of DNA-modified nanoparticles. Using a similar concept, a drop of nanoparticles can be cast on paper. As seen in Figure 12.22(A), when the sensor meets its target DNA, the color of the drop changes due to aggregation. This system does not require sophisticated equipment for analysis and is appropriate mainly to determine the presence of an analyte, not its concentration. Detecting small concentrations of target DNA is more challenging.

Figure 12.22 (B) presents gold nanoparticles modified with short, single-stranded DNA to indicate the presence of a particular DNA sequence that is hybridized on a transparent substrate in a three-component sandwich assay format. At high target DNA concentrations, many nanoparticles are attached to the transparent surface. The high concentration of nanoparticles causes a dramatic change in the color and the substrate looks pink. At lower target concentrations, the attached nanoparticles cannot be visualized with the naked eye. To facilitate visualization of nanoparticle labels that are hybridized to the array surface, plating silver ions on the gold nanoparticles is used for signal amplification. This way, very low surface coverage of nanoparticles can be visualized by a simple scanner or by the naked eye.

A similar principle is applied in the example presented in Figure 12.22 (C), where the attachment of nanoparticles to a substrate in a three-component sandwich assay format is detected by the change in scattered light. A more applicable approach than the sensing method in Figure 12.22 (C) is monitoring the electrical changes of a sensor while the sensing mechanism is similar to the previous examples. The setup in this sensing device (Figure 12.22 (D)) is an isolating substrate with two electrodes for measuring the electrical resistance. On this substrate, nanoparticles are assembled in a low concentration so there is no direct connection between the nanoparticles and, therefore, the electrical resistance is very high. If DNA is captured, silver will plate the nanoparticles. This plating will contact between the nanoparticles and the electrical resistance will dramatically decrease. By measuring the changes in the electrical resistance, a sensing signal can be obtained and monitored.

Different approaches that utilize similar sensing mechanisms, but use different characterization methods (or instrumentations) have been developed and tested to detect the attachment of nanoparticles to a substrate. One way is to monitor the surface-enhanced Raman spectroscopy (SERS) signal, as depicted in Figure 12.22 (E). Another option is to monitor the diffraction of a laser beam, when the laser beam is directed towards the nanoparticles that are attached to the surface (Figure 12.22 (F)).

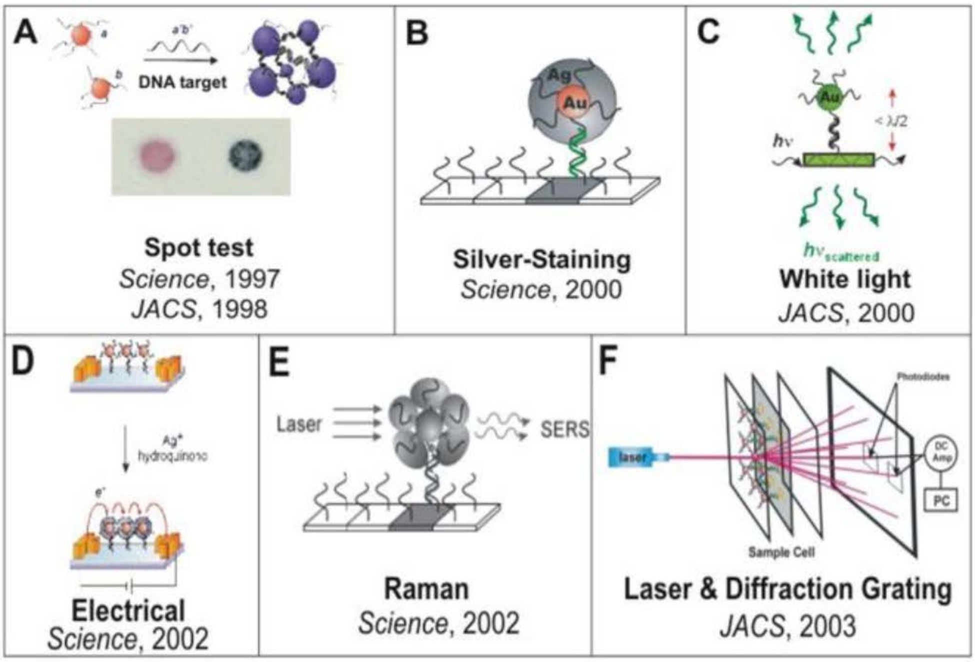


Figure 12.22 DNA-capped gold nanoparticles for the detection of DNA strands. (A) Single-stranded target oligonucleotide into a solution containing the appropriate probes resulted in the formation of a polymeric network of nanoparticles with a color change. (B-C) Very small changes in temperature permits discrimination of single base pair mismatches via either silver staining (Figure B) or light scattering (Figure C) to detect nanoparticles captured onto a surface by target DNA that has been immobilized by surface-bound oligonucleotides (a so-called "sandwich assay"). (D) Target DNA is captured in the gap between two electrodes using a sandwich assay. Following capture, silver is plated onto the nanoparticles, bridging the gap between the two electrodes and thus allowing current flow. The current flow is translated into DNA detection. (E) RAMAN dye labeled DNA-gold probes in conjunction with silver staining allows for multiplexed detection of DNA targets, because each target probe yields a distinct SERS signal that can be read via RAMAN spectroscopy. (F) Surface-bound nanoparticle probe-based assay with diffraction grating signal transduction scheme display a high sensitivity and selectivity afforded by nanoparticle probes and the experimental simplicity, wavelength-dependent resonant enhancement provided by the diffraction-based sensing technology.

REACTIVITY AND CATALYSIS

Most studies of nanoparticle reactivity have been performed on transition MNPs, in order to model and understand fundamental processes which occur in heterogeneous catalysis. In the case of finite MNPs, reactivity depends on the number of atoms and the overall geometry. Quite dramatic size-dependence is often observed in reactivity and catalytic studies (both in terms of activity and selectivity). For NAs, reactivity and catalytic behaviour can also depend on the composition and chemical ordering.

Core–shell bimetallic NAs are very promising for the design of new catalysts. It was shown that Ni core Pd shell NAs are efficient catalysts for Hiyama cross-coupling of phenyltrimethoxysilane and haloaryls, and are more active (per Pd atom) than monometallic Pd NPs and mixed Ni–Pd NAs, as illustrated in Figure 12.23.

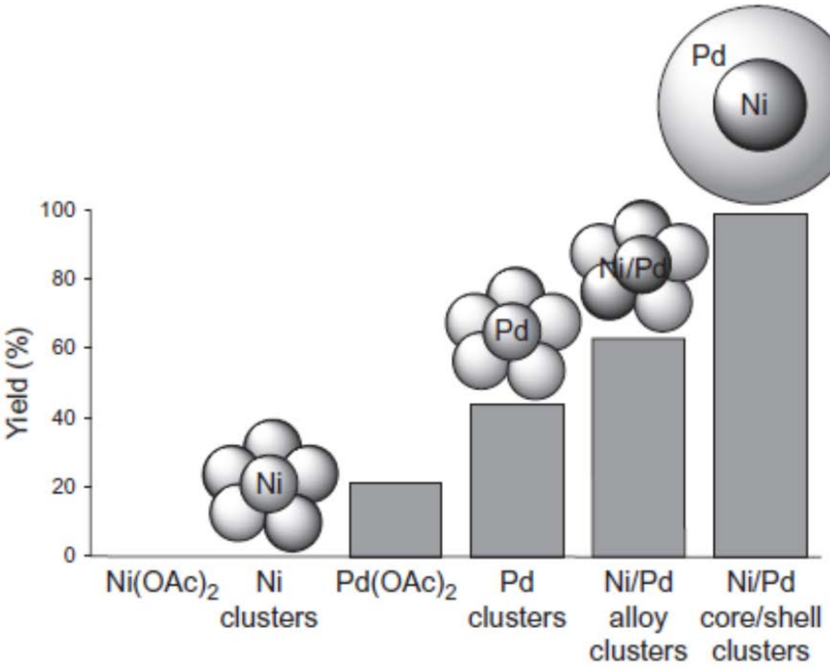


Figure 12.23 Comparison of the catalytic activity for six different systems in the Hiyama cross-coupling of iodotoluene and trimethoxyphenylsilane: Ni(OAc)2, Ni clusters, Pd(OAc)2, NiPd alloy clusters and Ni core Pd shell cluster.

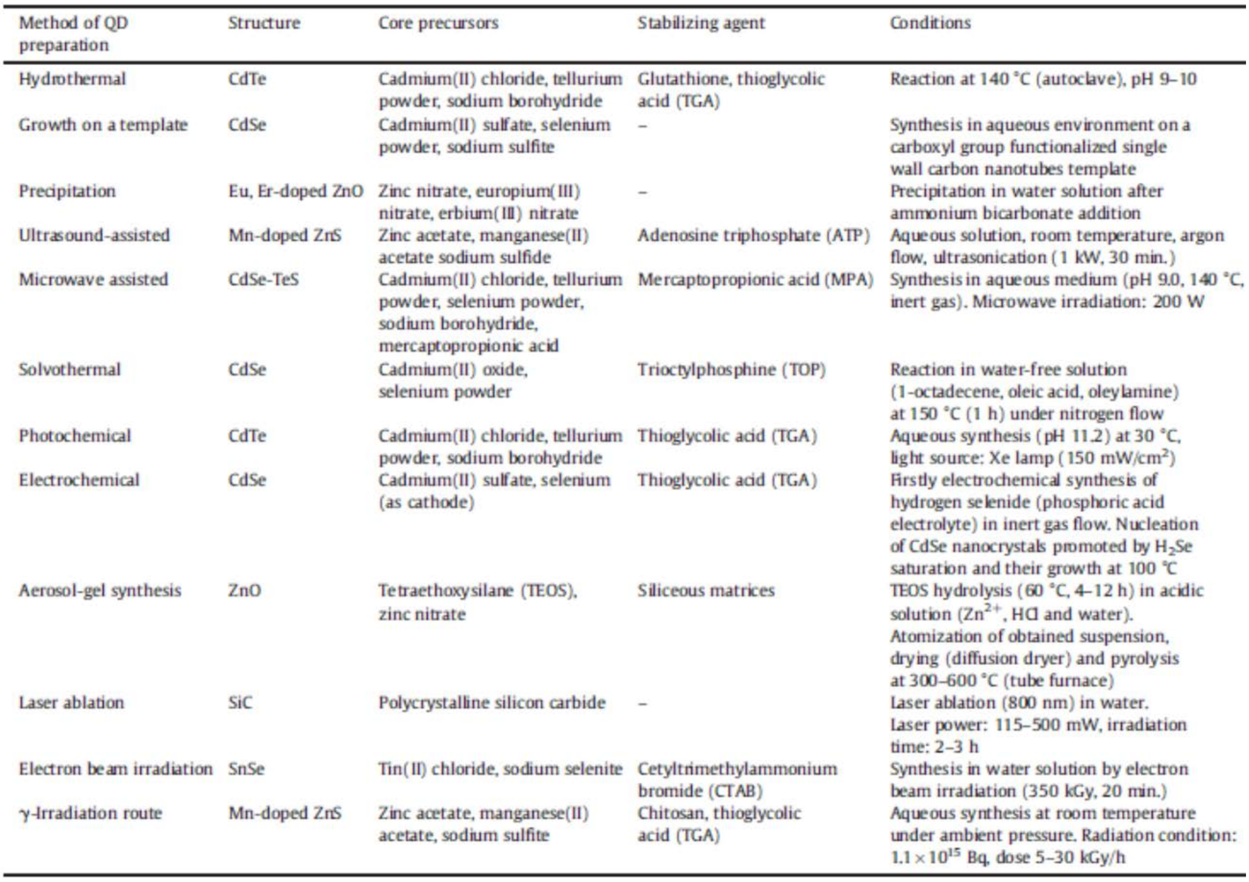
12. 6 INRODUCTION TO SEMICONDUCTOR NANOCRYSTALS

Syntesys of semiconductor nanocrystals and nanoparticles

In the last few years the development of semiconductor nanocrystals and nanoparticles (NPs) has been extended from the pure adjustment of the size of the particles to the control of more sophisticated properties as their shape and composition. The hope is that through these features different fields of applications might be opened for the use of these NPs.

NPs are fabricated by different methods such as electron beam irradiation , polyol-hydrolysis method , chemical precipitation method, photochemical synthesis , γ-radiation route or microwave-assisted aqueous synthesis . However, the two main routes of NP fabrication include organometallic and hydrothermal syntheses . The variety of NP synthesis methods is presented in Table 12.1.

Table 12.1
Methods of NP synthesis.



Method of QD preparation	Structure	Core precursors	Stabilizing agent	Conditions
Hydrothermal	CdTe	Cadmium(II) chloride, tellurium powder, sodium borohydride	Glutathione, thioglycolic acid (TGA)	Reaction at 140 °C (autoclave), pH 9–10
Growth on a template	CdSe	Cadmium(II) sulfate, selenium powder, sodium sulfite	–	Synthesis in aqueous environment on a carboxyl group functionalized single wall carbon nanotubes template
Precipitation	Eu, Er-doped ZnO	Zinc nitrate, europium(III) nitrate, erbium(III) nitrate	–	Precipitation in water solution after ammonium bicarbonate addition
Ultrasound-assisted	Mn-doped ZnS	Zinc acetate, manganese(II) acetate sodium sulfide	Adenosine triphosphate (ATP)	Aqueous solution, room temperature, argon flow, ultrasonication (1 kW, 30 min.)
Microwave assisted	CdSe-TeS	Cadmium(II) chloride, tellurium powder, selenium powder, sodium borohydride, mercaptopropionic acid	Mercaptopropionic acid (MPA)	Synthesis in aqueous medium (pH 9.0, 140 °C, inert gas). Microwave irradiation: 200 W
Solvothermal	CdSe	Cadmium(II) oxide, selenium powder	Trioctylphosphine (TOP)	Reaction in water-free solution (1-octadecene, oleic acid, oleylamine) at 150 °C (1 h) under nitrogen flow
Photochemical	CdTe	Cadmium(II) chloride, tellurium powder, sodium borohydride	Thioglycolic acid (TGA)	Aqueous synthesis (pH 11.2) at 30 °C, light source: Xe lamp (150 mW/cm^2)
Electrochemical	CdSe	Cadmium(II) sulfate, selenium (as cathode)	Thioglycolic acid (TGA)	Firstly electrochemical synthesis of hydrogen selenide (phosphoric acid electrolyte) in inert gas flow. Nucleation of CdSe nanocrystals promoted by H_2Se saturation and their growth at 100 °C
Aerosol-gel synthesis	ZnO	Tetraethoxysilane (TEOS), zinc nitrate	Siliceous matrices	TEOS hydrolysis (60 °C, 4–12 h) in acidic solution (Zn^{2+}, HCl and water). Atomization of obtained suspension, drying (diffusion dryer) and pyrolysis at 300–600 °C (tube furnace)
Laser ablation	SiC	Polycrystalline silicon carbide	–	Laser ablation (800 nm) in water. Laser power: 115–500 mW, irradiation time: 2–3 h
Electron beam irradiation	SnSe	Tin(II) chloride, sodium selenite	Cetyltrimethylammonium bromide (CTAB)	Synthesis in water solution by electron beam irradiation (350 kGy, 20 min.)
γ-Irradiation route	Mn-doped ZnS	Zinc acetate, manganese(II) acetate, sodium sulfite	Chitosan, thioglycolic acid (TGA)	Aqueous synthesis at room temperature under ambient pressure. Radiation condition: 1.1×10^{15} Bq, dose 5–30 kGy/h

NPs prepared by the organometallic route are typically stabilized by the hydrophobic ligands such as trioctylphosphine oxide (TOPO) or trioctylphosphine (TOP). The application of photoluminescent nanocrystals requires their water-dispersibility in order to be exploited as fluorescent labels or sensors . There are several strategies to make NPs obtained through the organometallic synthesis water-dispersible. One of them is the exchange of hydrophobic ligand at the NP surface by a hydrophilic one. Another possibilities include the silanization of the NP surface, the use of other polyfunctional hydrophilicmolecules such as coordinating oligomers (for example phosphine oligomers), and the use of dendrons. New strategies of NP solubilization consist on encapsulation of their hydrophobic surface by the addition of amphiphilic molecules. In this case the alkyl chains of amphiphilic agent interact with the n-octyl groups of TOP or TOPO through hydrophobic bonds. The hydrophilic endpoints of the amphiphilic molecules enable the dispersion in aqueous medium. Apart from NP encapsulation with amphiphilic oligomers (e.g. modified polimaleic acid), the surfactants such as cetyltrimethylammonium bromide (CTAB), gemini surfactants, and pegylated-phospholipids were reported in the literature. NPs can be also incorporated in silica nanospheres. In this method, NPs are exposed to tetraethyl orthosilicate (TEOS). TEOS is then gently hydrolyzed under controlled conditions to form the silica nanospheres resulting in modification of NP surface. NPs incorporated in these nanospheres are well isolated from the environmental medium and usually possess better photophysical properties than the starting NPs. Generally one nanosphere contains 3–10 NPs, however that number can be adjusted by the NPs/TEOS ratio and by the kinetic of TEOS hydrolysis.

Size Dependence

Especially in the case of semiconductor nanocrystals the effect of the size on the optical properties of the particles is very striking. The smallest CdSe nanocrystals with a diameter of less than 2 nm show for example a blue fluorescence. Larger nanocrystals of the same material (diameter ca. 6 nm) emit red light. In a simple model this can be understood through the confinement of the exciton, i.e. the light generated electron-hole pair whose recombination is responsible for the fluorescence emission, into the volume of the nanocrystal.

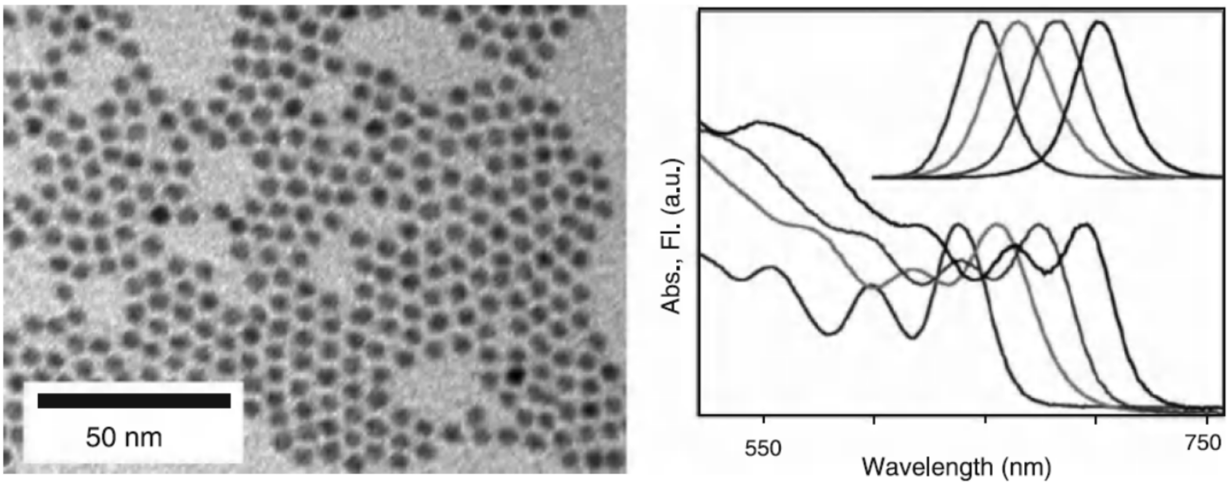


Fig. 12.24. Spherical semiconductor nanocrystals. The transmission electron micrograph (TEM) shows nanocrystals of CdTe. In the right panel absorption (Abs.) and fluorescence (Fl), shifted upwards for clarity) spectra of samples with different particle sizes are shown.

In a more detailed analysis, colloidal NPs can be considered as an intermediate species between atoms or molecules on the one hand and bulk material on the other hand. The energetic levels of a simple molecule can be calculated by the tight binding or LCAO (linear combination of atomic orbitals) approach. In this model the outer orbitals of the participating atoms are combined to new molecular orbitals. Generally one observes orbitals with energy lower than that of the atomic orbitals, the so-called binding orbitals, and orbitals of higher energy, termed antibinding. Extension of this model to a larger number of atoms leads to the formation of bands.

The binding orbitals form the valence band, whereas the antibinding orbitals are combined to the conduction band. The spacing between the bands, i.e. the bandgap, decreases with the number of atoms added to the molecule. The lower limit for the bandgap is the value of bulk material. This value is reached when the radius of the crystals is of the order of the Bohr radius of the exciton. This trend can be observed when comparing semiconductor nanocrystals of the same material, but of different sizes. The larger the nanocrystals are, the more the fluorescence colour is shifted towards the red, i.e. towards lower energies (Fig. 12.24).

Shape dependence.

Spherical nanocrystals can be considered as zero dimensional objects. Confinement is exerted in all three dimensions. For this the density of states is discontinuous. If one axis of the nanocrystal is extended, the density of states changes slightly, as confinement is now exerted only in two dimensions. In these dimensions the levels are still quantised. Levels attributed to wave vectors parallel to the long axis of the structure are quasi continuous. Therefore each of the quantised levels is effectively broadened by the absence of confinement in this long axis of the structure. The positions of the electronic levels are still dominated by the size of the smallest axes.

It can be shown that rod-shaped semiconductor nanocrystals, so-called nanorods, emit polarized light and that the degree of polarization scales with their aspect ratio. With increasing aspect ratio of the nanorods, only a slight shift of the bandgap is observed, which saturates at aspect ratios of the order of ten. This makes nanorods an appealing system for electronic devices. The long axis facilitates to contact the structure, while the short axes preserve the quantised nature of the electronic levels. For instance nanowires, i.e. nanorods with very high aspect ratios, of different materials can be embedded into electric circuits to act as transistors or other active elements. Electroluminescence is observed from CdSe nanorods, Their symmetry facilitates the alignment of nanorods. In some cases it is sufficient to slowly evaporate the solvent to obtain large areas of aligned nanorods. Better results are achieved when the nanorods are oriented by an electric field. Furthermore nanorods were proven to be of advantage in solar cells. Tetrapods, i.e. nanoparticles with four rod-shaped arms that are combined to the shape of a tetraeder, offer the possibility to expand this spectrum of applications. When three of the arms are contacted electrically, one of the arms can be used as a gate to control the current through the entire structure. Also, tetrapod-shaped semiconductor nanocrystals can be of advantage in solar cells. They exhibit a large surface on which charges can be separated, and still provide a pathway to transport charges to an electrode due to their complex and extended geometry. Furthermore, tetrapods exhibit an interesting structure of their excited states. Electrons and holes can be localised in the core region where the arms are linked and the overlap between the two wave functions differs considerably from one excited state to another.

12.7 GENERAL GROWTH MECHANISM OF SEMICONDUCTOR NANOCRYSTALS

Nucleation

The first step in the growth of any sort of nanocrystal is evidently the nucleation. Through a density fluctuation of the medium several atoms assemble to a small crystal that is thermodynamically stable, and thus does not decay to free atoms or ions. In that sense the nucleation can be understood as the overcoming of a barrier. This section will explain briefly the origin of this barrier. In the following we will distinguish between the crystalline phase, in which the atoms are bound to a crystal, and the solution phase, in which the atoms are dispersed freely in the solution. The nucleation in a solution at constant temperature and constant ressure is driven by the difference in the free energy between the two phases. At the simplest the driving forces in the nucleation event can be reduced to two, the gain in the chemical potential and the increase of the total surface energy. The gain in chemical potential can be understood as the energy freed by the formation of the bonds in the growing crystal. The surface term takes into account the correction for the incomplete saturation of the surface bonds. Upon formation of a spherical nucleus consisting of n atoms the total free energy of the system changes by the value

$$\Delta G = n(\mu_c - \mu_s) + 4\pi r^2 \sigma \tag{12.2}$$

(μ_c and μ_s are the chemical potentials of the crystalline phase and the solution phase, respectively, r is the radius of the nucleus and s the surface tension1.) In this equation the surface term constitutes the main difference between nanomaterials and bulk crystals. Bulk material is dominated by volume effects and thus the surface energy term in Eq. (12.2) can be neglected, whereas in nanocrystals a non-negligible portion of atoms might be situated on the surface and thus this term is of importance. In Eq. (12.2) the surface tension σ is assumed to be constant for any size and morphology of the crystal, which is a very rough approximation. This uniformity of the surface tension σ refers to the drop model of the nucleation.

We will assume a spherical shape for the nanocrystals and neglect any variation of the surface tension σ (i.e. isotropic crystal without facets). In this case the number of atoms n in the first term of Eq. (12.2) can be expressed by the radius r of the nanocrystals, taking into account the density dm of atoms in the nanocrystals. Then the equation reads:

$$\Delta G = \frac{4\pi d_m}{3} r^3(\mu_c - \mu_s) + 4\pi r^2 \sigma$$

(12.3)

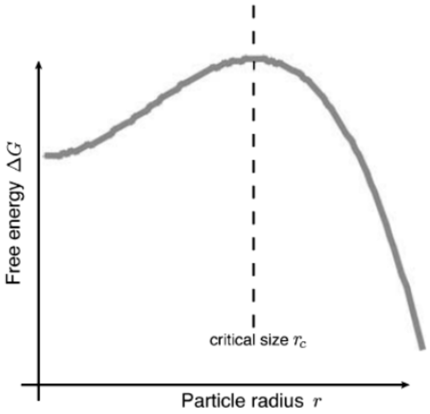


Fig. 12.25 Sketch of the potential landscape for the nucleation. Only at small values of the radius r in Eq. (12.2) the surface energy term outcompetes the r^3 contribution of the chemical potential, so that a barrier is imposed at the critical size r_c

In the case that the chemical potential of an atom in the solution is smaller compared to that of an atom within the crystal, the minimum of the free energy is given when all atoms are unbound, and so no stable crystals are formed. Here we are interested in the opposite case with the chemical potential of atoms in solution being larger than that of bound atoms. In that case the first term becomes negative, and thus the free energy reaches a maximum for a certain radius r_c, termed critical size, at which a nucleation barrier is imposed, see Fig. 12.25. For small nuclei the surface energy term dominates the free energy, whereas only for crystals much larger than r_c the growth is driven by the gain in chemical potential and in principle the crystals can grow to an infinite size. The amplitude of the nucleation barrier controls the rate at which crystals nucleate

Growth

The actual process of the deposition of monomers onto the growing nanocrystal can be split into two steps. First the monomers have to be transported towards the surface of the nanocrystal and in a second process they have to react with the nanocrystal. Generally the first process is accomplished through diffusion and thus at a rate dominated by the diffusion constant D, whereas the speed of the second process is given by the rate of reaction between freemonomers and the crystal surface.

In the following the dynamics of the growth process will be outlined briefly. The discussion will start from a mechanistic view of the growth in which the growth rate

$\dot{r} = \mathrm{d}r/\mathrm{d}t$ of a crystal of radius r depends only on the rate at which monomers are incorporated into the crystal. This latter rate is simply the time derivative of the number of monomers n in the crystal and it describes the number of monomers that go through the two processes mentioned above (diffusion and reaction) during a unit of time. We can write the growth rate $\dot{r}$ as

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\dot{n}}{4\pi r^2 d_\mathrm{m}} \tag{12.4}$$

The general dependence of the growth rate on the radius of the crystal is illustrated in Fig.12.26. It is interesting to note the presence of a maximum at a radius of $2r^*$. If all crystals present in the solution have a radius larger than this value, the smallest crystals grow fastest, and therefore the size distribution becomes narrower over time. The value of r^* depends mainly on the overall concentration of free monomers, but also on the reaction temperature and on the surface tension σ.

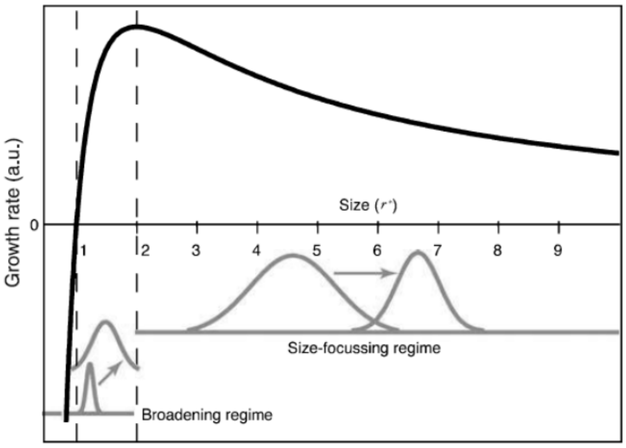


Fig. 12.26. Sketch of the growth rate dr/dt of the nanocrystals in units of the critical size r^*.As examples twosize distributions and their development with time are displayed. Note that in the example for the broadening regime, the smallest particles are larger than r^*. Therefore the mean size of the particles still increases. The situation would be different if the size distribution would comprise nanocrystals smaller than r^*

In the case of fluorescent nanocrystals, the width of the fluorescence spectrum is a good indicator for the quality of the size distribution. Samples of CdSe or CdTe nanocrystals generally have a band linewidth of ca. 30 nm or less, which in this case corresponds to ca. 100 meV. In Fig. 12.27 an example for the focussing of the size distribution is displayed. In this example the synthesis was carried out under conditions that favour size focussing. The nanocrystals were synthesised at a very high temperature, which reduces the nucleation event to a very short time span. Also, the concentration of monomers was sufficiently high to prevent the system to enter into the broadening regime.

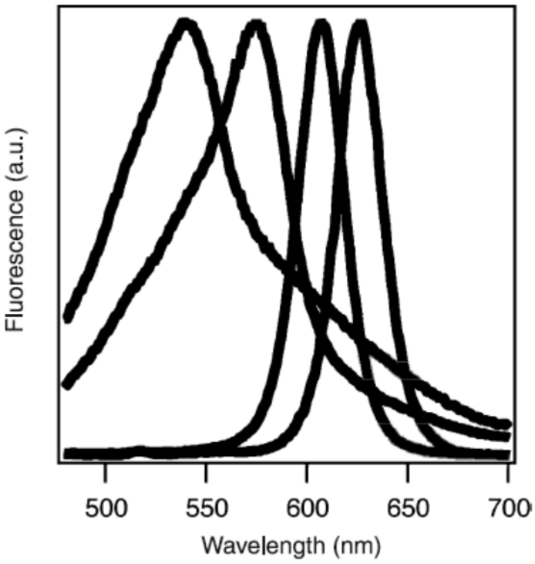


Fig. 12.27. Effect of size focussing in the synthesis of CdSe nanocrystals. The synthesis was performed at high temperature (370 ^{o}C). Aliquots were taken every 20 s during the run of the synthesis and fluorescence spectra were recorded. The first aliquot (leftmost spectrum) shows a broad fluorescence spectrum and thus a wide size distribution, whereas the fluorescence spectrum of the latest aliquot (rightmost spectrum) is very narrow with a FWHM of 28 nm

In some synthesis schemes it is inevitable to consume the reservoir of monomers and thus obtain a broadening of the fluorescence band. In these cases the system can be maintained in focussing regime by repeated injections of fresh monomers. Experimentally it is not always possible to obtain a perfect size distribution. In some cases the distribution is broadened or shows several distinct peaks. A possible, but somewhat laborious way to improve the size distribution is to perform a size selective precipitation after the synthesis is completed. In order to do so, a nonsolvent is added slowly in a controlled way. As an example the addition of the polar solvent methanol to nanocrystals dissolved in a non-polar solvent such as chloroform or toluene induces the precipitation of the nanocrystals. The larger particles become unstable in the solution at lower concentrations of the non-solvent than the smaller particles, and thus they precipitate first. This process has been successfully applied for instance in the synthesis of CdTe and CdSe . In Fig. 12.28 an example of a size-selective precipitation is shown.

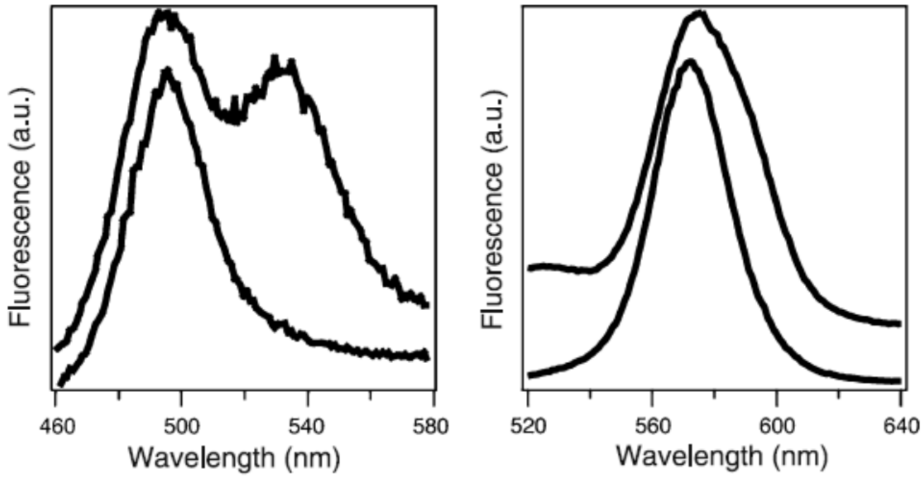


Fig. 12.28. Effect of the size-selective precipitation. By careful addition of a non-solvent to samples of CdSe nanocrystals with a bimodal (left) or simply broadened (right) size distribution the largest particles, i.e. those with an emission at higher wavelength, can be precipitated by centrifugation while the smaller particles remain in solution. The upper spectra show the fluorescence of the initial samples, whereas the lower spectra show the fluorescence of the supernatants. In the right example a further contamination of small, i.e. blue fluorescent, nanocrystals has been removed by a second precipitation

12.8 MAGIC SIZE CLUSTERS AND SHAPE CONROL OF SEMICONDUCTOR NANOPARTICLES

Magic Size Clusters

For many applications it is of interest to have materials at hand that exhibit a high fluorescence yield and at the same time can be packed very densely. Through this many light-emitting units could be packed into a small volume and the overall fluorescence yield of the material would be enhanced as an eventually low fluorescence quantum yield might be overcome by a larger number of light-emitting sites. Also, in applications as markers the size of the fluorophore is a limiting factor. Most techniques to produce water-soluble nanocrystals involve the generation of ashell of organic molecules around the nanocrystals.

To synthesise routinely nanocrystals with a smaller diameter, a different concept has to be employed. Already in the reported synthesis schemes of the CdSe nanocrystals the existence of highly stable and very small CdSe clusters was observed. In the size regime below 2 nm the stability of these clusters can be explained with the concept of magic sizes. Several discrete cluster configurations exhibit a higher stability as compared to other configurations. Generally their high stability coincides with a high symmetry in their structure. The magic size clusters could be considered as those clusters that possess complete shells of atoms.

In literature one can find protocols for the synthesis of several magic-size clusters. In these reaction schemes for each different cluster size a different technique is employed. For this reason also the structure and nature of the surfactants is different for each of the different clusters. The size of the cluster is determined only by the reaction time. By a careful treatment of the remaining supernatant it is even possible to eliminate the peak related to these large clusters almost completely, as shown in Fig. 12.29a. In the following the clusters are classified into six families, each of them characterised by a different peak in the absorption spectrum (Fig. 12.30).

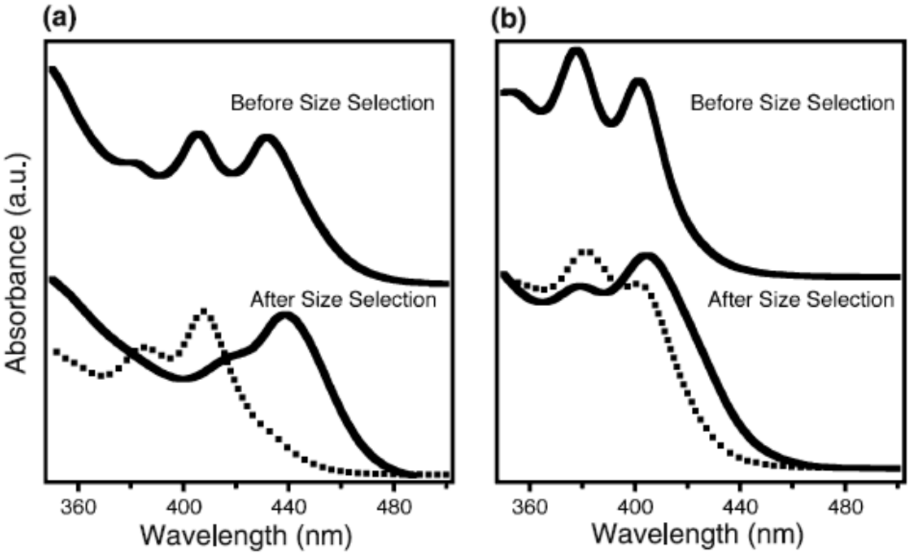


Fig. 12.29. Size-selective precipitation of two samples of magic size clusters. Methanol was added carefully to the cooled reaction solution (upper spectra) and the mixture was centrifuged. In the spectrum of the first precipitates (solid lines, lower spectra) the amplitude of the first peak is increased with respect to the other peaks. In the spectrum of the supernatants (dashed lines, lower spectra) the situation is inverted and the second peak is more pronounced than the first.

Another indication for the discontinuous size distribution can be found in the fluorescence spectra. As shown in Fig. 12.31 the fluorescence spectra at late stages of the reaction, i.e. when the families 4–6 are present, are composed of a series of distinct peaks, similar to the behaviour of the absorption spectra. The observation of the discontinuity in the fluorescence spectra is experimentally limited by the low fluorescence efficiency of the bare clusters. Especially the smallest clusters (families 1–3) show almost no sharp band edge fluorescence.

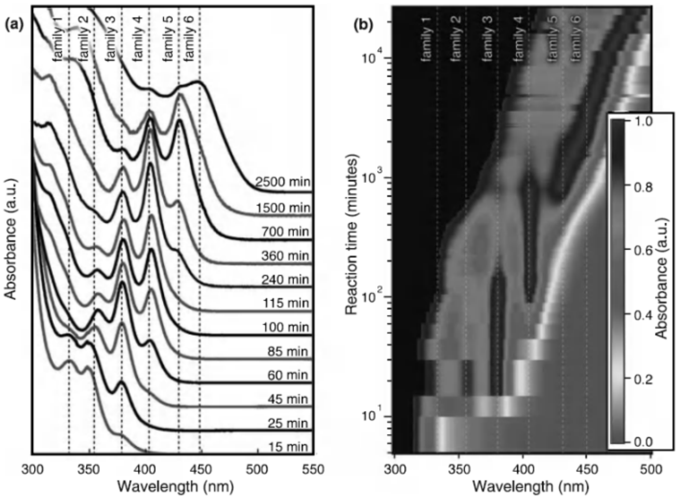


Fig. 12.30. Characteristic absorption spectra of magic size clusters of CdSe. Absorption spectra of aliquots taken during the synthesis of magic size clusters in two different representations (a, b). In both cases the spectra are normalised to the amplitude of the dominant peak. At the early stage of the reaction, only distinct peaks are visible at 330 (family 1), 360, 384, 406, 431 and 447 nm (family 6). In the upper right part of b the transition from the discrete growth dynamics to a continuous growth mode can be observed.

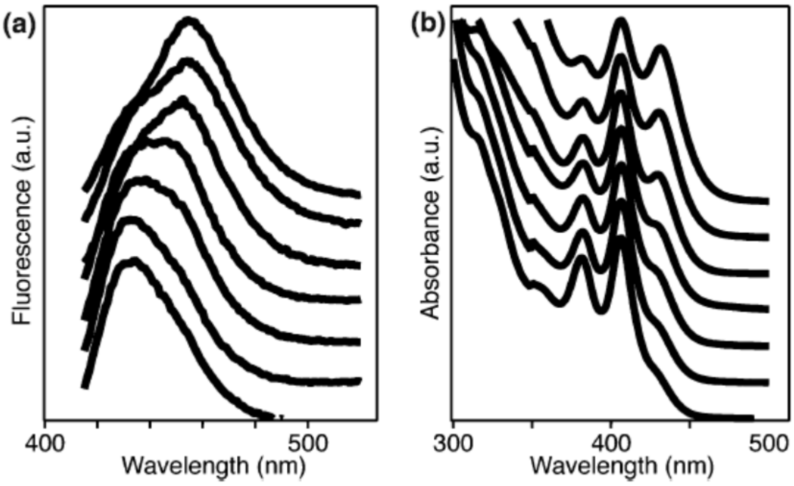


Fig. 12.31. Fluorescence spectra of the growing magic-size clusters. Fluorescence spectra (a) were recorded at different times during the growth process and for comparison the corresponding absorption spectra are shown in b. For clarity the spectra recorded at later stages of the growth are shifted upwards parallel the y axis. Even though it is not as clearly visible as in the case of absorption also the fluorescence spectra are composed of discrete peaks. The first (lowest) fluorescence spectrum shows a strong peak centred at ca. 430 nm with a small shoulder at higher wavelengths. The latest spectrum has a strong peak at ca. 450 nm and a shoulder at lower wavelengths. Intermediate spectra show a broadened peak. By comparison with the amplitude of the first two absorption peaks, one can infer that the quantum yield of the larger clusters, i.e. those fluorescing at 450 nm, is superior to that of the smaller clusters.

Shape control

The nanocrystals discussed so far were of almost spherical shape. However, in a more detailed analysis already in very small nanocrystals such as the magic-size clusters presented in the previous section different crystalline facets can be identified. On the other hand, due to the small size of these facets the deposition of only few atoms onto one of them changes its nature completely. With some nanocrystals the faceted shape becomes obvious, as for instance on octahedral PbSe nanocrystals, silver nanocubes or pencil-shaped Co nanocrystals (see Fig. 12.32).

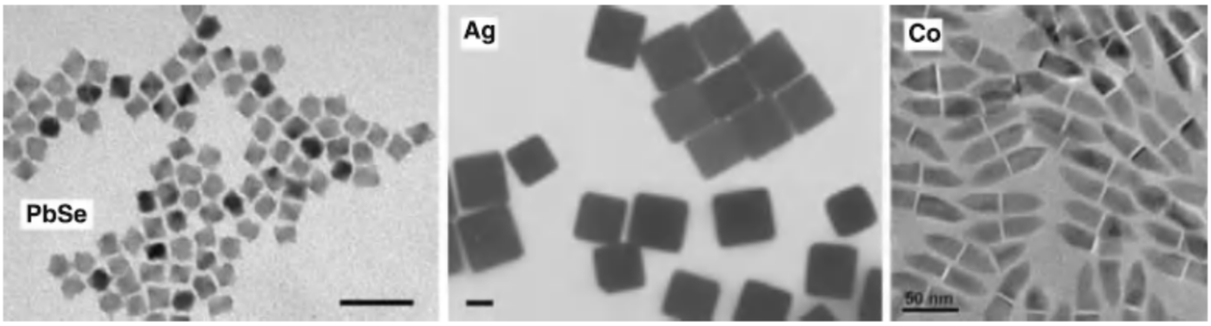


Fig. 12.32. Examples of faceting of nanocrystals. TEM images of differently shaped nanocrystals: octahedron-shaped PbSe, cube-shaped silver and pencil-shaped CoO. The scalebars represent 50 nm.

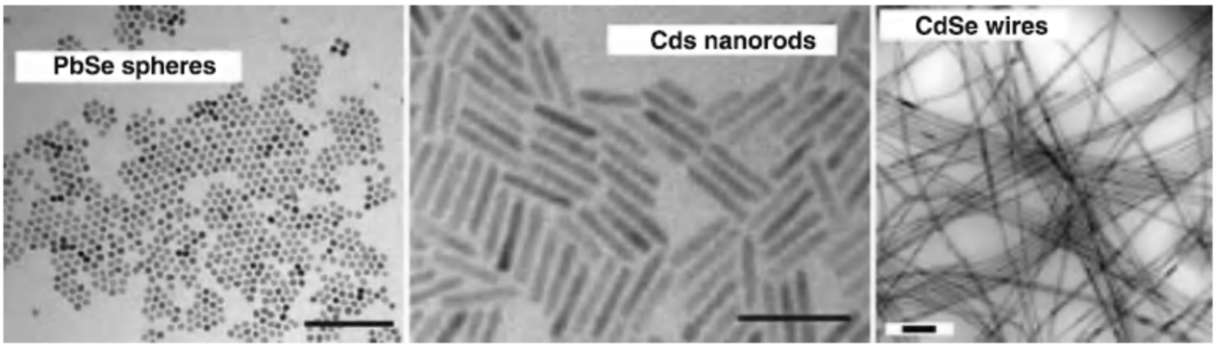


Fig. 12.33. TEM images of nanocrystals with different aspect ratios. Spherical PbSe nanocrystals (left), rod-shaped CdS and nanowires of CdSe (right). The scale bars correspond to 50 nm.

Indeed, the growth of spherical crystals is more an exception than the rule. For instance macroscopic quartz is usually found in an elongated shape with six large facets that intersect under an angle of 120°. The tip of the crystal shows a series of facets, all inclined with respect to the long axis of the crystal. Generally the shape of a crystal depends on the relative speeds at which the individual facets grow. Here, the speed of growth of a facet is measured as the speed at which the distance of its centre to the centre of the entire crystal increases. The faster a crystal grows on one facet the more likely to disappear is this facet A higher level of complexity is observed in branched nanocrystals (Fig. 12.33, 12.34).

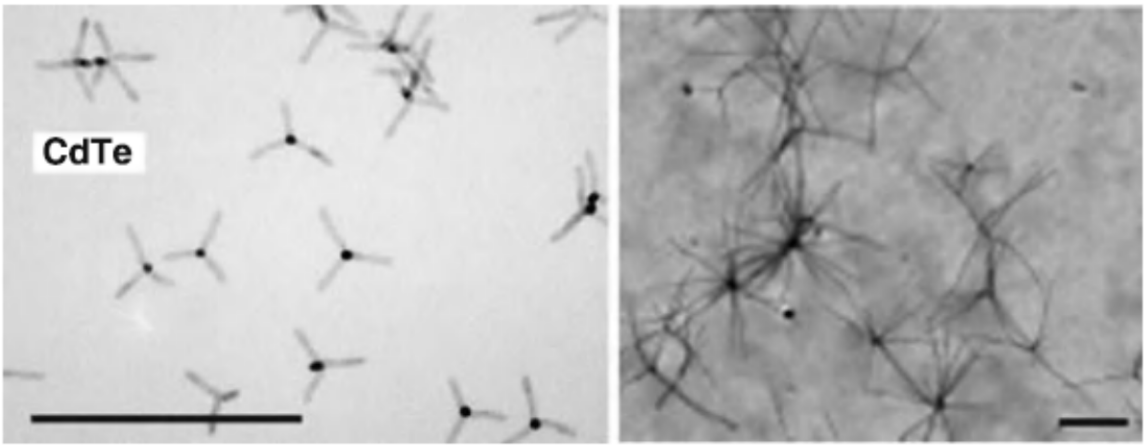


Fig. 12.34. TEM images of branched CdTe. (Left) The nanocrystals consist of four rods that are fused together in one point so that they span a tetrahedral angle between two legs. In the projection of the TEM, three legs can easily be identified, the forth legs points directly upwards. Due to the longer distance that the electron beam has to travel through the material this leg appears as a dark spot. (Right) Hyperbranched particles of CdTe. The scale bars represent 200 nm

12.9 THE MORE APPLICABLE SEMICONDUCTOR NANOCRYSTALS

CdTe nanocrystals

In a typical standard synthesis, $Cd(ClO_4)_2\ 6H_2O$ (or other soluble Cd salt) is dissolved inwater in the range of concentrations of 0.02M or less, and an appropriate amount of the thiol stabilizer is added under stirring, followed by adjusting the pH by dropwise addition of a 1 M solution of NaOH. The solution is placed in a flask fitted with a septum and valves and is deaerated by N_2 bubbling for 30 min. Under stirring, H_2Te gas is passed through the solution together with a slownitrogen flow. Alternatively, an excess of H_2Te may be passed through the deaerated solution of the NaOH with known concentration forming equimolar NaHTe solution. After this the estimated amount of NaHTe solution can be taken out and injected in the reaction flask. It should be noted that NaHTe solutions are inherently very unstable and become pink due to the oxidation. The use of only freshly prepared NaHTe solution is recommended. CdTeNCprecursors are formed at the stage of Te precursor addition (Reaction 12.5 or 12.5a); formation and growth of NCs (Reaction 12.6) proceed upon refluxing at 100°C under open-air conditions with a condenser attached.

$$Cd^{2+} + H_2Te \xrightarrow{HS-R} Cd-(SR)_xTe_y + 2H^+$$

$$Cd^{2+} + NaHTe \xrightarrow{HS-R} Cd-(SR)_xTe_y + H^+ + Na^+$$

$$Cd-(SR)_xTe_y \xrightarrow{100^\circ C} CdTe\ NCs$$

12.5, 12.5a, 12.6

A schematic drawing of the typically used experimental setup is shown in Fig. 12.35, although the specific design of this setup may vary from one group to another.

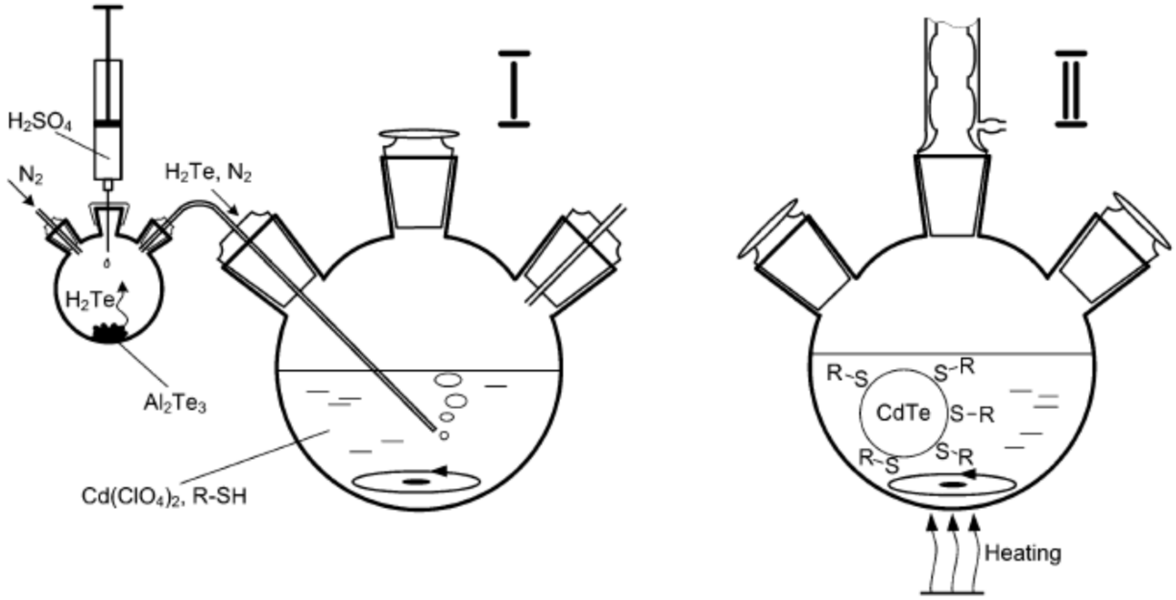


Fig. 12.35. Schematic presentation of the synthesis of thiol-capped CdTe NCs. First stage: formation of CdTe precursors by introducing H_2Te gas into the aqueous solution of Cd precursors complexed by thiols. Second stage: formation and growth of CdTe NCs promoted by reflux acid (MPA) and 2-mercaptoethylamine (or cysteamine).

Aqueous synthesis of semiconductor nanocrystals synthesis of the most stable (typically, for years) aqueous solutions of CdTe NCs (Fig.12.36) possessing negative charge due to the presence of surface carboxylic groups. Cysteamine-stabilized NCs possess moderate photostability (although they may be stable for years as well being kept in darkness) and attract an interest due to surface amino-functionality and positive surface charge in neutral and slightly acidic media. Other thiol stabilizers are mainly used when some specific functionalities are envisaged.

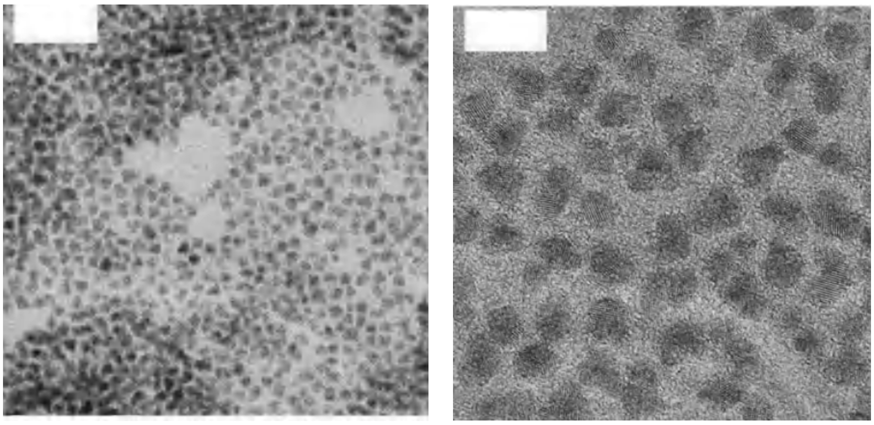


Fig. 12.36. TEM (top) and HRTEM (bottom) images of MPA-capped CdTeNCs, 5.5 nm average size,with PL maximum at 780nm.

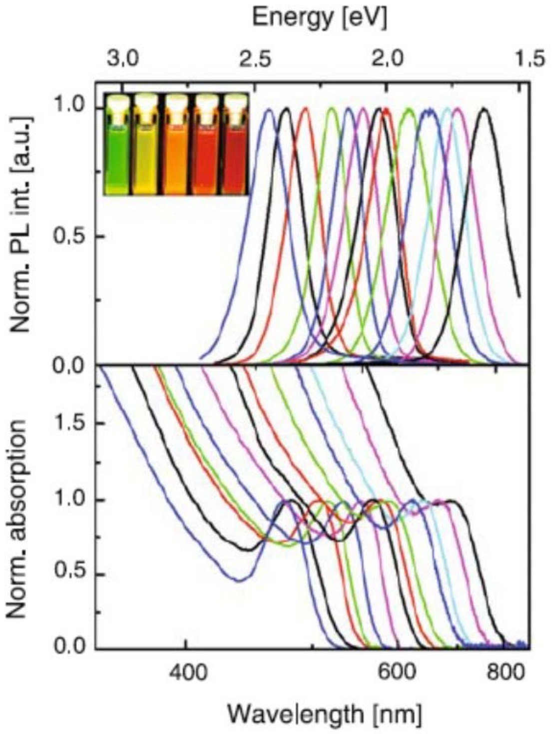


Fig. 12.37. A set of typical PL (top) and absorption (bottom) spectra of TGA-capped and MPA-capped CdTe NCs demonstrates their tuneability over a broad spectral range in the visible and near-infrared. Excitation wavelength is 450 nm. The inset shows a photograph of brightly emitting CdTe NCs of different sizes taken under UV lamp excitation.

Typical absorption and PL spectra of size-selected fractions of TGA- and MPA-capped CdTe NCs are shown in Fig. 12.37. PL spectra of TGA-capped CdTe NCs are tunable in the range of 500–700 nm, while those ofMPA-capped NCs are tunable between 530 and 800 nm. The MPA capping allows for a relatively quick and controllable growth of CdTe NCs up to 6 nm in diameter. The energy gap of bulk CdTe estimated from the absorption measurements at 300K is 1.43 eVor ca. 867 nm. The superior tunability of the absorption over the very broad spectral range is important for the use of thiol-capped CdTe NCs as absorbers in solar cells, for choosing optimal donor–acceptor pairs for FRET-based structures, as well as for tuning an optimal resonance condition in nanoplasmonics systems. Narrow PL spectra in combination with their tuneability and high PL QE are of a special interest for bio-labelling applications, imaging and LEDs based on thiol-capped CdTe NCs.

ZnSe nanocrystals

Currently, a lot of attention is paid to the safe handling of nanometer-sized materials, which demands among others the development of syntheses of colloidal semiconductor NCs from low-toxic materials and the use of environmentally friendly technologies. ZnSe NCs synthesized in water is one of the prominent examples of such kind of nanomaterials. The synthetic procedure for ZnSe NCs is very similar to the synthesis of CdTe NCs. In a typical synthesis $Zn(ClO_4)_2$ $6H_2O$ is dissolved in water in therange of concentrations of 0.02M or less, and an appropriate amount of the thiol stabilizer (1-thioglycerol, TGA or MPA) is added under stirring, followed by adjusting the pH by dropwise addition of 1M solution of NaOH to 6.5 in the case of TGA or MPA capping, or to 11.2–11.8 in the case of TG. The mentioned pH values were experimentally found to be optimal for the synthesis of stable colloids. The solution is deaerated by N_2 bubbling for 1 h. Under stirring, H_2Se gas (generated by the reaction of Al_2Se_3 lumps with an excess amount of 1N H_2SO_4 under N_2 atmosphere or electrochemically is passed through the solution together with a slow nitrogen flow. ZnSe precursors are formed at this stage. The further nucleation and growth of the NCs proceed upon refluxing at 100 ^oC under open-air conditions with a condenser attached growth of the NCs during reflux is indicated by a low-energy shift of the absorption. The PL efficiency of these solutions is negligible and shows mainly a broad trap-emission band (400–600 nm). An additional very weak band-edge emission appears only after long times of reflux. Among the capping agents used a relatively stronger trap-emission is found to be characteristic for TG-capped ZnSe NCs. The synthesis and characterization of this type of white-blue-emitting NCs was reported in details recently. Widely used ratio of the precursors Zn:Se: R–SH is near 1:0.5:2.5, which is similar to the traditional one for the CdTe NCs synthesis. The varying of this ratio towards lower amounts of the stabilizer allowed improvement of the photoluminescence in the case of 1-thioglycerol stabilizer resulting in strong whitish-blue trap-emission, but did not result in any considerable improvement in the cases of MPA and TGA.

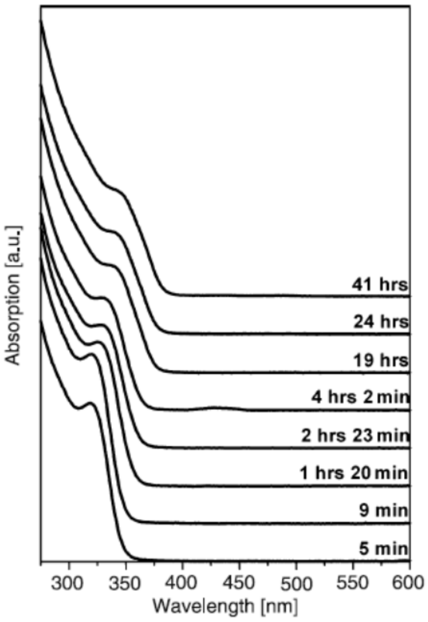


Fig. 12.38. Evolution of the absorption spectra of a crude solution of ZnSe NCs during the synthesis

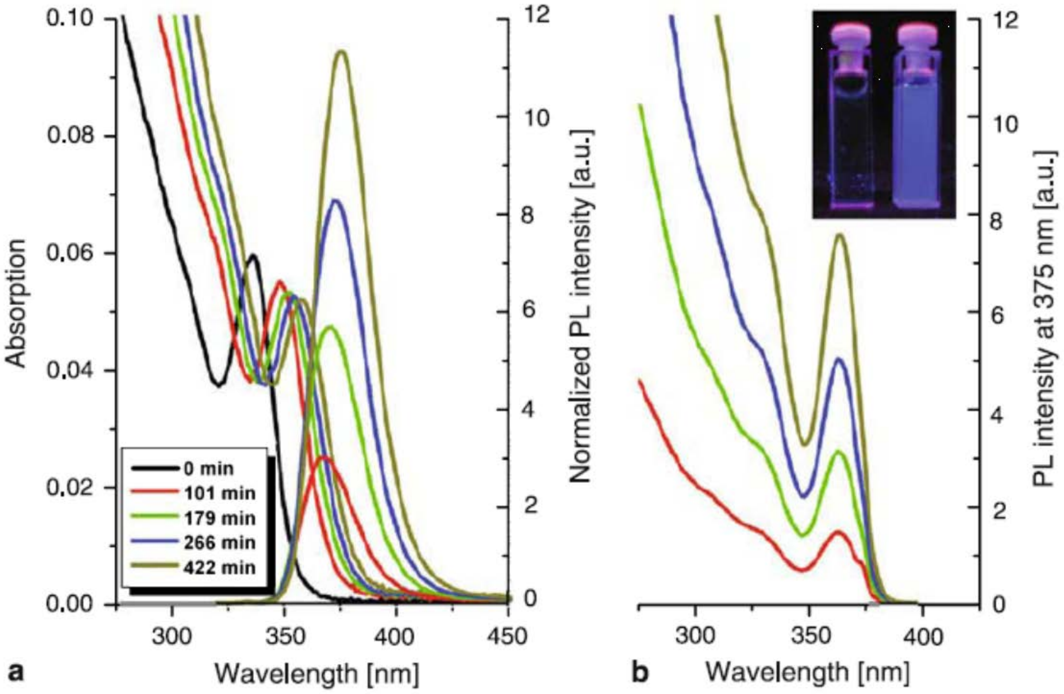


Fig. 12.39. Evolution of the absorption and PL (a) and of the PL excitation spectra (lobservation 375 nm) (b) of ZnSe NCs during illumination. Inset shows a true color fluorescence image (lex 366 nm) of the ZnSe NCs before (left) and after (right) the photo-treatment.

In order to improve the PL properties of the ZnSe NCs (enhancement of the bandedge and suppression of the trap-emission), the colloidal solutions can be irradiatedwith a xenon lamp or alternatively with aUV-lamp. The presence of excessof Zn2ю ions and TGA molecules was found to be necessary. The dependence of the absorption and PL spectra on the duration of the irradiation is shown in Fig. 12.38, 12.39. Under illumination, the PL QE increases from ca. 0.1% being characteristic for the as-prepared solutions up to 10–30% and even 50% if the pH during the treatment maintained to be 6.5 and 11, respectively.

HgTe and $Cd_xHg_{1-x}Te$ nanocrystals

In a typical aqueous synthesis of HgTe NCs metal Hg^{2+} ions react under N_2 atmosphere in aqueous solution with H_2Te gas in the presence of a thiol stabilizer.Adjustment of the pH to an appropriate value (11–12), as well as the judicious choice of absolute and relative concentrations of the reaction components allows thereaction rate, the quality of the product and its PL QE to be efficiently controlled. 1-Thioglycerol was found to be the best for controlling the synthesis at the precursors ratio of Hg:Te:R–SH being 1:0.25:2.5. The HgTe NCs growupon reaction of the precursors at room temperature; the reaction may be stopped by cooling the reaction solutions dow n in ice pad and keeping them later in a fridge. Reaction lasting for approx. 2 h produces highly concentrated solutions of HgTe NCs of approx. 4 nm in size, with a strong (40–50% PL QE) emission. The drawback of this reaction is the broad size distribution of the resulting samples, leading to a broad luminescence peak of the as-prepared NCs covering the spectral region from 800 to 1400 nm with a maximum located at 1080 nm. However, this broad spectral coverage is advantageous for telecommunication applications, as it coincides with the 1.3-mm telecommunications window. To obtain HgTe NC fractions with various mean sizes and narrower size distributions, a standard size-selective precipitation technique can be applied.

Alloyed $Cd_xHg_{1-x}Te$ NCs emitting in the spectral region between 800 and 1100 nm have been synthesized in water through the chemical modification of pre-synthesized thiol-capped CdTe NCs by Hg^{2+} ions. Due to the lower solubility of HgTe in comparison to CdTe in water, the Hg^{2+} ions substitute Cd^{2+} ions at the surface of the nanoparticles forming a $Cd_xHg_{1-x}Te$ alloy in the near surface region, as schematically depicted in Fig. 12.40. A layer of CdTe can then be grown on the surface of $Cd_xHg_{1-x}Te$ NCs by addition of more Cd^{2+} ions reacting with H_2Te gas, and the substitution process can be repeated further leading to an increase in both the NC size and the relative Hg content in the alloyed particles (Fig. 12.40). The substitution reaction occurs with a finite rate and starts at some thermodynamically preferred site. Alloying in bulk $Cd_xHg_{1-x}Te$ systems (MCT) is well-known and is in fact the basis of long-wavelength infrared photodetectors, such that with time Hg^{2+} ions first incorporate on a surface and then into the volume of the particles. The near-surface region of the NCs can therefore be described as a solid solution, alloy or mixed crystal of $Cd_xHg_{1-x}Te$, possibly with a concentration gradient decreasing towards the particle interior. The greater the amount of Hg^{2+} ions added and the longer the time period allowed, the greater the concentration of mercury in the interior of the particles until a real $Cd_xHg_{1-x}Te$ alloy results. The band gap of bulk $Cd_xHg_{1-x}Te$ alloy varies approximately linearly with the composition from 1.6 eV at x=1 (pure CdTe) to 0.3 eV for x=0 (pure HgTe). This has been observed experimentally for the $Cd_xHg_{1-x}Te$ NCs as a red shift in both the absorption and luminescence spectra with increasing Hg content resulting in emission wavelengths ranging from 700 to 1350 nm depending on the composition. PL quantum efficiencies as high as 50% have been measured for the alloy $Cd_xHg_{1-x}Te$ NCs which is similar to that of HgTe NCs.

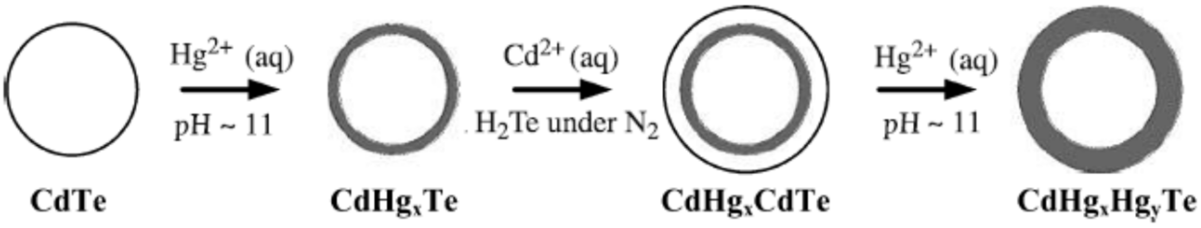
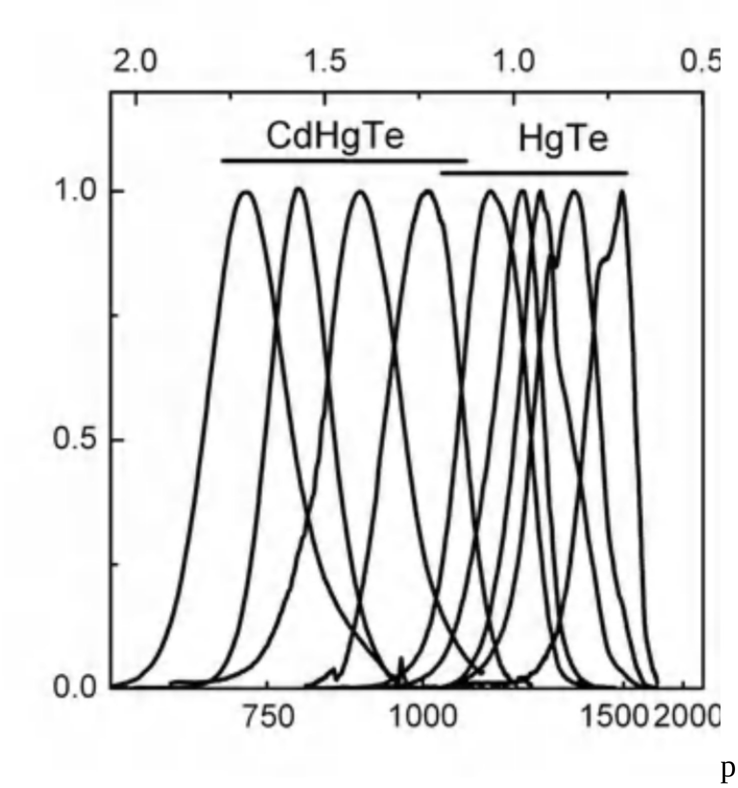


Fig. 12.40 Schematic diagram of the idealized synthetic routes to achieve mixed-phase $Cd_xHg_{1-x}Te$ nanoparticles



p

Fig. 12.41 . Typical photoluminescence spectra of the $Cd_xHg_{1-x}Te$ and HgTe NCs

Fig. 12.41 shows a series of normalized PL spectra of $Cd_xHg_{1-x}Te$ and HgTe NC fractions which cover the spectral region between 700 and 1900 nm. Both $Cd_xHg_{1-x}Te$ and HgTe

NCs are near-infrared-emitting materials whose recognition has grown rapidly all over the world in the last few years.

A potentially very significant application of these NCs is their deployment as optical amplifier media for telecommunications systems based on silica fibre technology which has optimal transmission windows around 1.3 and 1.55 mm, as discussed in several topical reviewsAnother evolving fields with a great potential are applications of near-infared-emitting NCs as fluorescent contrast agents for biomedical imaging in living tissue, electroluminescent devices, photodetectors and photovoltaics for solar energy conversion. It was found that $Cd_xHg_{1-x}Te$ NCs possess exceptionally high stability against photooxidation and degradation in typical biological buffer media like phosphate buffered saline. This fact, together with the strong emission in the biological spectral window (both haemoglobin and water show minimum absorption of light between 800 and 1100 nm) makes this kind of NCs especially attractive for bio-imaging and bio-labelling. Recently, $Cd_xHg_{1-x}Te$ NCs additionally protected by CdS shell were successfully used for the in vivo imaging.

Semiconductor nanocrystal–polymer composites: usingpolymers for nanocrystal processing

In addition of stabilization, hybridizing NCs with polymers can also bring out a number of merits. First of all, NCs are envisaged to inherit the good compatibility, the excellent processing capability, and the high-engineering performance from the polymer stabilizers, which are sought in most technical applications of NCs. In case of biomedical applications, the proper choice of hydrophilic polymer stabilizerssuch as polyethylene glycol (PEG) based (co-)polymers also render semiconductor NCs biocompatibility and reduce the cytotoxicity of the NCs. The merits of using polymers to stabilize NCs can be also foreseen from the intelligent response of polymers to a vast variety of external stimuli and their rich phase separation properties. Using polymers for directing the self-assembly of semiconductor NCs allows for the construction of tailor-made sophisticated arrays of the NCs with the desired properties In a typical preparation procedure, organometallic precursors of semiconductor NCs are first introduced into polymer matrices either through simple mixing or polymerization of monomers in presence of NC precursors or copolymerization of NC precursors and monomers if the precursors are polymerizable. Afterwards, the mixtures of polymers and NC precursors are exposed, for example, to chalcogenide solution or gas. Due to the polymer matrices preventing the infinite crystal growth, only nanometer-sized particles of semiconductors are obtained, thus yielding semiconductor NC–polymer composites. Well-established organometallic chemistry allows for design of NC precursors with good solubility in polymers, eliminating the incompatibility between NCs and polymers that is always problematic for hybridizing as-prepared NCs with polymers and even theirmonomers. Prior to synthesis ofNCs, one can easily process the mixtures of polymers and NC precursors into different device prototypes, bulk materials, thin films, microspheres, and so on, which prevents chemically deterioration of semiconductor NCs during the postprocessing, for instance oxidation.

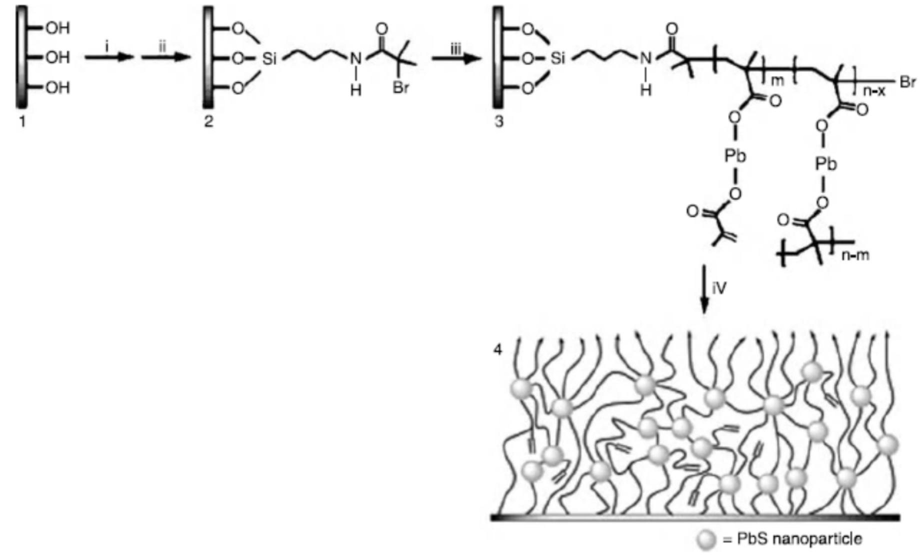


Fig. 12.42. Schematic illustration of the preparation procedure of PbS NC–polymer composites on a silicon wafer by surface-initiated ATRP of Pb-containing monomers, followed by reaction with H2S. Conditions: (i) 3-aminopropyltriethoxysilane, toluene; (ii) 2-bromo-2-methylpropionic acid, DCC, DMAP, CH_2Cl_2; (iii) lead dimethacrylate, p-toluenesulfonyl chloride, Cu(I)Cl, 2,20-bipyridine, DMF; (iv) H_2S gas.

After copolymerization of the Pb-containing monomers and cross-linkers via atom transfer radical polymerization (ATRP), as shown in Fig. 12.42, the resulting films were exposed to H2S gas, yielding PbS NCs embedded within polymer thin films, which should hold great promise in optoelectronics. Kumacheva and coworkers have used polymer microgels as in situ reactors for synthesis of semiconductor NCs, giving rise to micrometer-sized composite particles comprising polymer gels and semiconductor NCs (Fig. 12.43). They also found that the postheating treatment of the resulting composite particles may to some extent narrow the size distribution of the resulting semiconductor NCs and thus tune their photoluminescence properties.

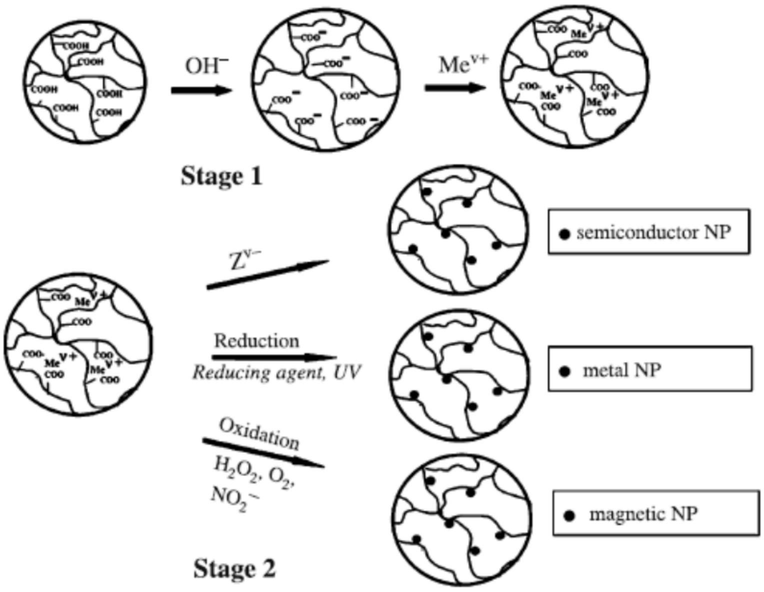


Fig. 12.43 Schematic of microgel-based synthesis of semiconductor, metal, and magnetic NCs.

It was fabricating transparent polymer composites of organic ZnS-overcoated CdSe NCs with little deterioration of the photoluminescence properties of the NCs by polymerization of laurylmethacrylate in the presence of the NCs. The long alkyl side branches of polylaurylmethacrylate (PLMA) efficiently prevented the NCs from aggregation and phase separation from the polymer matrices during polymerization A prototype of light-emitting device using the resulting composites was demonstrated (Fig. 12.44).

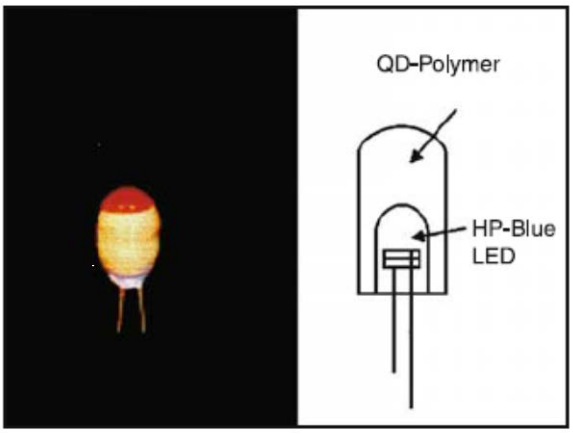


Fig. 12.44. A down-conversion light-emitting device derived from ZnS-overcoated CdSe NC–PLMA composites emitting at 590 nm. GaN light-emitting diode (LED) was used as the excitation source at 425 nm.

Suspension and emulsion polymerization in the presence of semiconductor NCs allowsculpture of semiconductor NC–polymer composites into microspheres, which should be very useful as fluorescence markers for biological detection. Nonetheless, the strategy of polymerization in the presence of semiconductor NCs has as a whole, the large problem that free radicals usually quench the photoluminescence of semiconductor NCs, lowering the optical properties of the resulting composites. It was demonstrated that the appropriate choice of initiators iscrucial to remain a better photoluminescence efficiency of the NCs when they were embedded in polymer matrices, AIBN may preserve the photoluminescence of the NCs embedded in polystyrene (PS) microspheres, thus leading to chemically robust fluorescent beads (Fig. 12.45).

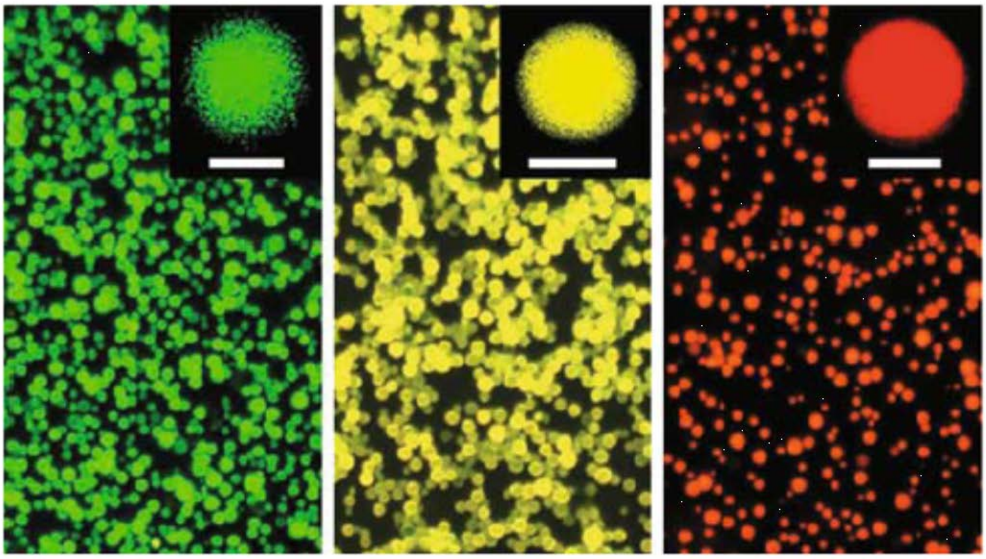


Fig. 12.45. Fluorescence images of PS beads loaded with green (left), yellow (middle), and red CdTe NCs (right). The insets are the corresponding confocal fluorescence images. The scale bars correspond to 2 mm.

Layer-by-layer assembly

Based on the strong interaction between NCs and polymers, alternatively depositing them on a substrate has paved a controlled way to tailor the chemical composition, thickness, and the vertical structure of NC–polymer composite films with a nanometer precision. The interaction commonly used for LbL growing semiconductor NC–polymer composites is electrostatic interaction. The LbL strategy can work well on different substrates, such as silica, glass, and wood, with different curvatures. (Fig. 12.46).

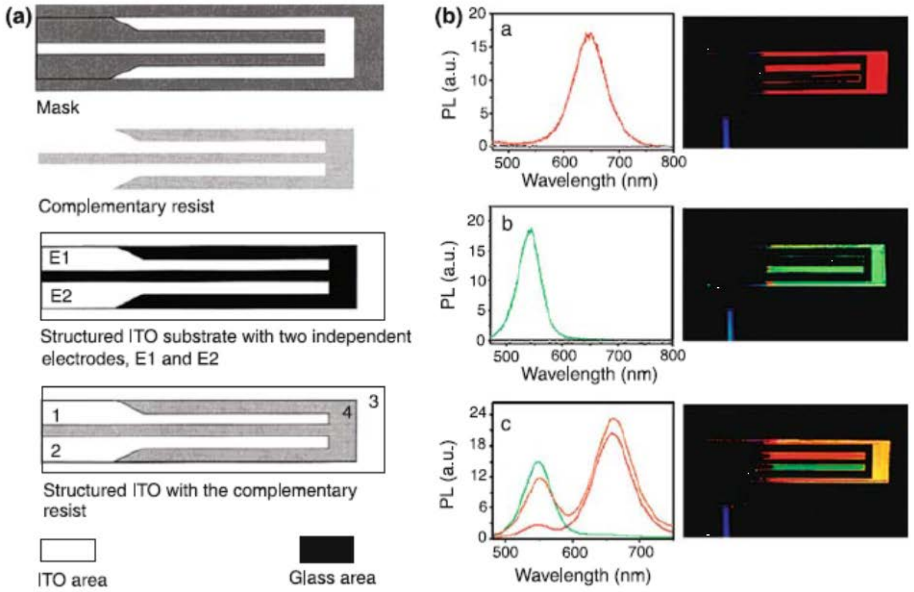


Fig. 12.46 a) Lateral structures on patterned ITO substrates. B) Photographs of the lateral structures formed by LbL deposition of green- and red-emission CdTe NCs and PDDA

12.10 SUMMARY OF POSSIBLE APPLICATIONS OF SEMICONDUCTOR NANOPARTICLES

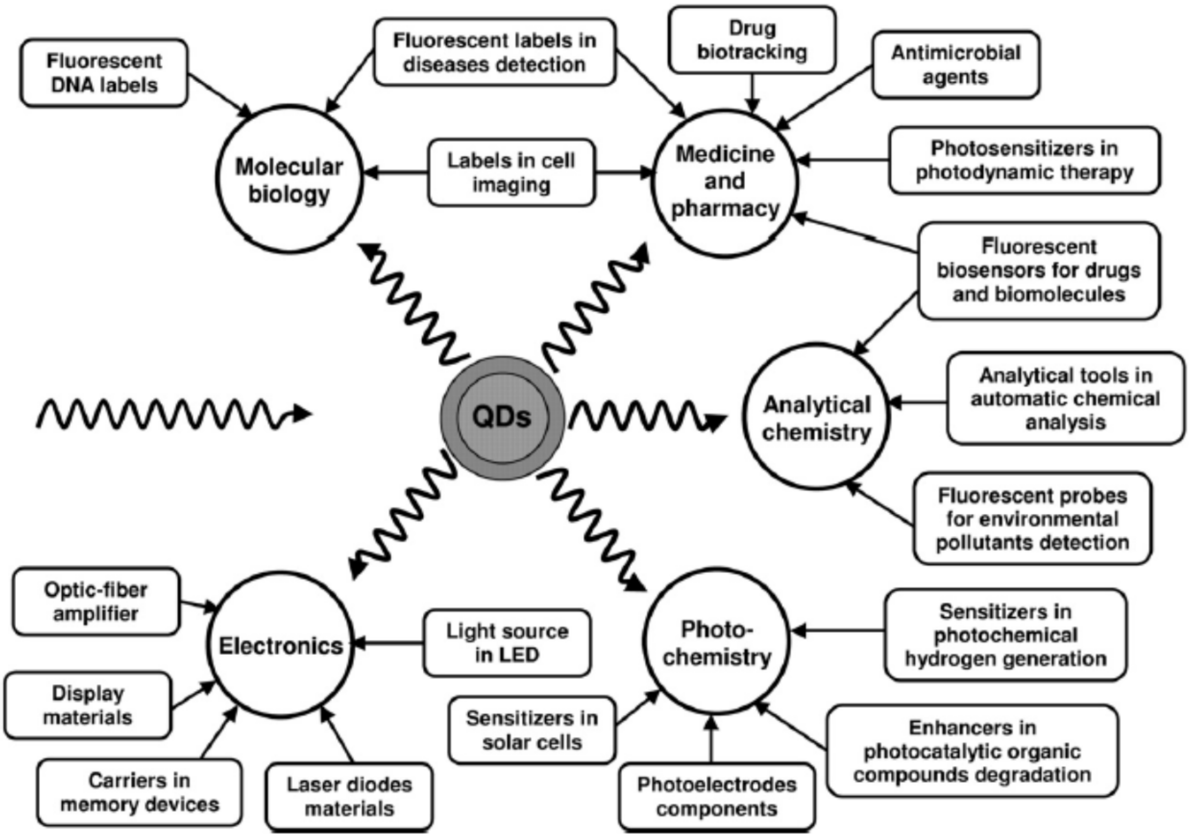


Fig. 12.47 The versatility of NP application in different branches of science.

Due to their excellent photophysical, magnetic, electronic and biological properties semiconductor Nanoparticles (NPs) are widely applied in medical field for biotracking of drug molecules, bioimaging biosensing , disease detection , photodynamic therapy, antimicrobial prevention and molecular biology.. In addition, NPs are also valuable tools in electronics and optoelectronics, photochemistry and modern analytical chemistry. The versatility of NPs application is presented in Fig. 12.47.

12.11 FORMATION OF METAL OXIDES NANOPARTICLES

Aqueous Sol-Gel Chemistry

The aqueous sol-gel process can shortly be defined as the conversion of a precursor solution into an inorganic solid via inorganic polymerization reactions induced by water. In general, the precursor or starting compound is either an inorganic (no carbon) metal salt (chloride, nitrate, sulfate, ...) or a metal organic compound such as an alkoxide. Metal alkoxides are the most widely used precursors, because they react readily with water and are known for many metals. Some alkoxides, which are widely used in industry, are commercially available at low cost (Si, Ti, Al, Zr), whereas other ones are hardly available, or only at very high costs (Mn, Fe, Co, Ni, Cu, Y, Nb, Ta, ...). In comparison to the preparation of silicates from silicon alkoxides, sol-gel processing of transition metal oxides has much less been studied, mainly due to the high reactivity of transition metal alkoxides. In general, the sol-gel process consists of the following steps (Figure 12.48): i) Preparation of a homogeneous solution either by dissolution of metal organic precursors in an organic solvent that is miscible with water, or by dissolution of inorganic salts in water; ii) conversion of the homogeneous solution into a sol by treatment with a suitable reagent (generally water with or without any acid/base); iii) aging; iv) shaping; and v) thermal treatment/ sintering. The first step in a sol-gel reaction is the formation of an inorganic polymer by hydrolysis and condensation reactions, i.e., the transformation of the molecular precursor into a highly crosslinked solid. Hydrolysis leads to a sol, a dispersion of colloidal particles in a liquid, and further condensation results in a gel, an interconnected, rigid and porous inorganic network enclosing a continuous liquid phase. This transformation is called the sol-gel transition. There are two possibilities to dry the gels. Upon removal of the pore liquid under hypercritical conditions, the network does not collapse and aerogels are produced. When the gel is dried under ambient conditions, shrinkage of the pores occurs, yielding a xerogel. One of the highly attractive features of the sol-gel process is the possibility to shape the material into anydesired form such as monoliths, films, fibers, and monosized powders, and subsequently to convert it into a ceramic material by heat treatment.

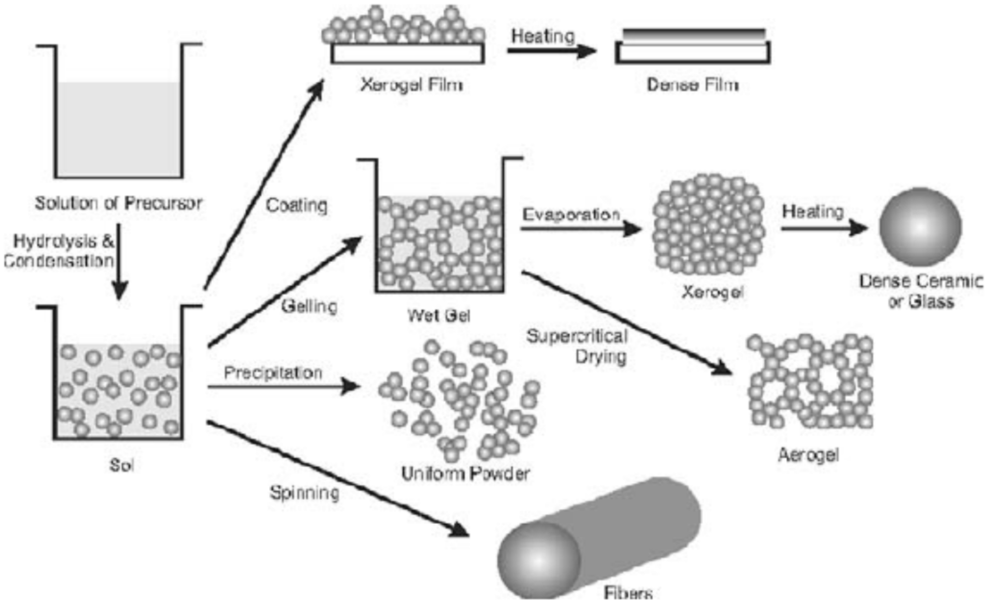


Fig. 12.48. Various steps in the sol-gel process to control the final morphology of the Product

Hot-Injection Method

Hot-injection method involves the injection of a room-temperature (“cold”) solution of precursor molecules into a hot solvent in the presence of surfactants (Figure 12.49).

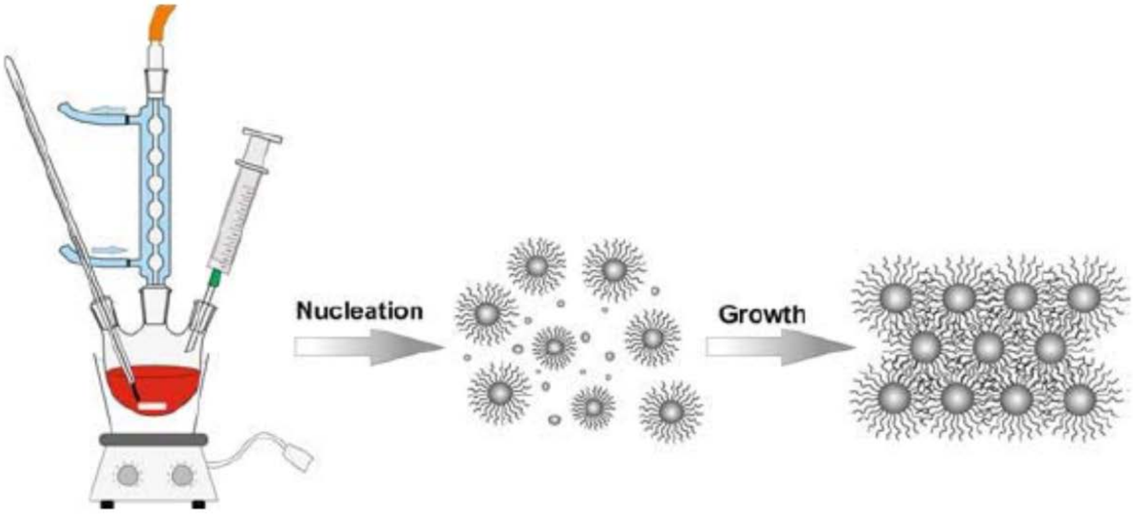


Fig. 12.49. Experimental set-up and reaction scheme for the hot-injection method.

The fast injection of the precursor induces a high degree of supersaturation, resulting in a short burst of nucleation. During the nucleation process the precursor concentration in the solution decreases abruptly. The drop in temperature, due to the injection of the “cold” reactants, and the low concentration of unreacted remaining precursor prevent any further nucleation events. In a next step, the temperature is carefully increased to a value which allows the slow growth of the nuclei to larger nanoparticles, however still suppresses further nucleation. The separation of nucleation and growth is a prerequisite to synthesizing monodispersed nanoparticles [99], and therefore the hot-injection method generally leads to nanoparticles with a narrow size distribution of 10%. If size distributions narrower than 5% are required, then a size-selection process has to be applied.

Heating-Up Method

In the so-called “heating-up” method the reaction solution is prepared at low temperature. Subsequent heating initiates the crystallization process, which finally leads to the formation of nanocrystals. The heating-up method is particularly advantageous for large-scale production, because of its simplicity. Nevertheless, the size uniformity that can be achieved by this approach is comparable to the best results obtained from the hot-injection method. Figure 12.50 gives an overview of selected magnetic nanoparticles synthesized by this technique.

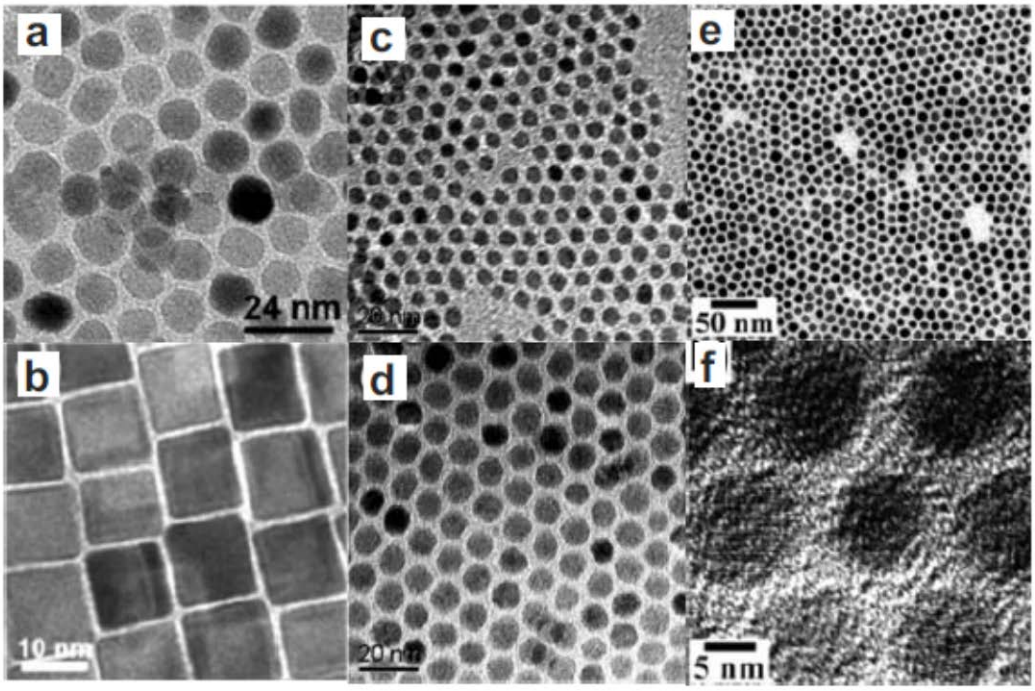


Fig. 12.50. a–f TEM images of magnetic nanoparticles: **a** spherical 12 nm sized $MnFe_2O_4$, **b** cubelike $MnFe_2O_4$, **c** spherical 5 nm-sized $MnFe_2O_4$, **d** spherical 10 nm-sized $MnFe_2O_4$, **e** spherical 10 nm-sized Mn_3O_4, and **f** HRTEM image of the nanoparticles displayed in **e**.

12.12. PROPERTIES AND APPLICATIONS OF METAL OXIDES NANOPARTICLES

Among all the functional materials to be synthesized on the nanoscale, metal oxides are particularly attractive candidates, from a scientific as well as from a technological point of view. The unique characteristics of metal oxides make them the most diverse class of materials, with properties covering almost all aspects of materials science and solid state physics. The great variety of structures and properties made them the primary target in solid state chemistry and still gives inspiration for designing new materials. This exceptionally broad spectrum of properties makes (bulk) metal oxides a vital constituent in technological applications like gas sensing, medical science, electronics, ceramics, energy conversion and storage, and surface coatings, just to mention a few (Figure 12.51).

Fig. 12.51. Selected applications of metal oxides

Magnetic Metal Oxides

Iron oxides and ferrites such as Fe_2O_3, Fe_3O_4, $CoFe_2O_3$ have a net magnetic moment due to the antiparallel arrangement of the moment of the atoms in the octahedral and tetrahedral sites of the inverse spinel structure (e.g. Fe^{2+}, Fe^{3+}, Co^{2+}). Due to their small size (usually <30–40 nm) magnetic oxide nanoparticles are generally single-domain and show a superparamagnetic behavior depending on the temperature, their size and their composition.

The magnetization of ferromagnetic and ferrimagnetic nanoparticles above a given temperature, called blocking temperature (TB), displays no hysteresis as a function of increasing and decreasing magnetic field (M-H). This is caused by the thermal energy kT that, in the superparamagnetic regime, is larger than the energy required to change the direction of the magnetic moment of a particle. Therefore, above TB the nanoparticles do not retain any magnetization in the absence of an externally applied magnetic field. Syntheses performed in benzyl alcohol from iron acetylacetonate lead to magnetite nanoparticles with tunable size from 12 to 25 nm . The magnetization at saturation is comparable to the one of the bulk magnetite (Figure 12.52), highlighting the high quality and purity of the as-synthesized nanocrystals

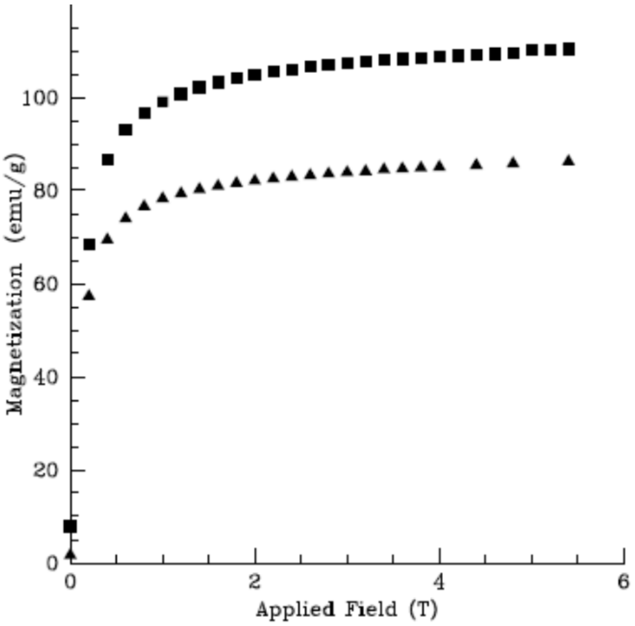


Fig. 12.52 Magnetization curves at 5 K of two samples of magnetite nanoparticles synthesized in benzyl alcohol having average sizes of 12 (▲) and 25 nm (■), respectively

Photoluminescent Metal Oxides *(Rare Earth-Based Nanostructures)*

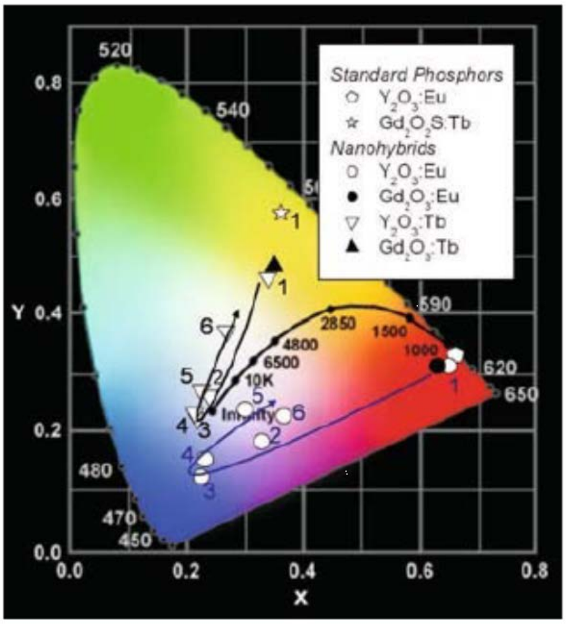


Fig. 12.53 CIE chromaticity diagram (1931) showing the (x,y) color coordinates of the roomtemperature emission of yttrium- and gadolinium based hybrid materials synthesized in benzyl alcohol doped with Tb $^{3+}$ and Eu $^{3+}$ ions and excited at different excitation wavelengths, 264-280 nm (1), 330 nm (2), 350 nm (3), 375 nm (4), 395 nm (5), and 420 nm (6). The black curve represents the color of the black body emission at different T (K)

Another interesting feature of these hybrid materials is due to their emission chromaticity that can be simply controlled by varying the excitation wavelength. For example, the mission color of the Eu $^{3+}$-doped samples may be tuned from the red spectral region to the white one crossing the purplish-pink, bluish-purple and the purple spectral areas of the diagram when the excitation wavelength varies from 280 to 420 nm (Figure 12.53). It was thus possible to precisely tune the emission color through the CIE chromaticity diagram by varying the excitation wavelength and without loosing the high radiance values.

Sensor Devices

Gas sensor devices having sensing layers made of metal oxide nanoparticles are generally fabricated by simple drop/spin coating or screen-printing of a nanoparticle suspension onto a patterned substrate equipped with electric contacts on the front and a heater on the back. Schematic representation of such a device is given in Figure 12,54. The deposited layer is then thermally treated at moderate temperature in order to eliminate the binder, and to form a porous structure attached to the substrate.

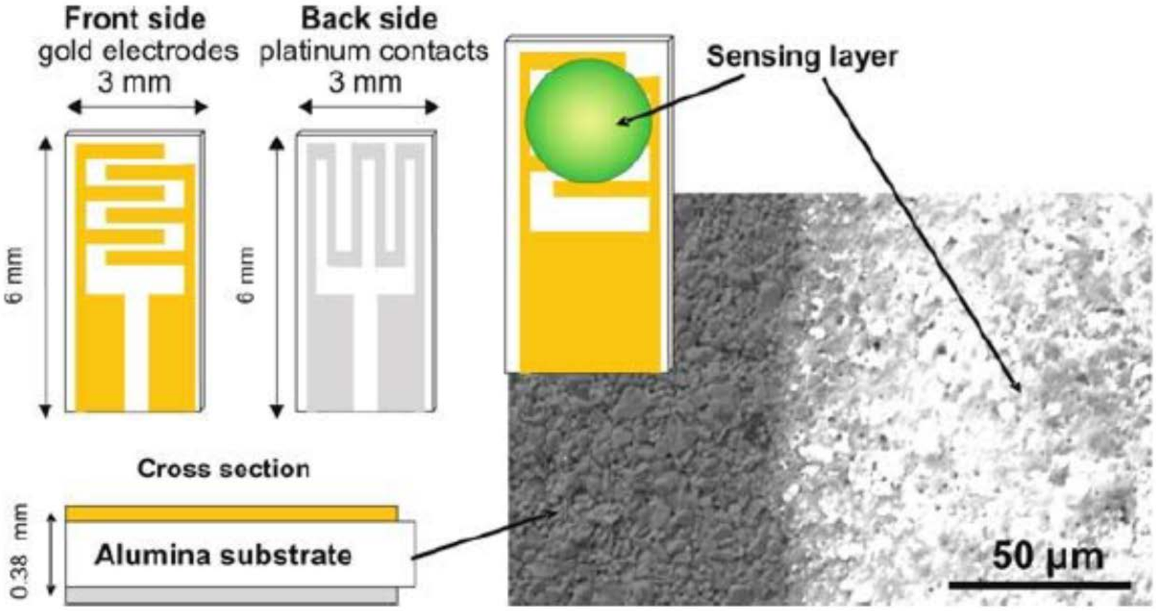


Fig. 12.54. Schematic representation of the MOS sensor device and micrograph of the In2O3 metal oxide sensing layer deposited on the ceramic substrate

The principles of sensing mechanism of resistive gas sensors are reported in many books and reviews and will not be considered here in details. In brief, the response of a n-type metal oxide semiconductor to the presence of a target gas relies on the surface reactions which occur between adsorbed oxygen species and the probed gas. Oxygen adsorbed on the surface traps free electrons because of its high electron affinity, forming a potential barrier at the grain boundaries. This potential barrier restricts the flow of electrons, causing the electric resistance to increase. When the sensor is exposed to an atmosphere containing reducing gases, e.g. hydrocarbons, CO, ethanol, etc., the gas molecules adsorb on the surface and react with active oxygen species. Reactions with surface oxygen species vary depending on the temperature and the reactivity of the sensing material. In any case, this lowers the potential barrier allowing electrons to flow more easily, thereby reducing the electrical resistance. On the other hand, the adsorption of oxidizing gases such as NO_2 and ozone on metal oxide surfaces increases the surface resistance. The opposite is true for p-type oxides, where electron exchange due to the gas interaction leads either to a reduction (reducing gas) or an increase (oxidizing gas) in electron holes in the valence band

The response of the In_2O_3 nanoparticles-based sensor to trace levels of NO_2 in function of the operating temperature is reported in Figure 12.55. The response to NO2 increases with increasing temperature, showing a maximum of sensitivity $S = I_{air}/I_{2ppmNO2} = 100$ around 200°C (Figure 12.55a). The dynamic response of the sensor at 250°C shows that the output signals essentially recover to the initial level, indicating the reversibility of the interactions between the sensing elements and NO_2, with fast (less than 60 sec.) response time (Figure 12,55b). The estimated detection limit of this sensor is about 1 ppb of NO_2 in dry air

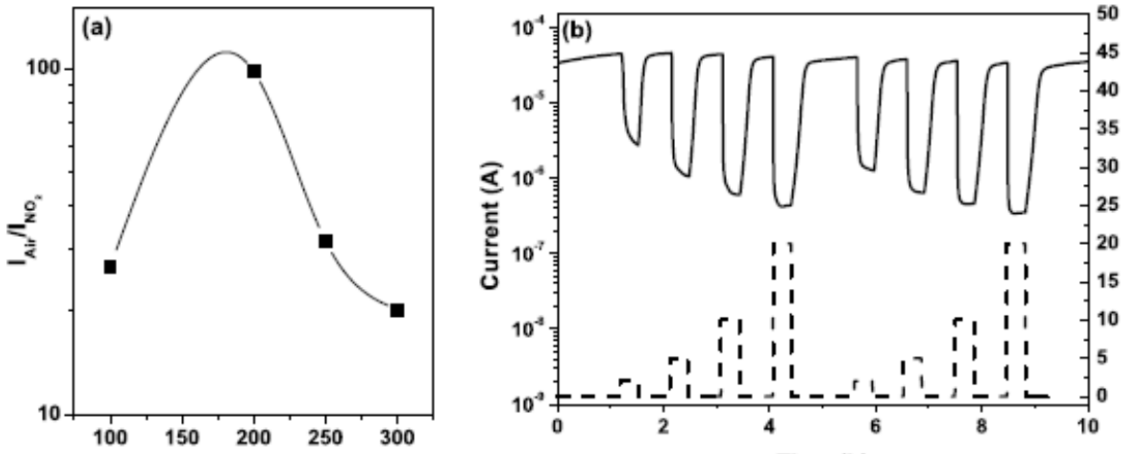


Fig. 12.55. a Gas responses as a function of working temperature to 2 ppm of NO_2, **b** dynamic response (*solid line*) of the $In_2 O_3$-based sensor to different NO_2 concentrations (*dashed line*) in dry air

Biomedical Applications

Nanometer-sized nanoparticles are promising for diverse biomedical applications as a result of their physical (especially optical and magnetic) properties.As a typical example, semiconductor nanoparticles are currently used as fluorescence probes for labeling and optical imaging of biological tissues. Magnetic nanoparticles find their application in magnetic resonance imaging (MRI) as contrast-enhancement agents. They can bring advantages over currently used MRI contrast agents (e.g. gadolinium complexes) such as better contrast and selectivity to a particular tissue or cell type.

CHAPTER 13 COLOIDAL SOLUTIONS OF QUANTUM DOTS OF A^{IV} SEMICONDUCTORS

13.1 INTRODUCTION

Quantum dots (QDs) are semiconductor nanoparticles in which “quantum confinement” effect is observed. Quantum confinement of electrons in semiconductor nanoparticles is observed when a crystallite radius less than the Bohr exciton radius (5 nm for silicon) and results in specific optical and electronic properties - size dependent bandgaps and thus a Stokes shift with high quantum yield photoluminescence.

Structurally, the central part of QDs consist of a core composed of a variety of substances such as A^{IV}, A^{2}B^6, A^{3}B^5 and more complex semiconductors which can be amorphous or crystalline, doped and undoped. The core are covered by other semiconductor shell and stabilizing layer. Depending on functional application of QDs this layer can be hydrophilic or hydrophobic. The linkers and ligands can be attached to these layers allowing formation of different complex. The examples of possible materials for core, doping ions, shell, stabilizing layer, linkers and ligands are presented in Fig.13. 1.

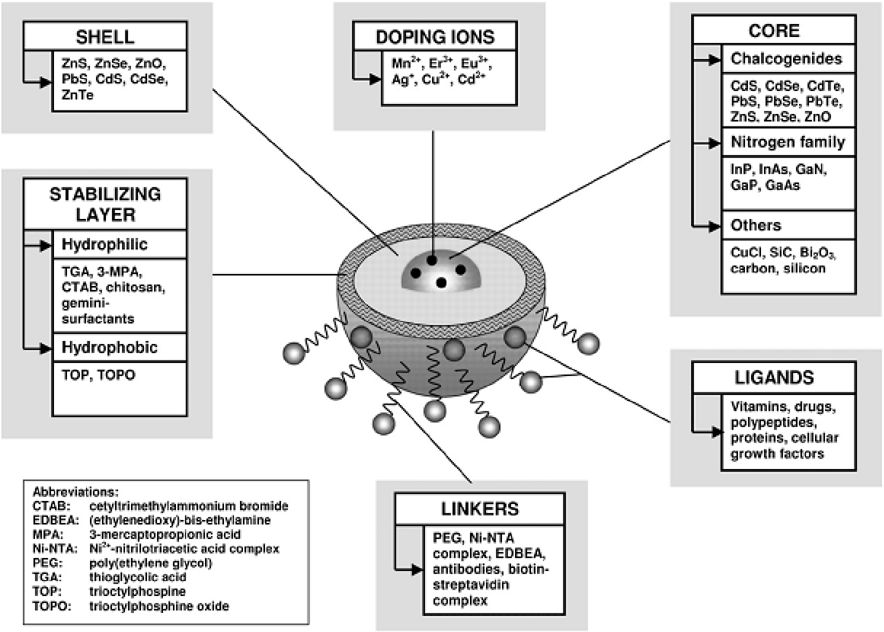


Fig. 13.1. Scheme of quantum dot structure

Due to their specific photophysical, magnetic, electronic and biological properties quantum dots are widely applied in microelectronics and nanolectronics (light emission diodes, memory devices, displays), photochemistry (hydrogen generation, sensitizes, photoelectrodes), analytical chemistry (chemical analysis), molecular biology and medicine (biosensors, fluorescent labels, photodynamic therapy, drug biotracking), etc.[1-4].

In the last few decades, the emergence of nanotechnology gave rise to exciting developments in the field of cell biology. Indeed, the ability to control matter at the nanoscale length paves the way for the sensing of biological systems at the level of single molecules by optical or electrical detection. The implementation of QDs for use in cell labeling has been one of the fastest growing areas in this field. The reason for this intense research activity, in particular fluorescence labeling, lies in the considerable advantages of semiconductor QDs over the classically used organic dye molecules or fluorescent proteins. Numerous reviews have been devoted to these QDs [1-9] including the method of QDs formation and possible application to bioimaging. Here we review the last achievements in the branch of Si and SiC QDs formation by electrochemical etching, their surface functionalization, analysis of luminescent properties of QDs colloidal solutions, and different bio-medical application.

13.2 PROPERTIES AND ADVANTAGES OF COLLOIDAL SOLUTIONS WITH QDS

The interest in colloidal solution with QDs lies in their fluorescence properties. Among the advantages of QDs, the principal one is the higher resistance of QDs to light-induced reactions such as photooxidation. This photochemical inertness known as the photobleaching phenomenon prevents the degradation of luminescence properties with time. Another important feature of QDs is their size and composition-induced tunable emission spanning from the ultraviolet to the near infrared (UV to NIR) with a narrow emission spectra and high quantum yield. The conjugation of a narrow emission with a large effective Stokes shift [8,10], a constant emission wavelength no matter what excitation wavelength is used [11], and a broad absorption spectrum are key parameters to achieve multiplexed imaging as the fluorescence signal of each QD can be readily separated and individually analyzed using a single excitation source. This multicolor imaging potentiality is particularly interesting for the monitoring of intracellular processes [3, 12]. As the last advantage, it is also important to note that QDs have orders of magnitude larger cross sections for two-photon excitation compared to the organic chromophore. Thanks to this property, infrared excitation in the transparency window of tissues (700–900nm) can be used with visible radiation to perform fluorescence detection.

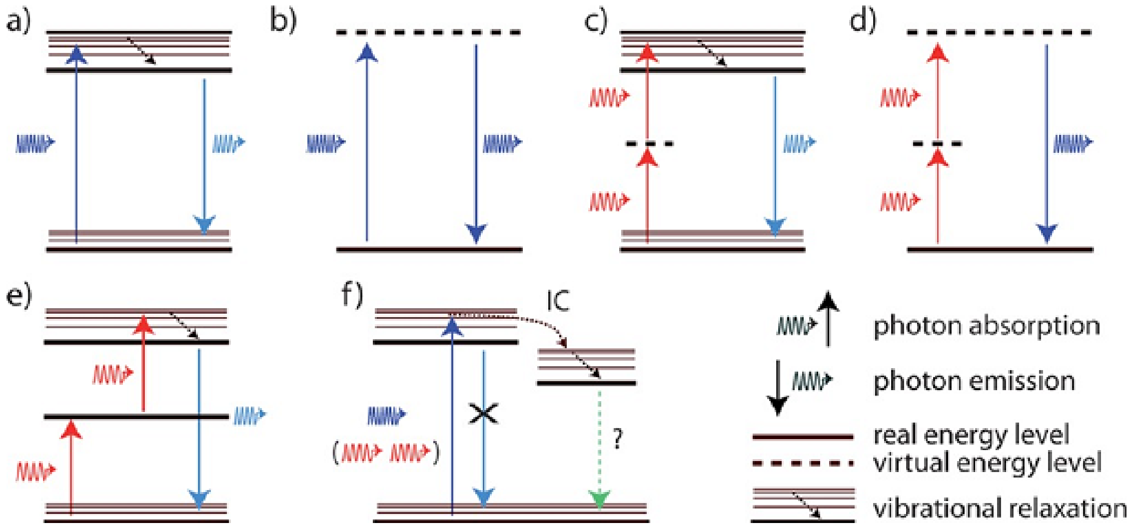


Fig. 13.2. (a) One-photon excited fluorescence. (b) Linear scattering. (c) Two-photon-excited fluorescence. (d) Second harmonic generation. (e) Sequential absorption and upconverted emission. (f) Principle of molecular bleaching showing intersystem crossing upon one- or two-photon absorption (from [7]).

Besides fluorescence (one photon and two-photon excited) we can note other effects like linear scattering, second harmonic generation, sequential absorption and upconverted emission. The schemes in Fig. 2 summarize emission effects in QDs that can be applied to bio-medical applications (see detailed explanation in [7]).

QDs present also two important drawbacks. The first one is that most QDs, as synthesized, have hydrophobic organic ligands coating their surface [13]. To make the QDs water soluble, their organophilic surface species have to be exchanged with more polar species. A complex multilayer [14] ligand shell has then to be realized, which leads in some cases of nonspecific optical absorption. The second and the major drawback concerns cytotoxicity. For example, the widely used II–VI semiconductor QDs (e.g. CdSe) were found to be cytotoxic through the release of free cadmium ions [15, 16]. As the result, these QDs strongly influence biological cell function and their application in efficient cell imaging is limited. Therefore, a protective shell (ZnS or a polymer, for example) must be systematically added to prevent cell death when using these types of QDs. To completely avoid this problem, one has to use QDs on the basis of more benign materials such as group IV semiconductors, first of all Si and SiC QDs.

13.3 TECHNOLOGY OF COLLOIDAL SOLUTION OF SI QDS

Different approaches have been proposed and described in literature to create Si nanoparticles and QD_S. The isolated particles were obtained from a gas-phase preparation using silanes or other Si sources via thermal decomposition, thermal annealing, reactive sputtering, thermal vaporization, microwave plasma, gas-evaporation, chemical vapour deposition [1,9,17-20].

"Wet chemistry" route to Si QDs is typically involve the reaction of Zintl salts (KSi/NaSi) or reductive organometallic Si compounds with $SiCl_4$ or NH_4Br [21,22]. Chemical reduction of silicon halogenides (SiX_4; X = Cl or Br) by $LiAlH_4$ in water-free reverse-micelle solutions gives Si QDs with narrow size distribution [23]. However, comparing to chemical route, electrochemical anodization is much more reproducible and simple technique for the fabrication of Si QDs emitting light in a large spectral range at room temperature under photoexcitation. Visible light emitting colloidal Si QDs are extensively studied [1]. Otherwise UV emitting Si QDs were only mentioned in a few papers [24-26], because it is relatively difficult to control fabrication of stable and homogeneously distributed Si NPs with dimensions less than 1.5 nm having the energy band gaps larger than 3 eV.

Commonly the anodization to prepare Si QDs as well as porous silicon (PS) takes place in the mixtures of HF (able to transform silicon into soluble H_2SiF_6 under anodization) and organic solvent (commonly ethanol) decreasing the wetting angle and thus allowing the penetration of electrolyte inside pores of PS. An experimental protocol for the preparation of Si QDs with desired emission wavelength (from NIR to UV) was described in details in the works [26, 27]. At the first stage, highly porous (>90 %) 300 μm thick PS layer was prepared by anodization (current density j = 55 mA/cm^2) of p-type (1-10 Ω cm) boron-doped (100)-oriented Si wafers with a backside Al ohmic contact in 1:1 (v:v) mixture of 48 % aqueous HF and absolute ethanol during 2 hours. To remove all HF traces the layer was washed several times in absolute ethanol and naturally dried, which resulted in its self-transformation into a strongly hydrogenated PS micropowder with about 6.5% mass content of H [27]. Finally the micropowder was mechanically grinded and dispersed in absolute ethanol, giving Si QDs saturated colloidal solution with 5 mg/ml concentration.

Transmission Electron Microscopy (TEM) pictures shown in Fig. 3 give a general view of quasi-spherical Si particles constituting the nanopowder suspension into the initial colloidal solution. As one can see, the general size distribution of the shown particles is relatively large. In particular, the initial nanopowder consists of large (>10 nm) porous particles

(Fig3 1-b)) and numerous small denser (<10 nm) particles (Fig. 3-c)). Total absence of clearly defined electron diffraction features indicates an amorphous-like state of the studied particles [27]. Clear and dark spots of the large porous particles (Fig. 3-b) correspond to the pits and solid Si phase, respectively. In general, their shape was already described in terms of fractal geometry [28]. As for the small (<10 nm) particles (Fig. 3-c)), they are well known to exhibit highly efficient luminescence in a large spectral range due to quantum and spatial confinements of photogenerated charge carriers and these particles can be considered as SiQDs.

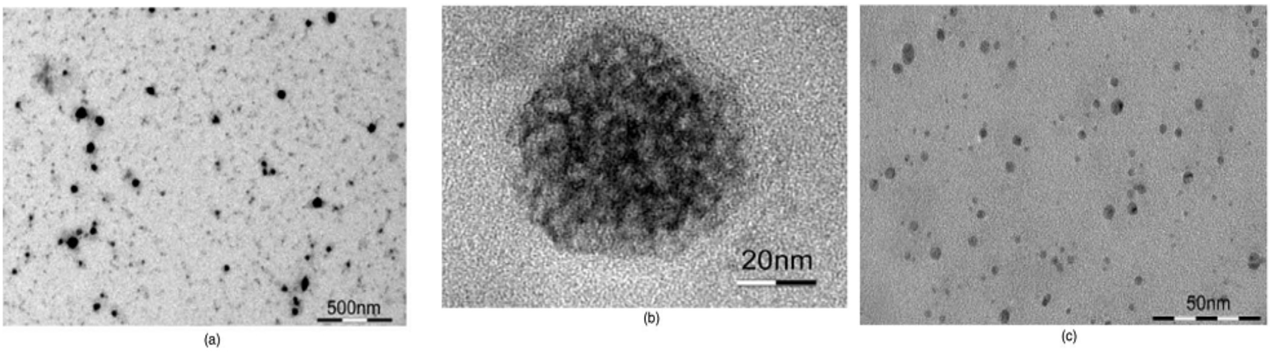


Fig.13.3 TEM pictures of Si NPs constituting the initial strongly hydrogenated PS nanopowder: (a) general view; (b) large porous NPs; (c) small dense Si NPs with dimensions less than 5 nm (from [26]).

Fig. 4 shows normalized photoluminescence (PL) spectra of the Si QDs in ethanol solutions [26]. First, the PL signal of the initial saturated solution is represented by spectrum 1. It is very similar to a usual PL spectrum of a red emitting PS layer. The spectra blueshifting in comparison to the initial PL is observed after sedimentation in the dark during 1 day (curve 2). Associating both PL spectra to particular size distributions of SiQDs and taking into account that the emission energy for the smaller QDs is higher, the observed sedimentation-induced evolution of the PL signal can be explained by the removal of the largest submicron PS QDs from the ethanol solution. Thus, the observed green-blue PL maxima of spectrum 2 may correspond to luminescence of the smaller porous QDs as those shown in Fig. 3b, for example. Further centrifugation leads to almost complete elimination of the QDs with dimensions larger than 2 nm from the colloidal solution. The PL spectrum 4 of this doubly centrifuged solution has a main maximum at 3.5 eV and a set of other UV peaks at the high energy wing situated in the spectral region between 3.5 and 4.5 eV.

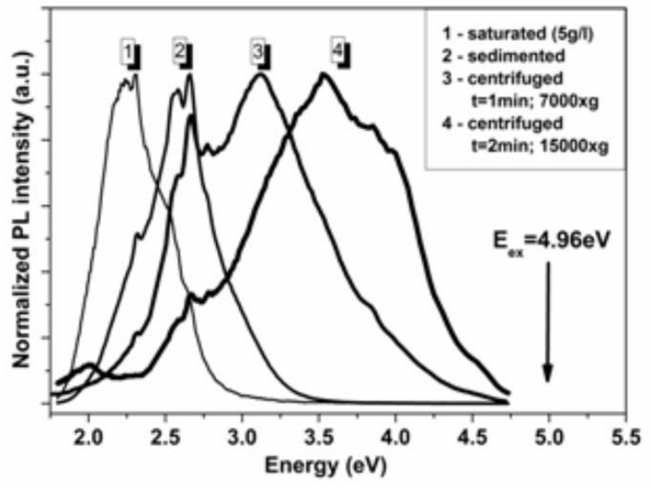


Fig.4 PL spectrum of the initial PS nanopowder in ethanol solution and its evolution upon sedimentation and moderate centrifugation (from[26])

In [29,30] we proposed a new method allowing more precise size selection of colloidal Si QDs. This approach is based on the use of meso-porous Si (meso-PS) free standing membranes as QDs filters (Fig.5). In particular, due to auto-filtration phenomenon an efficient filtering was achieved and size quantization of the amorphous Si QDs was brought to the fore. Via variation of the porosity and thickness of meso-porous layer it is possible to obtain PL spectra with maxima from 1.7 to 3.4 eV. Energy positions of these features are summarized in Fig. 6. The lowest (in terms of energy) feature is designated as *n* and all others as *n+i*, where *i* is the index number. A remarkable coincidence of the data for each PL peak number can be stated. In particular, one can deduce from Fig.6 that the PL energy values seem to vary discretely which is especially interesting taking into account total independence of the picked out data.

Since each spectral energy value assigned to a specific QD dimension, Fig. 6b shows this correlation according to the model of Delerue et al. [35] for which:

$$Eg\ (d) = 1.167 + 3.73 * d^{-1.39}\ , \qquad (1)$$

where $E_g(d)$ is the energy band gap of a spherical-like QD with diameter d. The QD diameter was estimated from Eq. (1) for each spectral PL feature n+i characterized by its proper energy value $E_g(d)$. Thus, in frame of experimental errors, Fig. 6-b shows that dimensions of the Si QDs undergone filtering vary discretely with a radius quantum equal to 0.12 nm. Existing of this quantum may correspond to step-like increasing of Si QDs radius on one new shell at the surface of Si QDs.

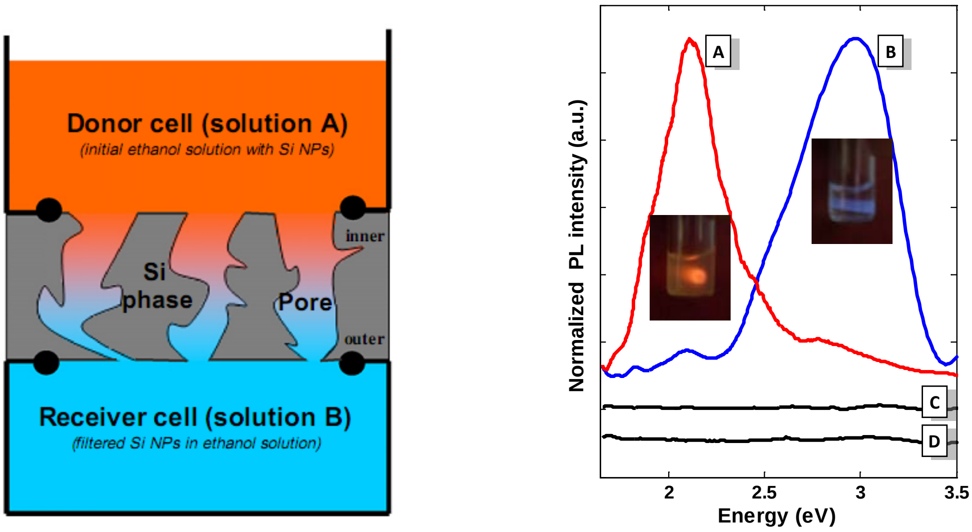


Fig.5 Scheme of a double-tank homemade hermetic filtering set-up with upper (donor) cell and bottom (receiver) cell (a) and (b) PL spectra with corresponding photos of initial (curve A) and filtered (curve B) ethanol solutions with dispersed Si QDs. PL signals from pure absolute ethanol and from pure ethanol without Si NPs passed through the meso-PS membrane are represented by non-normalized spectra C and D, respectively (from [29]).

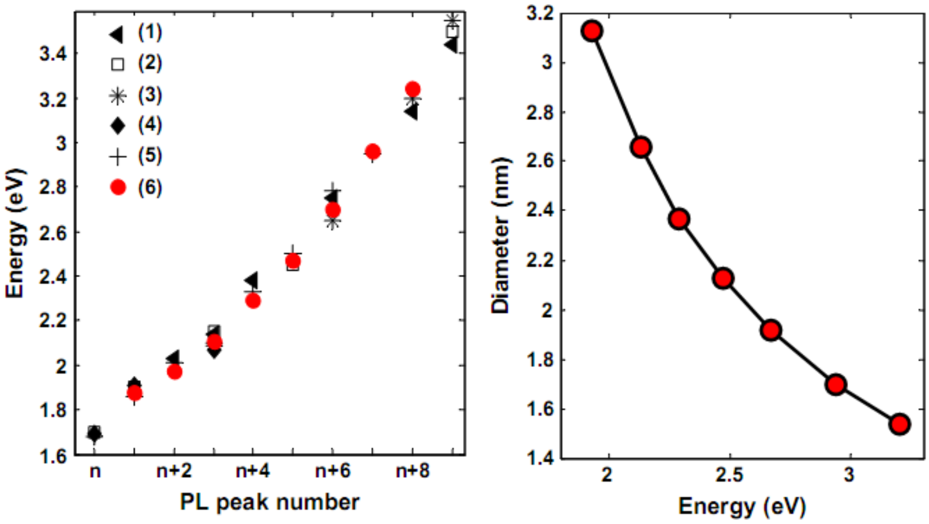


Fig.6(a): spectral positions of PL maxima/shoulders reported in (1) [31], (2) [32], (3) [26], (4) [33], (5) [34], and (6) [29]; (b): dependence of Si QDs size calculated according to the model described in [35] on the PL peak energies reported in [29].

The chromatography and other complex separation methods were applied too to separate electrochemically etched Si QDs into monodisperse fractions [1,9,36-38].

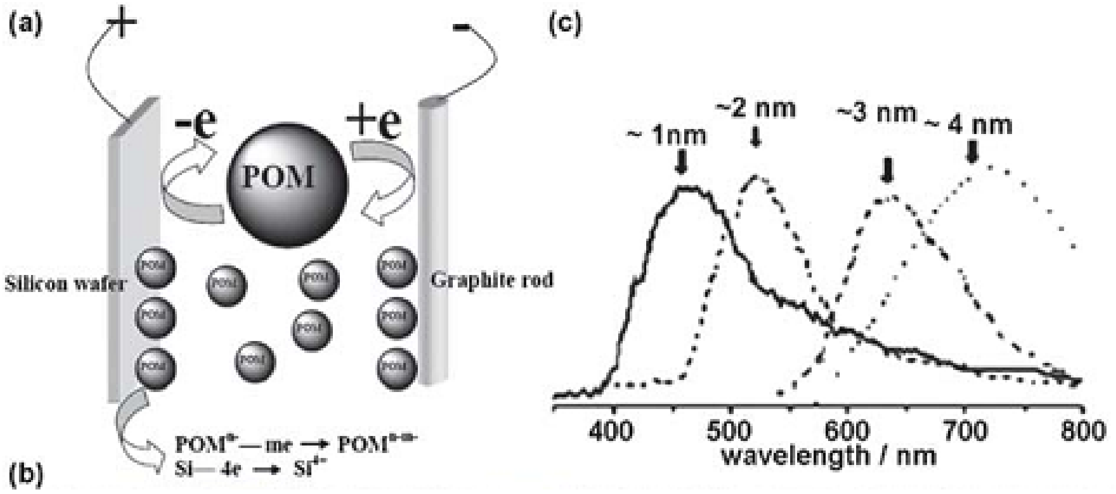


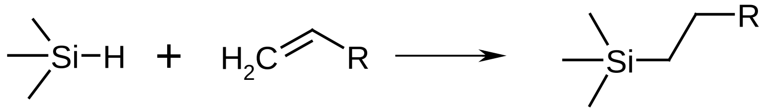
Fig.7 Scheme for the POM-assisted electrochemical etching process (a). Typical PL spectra of SiQDs with sizes ranging from 1 to 4 nm (c) from [9].

For particle size control, Kang et al. have developed a polyoxometalate (POM)-assisted electrochemical method, which yielded highly monodisperse Si dots (1, 2, 3, and 4 nm) with a narrow size distribution requiring no further separation (Fig.7) [36]. It represents a convenient and rational synthesis of a variety of Si nanostructures via the simple control of current density and choice of catalyst POMs. POMs clearly play a vital role incontrolling the size of Si dots and formation of Si nanostructures. The size control of Si dots may be attributed to the unique electronic characteristics of POMs in gaining and donating electrons simultaneously while keeping their structure unchanged.

13.4 GRAFTING OF SI QDS AND THEIR APPLICATION FOR BIOIMAGING

One of the major drawbacks of SiQDs for application for bioimaging is their instability over time in water or buffer solutions. It is well-known that PS surface slowly oxidizes in air which is resulting in degradation of optical and luminescent properties. Several methods have been used to achieve a change in the surface composition for improving the stability of materials. The PS passivation can be achieved by replacing the Si-H_x bonds with stronger S-C and Si-OC bonds [39,40]. Fortunately, the hydride species which cover the PS surface after Si etching provide a versatile starting point for various reactions, allowing the attachment of different species. An application of carbon grafting is prospective due to the fact that carbon directly bonded to silicon stabilizes porous Si against dissolution in aqueous solutions [41].

Si–C bonds are usually formed on hydride-terminated porous Si surfaces by hydrosilylation, i.e. an addition of Si–H fragments to unsaturated bonds of alkene or alkyne according to Scheme 1 [39, 41-46].



Scheme 1

Aryl and alkyl radicals of different length, aminogroups, thiols, carboxylic acid and other variable functionalities were successfully attached on the surface of PS as well as Si QDs by this approach. Differently from organosilicon compounds, the hydrosilylation on the surface of Si nanostructures does not require metalocomplex catalysts such as H_2PtCl_6; the $Si_{4-x}SiH_x$ groups can efficiently be activated to hydrosilylation by temperature (approx. 130 ^oC), visible light or Lewis acid catalysts [39, 44-46]. As an alternative to hydrosilylation, covalently attached organic layers can be formed on the porous Si surfaces under treatment with Grignard and alkyl- or aryllithium reagents, which is considered in the reviews [41, 47]. An interaction of surface SiH_x groups with alcohols at elevated temperature resulted in alkoxy-coated Si particles [1,48]. In general, alkyl coated Si QDs demonstrate superior qualities, such as better oxidation and hydrolytic stability as well as PL efficiency, comparing to unmodified SiH_x coated QDs [9].

In [49] the Si QDs with alkyl passivation of the surface were obtained through mechanical milling of Si wafers in a reactive liquid medium like alkynes and alkenes under an inert atmosphere. The reaction of the terminal triple or double bond with the reactive Si-Si and silicon surface radicals results in the formation of covalent Si–C bonds, which are preventing further oxidation of the silicon surface. This cycloaddition of unsaturated hydrocarbons resulted in organic passivated Si QDs. The milling in monofunctionalized organic liquids forms Si QDs that are soluble exclusively in organic solvents.

Very stable and bright emitting amine-terminated Si QDs with different alkyl chain lengths between the Si core and amine end-group were synthesized in [50]. The obtained Si QDs have a spherical shape and homogeneous size distribution (1.6 nm). Their emission can be tuned from the UV to the blue spectral region, in a controllable fashion, by only changing the alkyl spacer length. The emission quantum yields are 12% for all synthesized Si QDs. Excited state lifetimes are in the ns range and point to a direct band gap excitation. NH_2-terminated Si NPs exhibit an exceptional stability over a wide pH range (1–13) and high temperatures (120 ^{o}C).

Lee and co-workers recently presented an EtOH/H_2O_2 - assisted oxidation method to synthesize water-dispersed Si/SiO_xH_y core/shell QDs with a Si core of different controlled diameters [51]. Significantly, this method allows for fine tuning emission wavelengths of QDs, producing seven luminescent colors from blue to red, the fluorescent silicon nanospheres (Si NSs) containing several hundreds of SiQDs (Fig. 8). The as-prepared nanospheres possess excellent aqueous dispersibility, strong fluorescence (quantum yield: ~15%), robust photostability and favorable biocompatibility. They developed a new kind of water-dispersed oxidized SiNSs (O-SiNSs) prepared via thermal oxidation of the precursor SiNSs [52, 53]. The quantum yield of the O-SiNSs was dramatically increased to as high as 25%. More significantly, O-SiNSs are stable under high-power UV irradiation and in acidic-to-basic environments covering pH = 2—12. This extremely high pH stability leads to facile conjugation of nanospheres with antibodies, resulting in brightly luminescent silicon bioconjugates for immunofluorescent bioimaging (Fig. 9).

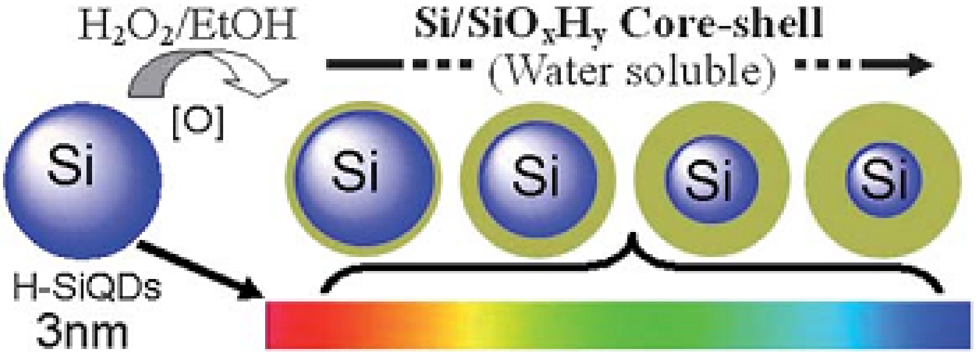


Fig. 8 3 nm H-SiQDs are oxidized in EtOH–H_2O_2; emission of the oxidized SiQDs ranges from salmon pink to blue (from [51]).

A production of stable silicon QDs with various surface functionalities through microwave-assisted hydrosilylation of hydride-terminated silicon QDs in the presence of various reactive compounds (decene, undecylenyl alcohol, and undecenoic acid) are considered in [54]. Reagents were selected from bifunctional compounds. One functional group should be the C=C which is involved in the hydrosilylation process and the other group could bear any other functionality to enable widespread applications of SiQDs.

A significant enhancement of the PL efficiency is observed for aqueous suspensions of PS particles coated by bioresorbable polymers, i.e., polylactic-co-glycolic acid (PLGA) and polyvinyl alcohol (PVA). The inner hydrophobic PLGA layer protects the QDs from the reaction with water, while the outer PVA layer makes the QDs hydrophilic [55]. UV-induced graft polymerization of acrylic acid on the surface of silicon nanoparticles was used to prepare a stable aqueous luminescent silicon QDs solution. By grafting a water-soluble polymer on the particle surface, the dispersions in water of the silicon nanoparticles became very stable and clear aqueous solutions could be obtained in [56]. Hydrogen-capped silicon QDs with strong blue PL were synthesized by the metathesis reaction of sodium silicide, NaSi, with NH_4Br. The hydrogen-capped Si QDs were further terminated with octyl groups and then coated with a polymer to render them water-soluble [57]. Silicon QDs were prepared and functionalized with alkene-terminated poly (ethylene oxide) to impart amphiphilic solution properties to the particles [58]. The alkyl-passivated Si nanocolloids were obtained via thermally-initiated hydrosilation of the hydrogenated Si nanoparticles, formed by electrochemical etching of Si wafer, with 1-undecene [59].

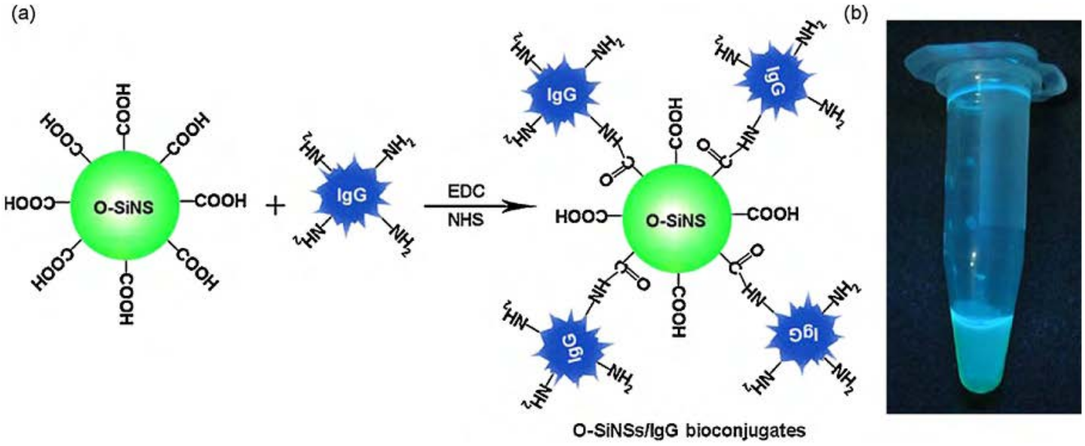


Fig.9 (a) Schematic illustration of silicon nanospheres conjugating with goat-anti-mouse IgG. The carboxylic acid groups of nanospheres readily reacted with the amino groups of IgG by using EDC and NHS as zero-length cross-linkers (Figure is not to scale). (b) Optical micrographs of the prepared silicon bioconjugates under UV (365 nm) irradiation. The prepared bioconjugates preserve stable and bright fluorescent intensity for over 1 month (right) (from [53]).

The carboxylic acid functionalized Si QDs (with the using of short alkyl chain organic peroxide, succinic acid peroxide) were applied for HeLa cells imaging [60] in several emission ranges. Hela or hela cell, is an immortalized cell type used in scientific research in vitro. The HeLa cells treated the carboxylic acid terminated Si QDs demonstrate the fluorescence with wide emission spectra arising from Si QDs (Figure 10 b–d). The bright blue fluorescence (450 nm) from the Si QDs is distributed uniformly and shows that the Si QDs were taken up into the cytoplasm (Figure 10 b). Remarkably, when the image for the 515 nm channel was collected, a bright green fluorescence from the internalized Si-QDs was clearly observed from the robust optical signal of the cells (Figure 10 c). Red fluorescence from the internalized Si QDs was also observed for the 605 nm channel (Figure 10 d). This behavior indicates that it might be possible to use hydrophilic Si QDs as chromophores for biological fluorescence imaging.

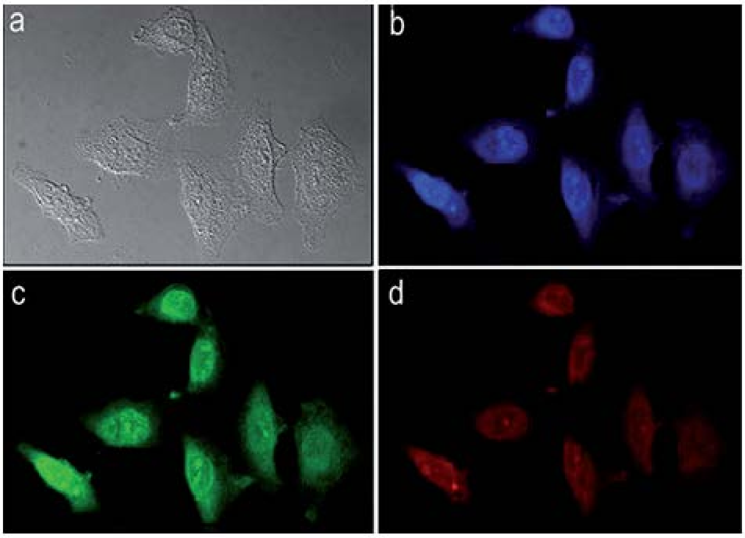


Fig.10. Confocal microscopic visualisation of HeLa cells treated with carboxylic acid terminated Si QDs collected at different channels: a) bright field image; and luminescence images collected at: b) 450 nm, c) 515 nm, and d) 605 nm (from[60]).

Application of functionalized luminescent Si QDs for bioimaging of tumors are considered in [18,61,62]. The micelle encapsulated SiQDs were injected intravenously into a Balb C mouse which was then sacrificed after 24 hours. The mouse was dissected and the organs were imaged to observe where the SiQDs were concentrating in the body (Fig. 11). The QDs concentrated preferentially in the spleen with some seen in the liver. The particles are not seen in the heart, lungs, kidneys, or the brain [62].

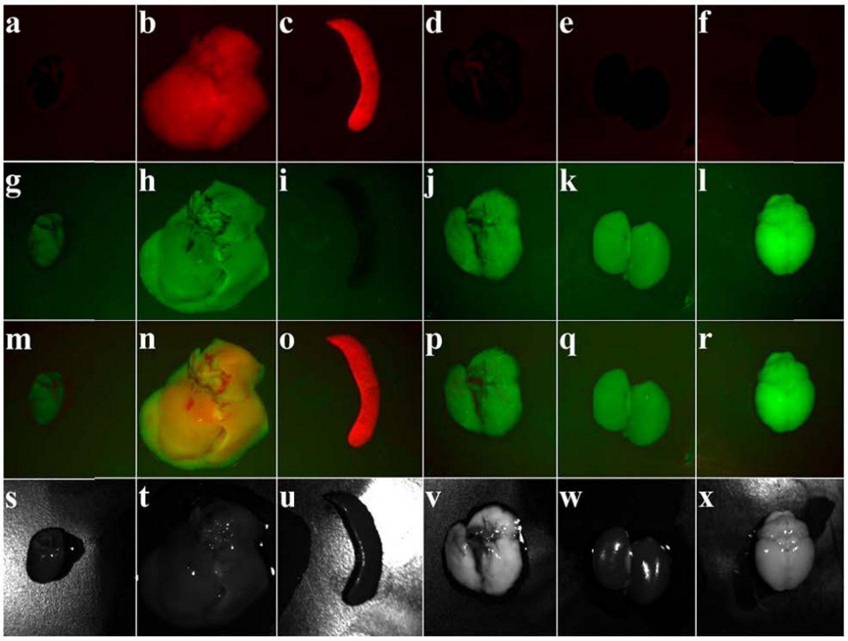


Fig. 11. Images of organs of a Balb C mouse 24 hours after injection of the PSiQDs. A, G, M, S: Heart; B, H, NN, T: Liver; C, I, O, U: Spleen; D, J, P, V: Lungs; E, K, Q, W: Kidneys; F, L, R, X: Brain. Top Row: Fluorescence; Second Row: Autofluorescence; Third Row: Overlay; Fourth Row: White light picture (from [62]).

13.5 FABRICATION AND PL PROPERTIES OF COLLOIDAL SOLUTIONS WITH SIC QDS

Luminescent properties of silicon carbide (SiC) nanostructures have been extensively studied during about last fifteen years. In comparison to extremely weak PL of bulk SiC substrates at room temperature, PL intensity of the SiC nanostructures with dimensions < 10 nm is significantly enhanced [63]. Particularly, room temperature PL of porous SiC nanostructures obtained by electrochemical etching has received a special attention from scientists. Radiative recombinations via surface states and impurity levels have been discussed in literature as main mechanisms of the observed PL signals coming from the SiC nanostructures [64, 65]. In [66] it was reported about ultraviolet PL from SiC nanoparticles with dimensions below 3 nm, which were extracted from a porous layer formed by electrochemical etching. In particular, quantum confinement effect of photogenerated charge carriers was discussed by the authors to explain an important PL signal detected at energies higher than 3 eV. In addition, it has been also shown that when the non-radiative surface states were efficiently passivated by solvent molecules and concentration of the nanoparticles in the solutions was sufficiently low reducing considerably physical interactions between the nanoparticles, significant above-gap PL spectral tails due to recombination of quantum confined photoexcited carriers could be clearly seen [67]. Several years ago, it has been reported on experimental evidence of quantum confinement effect in the PL spectra of colloidal suspensions of SiC nanoparticles with cubic crystalline symmetry (3C-polytype) obtained by an ultrasonic treatment of porous 3C-SiC layer in various solvents [68,69]. As the excitation wavelength was increased, the colloidal nanocrystals were found to give a strong above-gap emission band whose maximum was redshifted in the range from 2.8 eV to 2.2 eV. At the same time, as expected by the quantum confinement model, the emission intensity rises up to a maximum before decreasing which correlates to the size distribution supported by TEM measurements.

The method of forming SiC QDs obtained by electrochemical etching of bulk substrates of 3C polytype was described in [70,71]. Firstly, 3*C*-SiC nano-porous layer was formed by means of electrochemical anodization of a low resistivity grade (<1 Ω.cm) bulk 3*C*-SiC polycrystalline wafer. The etching process took place for 2-3 hours at a current density of 25 mA/cm^2 using a 1:1 HF(50%)/ethanol electrolyte. After the etching, a highly porous network constituted by numerous interconnected 3*C*-SiC nanocrystals was formed. The ultra-porous layer was washed with water, naturally dried in ambient air and then removed from the wafer. An intense mechanical dry grinding of the formed free nano-porous layer transformed it into a nano-powder state. The as-prepared 3*C*-SiC nanopowder can be dispersed in various polar solutions. To change surface charges, the suspended QDs are treated in different surfactants like cetyl-tri-methyl-ammonium bromide (CTMA) and sodium lauryl sulphate (SLS) which are added to the colloidal solutions to communicate positive and negative surface charges to the QDs, respectively.

FTIR spectra of the SiC QDs recorded in transmittance (pellet in KBr) and diffuse reflectance modes, are presented in Fig 12. The assignment of the bands coincides in general with results obtained for porous 6H-SiC and 3C-SiC [70-72]. Wide ν (OH) band centred at about 3500 cm^{-1} corresponds to the adsorbed water molecules as well as to C–OH and Si–OH groups. Multiple bands at 2985, 2943 and 2914 cm^{-1} (ν (CH_x)) and 1469, 1450 and 1375 δ (CH_x) indicate on presence of different types of aliphatic CH_x groups in the SiC QDs. Intense ν (C=O) band at 1743 cm^{-1} which is clearly seen in the spectrum of SiC sample, can be assigned not only to the presence of carboxylic acid groups, but also to their esters such as –$COOC_2H_5$, because the position of the band is shifted to higher wavenumbers comparing to common spectra of carboxylic acids (usually the C=O band of hydrogen bonded –COOH groups is observed at 1700 – 1725 cm^{-1}, however for –COOH group without H-bonds it can be found at 1770 cm^{-1}). Wide band at centred at 1213 cm^{-1} probably appeared due to the combination of ν (C–O) band of –COOH groups, ν (Si–O) band and some lattice bands of SiC

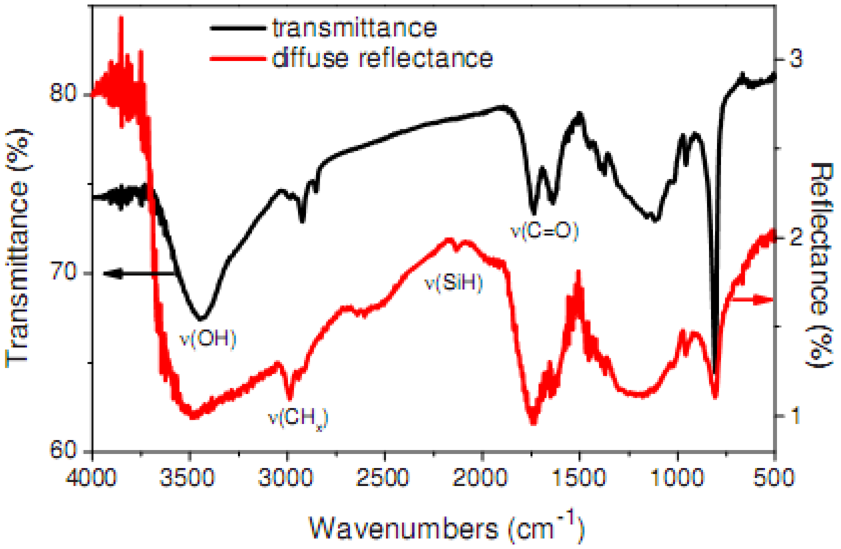
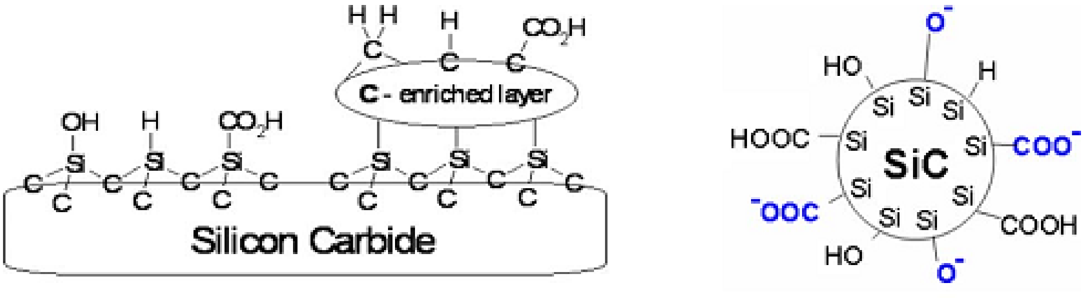


Fig.12 FTIR spectra (transmittance and diffuse reflectance) of palletized and ground 3C-SiC samples (from [70]).



(a) (b)

Fig.13 Schematic representation of: (a) general surface chemistry and (b) surface charges of 3C-SiC nanoparticles ([70]).

The surface groups of the 3C-SiC nano-powder are very similar to those of recently studied 6H-SiC porous nanostructures and schematically represented by a scheme in Fig.13a. According to [73], the 3C-SiC QDs surface contains carbon-enriched layer appeared mainly due to the preferential removing of Si atoms during the electrochemically assisted nano-powder formation. Dissociation of the carboxylic (Si-COOH) and/or silanol (Si-OH) acid groups in polar solvents leads to appearance of negative electric charges at the nanoparticle surface as it is illustrated by scheme in Fig.13b. In particular, these charges are responsible for extremely stable and optically homogeneous colloidal solutions obtained with the 3C-SiC QDs dispersed in solvents.

Electrostatic screening of the radiative band gap states by highly polar solvent media leads to a blueshift and a decrease in the width at half maximum of the PL spectra of the QDs. As for the surface charges, they govern band bending slope and thus influence strongly the radiative transitions via energy states in the band gap.

Transmission electron microscopy (TEM) image of the spherical-like SiC QDs is given in Fig.14a. An example of a daylight photo of the centrifuged optically homogeneous colloidal suspension of the 3C-SiC is shown by the inset in Fig.14a. Typical size distribution of the 3C-SiC QDs is presented in Fig.14b. As one can see, the majority of the QDs dimensions are found to be below 5 nm with the most probable size value being around 2.5 nm. The inset in Fig.14b shows typical PL spectrum of the 3C-SiC QDs dispersed in aqueous solutions under UV excitation. Since the mean size value of the SiC QDs are smaller then the Bohr's diameter of the exciton in bulk SiC substrate (~5.4nm), they exhibit highly efficient luminescence with energies higher than bandgap energy of bulk SiC due to spatial and quantum confinement effects [71].

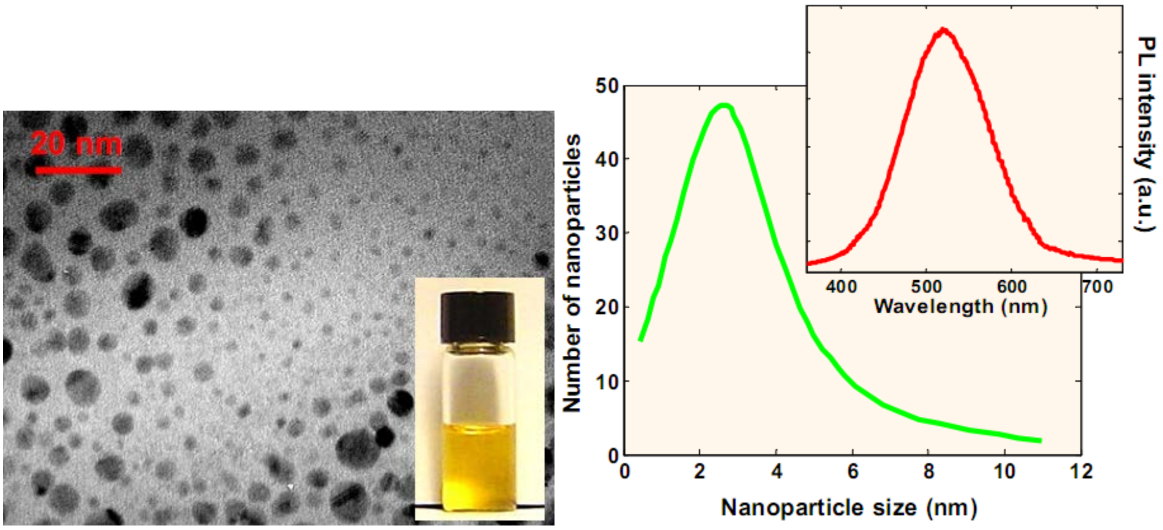


Fig.14 a) The general TEM image of the SiC QDs. Inset shows a colloidal solution with the dispersed SiC QDs; (b) size distribution of the SiC QDs. Inset shows a typical PL spectrum of the colloidal 3C-SiC QDs (from [81]).

Well defined spherical colloidal cubic SiC QDs with average diameter of 5 nm were synthesized using the reactive bonding and wet chemical etching method in [74,75]. These QDs show strong violet-blue PL emission. ATR-IR measurements revealed the surface structure of the SiC QDs which consists of Si-O-Si, C-O-C, CH, COOH, and COO-surface terminations as for electrochemically prepared QDs.

There are only few reports on surface functionalization of SiC nanoparticles [4,75-79]. Polymerization of the conducting polyaniline in the presence of camphorsulfonic acid and SiC nanoparticles results in the formation of polyaniline-camphorsulfonic acid shell with a thickness in the range from 0.5 nm to a few nanometers at the surface of SiC nanocrystals [79]. In [76] there are report PL and light reabsorption in the SiC QDs embedded in the binary poly(allylamine hydrochloride)-sodium poly(styrene sulfonate) polyelectrolytes solid matrix. The emission maximum wavelength can be tuned over a wide blue-violet spectral region indicating the quantum confinement as the origin of light emission.

13.6 SIC QDS AS FLUORESCENT BIO-LABELS

The highly luminescent 3C-SiC QDs was successfully explored as fluorescent agents for living cell imaging. By using fluorescent microscope we studied the cells of different plants, mouse fibroblast, healthy and cancer epithelial human cells with SiC QDs [8,71,80,81]. Covalent grafting of amino groups onto the carboxylic acid functionalities, naturally covering the surface of fluorescent nanoparticles produced from silicon carbide, allowed tuning of their surface charge from negative to highly positive (Fig.15). Grafting of amino-groups was performed by the reaction of SiC QDs surface carboxylic acid groups with ethylenediamine (En, $H_2N–CH_2CH_2–NH_2$), Fig.15.

Incubating 3T3-L1 fibroblast cells with differently charged SiC QDs demonstrates the crucial role of the charge in cell fluorescent targeting [71]. Negatively charged SiC QDs concentrate inside the cell nuclei. Close to neutrally charged SiC QDs are present in both cytoplasm and nuclei while positively charged SiC QDs are present only in the cytoplasm and are not able to move inside the nuclei (Fig.16). This effect opens the door for the use of SiC QDs for easy and fast visualization of long-lasting biological processes taking place in the cell cytosol or nucleus as well as providing a new long-term cell imaging tool. Moreover, here we have shown that the interaction between charged QDs and nuclear pore complex plays an essential role in their penetration into the nuclei. The importance of electrostatic interactions was more recently strengthened by authors studying the molecular structure of the yeast nuclear pore complex (NPC) and the translocation of model particles [82].

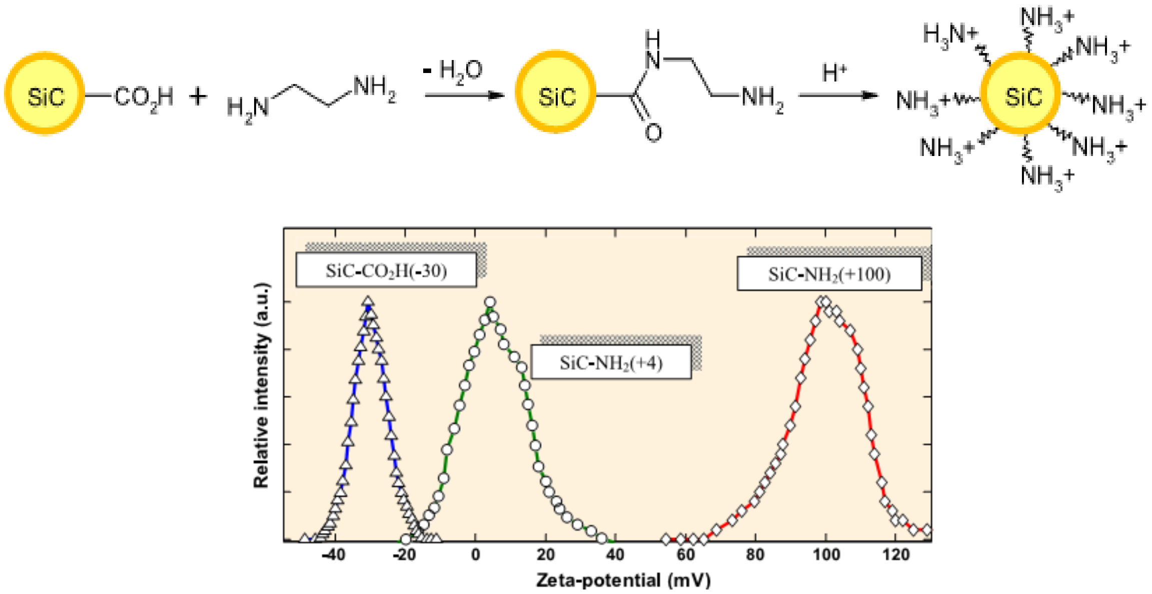


Fig.15.Grafting of amino groups on the surface of as-prepared SiC QDs (a) and Zeta-potential measurements performed on colloidal solutions of SiC–CO_2 H(-30), SiC–NH_2 (+4) and SiC–NH_2 (+100) (from [71]].

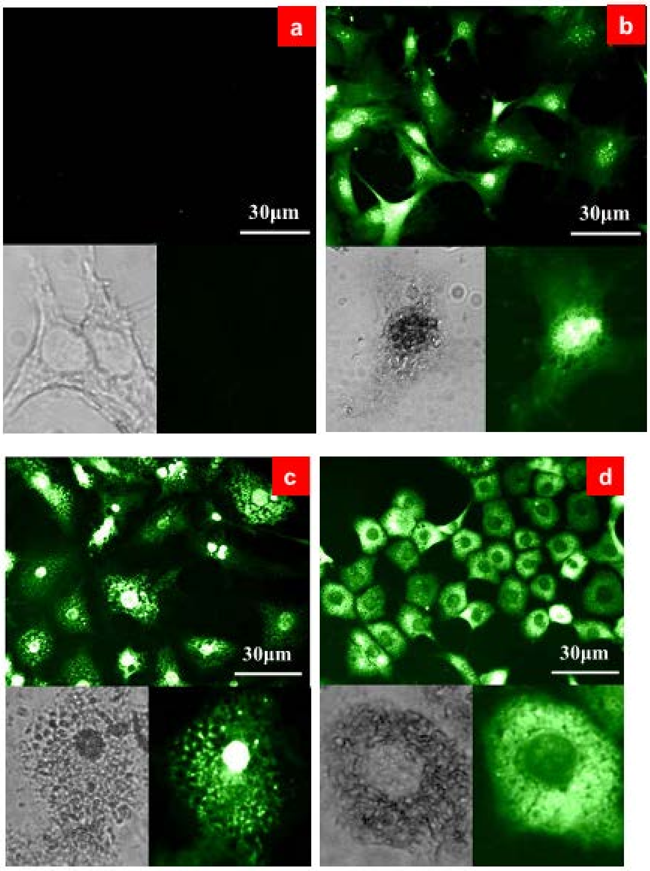


Fig.16 Fluorescence and visible microscopy images of 3T3-L1 cells: (a) without SiC QDs; (b) labeled with negatively charged SiC–CO_2H(-30) QDs; (c) labeled with quasineutral SiC–NH_2 (+4) QDs and (d) labeled with positively charged SiC–NH_2 (+100) QDs (from [71])

The uptake of QDs and its intra-nuclei concentration is dependant to cell proliferation (Fig.17). This has been shown by two different ways, first in healthy cells and second in cancer cells [81]. In healthy SG cells (a human immortalized gingival epithelioid cell line), the labeling is altered by the state of confluence, at confluence the labelling is low and absent in the nuclei (c) while it is strong and present in the nuclei when cells divide (a and b). In cancer cells HSC-2 (a human oral squamous carcinoma line), that proliferate even at confluence, the labeling is not influence by confluence (d, e and f). Although the mechanism linking intra-nuclear accumulation of SiC QDs and cell proliferation has not been yet identified, this observation is important for at least two reasons. Firstly, it brings evidence of a specificity of SiC QDs to be uptaken by proliferating cells that are promising to give specificity for cancer treatment or drug delivery with SiC QDs. Secondly, a majority of studies on QDs are done on proliferating cells, according to the present results, the behaviour of QDs may be not the same for confluent (proliferating) and confluent (non proliferating) cells.

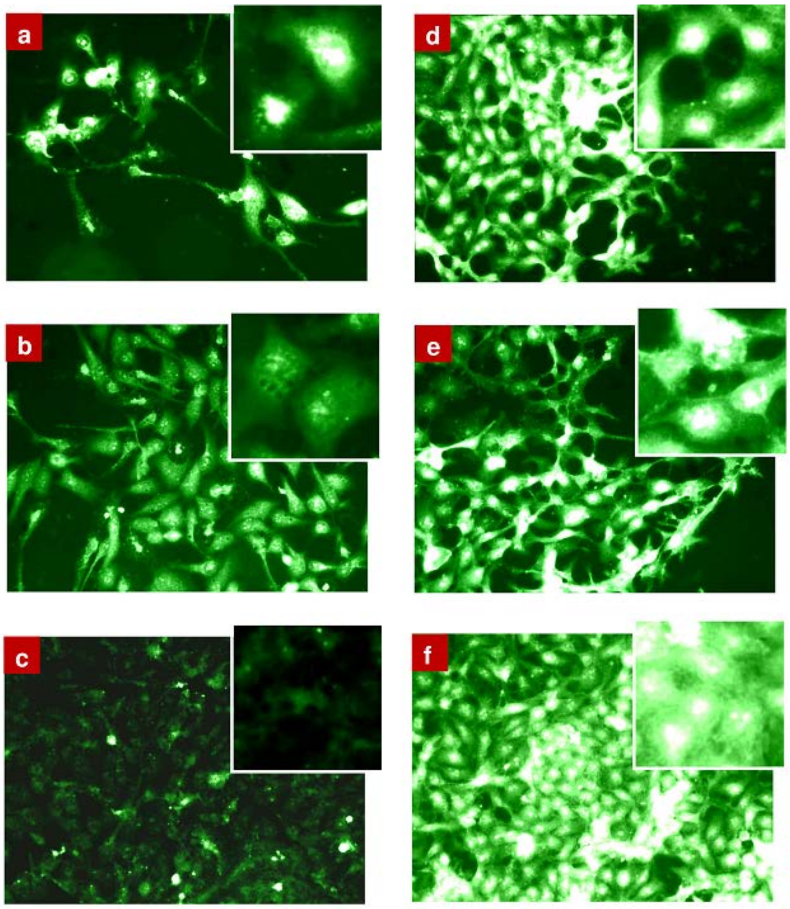


Fig. 17. Fluorescence microscopy images of SG and HSC-2 cells labeled by SiC NPs: a), d) low confluence b), e) intermediate confluence, c), f) total confluence (from[81]).

To test the cytotoxic potential of SiC QDs, we chose two oral squamous carcinoma and one immortalized oral epithelioid lines. Like the two cancer lines, the control cells we used possess a high growth rate and are used as a control in comparison to HSC-2 cells [83]. Higher concentrations of SiC QDs and longer exposure time are required to reduce S-G proliferation when compared to both human and murine tumoral cells. The differential toxicity of SiC QDs against rapidly dividing cancer cells suggests their potential use as antitumoral agents. The precise mechanism(s) of action of SiC QDs in inhibiting cell proliferation remains to be investigated. We considered the possibility of using NPs SiC for antitumor therapy. When SiC QDs concentration of 2 mg / mL kills 100% of the cancer cells HSC-2 is already at 6:00 incubation and almost no effect on the survival of the culture of healthy cells that have stopped cell division through a mechanism of contact inhibition. This means that toxic effects, as well as the total amount of nanoparticles in the cell decreases dramatically when the cage stops proliferation (division) [8].

SiC QDs have, however, few disadvantages preventing their use in complex bio-imaging applications. This includes low quantum yield and rather broad emission band, which makes multicolour imaging difficult. Low quantum yield also influences concentration at which labelling effect is detectable. Therefore, to compete with usual fluorophores, a strong PL enhancement of SiC QDs may be ensured by their nearfield coupling with multipolar plasmons [84]. In order to achieve a high PL enhancement factors, an approach based on: (1) localization of SiC QDs in the vicinity of randomly arranged silver (Ag) nanoparticles (NPs) chemically formed on silicon nitride (SiN_X)/glass substrates and (2) careful tuning of the multipolar plasmon bands of the nanoparticles to overlap the excitation and emission spectral bands of SiC QDs, has been developed.

Fluorescence images of fibroblast cells on glass and plasmonic substrates with or without SiC QDs can be seen in Figure 18. All images were obtained under the same observation conditions in terms of excitation (UV/violet excitation band (λ=350-460 nm) and an observation spectral range corresponding to the wavelengths λ >470 nm) and acquisition time. Natural fluorescence of the cells grown on the usual glass cover slip substrates is too weak to be observed in such experimental conditions (Figure 18a). On the other hand, the cells grown on the glass substrates and labelled with the SiC QDs can be better seen in Figure 18b. Strong auto-fluorescence enhancement of the fibroblast cells grown onto the nano-Ag/SiN_X/glass substrates is shown in Figure 18c. This label-free natural green cell auto-fluorescence is significantly increased in comparison with the cells grown on a simple glass substrate (Figure

18a). The cells with the up-taken SiC NPs grown onto the nano-Ag/SiN$_X$/glass substrates show a drastic fluorescence enhancement (Figure 18d) in comparison to the cells containing the same quantity of the SiC QDs and grown on a usually-used glass cover slips (Figure 18b).

Different mechanisms can be responsible for the observed fluorescence enhancement of the cells. Firstly, the localized plasmons appearing in the Ag NPs leads to an increase of excitation/absorption and photostimulated auto-emission of the cell membranes due to their close proximity to the Ag NPs. The fluorescence enhancement can be only observable when the plasmons are localized in very close vicinity to the cell membrane. Moreover, the cell membrane thickness is estimated to be around a few nanometers which allows plasmon-induced PL enhancement of the NPs up-taken by the cells and localized near the membrane after the dehydration and the subsequent fixation of the cells on the plasmonic nano-Ag/SiN$_X$/glass substrate. All these factors can explain the detected enhancement of cell auto-fluorescence and PL of the up-taken SiC QDs (see Figure 18c and 18d, respectively).

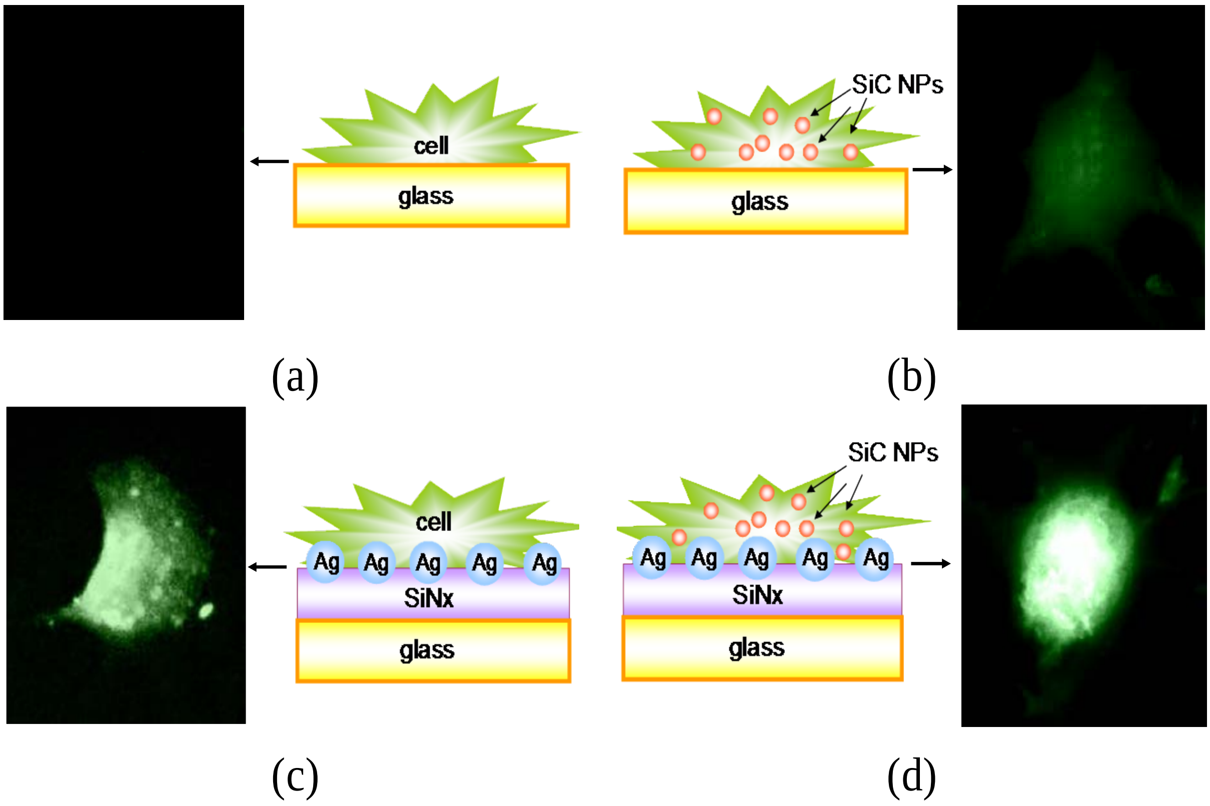


Fig. 18. Sample schematics and corresponding fluorescent images of fibroblast cells: (a) on glass substrate, (b) labelled with the SiC NPs on glass substrate, (c) on nano-Ag/SiNX/glass substrate, (d) labelled with the SiC NPs on nano-Ag/SiNX/glass substrate.

Secondly, the cell labeling with the SiC QDs can induce modification of the local effective dielectric constant which will lead to a better near-filed coupling (in terms of spectral overlap) with the localized plasmon modes under the same excitation conditions. In addition, the higher values of the local dielectric constant will also ensure the higher intensities of the plasmon-induced local electric field penetrating deeper inside the labelled cells. Finally, a certain degree of the excitation scattering within the nanostructured area and inside the cells may explain general PL enhancement taking place in the whole cell volume. Moreover, the use of a UV/violet filter selecting the exciting photons from the relatively large spectral range (λ= 350-460nm) of a white lamp leads to a simultaneous photo-excitation of a wide absorptive part of the extinction spectra, provoking significant increase of the absorption cross-sections of numerous cell fluorescent agents as well as of the big number of the SiC QDs with various sizes.

In addition, the auto-fluorescence cell signal and PL of the SiC QDs incorporated inside the cells (detected in the spectral range λ>470nm) are strongly enhanced by near-field interaction with the coupled dipole plasmon modes, as it is discussed in details in [84].

13.7. BIOSENSING BASED ON QUANTUM DOTS

Luminescent semiconductor QDs are a new class of nanomaterials whose specific photophysical properties are helping to create a unique generation of fluorescent biosensors . Due to the fact that the most common methods used for detection of biomolecules are based on fluorescence changes monitoring, the fluorescent probes have found wide-spread biosensing applications in immunoassays, nucleic acid detection, ion detection, clinical and diagnostic assays. Usually, in the presence of analyte the decrease of QDs fluorescence emission intensity is observed. This relationship is schematically presented in Fig. 19. The linear relationship between QD fluorescence emission changes and analyte concentration commonly follows Stern– Volmer function. As the luminescence phenomena of QDs is very sensitive to the QD surface states, it was expected that the interactions between a given analyte and the QD surface will result in the changes in the efficiency of the core electron–hole recombination.

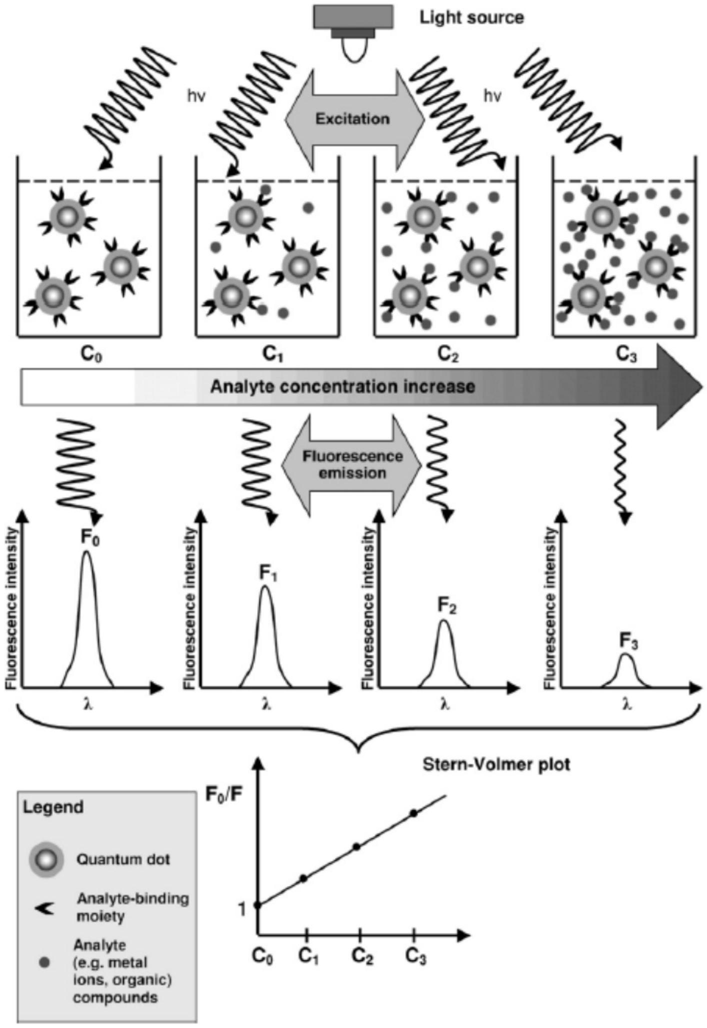


Fig. 19. General rule of analyte determination based on quantum dot fluorescence emission quenching.

13.8 CONCLUSION

This chapter focuses on the preparation of colloidal solutions of Si and SiC QDs by the method of electrochemical etching of bulk Si and 3C-SiC. It was shown that a broad size distribution of crystalline Si and SiC QDs are obtained. It is worth noting that the dimensions the dimensions of the Si QDs undergone filtering in colloidal solution vary discretely with a radius quantum equal to 0.12 nm. Existing of this quantum may correspond to step-like increasing of Si QDs radius on one new shell at the surface of Si QDs. The formed QDs show intense luminescent in visual region that is promising for fluorescent labeling of biomaterials. However, one of the major drawbacks of SiQDs for application for bioimaging is instability over time in water or buffer solutions. To overcome this drawback the several methods of surface functionalization are proposed and discussed. From the other side, the SiC QDs are stable in water solutions and does not required supplementary surface functionalisation for bioimaging. A strong fluorescence from the SiC QDs, which undoubtedly penetrate into the cell, has been observed. The studying of health and cancer cells using SiC QDs show that simple modification of surface charge of QDs give strong opportunity to target the same QDs in intracellular space with their preferential localisation inside or outside the cell nucleus. Electrostatic interaction between QDs and positively charged chromatin gives platform to completely switch-open penetration or it blocking corresponding to the surface charge of QDs that plays role of key to open/close nucleus membrane doors. These effects open wide perspectives for understanding and discovering nucleus transport mechanisms and give possibility for their bright imaging. Moreover, the uptake of QDs and its intra-nuclei concentration is strongly dependant on cell proliferation.

Chapter 14. Notubes. Fullerens and Carbon nanotubes. Graphene.

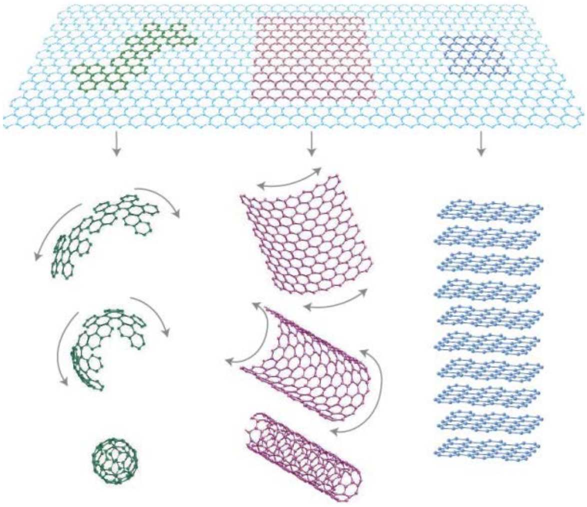
14.1. Introduction

Carbon is arguably the most fascinating element in the periodic table. It is the base for DNA and all life on Earth. Carbon can exist in several different forms. The most common form of carbon is graphite, which consists of stacked sheets of carbon with a hexagonal structure. Under high pressure diamond is formed, which is a metastable form of carbon.

The discovery of fullerenes [1], carbon nanotubes and graphene initiated a new era in carbon chemistry. In addition to the two well-known allotropes, graphite and diamond, a large number of carbon structures were synthesized [2-8]. The fullerene molecule (primarily the most abundant fullerene, C_{60}), the polymerized, and doped fullerenes will be first considered.

A related quasi-one-dimensional form of carbon, carbon nanotubes, have been known for several decades [9] and the single walled nanotubes since 1993 [10, 11]. These can be formed from graphene sheets which are rolled up to form tubes, and their ends are half spherical in the same way as the fullerenes (Fig. 14.1.). The electronic and mechanical properties of metallic single walled nanotubes have many similarities with graphene.

It was well known that graphite consists of hexagonal carbon sheets that are stacked on top of each other, but it was believed that a single sheet could not be produced in isolated form such that electrical measurements could be performed. It, therefore, came as a surprise to the physics community when in October 2004, Konstantin Novoselov, Andre Geim and their collaborators [12] showed that such a single layer could be isolated and transferred to another substrate and that electrical characterization could be done on a few such layers [13]. The single layer of carbon is what we call graphene.



``

Fig. 14.1. C_{60} fullerene molecules, carbon nanotubes, and graphite can all be thought of as being formed from graphene sheets, *i.e.* single layers of carbon atoms arranged in a honeycomb lattice [14].

14.2. Fullerens

Conceptually, the new structures [2 – 8, 15] can be derived from graphene, an sp^2-bonded honeycomb lattice of carbon atoms, which is the basic element in the structure of graphite [8]. Assume each carbon atom is connected to three neighbors, the bond lengths are fixed to be the same, but the bound angles are a bit variable. Fullerenes can be viewed as fully closed graphene sheets. The guiding principle is that each carbon atom is connected to three neighbors, and the bond angles must remain close to 120°. The perfect honeycomb lattice satisfies this requirement, as the building blocks are ideal hexagons. However, in order to get closed surfaces, one has to allow for an occasional pentagon in the structure, just as in the domes designed by the architect Buckminster Fuller (hence are the names of ‘‘fullerene’’ and ‘‘buckyball’’). There is a simple geometrical principle, Euler’s theorem, that connects the number of faces (f), vortices (v), and edges (e) of a closed three-dimensional polyhedron, $f+v-e=2$. If only hexagons and pentagons are allowed, one can easily prove that the number of pentagons is always 12.

The smallest fullerene, C_{36}, has the shape of a dodecahedron, with exactly 12 pentagons, and no hexagons. However, this molecule is very reactive [16], as the bond angles are quite stretched. If we look for surfaces where no two pentagons can be nearest neighbor we get the smallest and most abundant stable fullerene, C_{60} (Fig. 14.2). In this molecule the carbon nuclei reside on a sphere of about 0.7 nm diameter, with the electronic wavefunctions extending inside and outside by about 0.15 nm. The diameter of the molecule is approximately 1.0 nm, and there is a 0.4-nm-diameter cavity inside. This molecule still has the same symmetry properties as the dodecahedron (described by the icosahedral symmetry group, I_h), and the symmetry is preserved even if not all bond lengths are exactly the same. In fact, in C_{60} the edges that separate two hexagonal faces are slightly shorter (0.140 nm) than the edges separating a pentagonal and a hexagonal face (0.146 nm). The chemical bond along the shorter edge is sometimes considered double bond, whereas the others are single bonds. Notice, however, the single bond distance in diamond is 0.154 nm (much longer than 0.146 nm), and the typical double bond in molecules is 0.135 nm (much shorter than 0.140 nm). This already indicates that the simple bond picture does not work for C_{60}, and the π electrons are, at least partially, delocalized.

``

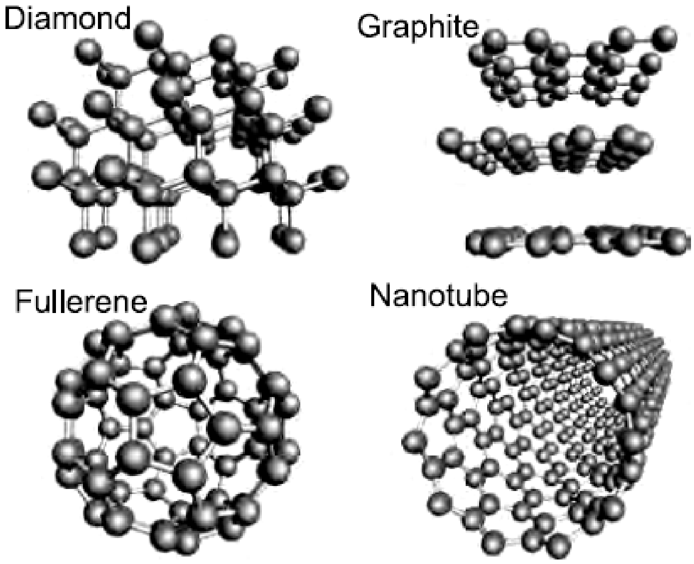


Fig. 14.2. Four carbon structures, representing diamond, graphite, a typical fullerene (C_{60}), and a typical nanotube (10,0 zigzag).

The most abundant fullerene higher than C_{60} is C_{70}. Its structure can be constructed by adding a ‘‘belt’’ of five hexagons in one of the equatorial planes of C_{60}. Other fullerenes, such as C_{76}, C_{78}, C_{80}, and up, have also been prepared and purified. The number of isomers increases rapidly with the increasing fullerene size, representing some difficulty for the determination of their structure (mainly by NMR measurements and theoretical calculations). It is interesting to note that one of the isomers of C_{80} has the exact shape of an icosahedron (there are also six other isomers with lower symmetry).

Nanotubes with closed ends can be obtained by capping the open ends with a fullerene fragment of appropriate size and shape. For example, the (5,5) nanotube can be capped at the two ends with the two halves of a C_{60} molecule. Adding more ‘‘belts’’ of hexagons to C_{70} yields the same structure. In this sense the distinction between nanotubes and fullerenes is somewhat artificial.

14.2.1. Synthesis of Fullerene

Once synthesized in the laboratory, fullerenes were also discovered in common sooth and in natural minerals, but it is still more productive to make them in the laboratory. The most common way of preparing fullerenes and nanotubes is the arc discharge method. The typical apparatus [3] consists of a high current source (e.g., and arc welding machine) and a vacuum jar, where a partial pressure (about 100–200 Torr) of He is maintained. Inside the jar two carbon rods are brought into contact to form an arc. A constant distance should be maintained between the rods as the material is gradually converted into carbon sooth. In one version of the apparatus the weight of one of the

``

carbon rods maintains a loose contact between the rods. The sooth is collected from the inside surface of the jar, and from other parts.

A more precisely controlled preparation method involves laser ablation of carbon [2, 17]. The graphite target is kept at high temperature (1200°C) in a furnace. A high power laser (typically a Nd:YAG laser operating at several hundred millijoules of power) is aimed at the target and produces a carbon plasma. A flow of inert gas carries the end products and deposits them in a colder part of the apparatus.

In both cases the reaction products are mixtures of various fullerenes, nanotubes, and amorphous carbon, although by properly setting and controlling the preparation conditions the relative quantity of the end products can be influenced greatly. Nevertheless, the separation and purification of the components is a critical part of the process. Various organic solvents (such as toluene or tetrahydrofurane) can be used to dissolve fullerenes or disperse the nanotubes. High-performance liquid chromatography (HPLC) is used to separate the different fractions. Once a solid phase is obtained fullerenes can be further purified by sublimation (e.g., C_{60} sublimes at 430°C). Smaller quantities of higher fullerenes can also be purified by gas-phase separation methods.

A third approach, primarily used in the production of SWNTs, is based on chemical vapor deposition (CVD). This method involves a carefully prepared catalyst, and it can be used to grow large quantities of ordered nanotubes [18]. It is generally believed that SWNTs are more readily obtained if the reaction temperature is not too high. Note that in the arc discharge method the reaction happens at temperatures larger than 3000°C, whereas in CVD the temperature of the reaction can be kept below 1000°C.

14.2.2. Molecular Properties

The unsaturated nature of the sp^2 bonds is responsible for several interesting properties of fullerenes and nanotubes. Functionalization (attaching other molecules with desirable properties) can be carried out easily. A pretty good description of the electronic properties can be obtained by assuming that three of the four valence electrons of each carbon are tied up in the three σ bonds, and by considering the behavior of the remaining one electron [7, 8]. We will focus on one of the fullerenes, C_{60}.

Electronic States in C_{60}. C_{60} has 240 valence electrons, but each carbon atom has three sigma bonds to its neighbors, using up a total of 180 electrons for this purpose. The energy of these electrons is well below the Fermi surface. They stabilize the structure, but they will not contribute to the conduction. The remaining 60 electrons are distributed around the molecule on orbitals that originate from the (much less tight) carbon–carbon π orbitals. There are two important differences

``

between these orbitals and the π electrons of a graphene plane. First, the three bonds around a carbon atom in C_{60} (or in any other fullerene) do not make a plane. Whereas in graphene the electrons had equal probability of being ‘‘below’’ and ‘‘above’’ the plane, in fullerenes the π electrons tend to spend more time outside of the ball than inside. Second, in C_{60} the C–C bond lengths are not uniform, the π electrons are not truly delocalized around the six-member carbon rings (like in benzene or graphene), but they are distributed over 30 ‘‘bulges’’ of electronic orbits that stick out of the C_{60} molecule. A somewhat lower electron density belongs to the other 60 orbits connecting the carbon pairs with longer bond lengths.

The first insight to the nature of the molecular orbitals can be obtained by borrowing ideas from the early days of nuclear physics, when the quantum mechanics of particles confined into a spherical potential well was first considered. In this case the states are still labeled by quantum numbers n, l, and m, just like in the hydrogen atom, but the peculiar degeneracy of the different angular momentum states, which characterizes the textbook treatment of the hydrogen atom, does not apply. In steep-walled potential well (like in a nucleus or in a C_{60} molecule) the lower l states will have lower energy.

Let us consider 60, noninteracting electrons, confined to a sphere on a high n orbit with various l’s. The first two electrons will fill up l=0 state. Proceeding to higher l, one has to count the number of available states, N=2(2l+1) (where the first factor of 2 stands for spin). It is easy to see that by reaching l=4 altogether 50 electrons are consumed. The remaining 10 electrons will all go to the l=5 state.

In reality the situation is much more complex. The potential does not have the spherical symmetry; each carbon atom creates its own potential well for the electrons. Nevertheless, a more exact treatment of the electrons (e.g., a Hu¨ckel molecular orbital calculation) shows that the above arguments work well for l=0,1,2,3,4. For higher energies the ‘‘spherical potential’’ approximation does not work anymore. In order to go any further we need to look at the energy splitting due to the true atomic potentials. In the real C_{60} l is not a good quantum number, and the electronic orbitals should be labeled according to the irreducible representations of the icosahedral symmetry group. The orbitals available for the remaining 10 electrons are, in the order of increasing energy, the h_u, the f_{1u}, and the f_{1g} levels. The degeneracy of these levels (including spin) is 10, 6, and 6, respectively. The h_u level is completely filled by the 10 remaining electrons, becoming the highest occupied molecular orbit (HOMO), and the f_{1u} level becomes the lowest unoccupied molecular orbit (LUMO). The HOMO–LUMO gap is about 2 eV.

The energy gap between the filled and empty electronic states causes the molecule to be nonreactive, and it accounts for the insulating nature of the C_{60} solid. The empty f_{1u} level can hold six electrons, and a proper addition of electrons to the molecule (sometimes called ‘‘doping’’, in

‘‘

analogy to adding electrons to a pure superconductor) should result in dramatic changes in the properties, including higher reactivity and metallic character. Two possible ways will be discussed here shortly: adding an atom inside the C_{60} cage (endohedral doping) and exchanging a carbon atom to another one (onsite doping; Fig. 14.3). The third (and most fruitful) method, exohedral doping, when the donor atom is placed in between the C_{60} molecules, will be considered later.

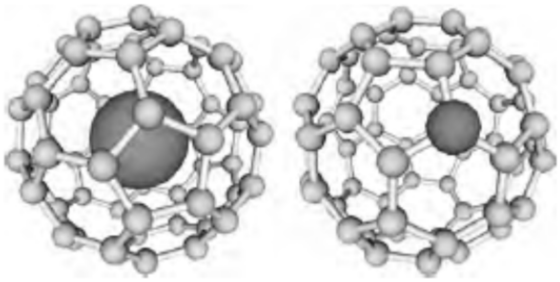


Fig. 14.3. Endohedral and on-site doping of a C_{60} molecule.

Endohedral C_{60} molecules can be prepared by a ‘‘brute-force’’ method, where ions of atoms are accelerated and implanted to the C_{60} cage [19]. The ions should have just enough energy to open up the cage and enter. The first collision should absorb and redistribute a good part of the initial kinetic energy so that the atom does not escape the cage. Endohedral molecules of M–C_{60} with M=N, P, Li, Ca, Na, K, Rb were produced this way in small quantities. A larger yield can be achieved by coevaporation [20] of the carbon and the metal is an arcdischarge chamber (typical for fullerene production). In this process mostly higher endohedral fullerenes, such as M–C_{82} (where M is a metal atom), can be extracted from the soot on a chromatographic column.

Electron spin resonance spectroscopy (ESR) studies and detailed quantum chemical calculations show that the endohedral dopant atom does not always transfer its charge to the cage. If the dopant remains neutral, it stays in the center of the cage, like in N–C_{60} [21]. Compounds such as La–C_{60} are very interesting, as complete charge transfer from La to the C_{60} cage results in a triply charged molecule.

On-ball doping has been achieved by replacing one carbon atom with a nitrogen atom obtaining $C_{59}N$ (azafullerenes) [22]. The replacement of N for C adds one extra electron to the cage, changes the structure locally, lowers the symmetry of the molecule, and splits the degeneracy of the electronic orbitals. Furthermore, this chemical substitution renders the molecule very reactive with a high electron affinity. At ambient temperature the $C_{59}N$ exists only in a dimerized phase.

``

14.2.3. Fullerene Solid

The vast majority of research on the fullerene-based solids was done on C_{60}. The molecule dissolves in various solvents. Crystallization from solution usually yields a structure that includes the solvent molecules [23]. However, sublimation in vacuum or inert gas atmosphere may yield pure C_{60} crystals of several millimeters or even centimeter size. In the solid form the centers of the fullerene molecules are ordered in a face-centered cubic (fcc) crystal lattice (Fig. 14.4).

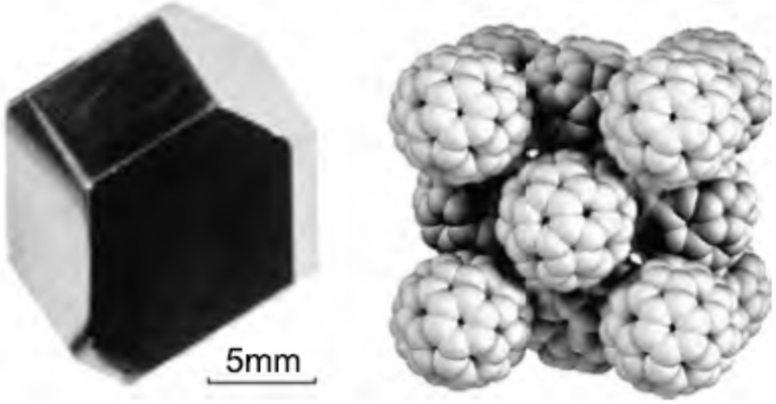


Fig. 14.4. A pure fullerene crystal, and the illustration of the structure of the fullerene solid [24].

At room temperature the fullerene molecules rapidly and randomly rotate between various orientations [25]. In time average the molecules look like spheres and all of the molecular sites are equivalent. (The well-defined molecular center and a nearly free molecular rotation are not a unique property of the fullerene molecule. It is characteristic to many organic solids, where the binding is due to van der Waals forces, and the molecule has ‘‘rounded’’ symmetric shape, like solid CH_4 and CCl_4.)

The nearest neighbor C_{60}–C_{60} distance corresponds to the diameter of the molecule. At room temperature the size of the cubic unit cell is 1.417 nm, and the nearest neighbor distance is 1.002 nm. As the temperature is lowered, the rotation of the molecules stops. The center of the molecules remains in about the same place, but the incompatibility of the icosahedral molecular symmetry and the cubic lattice symmetry makes it impossible to describe the structure in terms of the simple fcc lattice, with one C_{60} molecule of fixed orientation each unit cell. The freezing of the rotational motion actually happens in two stages. First, at 261 K the rotational axis of each molecule becomes constrained. The building block of the new structure consists of four C_{60} molecules arranged at the vertices of a tetrahedron, with each one of them spinning around a different, but well-defined axis [26 - 29]. Then, at lower temperatures, the rotation around these axes slows down and stops gradually. Below about 90 K the molecules are entirely frozen, but they never order perfectly.

``

Doped Fullerenes. The linear size of a fullerene molecule is 10 times larger than a typical atom. Even in a close-packed fcc structure, large empty spaces between the molecules are freely available for smaller atoms, ions, or molecules. Because of the unsaturated character of the C–C bonds on the fullerene, there are plenty of electronic states to accept electrons from appropriate donors. The combination of alkali and alkaline earth elements with fullerenes results in an unusual variety of A_nC_{60} materials that are mostly near-stoichiometric (A can stand for Na, K, Rb, Cs Ca, Sr, and Ba). Organic and inorganic molecules were also used to produce charge transfer salts with fullerenes. The term ‘‘doped fullerene’’ is often used to describe the product (although in its original sense, ‘‘doping’’ means a small, nonstoichiometric amount of charge transfer, typically in a semiconductor).

In the fcc structure the interstitial sites have either octahedral or tetrahedral symmetry. The tetrahedral sites are smaller, and there are twice as many of them as octahedral sites. A series of A_nC_{60} (with n=0,1,2,3,4,6,10) materials are based on the original fcc structure of pure C_{60}. Several other compounds are known where the C_{60} abandons the close-packed structure [30 - 35]. Two of these, the body-centered tetragonal A_4C_{60} and the body-centered cubic A_6C_{60}, are illustrated in Fig. 14.5. There is also a series of compounds with alkaline earth metals. Based on the simple electron band arguments, the HOMO level of the C_{60} can accept a total of six electrons. Anything less than that should result in a partially occupied electron band and lead to a metallic behavior in the solid state. Surprisingly, at room temperature and ambient pressure, only AC_{60}, and A_3C_{60} are metals [34].

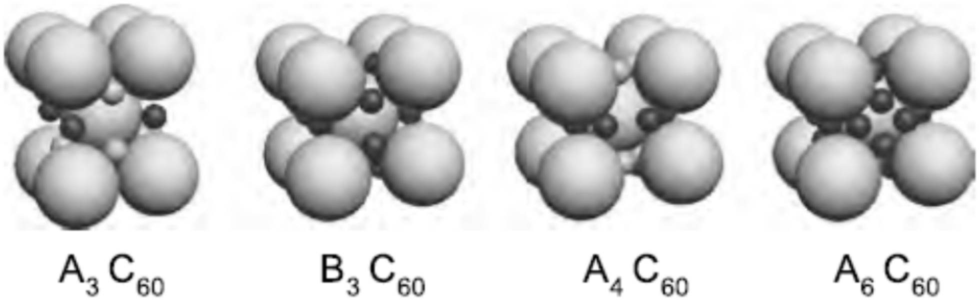


Fig. 14.5. The left most structure corresponds to A_3C_{60}, with A=K, Rb, Cs, (or a mixture of these). If the dopant is an alkaline-earth metal (denoted by B), the structure is B_3C_{60}. The A_4C_{60} and A_6C_{60} structures shown here are known to exist with A=K, Rb, and Cs.

Superconductivity and Magnetism in Fullerenes. A class of alkali-metal-doped fullerenes, with chemical composition A_3C_{60}, is metallic at room temperature [36] and turns into superconductor at low temperature [37]. The transition temperature is unusually large; for example, `` in Rb_3C_{60} it is 32 K [38]. In contrast to other exotic superconductors (such as the copper-oxide-based high temperature superconductors) the superconductivity in fullerides is well understood in terms of the theory developed in the 1960s by Bardeen, Cooper, and Schrieffer (BCS). The formation of Cooper pairs (two electrons that are held together by some attractive force) is the key element of the BCS theory. In the A_3C_{60} fullerides the conduction electrons are donated by the alkali metal atoms to the C_{60} molecules, the electrons easily hop between the molecules (because of the overlap of the molecular orbits), and the attractive interaction is mediated by the deformation of the molecule induced by the presence of the electron.

When C_{60} is combined with the strong electron acceptor tetrakis(diethylamino)ethylene (TDAE) each fullerene molecule loses an electron. It was expected that this material will also be a metal, and perhaps a superconductor, too. Instead, it turned out to be an insulator, but when it was cooled to low temperature a strong magnetism was detected. (TDAE)C_{60} is a ferromagnetic compound with a Curie temperature of 16 K [39, 40]. It is very rare to find ferromagnetic behavior with no transition metal (such as iron or nickel) in the chemical formula. (TDAE)C_{60} held the record of being the highest temperature truly organic ferromagnet.

Polymerization of Fullerenes. The ability of the C_{60} molecule to form covalently bonded polymers was first noticed in laser-irradiated thin films of C_{60} [41]. The polymerization happens in a Diels–Alder-type (2+2) cycloaddition reaction when the double bonds on two fullerenes are broken up and cross links are formed. This reaction has a large activation energy, and at room temperature it is extremely slow. It is believed that (2+2) cycloaddition also occurs in high-pressure/high-temperature polymerization of fullerenes [42]. Some of these high pressure synthesized disordered polymers are very attractive for applications, as they can exhibit mechanical properties stronger than diamond.

In doped fullerenes polymerization can happen without exposing the material to light, high temperature, or pressure. The best known examples are KC_{60} and RbC_{60}, where the fullerene molecules form linear chain structures (Fig. 14.6). The formation of these chains happens spontaneously as the material is cooled slightly below room temperature [43]. There are dramatic changes in the electrical conductivity and in the other properties; most notably the material becomes stable in air (otherwise all alkali-doped fullerene compounds decompose rapidly in air). This is a reversible transition, and the material turns back to the monomeric form if the temperature is raised. A slightly different polymer is formed in Na_4C_{60}. In this compound, sheets of C_{60} molecules are formed so that there is a single bond between the nearest neighbors.

``

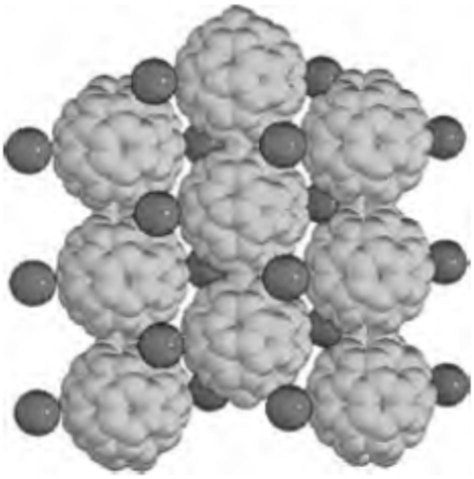
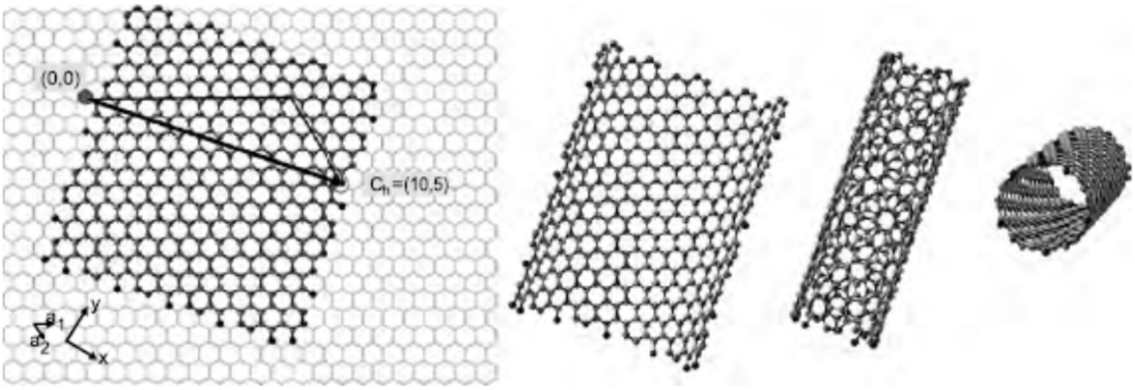


Fig. 14. 6. The low-temperature structure of RbC_{60}. The polymerized C_{60} chains are quite visible.

14.3. Carbon Nanotubes

14.3.1. Molecular structure

If a long strip of graphene is cut out and bent into a cylinder, so that loose carbon bonds match up, we get a nanotube. The matching carbon atoms can be labeled by a chiral vector $\mathbf{C_h}$. This vector is represented by two basic vectors ($\mathbf{a_1}$ and $\mathbf{a_2}$) and by two integers (n_1 and n_2); $\mathbf{C_h}=n_1\mathbf{a_1}+n_2\mathbf{a_2}$ (Fig. 14.7). Accordingly, the structure of every nanotube is unambiguously characterized by the $\mathbf{C_h}$ vector; in a shorthand notation $\mathbf{C_h} = (n_1, n_2)$, so that $n_1 \geq n_2$. Based on the patterns visible on the nanotube they may be called ‘‘zigzag’’ (for $n_2=0$; Fig. 14.2), ‘‘armchair’’ (for $n_1=n_2$), or ‘‘chiral’’ (for any other, see, for example, Fig. 14.7). The diameter of the nanotube is given by $d = (a/\pi)[3(n_1^2 + n_1 n_2 + n_2^2)]^{1/2}$, where a=0.144 nm is the carbon–carbon distance. There is no upper limit on the size of the nanotube, although at large diameter the structure becomes very soft (except for multiwalled nanotubes (MWNT)). For small diameter the carbon–carbon bonds are more strained, and the thinnest stable nanotube corresponds to $\mathbf{C_h}$=(5,5) (Fig. 14.8).



``

Fig. 14.7. Schematic illustration of deriving a nanotube structure from graphene. The unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, and the chiral vector $\mathbf{C}_h$ are indicated in the first panel. Bending the graphene strip results in the nanotube shown in the last panel.

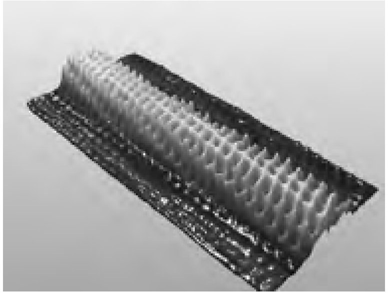


Fig. 14.8. Scanning tunneling microscope image of a nanotube. The peaks correspond to individual carbon atoms; the chirality of the structure is quite visible [44].

The nanotubes derived from a single graphene strip are called single-walled nanotubes (SWNT) [5, 6]. If a set of subsequently smaller diameter nanotubes are stacked like concentric cylinders, a MWNT is obtained (Fig. 14.9). The distance between the carbon layers in an MWNT is about the same as the distance between the graphene layers in graphite. MWNTs are produced more commonly, and the very first report of nanotubes by Iijima was about MWNTs [4].

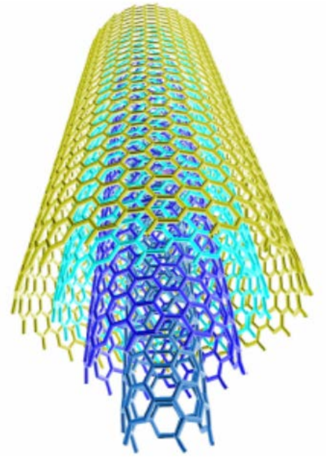


Fig. 14.9. Multiwalled nanotubes (MWNT).

Nanotubes generally have a length to diameter ratio of about 1000 so they can be considered as nearly one-dimensional structures. More detailed, a SWNT consists of two separate regions with different physical and chemical properties. The first is the sidewall of the tube and the second is the end cap of the tube. The end cap structure is similar to or derived from a smaller fullerene, such as C_{60}.

``

C-atoms placed in hexagons and pentagons form the end cap structures. It can be easily derived from Euler's theorem that twelve pentagons are needed in order to obtain a closed cage structure which consists of only pentagons and hexagons. The combination of a pentagon and five surrounding hexagons results in the desired curvature of the surface to enclose a volume. A second rule is the isolated pentagon rule that states that the distance between pentagons on the fullerene shell is maximised in order to obtain a minimal local curvature and surface stress, resulting in a more stable structure. The smallest stable structure that can be made this way is C_{60} the one just larger is C_{70} and so on. Another property is that all fullerenes are composed of an even number of C-atoms because adding one hexagon to an existing structure means adding two C-atoms.

The other structure of which a SWNT is composed is a cylinder. It is generated when a graphene sheet of a certain size that is wrapped in a certain direction. As the result is cylinder symmetric we can only roll in a discreet set of directions in order to form a closed cylinder (Fig. 14.7). Two atoms in the graphene sheet are chosen, one of which servers the role as origin. The sheet is rolled until the two atoms coincide. The vector pointing from the first atom towards the other is called the chiral vector and its length is equal to the circumference of the nanotube (Fig. 14.10). The direction of the nanotube axis is perpendicular to the chiral vector. SWNTs with different chiral vectors have dissimilar properties such as optical activity, mechanical strength and electrical conductivity.

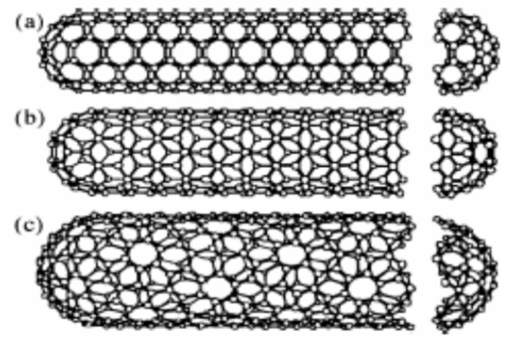


Fig. 14.10. Some SWNTs with different chiralities. The difference in structure is easily shown at the open end of the tubes: a) armchair structure, b) zigzag structure, c) chiral structure.

A nanotube rope [17] is a parallel arrangement of SWNTs in a crystalline order, as shown in Fig. 14.11. The rope is held together by weak van der Waals forces (just like graphite), but it is extremely strong along the long the direction. SWNT strands as long as a few centimeters have been produced [45].

``

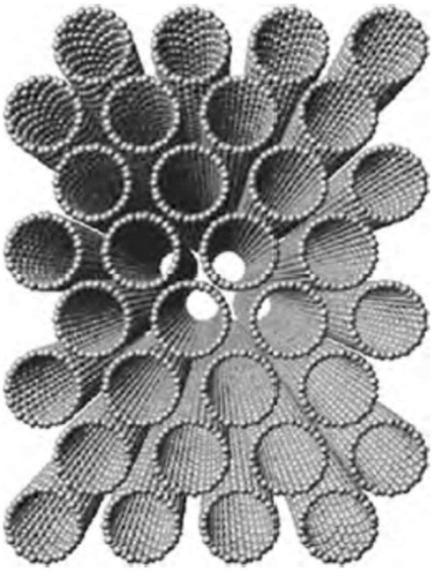
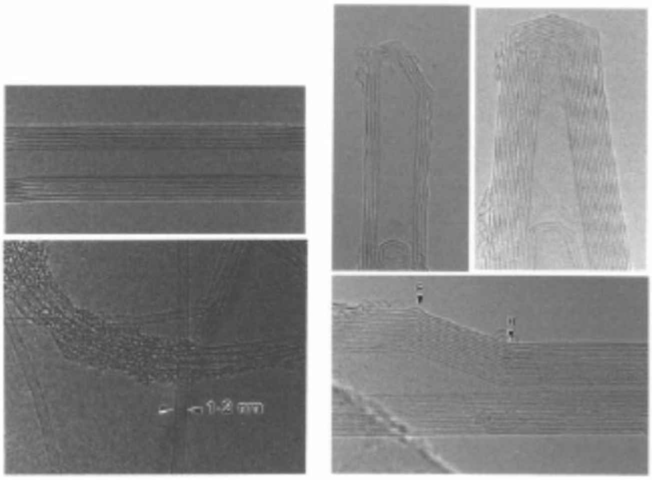


Fig. 14.11. Cross-sectional view of a nanotube rope.

After ideal structures without flaws, we discuss the possible desirable or undesirable defects. Deformations, such as bends and nanotube junctions, are introduced by replacing a hexagon with a heptagon or pentagon. Deformations can be inward or outward and, among others, electrical properties are seriously changed by these deformations. Another class of defects is caused by impurities that are built in during or after the nanotube growth process; Compounds that can be incorporated into the structure are for example catalyst particles.

Multi Walled Nanotubes (MWNT) can be considered as a collection of concentric SWNTs with different diameters. The length and diameter of these structures differ a lot from those of SWNTs and, of course, their properties are also very different (Fig. 14.12). In Fig. 14.12 carbon cones are also shown. It can be considered as a gradual transition from a large diameter to a smaller one without defects in the wall of the cone but with fewer pentagons in the end cap.



``

Fig. 14.12. Different structures of MWNTs. Top-left: cross-section of a MWNT the different walls are obvious, they are separated by 0.34nm. Rotation around the symmetry axis gives us the MWNT. Top-right: Symmetrical or non-symmetrical cone shaped end caps of MWNTs. Bottom-left: A SWNT with a diameter of 1.2nm and a bundle of SWNTs covered with amorphous carbon. Bottom-right: A MWNT with defects. In point P a pentagon defect and in point H a heptagon defect [46].

Introduction of defects can also result in various new structures such as Y-branches T-branches or SWNT junctions (Fig. 14.13). Under certain circumstances, these defects can be introduced in a 'controlled' way. These defects result in special structures that will have other, but even more interesting, properties than their original forms.

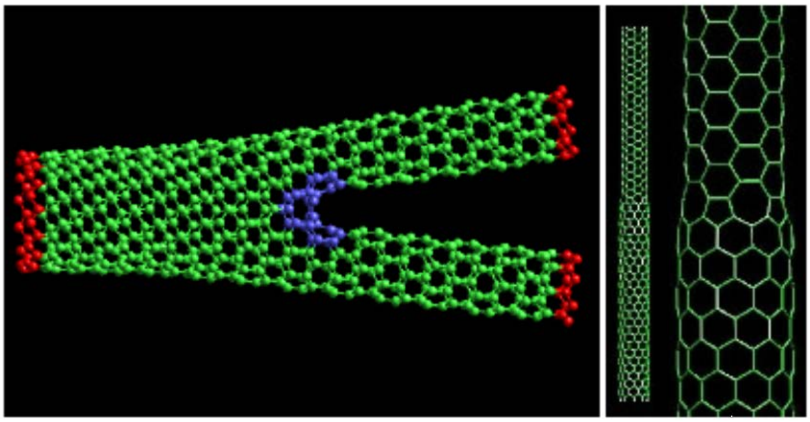


Fig. 14.13. Left: A Y-branch, the defects are marked in blue. Right: A transition from a metallic to a semiconducting SWNT. The change is made by insertion of pentagons and heptagons.

A final type of interesting structures is the so-called peapods, carbon nanotubes with C_{60} molecules enclosed in the nanotube.

14.3.2. Synthesis of CNT

Carbon nanotubes are generally produced by three main techniques, arc discharge, laser ablation and chemical vapour deposition. Scientists are researching more economic ways to produce these structures. In arc discharge, a vapour is created by an arc discharge between two carbon electrodes with or without catalyst. Nanotubes self-assemble from the resulting carbon vapour. In the laser ablation technique, a high-power laser beam impinges on a volume of carbon –containing feedstock gas (methane or carbon monoxide). At the moment, laser ablation produces a small amount of clean nanotubes, whereas arc discharge methods generally produce large quantities of impure material. In general, chemical vapour deposition (CVD) results in MWNTs or poor quality

``

SWNTs. The SWNTs produced with CVD have a large diameter range, which can be poorly controlled. But on the other hand, this method is very easy to scale up, what favours commercial production.

Growth mechanism. The way in which nanotubes are formed is not exactly known. The growth mechanism is still a subject of controversy, and more than one mechanism might be operative during the formation of CNTs. One of the mechanisms consists out of three steps. First a precursor to the formation of nanotubes and fullerenes, C_2, is formed on the surface of the metal catalyst particle. From this metastable carbide particle, a rodlike carbon is formed rapidly. Secondly there is a slow graphitisation of its wall. This mechanism is based on in-situ TEM observations (Fig. 14.14) [47].

The exact atmospheric conditions depend on the technique used later on these will be explained for each technique as they are specific for a technique. The actual growth of the nanotube seems to be the same for all techniques mentioned.

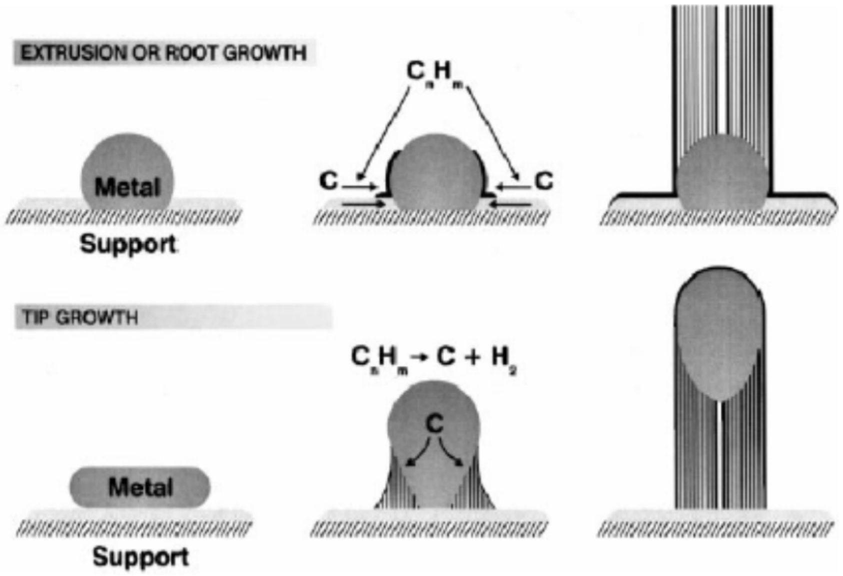


Fig. 14.14. Visualisation of a possible carbon nanotube growth mechanism.

There are several theories on the exact growth mechanism for nanotubes. One theory [48] postulates that metal catalyst particles are floating or are supported on graphite or another substrate. It presumes that the catalyst particles are spherical or pear-shaped, in which case the deposition will take place on only one half of the surface (this is the lower curvature side for the pear shaped particles). The carbon diffuses along the concentration gradient and precipitates on the opposite half, around and below the bisecting diameter. However, it does not precipitate from the apex of the hemisphere, which accounts for the hollow core that is characteristic of these filaments. For supported metals, filaments can form either by ‘extrusion (also known as base growth)’ in which the

``

nanotube grows upwards from the metal particles that remain attached to the substrate, or the particles detach and move at the head of the growing nanotube, labelled ‘tip-growth’. Depending on the size of the catalyst particles, SWNT or MWNT are grown. In arc discharge, if no catalyst is present in the graphite, MWNT will be grown on the C_2-particles that are formed in the plasma.

Arc discharge. The carbon arc discharge method, initially used for producing C_{60} fullerenes, is the most common and perhaps easiest way to produce carbon nanotubes as it is rather simple to undertake. However, it is a technique that produces a mixture of components and requires separating nanotubes from the soot and the catalytic metals present in the crude product.

This method creates nanotubes through arc-vaporisation of two carbon rods placed end to end, separated by approximately 1 mm, in an enclosure that is usually filled with inert gas (helium, argon) at low pressure (between 50 and 700 mbar). Recent investigations have shown that it is also possible to create nanotubes with the arc method in liquid nitrogen [49]. A direct current of 50 to 100 A driven by approximately 20 V creates a high temperature discharge between the two electrodes. The discharge vaporises one of the carbon rods and forms a small rod shaped deposit on the other rod. Producing nanotubes in high yield depends on the uniformity of the plasma arc and the temperature of the deposit form on the carbon electrode [50].

Insight in the growth mechanism is increasing and measurements have shown that different diameter distributions have been found depending on the mixture of helium and argon. These mixtures have different diffusions coefficients and thermal conductivities. These properties affect the speed with which the carbon and catalyst molecules diffuse and cool, affecting nanotube diameter in the arc process. This implies that single-layer tubules nucleate and grow on metal particles in different sizes depending on the quenching rate in the plasma and it suggests that temperature and carbon and metal catalyst densities affect the diameter distribution of nanotubes [50].

Depending on the exact technique, it is possible to selectively grow SWNTs or MWNTs, which is shown in Fig. 14.15. Two distinct methods of synthesis can be performed with the arc discharge apparatus.

``

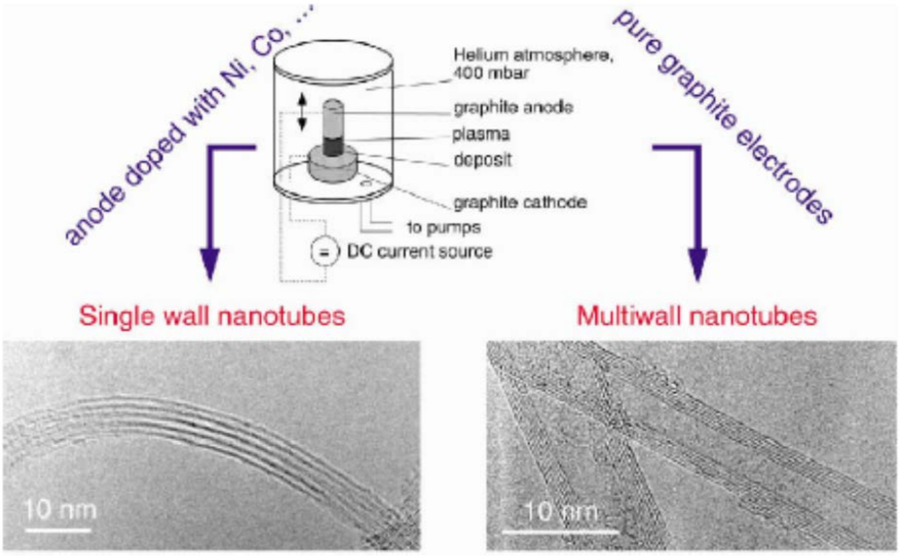


Fig. 14.15. Experimental set-up of an arc discharge apparatus.

Synthesis of SWNT. If SWNTs are preferable, the anode has to be doped with metal catalyst, such as Fe, Co, Ni, Y or Mo. A lot of elements and mixtures of elements have been tested by various authors [51] and it is noted that the results vary a lot, even though they use the same elements. This is not surprising as experimental conditions differ. The quantity and quality of the nanotubes obtained depend on various parameters such as the metal concentration, inert gas pressure, kind of gas, the current and system geometry. Usually the diameter is in the range of 1.2 to 1.4 nm. A couple of ways to improve the process of arc discharge are stated below.

The most common problems with SWNT synthesis are that the product contains a lot of metal catalyst, SWNTs have defects and purification is hard to perform. On the other hand, an advantage is that the diameter can slightly be controlled by changing thermal transfer and diffusion, and hence condensation of atomic carbon and metals between the plasma and the vicinity of the cathode can control nanotube diameter in the arc process. This was shown in an experiment in which different mixtures of inert gases were used [52]. It appeared that argon, with a lower thermal conductivity and diffusion coefficient, gave SWNTs with a smaller diameter of approximately 1.2 nm. A linear fit of the average nanotube diameter showed a 0.2 nm diameter decrease per 10 % increase in argon helium ratio, when nickel/yttrium was used (C/Ni/Y was 94.8:4.2:1) as catalyst.

A second way of control is plasma control by changing the anode to cathode distance (ACD). The ACD is adjusted in order to obtain strong visible vortices around the cathode. This enhances anode vaporisation, which improves nanotubes formation. Combined with controlling the argon-helium mixture, one can simultaneously control the macroscopic and microscopic parameters of the nanotubes formed. With a nickel and yttrium catalyst (C/Ni/Y is 94.8:4.2:1) the optimum

``

nanotube yield was found at a pressure of 660 mbar for pure helium and 100 mbar for pure argon. The nanotube diameter ranges from 1.27 to 1.37 nanometre.

Knowing that chemical vapour deposition (CVD) could give SWNTs with a diameter of 0.6–1.2 nm, researchers tried the same catalyst as used in CVD on arc discharge. Not all of the catalysts used appeared to result in nanotubes for both methods. But there seemed to be a correlation of diameter of SWNTs synthesised by CVD and arc discharge As a result, the diameter can be controllably lowered to a range of 0.6-1.2 nm with arc-discharge. Using a mixture of Co and Mo in high concentrations as catalyst resulted in this result. These diameters are considerably smaller than 1.2-1.4 nm [51], which is the usual size gained from arcdischarge.

There is also progress in developing methods to improve the oxidation resistance of the SWNTs, which is a consequence of the defects present in nanotubes. A strong oxidation resistance is needed for the nanotubes if they have to be used for applications such as field emission displays. Recent research has indicated that a modified arc-discharge method using a bowl-like cathode (see Fig. 14.16), decreases the defects and gives cleaner nanotubes, and thus improves the oxidation resistance [53]. The Raman spectrum of the newly synthesised nanotubes shows that the nanotubes formed are cleaner and less defective compared with those synthesised by conventional methods. The anode rod contained Ni and Y catalyst (C /Ni/Y is 94.8:4.2:1). No information is given about the diameter size.

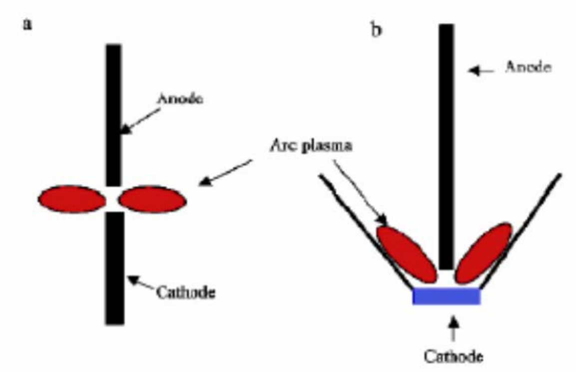


Fig. 14.16. Schematic drawings of the electrode set-ups for (a) the conventional and (b) the new arc discharge electrodes.

The researchers discovered that it was possible to form MWNTs in open air [54]. A welding arc torch was operated in open air and the process was shielded with an argon gas flow. The anode and cathode were made of graphite containing Ni and Y (Ni/Y is 4.2:1 at. %). This method was modified for preparing SWNTs [55]. A plate target made of graphite containing metal catalyst Ni and Y (Ni/Y is 3.6:0.8 at. per cent), was fixed at the sidewall of a water–cooled, steel based electrode. The torch arc aimed at the edge of the target and the soot was deposited on the substrate behind the target (see Fig. 14.17). The arc was operated at a direct current of 100 A and shielding argon gas flowed through the torch, enhancing the arc jet formation beyond the target. In the soot, carbon nanohorns (CNHs) and bundles of SWNT with an average diameter of 1.32 nm were found. However, the yield was much lower than for the conventional low-pressure arc discharge method. There are two reasons for this fact. At first, because of the open air, the lighter soot will escape into the atmosphere. Secondly, the carbon vapour might be oxidised and emitted as carbon dioxide gas. In order to improve the yield in this method, contrivances for collecting soot and development of an appropriate target are required. This method promises to be convenient and inexpensive once the conditions for higher yield are optimised. With a Ni/Y catalyst (Ni/Y is 3.6:0.8), SWNT bundles and CNHs are formed. In this case the SWNTs have a diameter of approximately 1.32 nm [55].

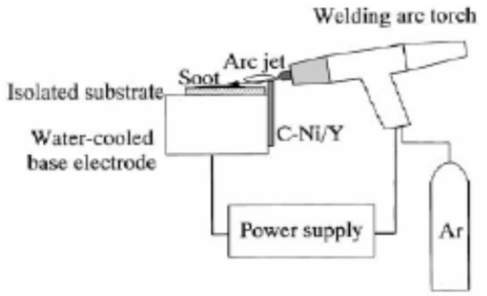


Fig. 14.17. Experimental set-up of the torch arc method in open air.

Synthesis of MWNT. If both electrodes are graphite, the main product will be MWNTs. But next to MWNTs a lot of side products are formed such as fullerenes, amorphous carbon, and some graphite sheets. Purifying the MWNTs, means loss of structure and disorders the walls. However scientist are developing ways to gain pure MWNTs in a large-scale process without purification. Typical sizes for MWNTs are an inner diameter of 1-3 nm and an outer diameter of approximately 10 nm. Because no catalyst is involved in this process, there is no need for a heavy acidic purification step. This means, the MWNT, can be synthesised with a low amount of defects.

A first, possibly economical route to highly crystalline MWNTs is the arc-discharge method in liquid nitrogen [49], with this route mass production is also possible (Fig. 14.18). For this option low pressures and expensive inert gasses are not needed. The content of the MWNTs can be as high as 70 % of the reaction product. Analysis with Auger spectroscopy revealed that no nitrogen was incorporated in the MWNTs. There is a strong possibility that SWNTs can be produced with the same apparatus under different conditions.

``

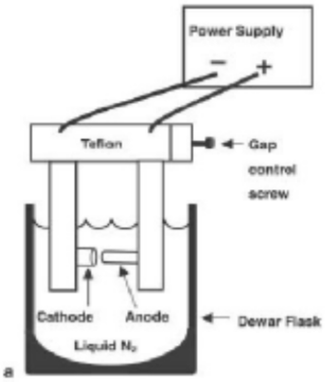


Fig. 14.18. Schematic drawings of the arc discharge apparatus in liquid nitrogen.

Synthesis of MWNTs in a magnetic field [56] gives defect-free and high purity MWNTs that can be applied as nanosized electric wires for device fabrication. In this case, the arc discharge synthesis was controlled by a magnetic field around the arc plasma. Extremely pure graphite rods (purity > 99.999 %) were used as electrodes. Highly pure MWNTs (purity > 95 %) were obtained without further purification, which disorders walls of MWNTs.

A second possibly economical route to mass production of MWNTs is synthesis by plasma rotating arc discharge technique (Fig. 14.19) [57]. The centrifugal force caused by the rotation generates turbulence and accelerates the carbon vapour perpendicular to the anode. In addition, the rotation distributes the micro discharges uniformly and generates a stable plasma. Consequently, it increases the plasma volume and raises the plasma temperature. At a rotation speed of 5000 rpm a yield of 60 % was found at a formation temperature of 1025 °C without the use of a catalyst. The yield increases up to 90% after purification if the rotation speed is increased and the temperature is enlarged to 1150 °C. The diameter size was not mentioned in this publication.

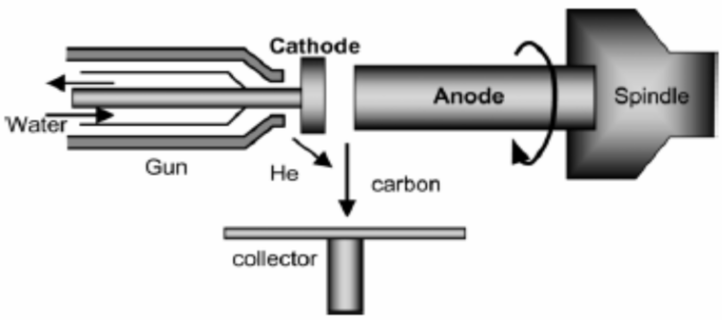


Fig. 14.19. Schematic diagram of plasma rotating electrode system.

Laser ablation. In 1995, Smalley's group [58] at Rice University reported the synthesis of carbon nanotubes by laser vaporisation. The laser vaporisation apparatus used by Smalley's group is shown in Fig. 14.20. A pulsed [59, 60], or continuous [61, 62] laser is used to vaporise a graphite target in an oven at 1200 °C. The main difference between continuous and pulsed laser, is that the

``

pulsed laser demands a much higher light intensity (100 kW/cm^2 compared with 12 kW/cm^2). The oven is filled with helium or argon gas in order to keep the pressure at 500 Torr. A very hot vapour plume forms, then expands and cools rapidly. As the vaporised species cool, small carbon molecules and atoms quickly condense to form larger clusters, possibly including fullerenes. The catalysts also begin to condense, but more slowly at first, and attach to carbon clusters and prevent their closing into cage structures [63]. Catalysts may even open cage structures when they attach to them. From these initial clusters, tubular molecules grow into single-wall carbon nanotubes until the catalyst particles become too large, or until conditions have cooled sufficiently that carbon no longer can diffuse through or over the surface of the catalyst particles. It is also possible that the particles become that much coated with a carbon layer that they cannot absorb more and the nanotube stops growing. The SWNTs formed in this case are bundled together by van der Waals forces [63].

There are some striking, but not exact similarities, in the comparison of the spectral emission of excited species in laser ablation of a composite graphite target with that of laser-irradiated C_{60} vapour. This suggests that fullerenes are also produced by laser ablation of catalyst-filled graphite, as is the case when no catalysts are included in the target. However, subsequent laser pulses excite fullerenes to emit C_2 that adsorbs on catalyst particles and feeds SWNT growth. However, there is insufficient evidence to conclude this with certainty.

Laser ablation is almost similar to arc discharge, since the optimum background gas and catalyst mix is the same as in the arc discharge process. This might be due to very similar reaction conditions needed, and the reactions probably occur with the same mechanism.

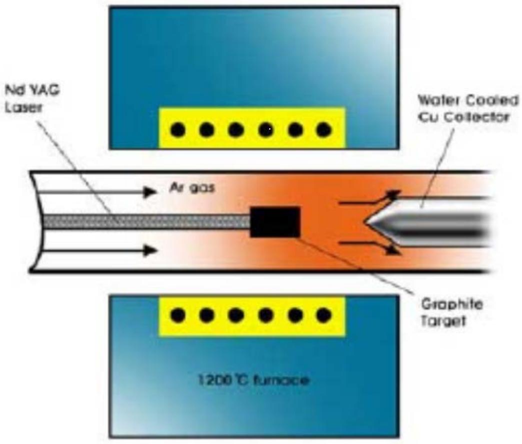


Fig. 14.20. Schematic drawings of a laser ablation apparatus.

SWNT versus MWNT. The condensates obtained by laser ablation are contaminated with carbon nanotubes and carbon nanoparticles. In the case of pure graphite electrodes, MWNTs would be synthesised, but uniform SWNTs could be synthesised if a mixture of graphite with Co, Ni, Fe or

``

Y was used instead of pure graphite. SWNTs synthesised this way exist as 'ropes' (Fig. 14.21) [61, 63]. Laser vaporisation results in a higher yield for SWNT synthesis and the nanotubes have better properties and a narrower size distribution than SWNTs produced by arc-discharge. Nanotubes produced by laser ablation are purer (up to about 90 % purity) than those produced in the arc discharge process. The Ni/Y mixture catalyst (Ni/Y is 4.2/1) gave the best yield. The size of the SWNTs ranges from 1-2 nm, for example the Ni/Co catalyst with a pulsed laser at 1470 °C gives SWNTs with a diameter of 1.3-1.4 nm26. In case of a continuous laser at 1200 °C and Ni/Y catalyst (Ni/Y is 2:0.5 at. %), SWNTs with an average diameter of 1.4 nm were formed with 20-30 % yield, see Fig. 14.21 [61].

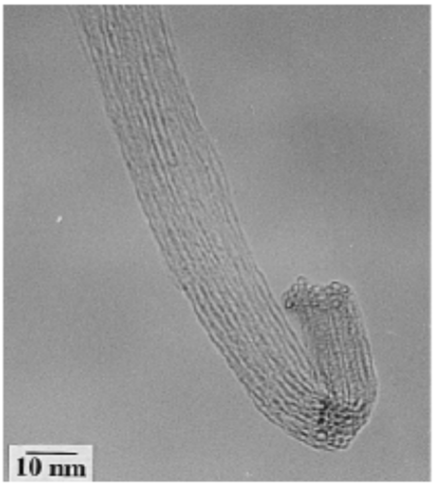


Fig. 14.21. TEM images of a bundle of SWNTs catalysed by Ni/Y (2:0.5 at. %) mixture, produced with a continuous laser.

Chemical vapour deposition. Chemical vapour deposition (CVD) synthesis is achieved by putting a carbon source in the gas phase and using an energy source, such as a plasma or a resistively heated coil, to transfer energy to a gaseous carbon molecule. Commonly used gaseous carbon sources include methane, carbon monoxide and acetylene. The energy source is used to “crack” the molecule into reactive atomic carbon. Then, the carbon diffuses towards the substrate, which is heated and coated with a catalyst (usually a first row transition metal such as Ni, Fe or Co) where it will bind. Carbon nanotubes will be formed if the proper parameters are maintained. Excellent alignment [64], as well as positional control on nanometer scale [65], can be achieved by using CVD. Control over the diameter, as well as the growth rate of the nanotubes can also be maintained. The appropriate metal catalyst can preferentially grow single rather than multi-walled nanotubes [48].

``

CVD carbon nanotube synthesis is essentially a two-step process consisting of a catalyst preparation step followed by the actual synthesis of the nanotube. The catalyst is generally prepared by sputtering a transition metal onto a substrate and then using either chemical etching or thermal annealing to induce catalyst particle nucleation. Thermal annealing results in cluster formation on the substrate, from which the nanotubes will grow. Ammonia may be used as the etchant [60 - 63]. The temperatures for the synthesis of nanotubes by CVD are generally within the 650–900 oC range [64 - 67]. Typical yields for CVD are approximately 30%.

These are the basic principles of the CVD process. In the last decennia, different techniques for the carbon nanotubes synthesis with CVD have been developed, such as plasma enhanced CVD, thermal chemical CVD, alcohol catalytic CVD, vapour phase growth, aero gel-supported CVD and laser assisted CVD.

Plasma enhanced chemical vapour deposition. The plasma enhanced CVD method generates a glow discharge in a chamber or a reaction furnace by a high frequency voltage applied to both electrodes. Figure 14.22 shows a schematic diagram of a typical plasma CVD apparatus with a parallel plate electrode structure.

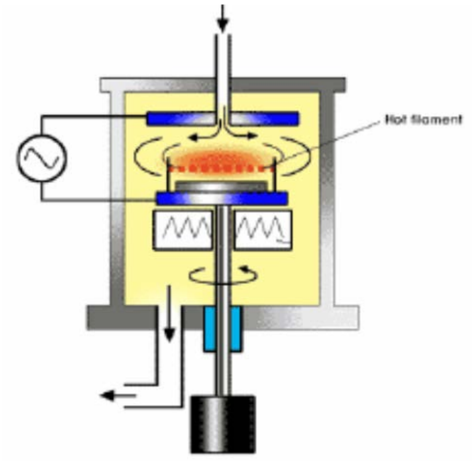


Fig. 14.22. Schematic diagram of plasma CVD apparatus.

A substrate is placed on the grounded electrode. In order to form a uniform film, the reaction gas is supplied from the opposite plate. Catalytic metal, such as Fe, Ni and Co are used on for example a Si, SiO_2, or glass substrate using thermal CVD or sputtering. After nanoscopic fine metal particles are formed, carbon nanotubes will be grown on the metal particles on the substrate by glow discharge generated from high frequency power. A carbon containing reaction gas, such as C_2H_2, CH_4, C_2H_4, C_2H_6, CO is supplied to the chamber during the discharge [68].

The catalyst has a strong effect on the nanotube diameter, growth rate, wall thickness, morphology and microstructure. Ni seems to be the most suitable pure-metal catalyst for the growth of aligned multiwalled carbon nanotubes (MWNTs) [69]. The diameter of the MWNTs is approximately 15 nm. The highest yield of carbon nanotubes achieved was about 50% and was obtained at relatively low temperatures (below 330 °C) [68].

Thermal chemical vapour deposition. In this method Fe, Ni, Co or an alloy of the three catalytic metals is initially deposited on a substrate. After the substrate is etched in a diluted HF solution with distilled water, the specimen is placed in a quartz boat. The boat is positioned in a CVD reaction furnace, and nanometre-sized catalytic metal particles are formed after an additional etching of the catalytic metal film using NH_3 gas at a temperature of 750 to 1050°C. As carbon nanotubes are grown on these fine catalytic metal particles in CVD synthesis, forming these fine catalytic metal particles is the most important process. Figure 14.23 shows a schematic diagram of thermal CVD apparatus in the synthesis of carbon nanotubes.

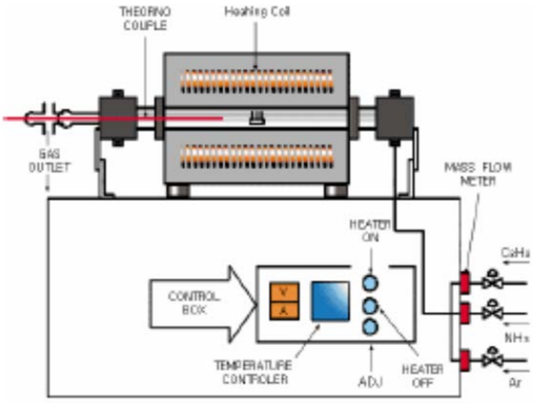


Fig. 14.23. Schematic diagram of thermal CVD apparatus.

When growing carbon nanotubes on a Fe catalytic film by thermal CVD, the diameter range of the carbon nanotubes depends on the thickness of the catalytic film. By using a thickness of 13 nm, the diameter distribution lies between 30 and 40 nm. When a thickness of 27 nm is used, the diameter range is between 100 and 200 nm. The carbon nanotubes formed are multiwalled [70].

Flame synthesis. This method is based on the synthesis of SWNTs in a controlled flame environment, that produces the temperature, forms the carbon atoms from the inexpensive hydrocarbon fuels and forms small aerosol metal catalyst islands [71 - 73]. SWNTs are grown on these metal islands in the same manner as in laser ablation and arc discharge.

These metal catalyst islands can be made in three ways. The metal catalyst (cobalt) can either be coated on a mesh [71], on which metal islands resembling droplets were formed by physical vapour deposition. These small islands become aerosol after exposure to a flame. The second way [73], is to create aerosol small metal particles by burning a filter paper that is rinsed with a metal-ion (e.g. iron nitrate) solution. The third way is the thermal evaporating technique in which metal powder (e.g. Fe or Ni) is inserted in a trough and heated [72].

In a controlled way a fuel gas is partially burned to gain the right temperature of ~800 °C and the carbon atoms for SWNT production. On the small metal particles the SWNTs are than formed (Fig. 14.24). As optimisation parameters the fuel gas composition, catalyst, catalyst carrier surface and temperature can be controlled [71].

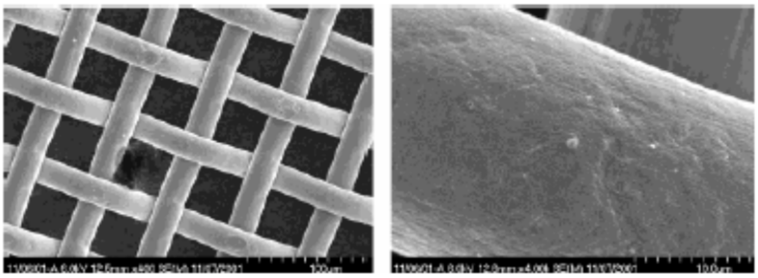


Fig. 14.24. Meshes on which the metal catalyst is coated, used in flame synthesis.

In Table 14.1, a short summary of the three most common CNT obtaining techniques used nowadays is given.

Table 14.1. A summary of the major CNT production methods and their efficiency.

Method	**Arc discharge method**	**Chemical vapour deposition**	**Laser ablation (vaporization)**
Who	Ebbesen and Ajayan, NEC, Japan 1992 15	Endo, Shinshu University, Nagano, Japan 53	Smalley, Rice, 199514
How	Connect two graphite rods to a power supply, place them a few millimetres apart, and throw the switch. At 100 amps, carbon vaporises and forms a hot plasma.	Place substrate in oven, heat to 600 °C, and slowly add a carbon-bearing gas such as methane. As gas decomposes it frees up carbon atoms, which recombine in the form of NTs	Blast graphite with intense laser pulses; use the laser pulses rather than electricity to generate carbon gas from which the NTs form; try various conditions until hit on one that produces prodigious amounts of SWNTs
Typical yield	30 to 90%	20 to 100 %	Up to 70%

``

SWNT	Short tubes with diameters of 0.6 - 1.4 nm	Long tubes with diameters ranging from 0.6-4 nm	Long bundles of tubes (5-20 microns), with individual diameter from 1-2 nm.
MWNT	Short tubes with inner diameter of 1-3 nm and outer diameter of approximately 10 nm	Long tubes with diameter ranging from 10-240 nm	Not very much interest in this technique, as it is too expensive, but MWNT synthesis is possible.
Pro	Can easily produce SWNT, MWNTs. SWNTs have few structural defects; MWNTs without catalyst, not too expensive, open air synthesis possible	Easiest to scale up to industrial production; long length, simple process, SWNT diameter controllable, quite pure	Primarily SWNTs, with good diameter control and few defects. The reaction product is quite pure
Con	Tubes tend to be short with random sizes and directions; often needs a lot of purification	NTs are usually MWNTs and often riddled with defects	Costly technique, because it requires expensive lasers and high power requirement, but is improving

14.3.3. Electrical Conduction in Nanotubes

Depending on their chiral vector, carbon nanotubes with a small diameter are either semi-conducting or metallic. The differences in conducting properties are caused by the molecular structure that results in a different band structure and thus a different band gap. The differences in conductivity can easily be derived from the graphene sheet properties [74]. It was shown that a (n,m) nanotube is metallic as accounts that: $n=m$ or $(n-m) = 3i$, where i is an integer and n and m are defining the nanotube. The resistance to conduction is determined by quantum mechanical aspects and was proved to be independent of the nanotube length [75, 76].

In order to understand the electrical properties of a nanotube, we first have to look at a single graphene sheet. Graphene is on the borderline of being metallic: the band structure calculation results in a ‘‘zero-gap semiconductor’’ behavior [77]. Fig. 14.25 shows an illustration of the electron energy as a function of the two-dimensional wave vector. The Brillouin zone of graphene is a hexagon, and the allowed electronic states (wave vectors) correspond to the points inside the hexagon; the electronic bands reach the Fermi level at the six corners of the hexagon. Therefore the Fermi ‘‘surface’’ consists of six points, as indicated in the two panels in Fig. 14.26. (In graphite the influence of the neighboring layers causes the bands to cross the Fermi level, and there will be finite Fermi surface, yielding a semimetallic behavior.)

``

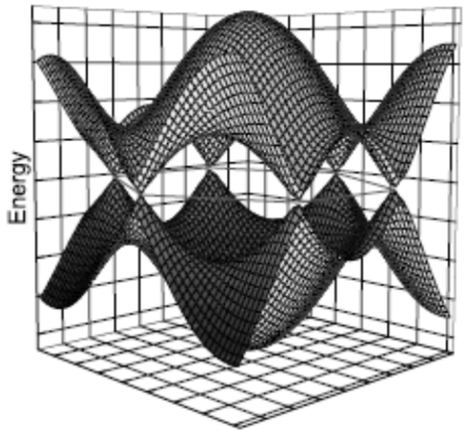


Fig. 14.25. Schematic representation of the energy bands in graphene. The possible electron states are constrained in a hexagon in the base plane. The lower band is full, the upper band is empty. The energy of the electrons is zero (equal to the Fermi energy) at the six corners of the hexagon.

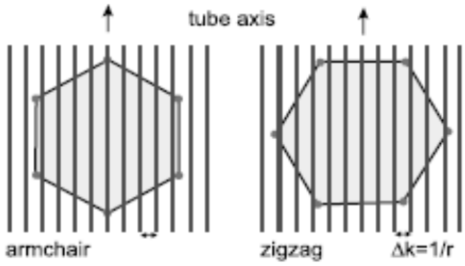


Fig. 14.26. The six corners of the hexagon correspond to electronic states available for conduction. The lines represent the allowed electronic states in a zigzag and an armchair nanotube.

By bending the graphene sheet into a nanotube the most drastic change will be the constraint introduced due to the periodic boundary condition along the circumference of the cylinder. Accordingly, in the direction perpendicular to the tube axis the allowed wave vectors will be constrained to lines. The larger the diameter of the nanotube the more dense are the lines. Naturally, for a very large diameter nanotube the allowed wave vectors are just about the same as for the graphite, and the properties are also similar. However, for smaller diameters the constraint is really important. As shown in the two panels of Fig. 14.26, the relative orientation of the lines corresponding to the allowed states is sensitive to the actual structure of the nanotube [8, 78]. For armchair nanotubes one of the lines always intersects the corner of the hexagon, therefore these nanotubes are metallic; for the zigzag structure it is possible that the lines miss the corners, and in those cases the nanotube is semiconducting. By carrying out this analysis in more detail, we obtain that a nanotube is metallic if the integers in the chiral vector describing the structure $\mathbf{C_h}=(n_1,n_2)$ satisfy $n_1-n_2=3j$, where j is an arbitrary integer. More detailed calculations, taking into account the

``

curvature of the graphene sheet, lead to the conclusion that for $n_1=n_2$ (armchair) the nanotubes are metallic, for $n_1-n_2=3j$ there is a tiny gap at the Fermi surface, and for all other cases the gap is large.

Much progress has been made in the experimental test of these predictions [79]. One prefers to work with a SWNT, because for a MWNT it is unlikely that all of the component nanotubes are of the same type (metal or semiconductor). The first conductance measurements were done on a nanotube rope, a collection of similar SWNTs. Several single SWNTs of various chirality have been tested later. The measurements largely support the conclusions drawn above. For the complete interpretation of the experimental results one needs to step beyond the simple theory of bulk metals: the electrical leads to the nanotube must be treated as tunneling contacts. The resonance between higher quantum levels and the Coulomb repulsion of electrons on the nanotube play important roles [80, 81].

Electronic property measurements of single SWNTs. Studying the electronic properties of SWNTs, scientists have been able to calculate models for energy bands and Density of States (DOS) in single walled carbon nanotubes. If these models have to be experimentally confirmed, a way has to be found in which atomic structures and electron densities can be visualised. Scanning Tunnelling Microscopy (STM) can do this job as it can display atomic structures as well as measure the DOS.

As is shown in Fig. 14.27 and Fig. 14.28A, chirality can clearly be determined from STM measurements. Combining this fact with the ability to measure electronic properties allows studying the effect of nanotube chirality on electronic properties. The spectroscopic image (Fig. 14.28B) shows the density of states as measured by a STM. As these two images show, this technique can also be used to determine properties of intra-molecular SWNT junctions. In addition, it can also be used for measuring influence of symmetry, defects, and doping, electronic contacts and so on.

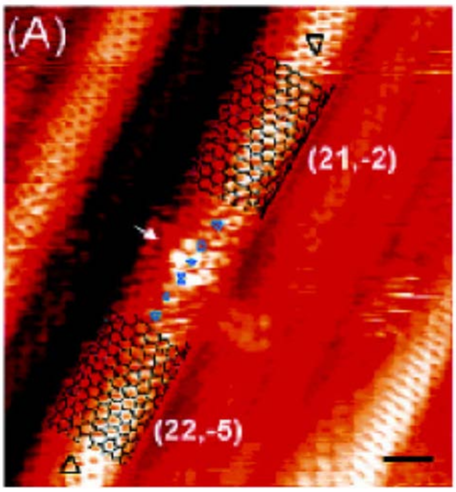


Fig. 14.27. STM image of a SWNT. The chirality can be determined easily.

``

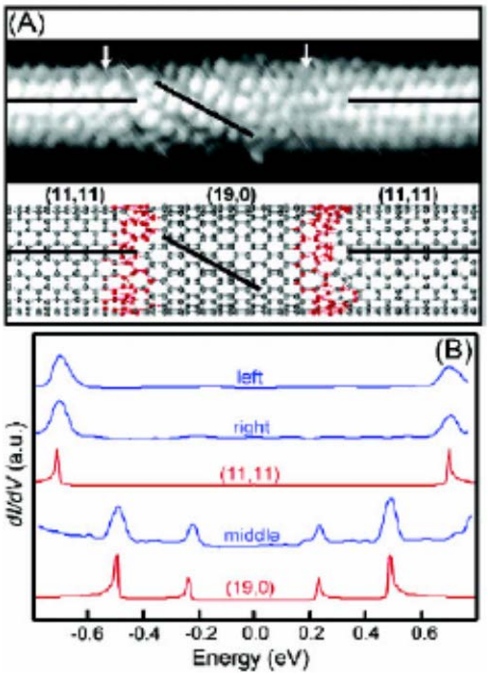


Fig. 14.28. A) STM atomic resolved image of Intra Molecular Junction of SWNT. B) STM image measurement representing the Density of States corresponding to lines in A [82].

14.3.4. Potential Applications of CNTs

Energy storage. Graphite, carbonaceous materials and carbon fibre electrodes are commonly used in fuel cells, batteries and other electrochemical applications. Advantages of considering nanotubes for energy storage are their small dimensions, smooth surface topology and perfect surface specificity. The efficiency of fuel cells is determined by the electron transfer rate at the carbon electrodes, which is the fastest on nanotubes following ideal Nernstian behaviour [76].

Two elements that can be electrochemically stored in CNTs are hydrogen and lithium. Hydrogen can also be stored in CNTs by gas phase intercalation. Three units are commonly used to describe the hydrogen and lithium contents of storage materials with:

1. [H/C] ([Li/C]) as the ratio of hydrogen (lithium) atoms per atom of storage material, in this case carbon;

2. [wt%] as the ratio of the mass of hydrogen (lithium) to the mass of storage material (the gravimetric density);

``

3. [kgH_2m^{-3}] as the ratio of the mass of molecular hydrogen to the volume of storage material (volumetric density) [83].

Electrochemical storage of hydrogen. There are two methods to store hydrogen atoms reversibly in CNTs. One method is called gas phase intercalation and it is explained in section 5.3. The second method described in this section is based on a electrochemical charge-discharge process, in which the hydrogen absorption is controlled by the potential [84].

The hydrogen storage capacity of the CNT samples is analysed by means of electrochemical galvanostatic measurement in a 6 M KOH electrolyte. There are commonly three electrodes in the setup: (*i*) a work electrode (negative), often made of a mixture of gold or nickel with the nanotube material pressed into a pellet, (*ii*) a reference electrode ($Hg/HgO/OH^-$) and (*iii*) a counter electrode, usually made of nickel. In Fig. 14.29, the reference electrode is left out. Instead, a polymer separator separates the working and the counter electrode.

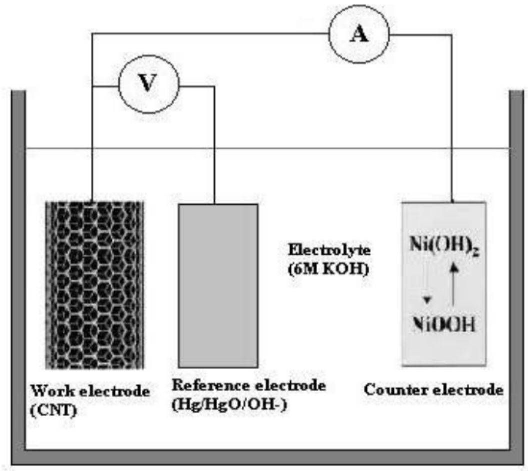


Fig. 14.29. A schematic diagram of a charge-discharge cycling apparatus [84].

During the charging process, the water in the electrolyte dissociates at the work electrode. The adsorbed atomic hydrogen may be absorbed or intercalate in the electrode or recombine at the surface to molecular hydrogen and diffuse into the electrode or form gas bubbles at the surface of the electrode. During the discharge process, the hydrogen in the electrode recombines with the OH^- ions in the electrolyte to form water molecules. The amount of hydrogen desorbed from the electrode can therefore be measured by measuring the electric charge, which is equal to the product of current and time in a galvanostatic setup [85].

The following equation gives the reaction in the working electrode:

reduction

``

$$\mathrm{CNT} + x\mathrm{H_2O} + x\mathrm{e^-} \underset{\text{oxidation}}{\leftrightarrow} \mathrm{CNT} + x\mathrm{H^+} + x\mathrm{OH^-} \qquad (14.1)$$

The reaction at the counter electrode, in the case of a Ni-electrode is given by:

$$\mathrm{Ni(OH)_2} \underset{\text{oxidation}}{\overset{\text{reduction}}{\leftrightarrow}} \mathrm{NiOOH} + \mathrm{H^+} + \mathrm{e^-} \qquad (14.2)$$

When measuring the electrochemical hydrogen storage capacity of a CNT sample, a double-layer capacitance is measured too. This effect has to be taken into account when looking at the experimental results of the capacity measurements.

Measurement methods are charge-discharge measurements, cyclic voltammetry, constant current measurements, electrochemical impedance measurements and chrono amperometry. The charge/discharge experiments are commonly performed at a constant current. During the discharge, the cell resistance leads to an overpotential. With cyclic voltammetry, different voltages are applied at a certain sweep rate and the current is measured. From the thus obtained oxidation and reduction peaks it can be speculated if the hydrogen adsorption is due to the CNTs or not. With chrono amperometry the current behaviour is studied after applying a voltage step. From impedance spectra, performed by applying sinusoidal voltage over a certain frequency range, a ‘complex’ resistance of the CNT-sample can be calculated. These spectra commonly consist of two semicircles and a slope, similar to a metal hydride [86, 87]. The smaller semicircle in the high frequency region is probably due to the double-layer capacitance. The semicircle in the low frequency region is attributed to an electrochemical reaction. These spectra of CNTs decorated with metallic nickel as a function of cycling have been studied and found that the smaller semicircle hardly changed upon cycling, while the larger semicircle remarkably increased, indicating an increase in surface reaction resistance upon cycling [86].

Several experiments have been performed in this research area with often differing conclusions. It has been demonstrated that SWNTs could electrochemically store hydrogen, in which SWNT soot containing a few percent of 0.7 – 1.2 nm diameter SWNTs was mixed with gold as a compacting powder (to stabilise the electrodes) in a 1:4 ratio to form electrodes. Gold was used because it is noble and does not participate in any electrochemical reaction. In the nanotubes-sample, nickel, iron and C_{60} were present too. Pure samples of both nickel and iron did not hydride during cycling and C_{60} only showed a low capacity of 14 mAh/g. For the nanotubes-sample they found a maximum capacity of 110 mAh/g, indicating that the nanotubes are responsible for the hydrogen uptake. This result corresponds to ~0.39 wt% hydrogen. Note that 1 Ah/g corresponds to 3.54 wt% hydrogen stored in carbon. Only slight capacity loss was observed after many cycles [88, 89].

``

The electrochemical hydrogen storage behaviour of ropes of aligned SWNTs (up to 100 mm in length and 50 mm in diameter) with a large sample quantity of 200 mg (with a purity in the range 60-70%) at 298 K under normal atmosphere has been investigated [90]. The work electrode was made without compacting powder by pressing the sample to a sheet of nickel foam. With a constant current density of 25 mA/g, the plateau of the discharge potential was observed around –0.6 V (vs Hg/HgO) and a discharge capacity of 503 mAh/g was obtained, corresponding to a hydrogen storage capacity of 1.84 wt% in SWNTs. They observed the same maximum discharge capacity for three different samples, so the results appeared reproducible. After 100 cycles the electrode still retained 80% of the maximal capacity. The loss is believed to be due to the mechanical instability of the electrode. The investigators explain the higher capacity (in comparison to the 110 mAh/g obtained in work [89]) as due to the purity and larger mean diameter of the SWNTs (1.72 nm, resulting in larger micropores). Furthermore, no compacting powder was used, resulting in a higher percentage of SWNTs in unit volume in the work electrode [90].

The electrochemical hydrogen storage behaviours of CNTs grown by chemical vapour deposition (CVD), arc-discharge (AD) and laser ablation (LA) have been compared [91]. The latter were purified (90 wt%). The charge capacities of the electrodes remained unchanged with increasing cycle number, but the reversible capacity behaviour decreased with cycling. The samples with the LA grown CNTs stored the highest amount of hydrogen, 1.6 wt% (corresponding to 400 mAh/g), which was 16 times that of CVD and AD grown CNTs. During purification, besides the elimination of metals and amorphous carbon, CNTs may separate partially from their bundles, resulting in an increase in active sites for hydrogen storage for the LA grown CNTs. They also investigated the charge-discharge behaviours for Li-, Na- and K-doped sample electrodes. Each sample obtained a higher hydrogen capacity. The Li-doped CVD electrode showed a unique discharge behaviour during cycling. With increasing cycle number, the sample showed a large capacity reversibility. The investigators concluded from cyclic voltammetric measurements that the high hydrogen storage capacity of the alkali metal-doped CNTs originated from an increase in hydrogen adsorption sites. This is suggested to be due to the introducing of the metals in the CNT bundles and separation of tubes, but not from their chemical effect [91].

Several nanotubes samples at room temperature, fabricated of a mixture of 10 mg of SWNT material with 90 mg of gold powder to form a mechanically stable and conducting electrode have been investigated [89]. Fig. 14.30 shows the charge-discharge equilibrium curves measured with the SWNT/gold electrode. The shape of the curves is different than the shape of an equilibrium curve for a metal hydride, in which there is a wide plateau due to a phase transition occurring at a specific potential. In nanotubes, there are no well-defined interstitial sites and no phase transition is expected [89].

``

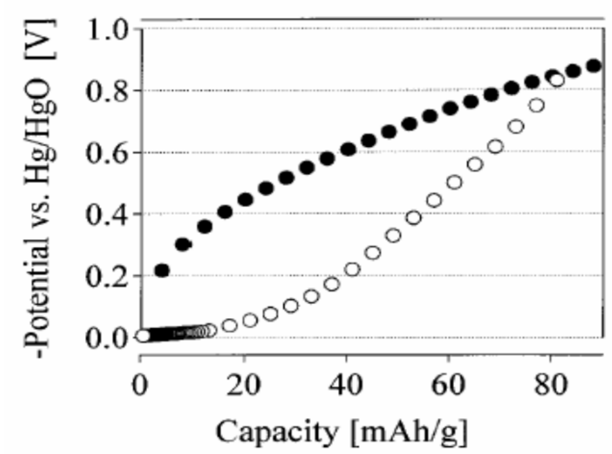


Fig. 14.30. Equilibrium charge-discharge curves of an electrode of SWNTs with gold as compacting powder (full circle: charge; open circle: discharge). The shape of the curve is different form that for metal hydrides where a plateau occurs [89].

A few SWNT samples showed hydrogen storage capacities in the range of 0.7-0.9 wt%. The fact that this reaction was reversible is a good indication that the measured capacity is due to hydrogen storage. An additional discharge from the oxidation of impurities in the nanotubes would not necessarily be reversible. The electrochemically measured discharge capacity at 293 K of the nanotube samples correlated with the surface area. The linear relationship is shown in Fig. 14.31. Note that only the round markers in the graph represent the surface area of CNT-samples, the fitted line corresponds to this data (the data from gas phase experiments, indicated by triangular markers, are not included in this fit). The line intercepts the axis at the origin and the slope is 1.5 wt%/1000 m^2/g. Theoretically, the maximum specific surface area is 2630 m^2g^{-1}. From this, they extrapolated a maximum discharge capacity of 2 wt% [85].

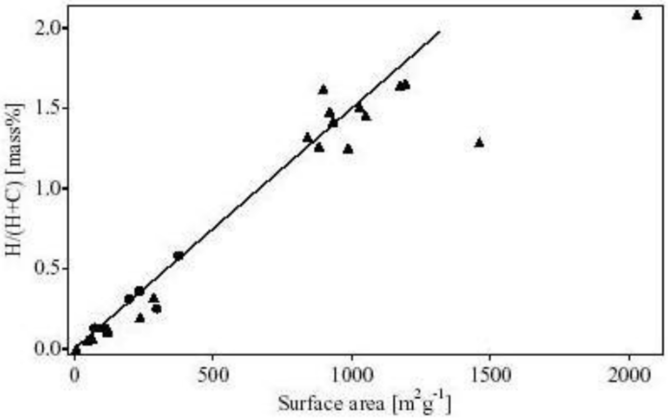


Fig. 14.31. Desorbed amount of hydrogen versus the surface area (round markers, corresponding to the electrochemical experiments) of several CNT samples with a linear fit (corresponding to the data of the round markers). The triangular markers correspond to gas phase experiments [85].

``

Less encouraging results have been obtained in work [92] where the accumulation of hydrogen by the electrochemical decomposition of aqueous alkaline medium on a negatively polarised carbon electrode under ambient conditions has been investigated. For several SWNTs the reversible hydrogen storage capacity never exceeded 0.5 wt%. They compared several published data and found no systematic relationship between the indicated purity and the maximum discharge capacity, suggesting that SWNTs are probably not responsible for the values of sorption observed. Their final conclusion is that hydrogen storage capacity of SWNTs under a high pressure at room temperature is negligible [92].

Gas phase intercalation of hydrogen. SWNTs may be used as miniaturised chemical sensors. On exposure to environments, which contain NO_2, NH_3 or O_2, the electrical resistance changes.

Gas phase intercalation of hydrogen in CNTs concerns the adsorption of H_2, called physisorption instead of chemisorption (involving H^+ and chemical bonds). This adsorption of H_2 (other gases are possible too) on the surface of CNTs is a consequence of the field force at the surface of the solid, called the adsorbent, which attracts the molecules of the gas or vapour, called adsorbate. The forces of attraction emanating from a solid can be either physical (Van der Waals) or chemical (thus chemisorption, involving the electrochemical storage of hydrogen). This section is about the storage due to the physical forces.

Carbon nanotubes have attracted considerable interest due to several reports of high hydrogen storage capacities at room temperatures, even higher than the goals set for vehicular storage by the Department of Energy (being an H_2-storage capacity of 6.5 wt% and 62 kg H_2/m^3). However, conflicting reports indicate that room temperature storage capacities do not exceed ~ 0 wt%. Reporting capacities range from 0 to 60 wt%. Several reports on hydrogen storage exist, which can be divided according to their results. The most important results found are:

1. Hydrogen storage with a H/C ratio larger than 2, thought to be incorrect (> 14 wt%). For example, a storage capacity of 67 wt% in GNFs is reported, but this seems highly irrational because this corresponds to an average of about 24 hydrogen atoms for each carbon atom;
2. Hydrogen storage consistent with expectations based on findings for activated and other, conventional high-surface area carbons (0 -2 wt%);
3. Hydrogen storage results are in the intermediate range and are not obviously incorrect (2-14 wt%).

There are several possibilities for the hydrogen to be stored in the CNT samples. For hydrogen storage in closed CNTs the structure has two possible sites: inside the tubes and in the

``

interstitial sites between the tubes array (Fig. 14.32). In case of a long SWNT closed with fullerene-like end caps, the hydrogen can only get access to the tube interior via the hexagons of the graphite-like tube wall. An opened tube with removed caps gives an easier access for the hydrogen molecules to the tube. Typically, the tubes are very long and therefore a good diffusivity of the hydrogen inside the tube will be required in order to fill the whole tube volume. Cutting the rope in shorter pieces may therefore help to improve hydrogen storage and its kinetics [93].

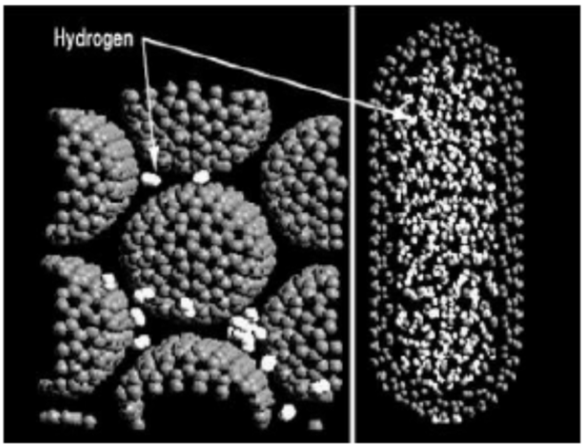


Fig. 14.32. Schematic representation showing potential sites for hydrogen adsorption within a nanotubes - bundle: (left) hydrogen atoms occupying the interstitial spaces between the tubes, and (right) hydrogen atoms inside the tube interior [94].

Mainly, three different techniques are applied to study the hydrogen storage in solids: (1) the volumetric method, (2) the gravimetric method and (3) thermal desorption spectroscopy.

(1) The volumetric method measures the pressure drop owing to hydrogen adsorption after applying a hydrogen pressure to the specimen, contained in a constant volume. Similarly the pressure increase due to desorption can be measured. For good accuracy and reliable results, this method requires typically specimen masses of 500 mg or higher. Leakage or temperature instability of the apparatus may give rise to large experimental errors. The advantage of this technique is that both adsorption and desorption can be measured.

(2) The gravimetric method measures the weight change of the specimen due to absorption or desorption of hydrogen. A high accuracy can be achieved even at sample masses of 10 mg. However, this technique is sensitive to all gasses adsorbed or desorbed since it is only based on weighing.

(3) Thermal desorption spectroscopy (TDS) measures only the hydrogen desorption in high vacuum utilising mass spectrometry. This method is selective and highly sensitive, which can be improved by using deuterium-loaded specimens. In this case no disturbing background from water or other hydrogen containing adsorbents occurs [93].

``

None of the experimental results have been confirmed by independent research groups. Obtaining activated CNT hydrogen storage materials with highly reproducible adsorption capacities has not yet been achieved. A possible reason is that hydrogen storage is only optimised for a very specific and narrow distribution of CNTs of distinct types and diameters. Although theoretical calculations predict that a range of 4-14 wt% hydrogen adsorption in carbon nanotubes is possible, they do not clearly distinguish between chemisorption and physisorption. Chemisorption (the covalent bonding of hydrogen) would require a high temperature and high activation energy for hydrogen release, whereas any practical fuel cell application would require low adsorption and desorption energies [94].

SWNTs are highly promising for H_2 adsorption, even at room temperature [96]. They measured the H_2 storage capacity of SWNTs synthesised by a hydrogen arc discharge method, with a relatively large sample quantity (about 500 mg) at ambient temperature under a modestly high pressure, which was soaked in hydrochloric acid and then heat-treated in vacuum. A H_2 uptake of 4.2 wt%, which corresponds to a H/C atom ratio of 0.52, was obtained by these SWNTs with an estimated purity of 50 to 60 wt%. Also, 78.3 % of the adsorbed hydrogen (3.3 wt%) could be released under ambient pressure at room temperature, while the release of the residual stored hydrogen (0.9 wt%) required some heating of the sample. Furthermore, they found that after four cycles of adsorption/desorption, the H_2 uptake capacity of the SWNTs remained unchanged and that less than 1 wt% could be stored at pressures lower than 5 MPa. The researchers argued that this relatively high H_2 adsorption capacity of their SWNTs could be related to their larger mean diameter of 1.85 ± 0.05 nm, while typical SWNT diameters are in the range of 1.2 to 1.4 nm [95].

Careful work at NREL (National Renewable Energy Laboratory) indicates a maximum capacity for adsorption of hydrogen on SWNTs is ~8 wt%. Samples displaying this maximum value were prepared by sonicating purified SWNTs in a dilute nitric acid solution with a high-energy probe. In this process SWNTs are cut into shorter segments and this introduces a Ti_6Al_4V alloy due to the disintegration of the ultrasonic probe. The hydrogen adsorption on many of the NREL samples can be explained fully by adsorption on the incorporated alloy. However, one-half of the samples show capacities that are too high to be explained by the presence of the alloy alone. Assuming that the alloy in the SWNT sample behaved like the pure probe-generated alloy sample of Ti-6AI-4V, the hydrogen uptake on the SWNT fraction still reached ~ 8 wt%. Cross-calibration of the experimental apparatuses with three different standards establish the validity of the measurements, and repeated measurements on a given sample yield reproducible results, but nominally similar sample preparation procedures do not repeatedly produce samples that exhibit the same hydrogen storage capacities. A possible explanation is that the degree of tube cutting, and how, where and in what form the metal particles are incorporated can vary dramatically even if

``

identical sonication parameters are employed. Earlier, they found that the desorption of hydrogen fitted first order kinetics and the activation energy for desorption was measured to be 19.6 kJ/mol. Also, adsorption measurements on samples having differing distributions indicate a link between SWNT size and type and hydrogen capacity. More recently, they developed a controlled dry cutting technique that does not employ ultrasonication (and so does not simultaneously incorporate a metal hydride alloy). Transmission electron microscopy analyses and Raman spectroscopy showed that significant cutting occurred without extensive damage to the SWNTs.

Electrochemical supercapacitors. Supercapacitors have a high capacitance and potentially applicable in electronic devices. Typically, they are comprised two electrodes separated by an insulating material that is ionically conducting in electrochemical devices. The capacity of an electrochemical supercap inversely depends on the separation between the charge on the electrode and the counter charge in the electrolyte. Because this separation is about a nanometre for nanotubes in electrodes, very large capacities result from the high nanotube surface area accessible to the electrolyte. In this way, a large amount of charge injection occurs if only a small voltage is applied. This charge injection is used for energy storage in nanotube supercapacitors [96]. Generally speaking, there is most interest in the double-layer supercapacitors and redox supercapacitors with different charge-storage modes.

In the future, supercapacitors might become an excellent means of certain kinds of energy storage. These electrochemical capacitors have a long durability (over 10^6 cycles), don't suffer from short circuit conditions, have a complete discharge and possess a high power density. Loading of a supercap can be performed at high current densities, which decreases the loading time needed. However, their energy density is lower than for conventional batteries, which is a possible drawback for possible applications. Typical electrochemical accumulators, in which compounds only take place in redox reactions, cannot fulfil these good characteristics that electrochemical capacitors have. Supercapacitors have already been applied in small-scale energy storage devices, such as in memory backup devices. Now the capability of supercapacitors with a high power density is increasing, potential applications extend to hybrid battery/supercapacitor systems [97].

Carbon in general, and especially nanotubes, form an attractive material for electrochemical applications as they have a large active surface area. In addition, carbon is a relatively cheap, low density, environmentally friendly and highly polarisable material which makes application even more attractive [98].

At first this section deals with basic processes in supercapacitors. Then, the determination of supercapacitor properties is explained briefly. Thereafter, supercapacitors based on CNTs are

``

investigated after which attention is given to the modification of CNTs in supercapacitors to improve their properties.

The basic principle of energy storage in a supercapacitor is based on creating a charge-separated state in an electrochemical double layer (Fig. 14.33). In this case, energy storage is based on the separation of charges in the double layer across the electrode/electrolyte interface. The positive electrode is electron deficient whereas the negative electrode contains a surplus of electrons. The energy (W) stored in a capacitor as a function of the voltage applied (U) and the capacity (C) is equal to:

$$W = \frac{CU^2}{2}. \qquad (14.3)$$

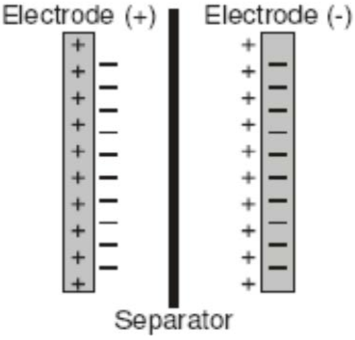


Fig. 14.33. Scheme of an electrochemical capacitor.

The electrodes of a supercapacitor must be electrochemically stable, which is the case for chemically unmodified carbon. The decomposition voltage of the electrolyte determines the maximum operating voltage of a supercapacitor. For the generation of high voltages, aprotic electrolytes with a decomposition range between 3 and 5 V should be used. However, these liquids only have a fraction of the conductivity that water has. In addition, the use of an aprotic electrolyte has technological, economical and safety barriers [98]. The final electrolyte choice depends on the demanded specific power and energy values.

Electrochemical capacitors based on carbon are of two different types depending on the type of energy storage. The first type is the electrical double layer capacitor (EDLC) where only a pure electrostatic attraction between ions and the charged surface of an electrode takes place. The second type is a supercapacitor (SC), which is additionally based on faradaic supercapacitance reactions. The total capacitance C is determined by the series capacitances of the anode (C_A) and cathode (C_C) according to the equation [98]:

$$\frac{1}{C} = \frac{1}{C_A} + \frac{1}{C_C}. \qquad (14.4)$$

``

In the EDLC, the contact between the electrode surface and the electrolyte plays an important role and determines the amount of charge stored. The capacitance C is proportional to the surface area S and the permitivity ε of the electrolyte and reciprocally dependent on the distance of charge separation:

$$C = \frac{S\varepsilon}{d}. \tag{14.5}$$

In practice, the surface area determined by gas adsorption differs from the active surface area available for charged species. When ions are solvated by water molecules, their mean diameter is approximately 15Å. This explains the need for a relatively large pore size of the electrode material for good interaction of ions with the electrode. Thermal treatment of the electrodes results in significant alteration of the pore size distribution and thus to enhanced interaction [2]. Increasing the capacitance values 10 to 100 times is possible by using pseudocapacitance effects. These depend on the surface functionality of carbon and/or on the presence of electro-active species [98].

As said before, in supercapacitors faradaic reactions similar to processes in accumulators occur. Pseudocapacitances arise when the charge q, required for the progression of an electrode process, is a continuously changing function of potential U. These pseudocapacitance effects, for example electrosorption of H or metal ad-atoms and redox reactions of electroactive species, strongly depended on the chemical affinity of carbon materials to the ions sorbed on the electrode surface [98].

An ideal double layer capacitance results in ideally rectangular shaped cyclovoltammetry diagrams. This phenomenon is ideal if the current density is independent of the potential applied and if this effect is purely electrostatic of nature. Due to redox peaks, pseudocapacitances result in deviations from this ideal shape as is shown in Fig. 14.34.

ECDL capacitors can be represented in an equivalent electrical circuit as is shown in Fig. 14.35. This circuit consists of a capacitance C, a parallel resistance R_F which is responsible for the self-discharge, an inductance L and an equivalent series resistance R_S that models the internal resistance. A maximum value of C and a minimum value of R_S result in a high value for the power density and energy value.

``

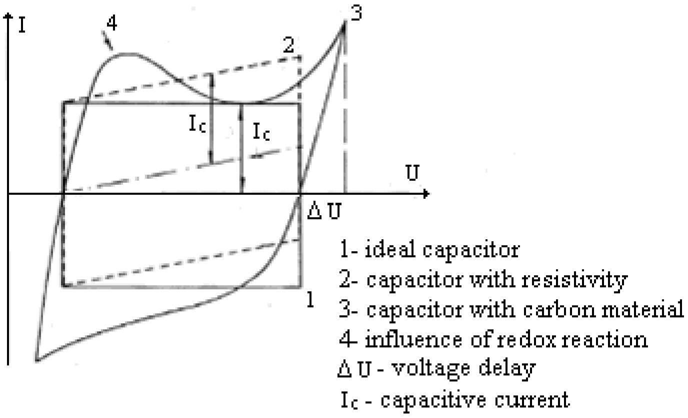


Fig. 14.34. Typical charge/discharge voltammetry characteristics of an electrochemical capacitor [98].

The time constant for the charging and discharging cycles is equal to R_S C, while the self-discharge time constant is equal to $R_F C$. In order to minimise self-discharge the value for R_F should be as large as possible.

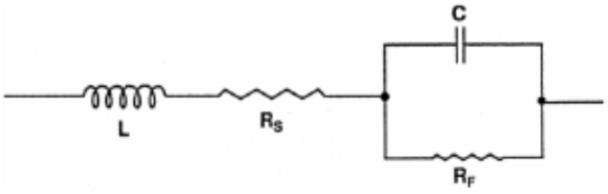


Fig.14.35. Equivalent circuit for a real ECDL capacitor.

Determination of supercapacitor properties. Key techniques to determine capacities are cyclic voltammetry, galvanostatic charge/discharge, external resistor discharge and impedance spectroscopy. Each technique reveals other specific information about the capacitor performance. Usually, a three-electrode set-up is used for material characterisation. In the resistor discharge method, the time needed to completely discharge through a certain resistor is measured. The term “impedance spectroscopy” implies the dependence of impedance, or a kind of ‘generalised’ resistance (V / I), on a certain frequency. This frequency is that of the power source used. For various reasons, the current might have a phase delay with regard to the voltage. Both this phase difference as the magnitude of the impedance (Z) play a key role in determining electrochemical mechanisms and electrode characteristics. In impedance spectroscopy, these parameters are determined as a the logarithm of the angular frequency (*log(w)*).

Periodic processes, such as those in impedance spectroscopy, can be described in terms of real and imaginary parts of the parameters investigated. These complex plots, also called Cole-Cole

``

plots, provide information about key parameters such as internal resistances and the kinetics of different processes.

Modification of CNTs. Carbon nanotubes have been proposed as electrodes for supercapacitors. Different values of capacitance mainly depend on the kind and purity of the samples. For purified nanotubes specific capacitance varies from 5 to 80 F/g. Pure carbon nanotubes have a moderate surface area (120 to 400 m^2/g) because of their highly mesoporous character. The more graphitised nanotubes show smaller values of capacitance. However, presence of defects causes an increase of ability for accumulation of charges [99].

There is a great difference in capacitance between SWNTs and MWNTs. However, the many different ways of producing the nanotubes also create differences in capacitance values. This can be seen in Table 14.2. Surface groups noticeably enhance the measured capacitance value, as is the case with MWNTsCo/700°C modified by HNO_3.

Table 14.2. Specific capacitances of differently produced carbon nanotubes in F/g [100].

	SWNTs Rice	SWNTs Rice, 1650°C	MWNTs, templ/800°C	MWNTs, Co/700°C	MWNTs, Co/900°C	MWNTs, Co/700°C, mod. HNO_3
6 mol/l KOH	40	18	36	80	62	137

To increase the capacitance of nanotubes it is possible to increase the electrode surface area or to increase the pseudocapacitance effects obtained by addition of special oxides or electrically conducting polymers (ECP) like polypyrrole (PPy). The ECPs have the advantage of lower costs compared to oxides. Another advantage is that the pseudocapacitance effects of ECPs are quite stable [101].

Charging of the electrical double layer proceeds mainly in the micropores. These micropores are just limited present in the pure CNTs. So, we have to modify them to get mesopores as well as micropores. This is possible by activation of the nanotubes by KOH. The exact mechanism of this activation by KOH is not clear yet. Both activation and coating increases the capacitance of the nanotubes [102, 103].

The mesoporous character of as-produced nanotubes essentially determines their electrochemical properties. For charging of the double layer a developed surface area is needed, thus the presence of micropores is necessary. With chemical activation of pure MWNTs by KOH, microporosity is introduced (Fig. 14.36). This KOH activation is performed in [102] at 800 °C under argon flow. The KOH:C weight ratio is 4:1 during the entire process. Afterwards, the samples are washed with demineralised water. Specific surface area is measured by nitrogen adsorption at 77 K after the samples are outgassed at 350 °C for 24 hours until pressure reaches 10-6 mbar. The electrodes are prepared in the form of pellets with 85 wt% nanotubular material, 5 wt% acetylene black and 10 wt% polyvinylidene fluoride. The mass of the electrodes varies from 4 to 8 mg. Three different types of electrolytic solutions are used: aqueous 6 M KOH, aqueous 1 M H_2SO_4 and organic 1.4 M $TEABF_4$ in acetonitrile. Two different types of MWNTs are used. One type is obtained by decomposition of acetylene at 700 °C on cobalt (12.5 %) supported on silica (A/CoSi700). The other type is prepared by decomposition of acetylene at 600 °C on Co particles from a solid solution of cobalt oxide and magnesium oxide (A/$Co_xMg_{1-x}O$). The values of capacitance are estimated by voltammetry, galvanostatic charge/discharge cycling and impedance spectroscopy.

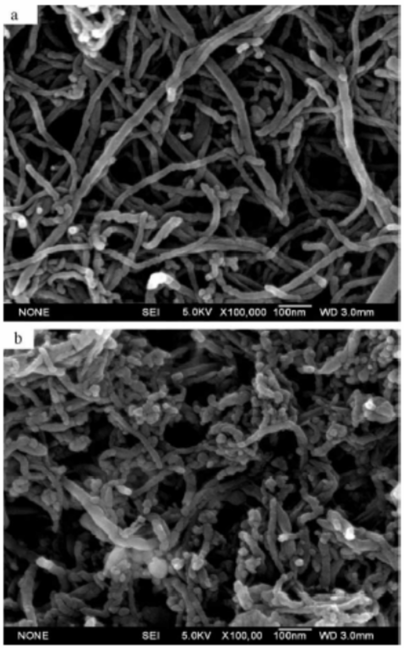


Fig.14.36. SEM images of normal CNTs (a) and activated CNTs (b) [103].

The specific surface of the A/CoSi700 nanotubes increased from 430 to 1035 m^2/g and the micropore volume changed from nearly zero to 0.47 cm^3/g after activation. For A/$Co_xMg_{1-x}O$ nanotubes the specific surface area increased from 220 to 885 m^2/g and the micropore volume increased to 0.40 cm^3/g from almost zero. These are significant changes. For the latter type an additional effect is that the tips are opened by KOH activation. Microporosity is now enhanced by activation, but the mesoporous character is still present because of the entanglement and presence of a central canal. The results of the capacitances of the different nanotubes in three different electrolytes are given in Table 14.3. The capacitances for the unactivated nanotubes are between 10 and 15 F/g. The values of capacitance were estimated by voltammetry with a scan rate of potential from 1 to 10 mV/s and galvanostatic charge/discharge cycling from 0 to 0.6 V or higher voltage limitation until 1.2 V.

``

Table 14.3. Capacitances of the activated MWNTs. The unactivated nanotubes show capacitances between 10 and 15 F/g. The surface area of the activated and unactivated nanotubes is also given.

Electrolyte	A/CoSi700 activated		$A/Co_xMg_{1-x}O$ activated	
1 M H_2SO_4	95 F/g		85 F/g	
6 M KOH	-		90 F/g	
1.4 M TEABF4	65 F/g		65 F/g	
	not act.	act.	not act.	act.
Surface area (m^2/g)	430	1035	220	885

The modification of carbon material by a specific additive providing quick pseudo-capacitance redox reactions is another way to enhance capacitance. This is possible with metal oxides, but in this case the addition of ECP is used. ECP itself has a capacitance of about 90 F/g. Pseudocapacitance effects of ECP are relatively stable. If we coat a nanotube with, for instance, polypyrrole we take the profit of the good electronic conducting properties and keep the advantage of ionic conductivity in the opened mesoporous network of the nanotube. These are perfect conditions for a supercapacitor.

Five nanotube samples have been taken in [101], all of them made in a different way and measured the capacitance of them with and without a layer of PPy coated on the surface. The specific surface area of the nanotube materials was measured by nitrogen adsorption at 77 K. The chemical polymerisation of pyrrole on the nanotubes was performed with ammonium persulfate as an oxidant. The thickness of the PPy layer was about 5 nm. Electrodes were either from bucky paper or pellets of a mixture of MWNTs (85 wt%), acetylene black (5 wt%) and polyvinylidene fluoride (10 wt%). The aqueous electrolyte was 1 M H_2SO_4. The results of capacitance measurements on the different nanotubes are given in Table 14.4. It can be concluded that the nanotubes with electrochemically deposited polypyrrole give much higher values of capacitance than the untreated samples. This proofs that the properties of both materials are used in a nice way. In Fig. 14.37 the difference in voltammograms of Hyperion with and without coating can be seen.

``

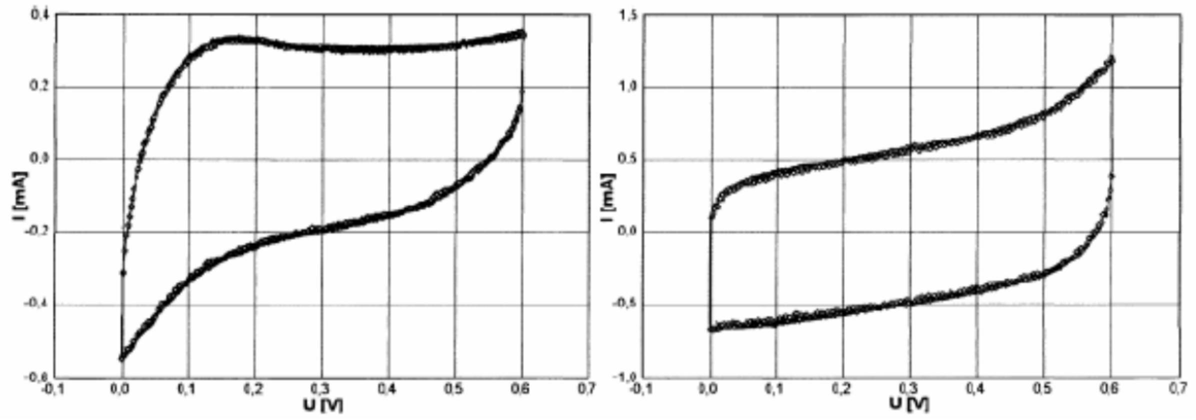


Fig. 14.37. Potentiodynamics characteristics (2 mV/s) of a capacitor assembled in 1 M H_2SO_4 from Hyperion nanotubes without PPy (left) and with PPy (right). Mass of each electrode was 3.5 mg [100].

Striking is the effect with the coated nanotubes of the type P/800Al. In this case it might be expected that a thin PPy film also covers the inner core which is also accessible for the electrolyte. A couple of the capacitors have been cycled over 2000 cycles and the charge loss never exceeded 20%. So, coating of the nanotubes with polypyrrole seems a nice way to enhance capacitance and efficiency for a long durability.

Table 14.4. Capacitance values (F/g) of the nanotubular material in acidic medium (1 M H_2SO_4) with and without PPy [101].

Kind of nanotubes	*C* (F/g) without PPy	*C* (F/g) with PPy
A/CoSi700	65	141
Hyperion	78	146
P/800Al	5	123
Pyrograf III	16	130
A/C0NaY600	50	163

Field emitting devices. If a solid is subjected to a sufficiently high electric field, electrons near the Fermi level can be extracted from the solid by tunnelling through the surface potential barrier. This emission current depends on the strength of the local electric field at the emission surface and its work function (which denotes the energy necessary to extract an electron from its highest bounded state into the vacuum level). The applied electric field must be very high in order to extract an electron. This condition is fulfilled for carbon nanotubes, because their elongated shape ensures very large field amplification [76].

``

For technological applications, the emissive material should have a low threshold emission field and large stability at high current density. Furthermore, an ideal emitter is required to have a nanometer size diameter, a structural integrity, a high electrical conductivity, a small energy spread and a large chemical stability. Carbon nanotubes possess all these properties. However, a bottleneck in the use of nanotubes for applications is the dependence of the conductivity and emission stability of the nanotubes on the fabrication process and synthesis conditions [104].

Examples of potential applications for nanotubes as field emitting devices are flat panel displays, gas discharge tubes in telecom networks, electron guns for electron microscopes, AFM tips and microwave amplifiers.

Fabrication of CNT electron field emitters. Before we go into field emitting properties we have to discuss the different production techniques. Two different kinds of electron sources can be made from CNTs, single and multiple electron beam devices. For example single beam as a electron bundle in an electron microscope and multiple beam in a flat panel display (Fig. 14.38).

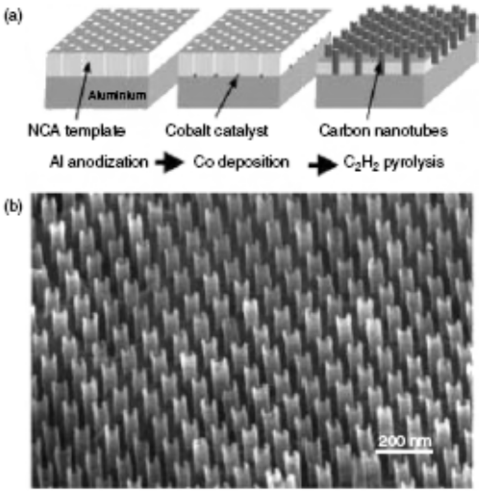


Fig. 14.38. Vertically aligned CNTs.

Single nanotube emitters.

- Single MWNT emission source:

Individual MWNTs can be mounted on an etched wire under an optical microscope with two, independently moving, 3-axis micromanipulators. The individual nanotubes stick to the tip only by van der Waals forces or by applying a bit of conductive adhesive to the tip. Because of the too low resolution of an optical microscope SEM isnecessary to observe the single MWNT.

- SWNT emission from raw material:

It is more difficult to obtain a single SWNT from ropes or unpurified material, so a macroscopic fragment of raw material is picked up with a pair of tweezers and attached to a tungsten tip so you

``

actually have many emitters but for measurements one can use a probe hole in the counter electrode to study the behaviour of only a single SWNT as field emitter. Use of SEM is necessary mandatory.

- SWNT emission from rope of SWNTs:

A rope is picked out of a mat of SWNT material and detached from it by making electrical contact and applying a voltage difference (~10 V) over the SWNT and the mat. Use of SEM is mandatory.

- SWNT emission from purified material:

Commercially available material with well separated tubes in a solution allows mounting of tubes with very little contaminating material. The disadvantage are the unknown effects of the purification process on the morphology of the CNTs.

- A SWNT attached to an AFM tip:

By using the electron bundle of a SEM for electron beam irradiation one can weld the SWNT to the AFM tip. The welds are mechanically strong and have a good electronic contact.

- Single SWNT growth on AFM tip:

Direct growth of a SWNT by CVD on an AFM tip can be done and looks promising.

Continuous film emitters.

Planar films of emitting nanotubes can be made rather easily:

- Film from MWNTs in a colloidal suspension:

A film by drawing a colloidal suspension of MWNTs through a nanopore alumina filter has been produced. Next, the film was transferred by pressing the filter face down on a Teflon or Teflon-coated metal surface. This technique is not restricted to MWNTs but can be used for all types of nanotubes. The same result can be obtained by spraying the suspension on a heated substrate.

- Purified MWNTs dispersed in a matrix:

The nanotubes have been dispersed in an epoxy matrix and prepared emitters of (50×50) μm^2 under pressure and with a polishing to obtain a reproducible and macroscopically flat surface.

- Crushed and mixed MWNTs:

Crushed MWNTs are mixed with products to form a paste which can be applied to a metallic cathode and cured.

Patterned films using post deposition techniques.

For many applications it is necessary for the emitting film to have a certain pattern. Different post deposition techniques to obtain patterned films are presented here:

- A nanotube/epoxy paste is pressed into channels etched in a glass substrate. Cathode surface is polished to expose the micro channels.

- Nanotubes mixed in an organic mixture of nitrocellulose, the resulting paste is squeezed onto a metal-patterned sodalime glass through a metal mesh. Subsequently the emitter was heated to remove the organic binder.

``

- A resist layer is patterned by e-beam lithography on a substrate. After this a nanotube dispersion is applied to the substrate. Nanotubes precipitate in the resist openings and after lift-off a patterned film is obtained.

- It is also possible to position individual nanotubes at specific locations and with specific orientations by applying a nanotube suspension on self-assembled monolayer templates defined by lithography.

Patterned films using direct growth on catalyst support.

Predefined catalyst patterns can be used to grow CNTs on specific locations. The patterns are made from i.e. cobalt films that are deposited from laser ablation. The obtained film is etched with single laser pulses using a cylindrical lens, while being exposed to air. The MWNTs only grow on the Co particles along the edges of the eroded tracks.

A second technique is based on standard photolithography or e-beam lithography. Basically a resist is coated on the substrate, exposed and developed. The resist can be used in two ways:

- Resist as mask over the catalyst and CNT growth over resist after which the resist is removed and a pattern is left over

- Resist as protection, unprotected catalyst is evaporated and after removal of resist the CNTs are grown

The techniques to grow the CNTs after making a catalyst pattern also differ and are much alike described in the continuous film production.

One method that is different from the others is the use of a stamp with ink, called microcontact printing. Liquid phase catalyst precursor is dropped on a patterned elastomer stamp. The stamp is then pressed against a substrate and afterwards MWNTs are grown by CVD on the pattern of catalyst. Important is the ability to vary the catalyst precursor concentration, hydrocarbon gas and the temperature. All these parameters can be optimised for the desired MWNT density [104].

Field emission from CNT films. Field emission is the emission of electrons from a solid by tunnelling through the surface potential barrier. The emission current depends on the electric field, E, and the workfunction, Φ. The Fowler-Nordheim model shows that the dependence of the emitted current on the local electric field and the workfunction is exponential. As a consequence small variations in the shape or surrounding of the emitter have strong impact on the emitted current. This sensitivity brings along many problems in comparing measurements because of the different shapes of set-ups. There is also a difference between integrated (macroscopic) , millimetre, measurements and microscopic, micrometer, measurements.

``

Most reports on field emission describe the fabrication method and a typical resulting I-V curve. All types of nanotubes have excellent field emission properties with turn on fields that produce 10 μA/cm^2, of only 1 V/μm and threshold fields that produce 10 mA/cm^2, around 5 V/μm. Nanotube films are also able to emit a current density up to a few A/cm^2 at fields below 10 V/μm. Because of the large differences in results for slightly differing parameters a comparison of films under the same experimental conditions has been also provided (Table 14.5) [104].

Table 14.5. The emission characteristics of CNT films under the same experimental conditions [104].

Influence	Emitter	d (μm)	S (cm^2)	E_{on} (V/μm)	E_{th} (V/μm)	Remarks
Density and geometry	MWNT	25	n.a.	<2.7	~4.8	Random alignment
	MWNT	25	n.a.	~40	n.a.	Short vertical tubes
Geometry	Arc MWNT	125	0.07	2.6	4.6	Average over 15 emitters
	SWNT	125	0.07	2.8	5.2	Average over 12 emitters
	Open MWNT	125	0.07	4.5	30	Average over 6 emitters
	Graphitic fibers	125	0.07	5.6	14	Average over 5 emitters
Surface treatment	MWNT t-a coated	125-400	0.01	1.6	n.a.	
	MWNT as produced	125-400	0.01	2.4	n.a.	
Density and geometry	SWNT	10-500	10^{-5}	n.a.	2.4	Random aligned
	CVD MWNT	10-500	10^{-5}	n.a.	3.5	Dense aligned arrays
Density	CVD MWNT	125	0.007	9.8	14.4	Low density, patterned films
	CVD MWNT	125	0.007	2.2	3.3	Medium density, patterned films
	CVD MWNT	125	0.007	3.6	5.3	High density, patterned films

Differences occur due to structural differences and differences in treatment, diameter size, open/closed cap, surface treatment etc. Density and orientation of the tubes on the film also influences the emission. When comparing samples with different densities as in Fig.14.39 the influence can be seen in the I-V curves see Fig. 14.40.

``

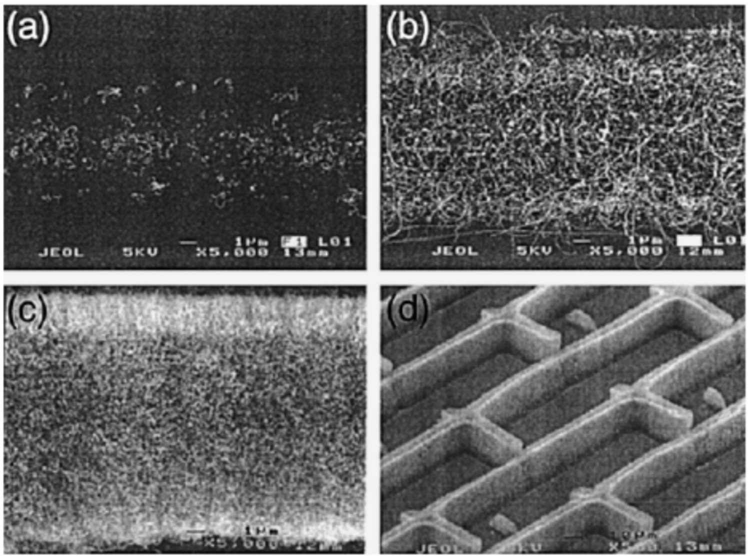


Fig. 14.39. SEM micrographs of patterned films of MWNTs prepared by μCP of a catalyst followed by CVD growth. The concentration of catalyst in the ink solution was (a) 1 mM, (b) 5 mM, (c) 40 mM, (d) 50 mM [104].

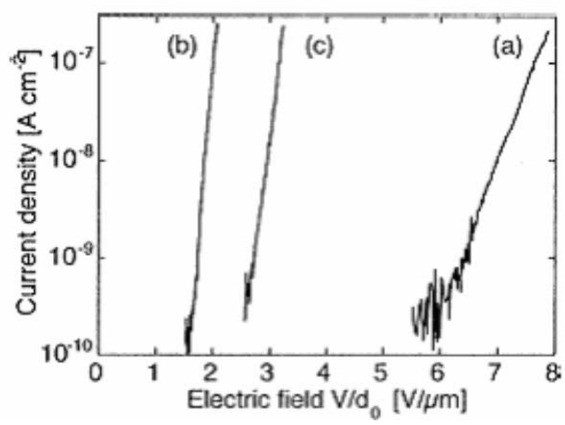
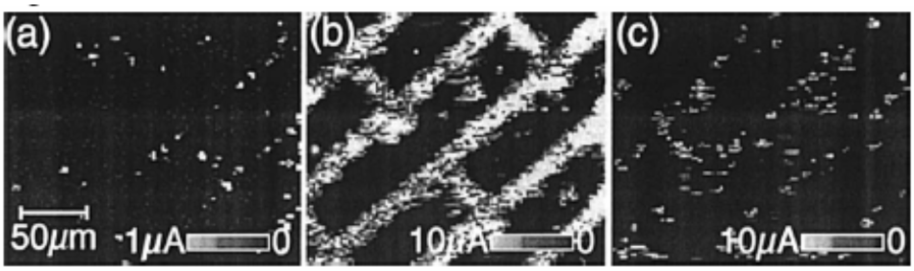


Fig. 14.40. I-V curves for different MWNT densities as seen in Fig. 14.38. The best result is (b) [104].

Analysis of eleven samples shows that films of medium densities with nanotubes protruding over the film surface show emission at the lowest fields. Which also can be seen in the scanning field emission images (Fig. 14.41). Here it is also clear that the medium density film has the highest current density.



``

Fig. 14.41. Scanning field emission images of MWNT films of different densities acquired at constant voltage and tip height. The grey scale represents current intensity. Images were taken from samples similar to those from Fig. 14.39 [104].

Degradation of CNT films. Several different processes can lead to degradation of the CNT film, resistive heating, bombardment from gas molecules, arcing electrostatic deflection and mechanical stresses etc. Also chemical degradation phenomena such as adsorption and desorption of molecules on the emitter surface can occur.

Stable emissions over more than 20 h have been observed and even an emission of 8000 h of 10 mA/cm^2 with an 11% field increase has been possible. The origin of degradation is not clear but it seems that residual gasses have a significant influence as well as the emitted current density. In addition the intrinsic properties of the CNTs have an importance. Tests have shown that degradation of SWNTs was much faster than of MWNTs.

Field emission from single CNTs. Experiments on single nanotube emitters give a reliable way to investigate the emission mechanism and workfunction. First measurements were performed at study an arc produced MWNT mounted on a carbon fibre. It was shown that the emission followed roughly a Fowler-Nordheim behaviour. The currents emitted were about 100 nA at 0.12 V/μm field strength. Other studies show Fowler-Nordheim behaviour up to 5-20 nA for closed and open tubes. As the current increased, saturation effects were observed. It was noted that all emitters were able to emit over an incredible current range up to 0.1 mA or for some even 0.2 mA per tube. It was also noticed that open tubes needed twice the voltage to start emitting. But there are contradictory results [105]. They found that open MWNTs started emitting electrons at the lowest fields, followed by MWNTs produced in a hydrogen atmosphere and finally the closed arc discharge MWNTs started emitting electrons at the higher fields. The differences observed are probably due to different measurement set-ups, growing methods, purity and emitter geometry.

<u>Emitter stability and failure.</u> FEM measurements on single SWNTs showed two regions in the I-V curve. Step like fluctuations at low emitted currents with a switching frequency that increased with the current and became maximal at the onset of saturation, followed at higher currents by stable emission with flicker noise.

Failure of single nanotubes is in almost all cases abrupt. For a single closed MWNT stable emission was observed at 2 μA for more than 90 h at 10^{-7} mbar. Failure occurred on most emitters in a catastrophic irreversible way in less than 10 ms.

However one MWNT has been measured during more than two months at 0.4 μA at 10^{-9} mbar without any observable degradation. The mechanism that leads to catastrophic failure is not

``

yet completely understood but it appears that the failure occurs at a very short time scale (<1 ms) at currents above 0.1 mA and that it involves an irreversible damage to the tube, in all cases the failure occurs on the defects of the nanotube. It still is unclear whether the degradation in CNT films is abrupt or gradual.

Transistors. The field-effect transistor – a three-terminal switching device – can be constructed of only one semiconducting SWNT. By applying a voltage to a gate electrode, the nanotube can be switched from a conducting to an insulating state [93]. A schematic representation of such a transistor is given in Fig. 14.42.

Such carbon nanotube transistors can be coupled together, working as a logical switch, which is the basic component of computers [106].

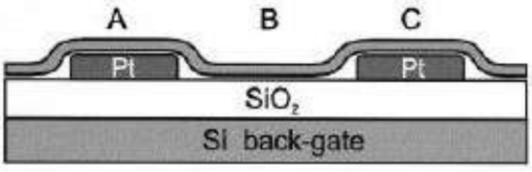


Fig. 14.42. A single semi-conducting nanotube is contacted by two electrodes. The Si substrate, which is covered by a layer of SiO_2 300nmthick, acts as a back-gate.

Smaller silicon based integrated circuits result in both a higher speed and device density. As a result, down scaling of these devices has been very important since their first implementation. However, at the moment it is generally accepted that silicon devices will reach fundamental scaling limits within a decade or so. This limit is caused by the minimum wavelength of light used in lithographic techniques used for integrated circuit production nowadays. For this reason a quest for alternative, integrated circuits with smaller dimensions has started. A major step in down scaling would be the application of single molecules in electronic devices [107]. Carbon nanotubes have already shown promising results in single molecular transistors.

For successful implementation of molecular transistors in large and complex logic systems they must show signal amplification. Signal amplification makes it possible to reference separate signals along a chain of logical operations. In addition, noise caused by thermal fluctuations and environmental disturbances is also reduced [107]. Three terminal nanotransistors, in special, field-effect-transistors show amplifying behaviour and have recently been investigated for this reason.

CNT based transistors can be divided into two main categories: CNT field-effect-transistor (CNTFET) and CNT single-electron-transistor (CNTSET). CNTFET behaviour and performance is

``

also compared with conventional the Metal-Oxide-Semiconductor (MOS) Field Effect Transistor (FET).

Principles of the MOSFET and CNTFET. A basic MOSFET involves two main electrodes, designated as "source" and "drain", connected by a semi-conducting channel (Fig. 14.43). In conventional devices this channel is made of silicon. In CNTFETs the channel is formed by a single, semi-conducting CNT. A third electrode, the "gate" is separated from the channel by a thin insulator film and can also be placed on top of the nanotube.

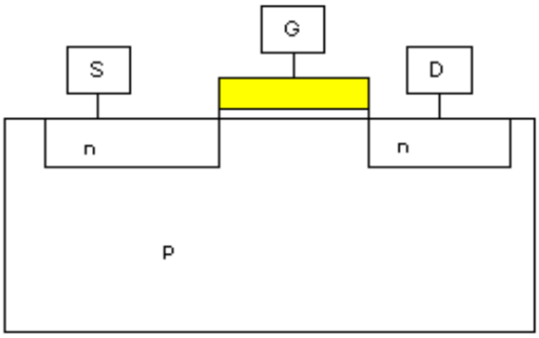


Fig. 14.43. Set-up of a MOSFET.

If no charge is present on the gate, the channel is not conducting. The channel of a p-type (hole conducting) MOSFET becomes conductive when a negative charge is placed on the gate and a certain threshold voltage is exceeded. This is the very basic principle of this type of transistor. Almost the same accounts for n-type MOSFETs in which a positive charge must be put on the gate. A CNTFET works in a slightly different way as will be explained further on.

Manufacturing of a CNTFET. In 1998 Semi-conducting SWNTs [108] and MWNTs [109] were used in CNTFETs for the first time. Due to the basic principles of a CNTFET a metallic CNT cannot be used since it is always conducting. An example of the manufacturing of one of the first CNTFETs is given below (Fig. 14.44).

As grown CNTs are purified by sonication in a H_2SO_4 /H_2O_2 to remove catalyst particles. The purified nanotubes are dispersed in dichloroethane and spread over a silicon substrate, which is covered with silicon oxide, an insulating compound. On top of this layer, predefined electrode areas are created by conventional lithographic techniques. The nanotubes are randomly spread over the surface this way and in some cases a nanotube connects two electrodes [110]. A slightly different approach is to spin coat the CNTs on a substrate first, after which the electrodes are deposited with electron beam lithography. The latter methodology can also be used to create top gated CNTFETs, which is schematically shown in Fig. 14.44 [110].

``

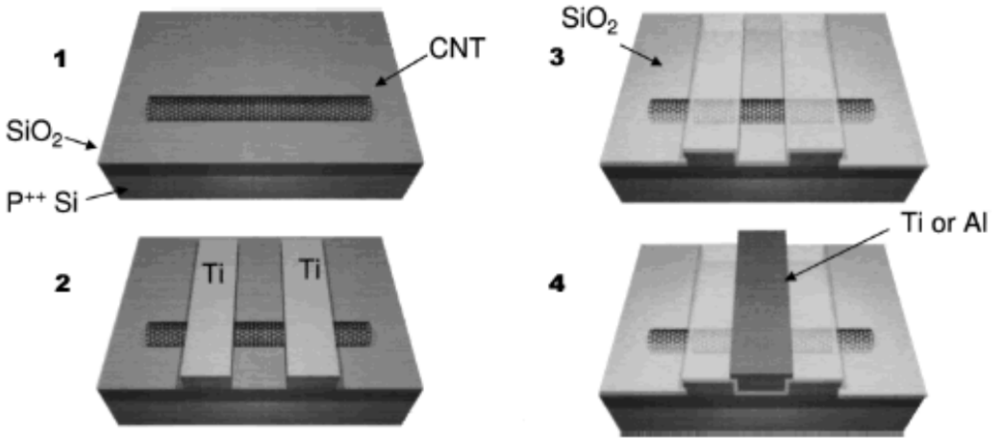


Fig. 14.44. Fabrication of a top gate CNTFET.

It is also possible to position a nanotube at a certain position with AFM-like techniques. Semiconducting and metallic nanotubes in a single rope can be separated by ‘constructive destruction’. A current is put through the nanotube rope and the metallic tubes break down selectively, leaving the semi-conducting nanotubes over. The high stability of metallic nanotubes is maintained as long as moving electrons have insufficient energy to excite optical phonons [74]. Furthermore, the environment also plays an important role as adsorbed gasses influence the nanotubes behaviour. So adsorption of gasses should be prevented as much as possible.

CNTFET physics.

Conductivity in CNTs

To develop a better understanding of the switching process, a detailed model of charge transport in the channel must be set up. Electron transport in metallic CNTs is ballistic of nature, which means that electrons move with a much higher speed than expected on the ground of thermal equilibrium. Ballistic transport can only occur in strictly confined semi-conductor regions and is totally determined by material properties. Electron transport in semi-conducting CNTs is far more complicated due to interband scattering, influences of charges at the electrode interface and adsorbed gasses. Therefore, conductivity in semi-conducting CNTs is usually described in classic physical laws of diffusivity [74]. Diffusive charge transport means that electron motion can be described with an effective mobility μ. In case of diffusion transport, the average free path l of an electron is relatively short (~ 2 nm) compared with the ballistic transport situation (l ~ 10 μm) [111].

Doping of CNTs

Classical semiconductor physics is mainly based on two types of charge transport: either by negative or positive charge carriers. In the first case the material is n-type semiconducting while in the latter case is spoken of a p-type semiconductor. Some times it is even possible to have nanotube conductivity in a transistor either by n-type or p-type charge carriers, dependent on the gate-source voltage applied [107]. This behaviour is called *ambi-polar*.

As grown, single walled CNTs show p-type semi-conducting behaviour which is explained by several theories. The most commonly accepted theory states that the p-type conducting behaviour is caused by charge transfer from the nanotube to the electrode material with a relatively low Fermi level. Another explanation for the p-type behaviour is doping of the nanotubes by foreign atoms during synthesis or purification or the absorption of gasses [108, 110]. However, adsorption of oxygen is not true doping for several reasons.

The large charge carrier density, which is approximately one hole per 250 Carbon atoms, in semiconducting nanotubes is remarkable [109]. CNTS can be (partially) doped to an n-type semi-conductor in several ways. p-Type nanotubes can be exposed to electron donating compounds such as alkali metals to form n-type semiconductors [112]. An easier way to perform the p-n transformation is by partially heating the nanotube, which transformation has shown to be reversible of nature [114]. However, doping of NTs by potassium and adsorption of oxygen resulted in different behaviour which indicated a different doping mechanism [113, 114].

Switching mechanism of a MOSFET and CNTFET. A MOSFET is indeed a field effect transistor. Directly after the discovery of the CNTFET, mechanisms for the functioning of this type of devices were developed. A link to the MOSFET mechanism was made rapidly which meant that influence on the channel was considered of key importance for this type of devices.

Later investigations into CNTFETS have shown the partial incorrectness of this mechanism. It became clear that the transistor functioning is mainly based on the Schottky barriers (SB) present at the metal / semiconducting CNT junctions [107, 113]. This type of transistor is very uncommon in conventional semiconductor physics. Switching occurs rather by modulation of the contact resistance than by altering the channel resistivity. SB transistors can have very small dimensions which possibly allows down scaling of silicon devices [115]. For very small Schottky barriers the device functions as a conventional channel limited field-effect-transistor. Adsorption of gasses results in other properties of the CNT transistors by changing the electrode properties.

The Schottky barrier present at the metal /semiconductor interface prevents charges from being transported through this interface. At low voltages there is substantial charge density in the channel, but these are blocked by the Schottky barriers. Applying a larger gate-source voltage results in a stronger electric field at the electrodes and as a result the Schottky barrier becomes smaller. Now, the charge carriers can tunnel in a thermally assisted way and the transistor becomes conductive [114].

``

Characteristics of a CNTFET. Even at their first investigations, CNT based transistors had excellent properties that were comparable with silicon based MOSFET transistors. The high stability and conductivity of single carbon nanotubes result in excellent transistor behaviour. High current densities for the 'on'state (2100 μA/μm) were achieved together with excellent on/off ratios of approximately 10^5. In a similar MOSFET the maximum current density is only 650 μA/μm under the same bias conditions.

Channel length of CNTFETs can be reduced to 5 nm, before tunnelling at room temperature results in unacceptable leakage currents. This means that very high device densities could be achieved. Top gate devices even show better performance than transistors in which the substrate is the gate electrode. Switching in top gated CNTFET happens at much lower gate voltages. Another advantage of top gate transistors is the ability to switch different transistor individually. On / off ratios as high as 10^6 can be achieved with top gate CNTFETs.

Optimisation of CNTFETs. CNTFET devices can be improved in several ways, some of which will be described here. The contact resistivity can be reduced by annealing the sample at a high temperature in an inert atmosphere. This will improve the device characteristics enormously. The geometry of the electrodes is also very important, a needle-like electrode results in a stronger electric field at the electrode/nanotube interface and thus in smaller Schottky barriers which improves the performance of the transistor.

Future outlook. Only a single CNT FET does not make a completely functional electronic circuit. Several transistors can be combined with each other to form different logic operators. With these operators, fully functional devices can be created with much faster transistors. But new technologies for accurate placement of nanotubes should be developed just as methods for smaller electrode contacts. Else the size limit of molecular transistors is still determined by the contact dimensions.

Single electron transistor. Due to the ability to define nano-scale regions of p or n material on a SWNT scientist have been able to make such a small quantum dot (QD). The properties of such a small doped region are influenced by the quantum mechanical effects of a single electron. The Coulomb force of a single electron can prevent the entrance of a second electron and tunnelling effects also play a role in the process. The energy for one electron to move into the system is:

$$E_C = \frac{e^2}{2C} \qquad (14.6)$$

``

E_c is called the Coulomb blockade energy, which is the repelling energy of the previous electron to the next electron. So charging and discharging becomes a discontinuous process, electrons have to move through the system one by one. If two QDs are joined at a point and form a channel, it is possible for an electron to pass from one dot to the other over the energy barrier, this phenomenon is called tunnelling. In order to overcome the Coulomb blockade energy E_c, the applied voltage over the QDs should be

$$V > \frac{e}{C} \tag{14.7}$$

In order to observe Coulomb blockade and tunnelling, the energy of the electron must be higher than the thermal scattering energy,

$$\frac{e^2}{2C} > k_B T \tag{14.8}$$

If these QDs are connected to other conductive materials (metal wires, polymers, CNTs) a SET is formed.

Construction of SET on SWNT.

Chemical profiling

As with normal CNT transistors the CNTs are doped in an uncovered region with K to make it n-type, the neighbouring regions are covered with PMMA to keep them p-doped. This p-n-p nanotube system behaves as a single QD, investigations showed that the QD are confined in the n region with the two pn junctions as barriers for the dots (Fig. 14.45). Chemical profiling a single SWNT can lead to on-tube QDs much smaller than the lengths of the nanotubes between metal electrodes (3-5 μm). In order to function properly the system has to be cooled to <100K [116].

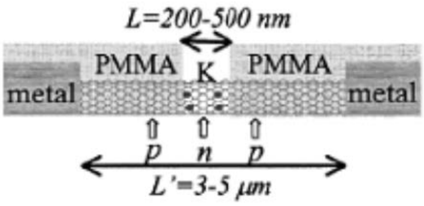


Fig. 14.45. Schematic p-n-p nanotube device.

Bending CNTs

Another technique found earlier [117] showed the fabrication of a room temperature SET (RTSET). They were able to produce a region between two tunnelling barriers formed by bending a metallic SWNT with an AFM. The nanotube is placed between Au electrodes on a Si/SiO$_2$ substrate. In Fig. 14.46 A and B are the bending steps with an AFM and C and D show the results.

``

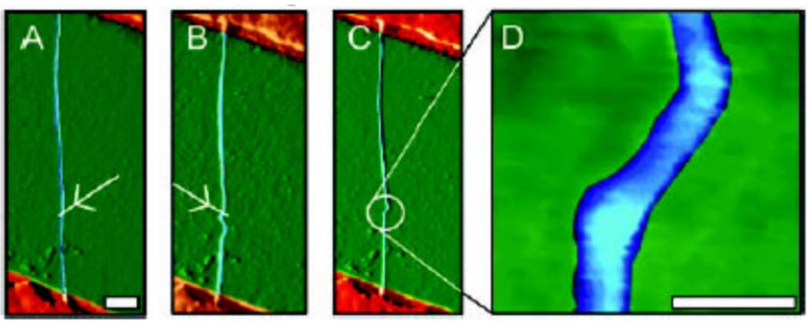


Fig. 14.46. Fabrication of a room-temperature SET using AFM. Step A and B: bending with AFM. C and D is the resulting RTSET [118].

Working mechanism of a SET. A set consists out of a conducting island that is connected by tunnelling barriers to two metallic leads. For temperatures and bias voltages that are low relative to a characteristic energy required to ad an electron to the island, electrical transport through the device is blocked. Conduction is possible however by tuning a voltage on a close-by gate, this makes the device into a three terminal transistor [116, 117].

``

14.4. Graphene

14.4.1. Molecular Structure

Before monolayer graphene was isolated in 2004, it was theoretically believed that two dimensional compounds could not exist due to thermal instability when separated. However, once graphene was isolated, it was clear that it was actually possible, and it took scientists some time to find out exactly how. After suspended graphene sheets were studied by transmission electron microscopy, scientists believed that they found the reason to be due to slight rippling in the graphene, modifying the structure of the material. However, later research suggests that it is actually due to the fact that the carbon to carbon bonds in graphene are so small and strong that they prevent thermal fluctuations from destabilizing it.

Graphene is a single layer of carbon packed in a hexagonal (honeycomb) lattice, with a carbon-carbon distance of 0.142 nm (Fig. 14.47). It is the first truly two-dimensional crystalline material and it is representative of a whole class of 2D materials including for example single layers of Boron-Nitride (BN) and Molybdenum-disulphide (MoS_2), which have both been produced after 2004 [13].

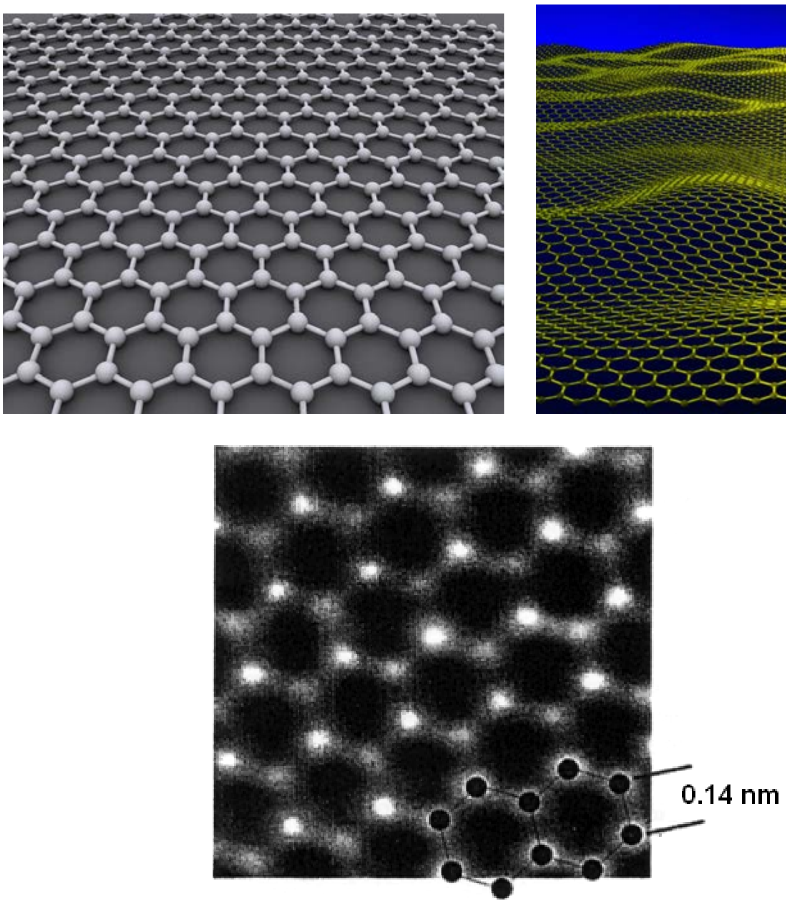


Fig. 14.47. Schematic and SEM images of graphene.

``

Graphene's stability is due to a tightly packed, periodic array of carbon atoms and an sp^2 orbital hybridization - a combination of orbitals p_x and p_y that constitute the σ-bond. Graphene has three σ-bonds and one π-bond. The final p_z electron makes up the π-bond, and is key to the half-filled band that permits free-moving electrons.

Graphene sheets in solid form usually show evidence in diffraction for graphite's (002) layering. This is true of some single-walled nanostructures [118]. However, unlayered graphene with only (hk0) rings has been found in the core of presolar graphite onions [119]. TEM studies show faceting at defects in flat graphene sheets [120] and suggest a role for two-dimensional crystallization from a melt.

Graphene can self-repair holes in its sheets, when exposed to molecules containing carbon, such as hydrocarbons. Bombarded with pure carbon atoms, the atoms perfectly align into hexagons, completely filling the holes [121].

The atomic structure of isolated, single-layer graphene was studied by transmission electron microscopy (TEM) on sheets of graphene suspended between bars of a metallic grid [122]. Electron diffraction patterns showed the expected honeycomb lattice. Suspended graphene also showed "rippling" of the flat sheet, with amplitude of about one nanometer. These ripples may be intrinsic to the material as a result of the instability of two-dimensional crystals [123, 124], or may originate from the ubiquitous dirt seen in all TEM images of graphene. Atomic resolution real-space images of isolated, single-layer graphene on SiO_2 substrates are available [125, 126] via scanning tunneling microscopy. Photoresist residue, which must be removed to obtain atomic-resolution images, may be the "adsorbates" observed in TEM images, and may explain the observed rippling. Rippling on SiO_2 is caused by conformation of graphene to the underlying SiO_2, and is not intrinsic [125].

Chemical. Graphene is the only form of carbon (and generally all solid materials) in which each single atom is in exposure for chemical reaction from two sides (due to the 2D structure). It is known that carbon atoms at the edge of graphene sheets have special chemical reactivity, and graphene has the highest ratio of edgy carbons (in comparison with similar materials such as carbon nanotubes). In addition, various types of defects within the sheet, which are very common, increase the chemical reactivity [127]. The onset temperature of reaction between the basal plane of single-layer graphene and oxygen gas is below 260 °C (530 K) [128] and the graphene burns at very low temperature (e.g., 350 °C (620 K)) [129]. In fact, graphene is chemically the most reactive form of carbon, owing to the lateral availability of carbon atoms. Graphene is commonly modified with oxygen- and nitrogen-containing functional groups and analyzed by infrared spectroscopy and X-ray photoelectron spectroscopy. However, determination of structures of graphene with oxygen- and nitrogen- containing functional groups is a difficult task unless the structures are well controlled

``

[130, 131]. It has been reported that sheets of graphene one atom thick are a hundred times more chemically reactive than thicker sheets.

14.4.2. Graphene Fabrication

There are currently several methods to produce graphene described in the scientific and commercial literature showing the intensity of development work globally (Fig. 14.48).

Exfoliation techniques. The most common method of graphene fabrication is exfoliation which finds its roots with a technique that has been around for centuries – writing with a graphite pencil. By writing with a pencil you create many graphene sheets spread over your paper. Unfortunately this method is uncontrollable and you are typically left with many sheets of varying thicknesses. If you want to study a single graphene sheet you need to locate it. The problem amounts to trying to find a needle in a haystack. A way around this problem was solved by Andre Geim's group in Manchester [12]. By gently rubbing or pressing a freshly cleaved graphite crystal on an oxidized silicon wafer graphene flakes with the correct thickness of oxide, single atomic layers are visible under an optical microscope due to thin film interference effects [13, 132]. This technique simplifies the process of finding single graphene sheets but obviously limits this fabrication scheme to devices for research purposes. For the case of suspended graphene sheets this process may take ~ 1 hour to find relatively thin ~ 1-5 nm thick suspended graphene devices but could take several days or weeks to find a suitable single suspended layer. There are recent attempts to improve the quality and yield of exfoliation techniques. These include stamping methods which use silicon pillars to transfer graphene flakes and electrostatic voltage assisted exfoliation which uses electrostatic forces to controllably separate graphene from bulk crystals [133, 134]. These are very recent developments and only time will tell whether they yield significant improvement over standard exfoliation.

Disperse from solution. Graphene platelets can also be created by chemically cutting open carbon nanotubes, one method describing how the nanotubes are cut open in solution by action of potassium permanganate and sulfuric acid. Chemical, solvent or sonic exfoliation (separation) of graphene layers from graphite has also been developed. common graphene fabrication technique is to disperse graphene from solution [135]. In this method graphite flakes are sonicated in a solution and then dispersed onto the wafer.

``

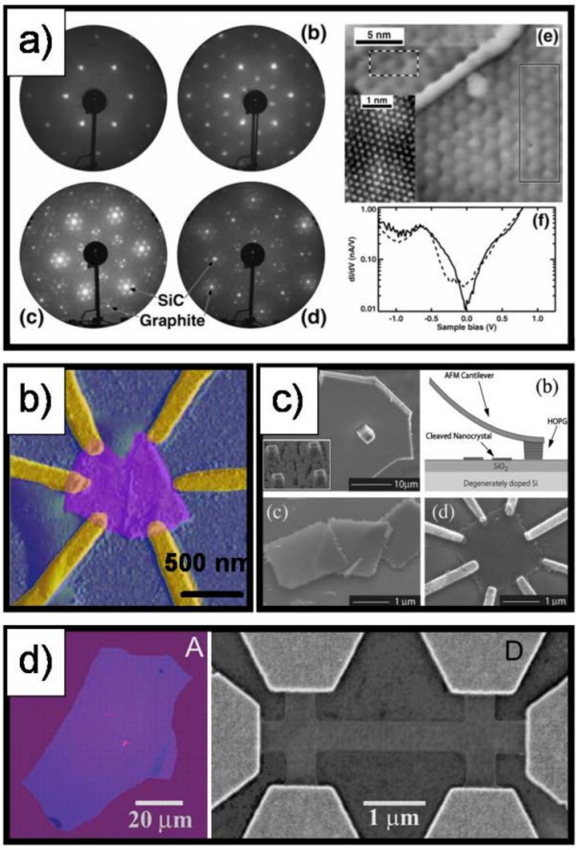


Fig. 14.48. a) Growing few layer graphene on SiC [136]. b) AFM image of a few layer graphene quantum dot fabricated by dispersion from solution [135]. c) Nanopencil used to extract few layer graphene flakes from HOPG [137]. Optical image of a few-layer graphene sheet [12].

An AFM is used to locate individual sheets making this technique very time consuming relative to the optical detection scheme. Long sonication times are needed to break the graphite down and this typically results in small flakes. Recently a similar technique was used to fabricate graphene ribbons with nm-scale widths [138]. One of the difficulties in dispersing graphene from solution is separating the layers without breaking them. A way around this is to intercalate the graphite and dissolve it in a solvent. When the intercalant dissolves it separates the graphene sheets. This technique was shown to work effectively for graphene oxide**.** However, the success of similar techniques on graphene is limited due to the chemistry required to keep individual graphene sheets from aggregating in solution.

Direct growth of graphene. The technique which currently seems to have the greatest potential for mass production is the direct growth of graphene. Typically this is accomplished by heating a SiC wafer which results in the partial graphitization of the upper layer [139]. However, controlling the number of layers as well as the grain sizes is difficult with this technique limiting the mobilities achieved so far with this form of graphene [140]. Furthermore, isolating single sheets is

``

problematic and additional lithography is required to pattern electrostatic gates on top of the graphene. Making suspended mechanical structures from grown graphene has yet to be demonstrated.

Chemical vapor deposition. Chemical vapor deposition (CVD) and molecular beam epitaxy (MBE) are two other potential routes to graphene growth. Carbon nanotubes and diamond are successfully grown using CVD and the preferred method of growth for high quality GaAs/AlGaAs heterostructures is MBE. For the time being, exfoliation remains the preferred method for most of the experimental research groups around the world. However as in diamond, wide spread applicability of graphene is limited by the crude and time consuming methods currently used to fabricate and isolate single graphene sheets.

The research community is currently in need of a reliable and reproducible graphene fabrication method if graphene is ever to move beyond being a laboratory curiosity. Plasma deposition techniques, the reduction of graphene oxides (RGO) and other synthetic methodologies are being introduced as route to scale up manufacture. Large scale manufacture of graphene is beginning to take place in locations where the chemistry based process industries have the infrastructure needed and can manage the materials, processes and technologies.

14.4.3. Electrical Properties of Graphene

Most of the experimental research on graphene focuses on the electronic properties. The most notable feature about the early work on graphene transistors was the ability to continuously tune the charge carriers from holes to electrons. An example of the gate dependence in single layer graphene is shown in Fig. 14.49. This effect is most pronounced in the thinnest samples whereas samples from multiple layers show much weaker gate dependence due to screening of the electric field by the other layers.

For more practical applications one would like to utilize the strong gate dependence of graphene for either sensing or transistor applications. Unfortunately, graphene has no band gap and correspondingly resistivity changes are small.

Therefore, a graphene transistor by its very nature is plagued by a low on/off ratio. However one way around this limitation, is to carve graphene into narrow ribbons. By shrinking the ribbon the momentum of charge carriers in the transverse direction becomes quantized which results in the opening of a band gap. This band gap is proportional to the width of the ribbon. This effect is pronounced in carbon nanotubes where a nanotube has a band gap proportional to its diameter. The

``

opening of a band gap in graphene ribbons has recently been observed in wide ribbon devices lithographically patterned from large graphene flakes [141] and in narrow chemically synthesized graphene ribbons [138].

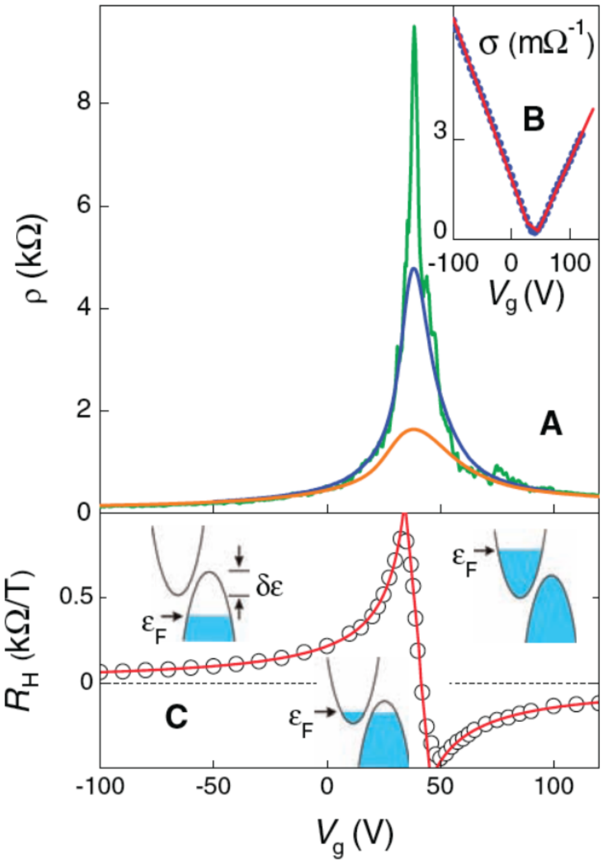


Fig. 14.49. A) The (longitudinal) resistivity of a few layer graphene sample for three different temperatures (5K green, 70K blue, 300K orange), note the ambipolar dependence on gate voltage. B) The conductance as a function of gate voltage at 77K. C) The Hall resistivity as a function gate voltage for the same sample [13].

Graphene is a zero-gap semiconductor. Four electronic properties separate it from other condensed matter systems.

Electronic spectrum. Electrons propagating through graphene's honeycomb lattice effectively lose their mass, producing quasi-particles hat are described by a 2D analogue of the Dirac equation rather than the Schrödinger equation for spin-½particles [142, 143].

Dispersion relation. Using a conventional tight-binding model the dispersion relation produces energy of the electrons with wave vector *k* is [144, 145].

$$E = \pm\sqrt{\gamma_0^2(1 + 4\cos^2\frac{k_y a}{2} + 4\cos\frac{k_y a}{2}\cos\frac{k_x\sqrt{3}a}{2})}$$

``

with the nearest-neighbor hopping energy $\gamma_0 \approx 2.8$ eV and the lattice constant $a \approx 2.46$ Å. The conduction and valence bands, respectively, correspond to the different signs; they touch each other at six points, the "*K*-values" of the two-dimensional hexagonal Brillouin zone (two of these six points are independent, while the rest are equivalent by symmetry (Fig. 14. 50). In the vicinity of the *K*-points the energy depends *linearly* on the wave vector, similar to a relativistic particle [144, 146]. Since an elementary cell of the lattice has a basis of two atoms, the wave function has an effective 2-spinor structure.

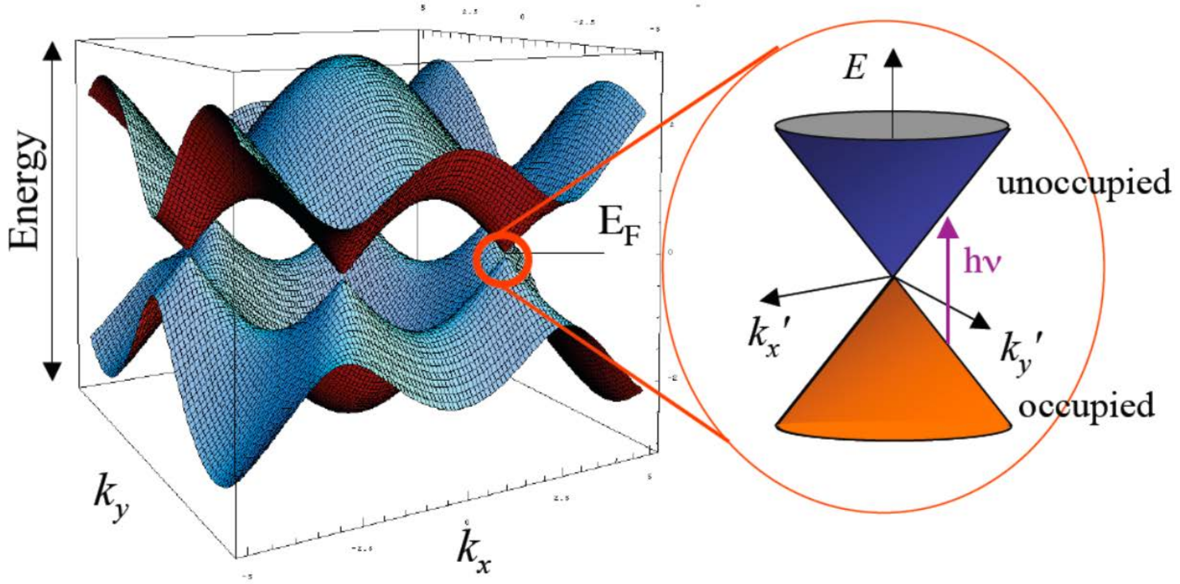


Fig. 14.50. The energy, *E*, for the excitations in graphene as a function of the wave numbers, k_x and k_y, in the *x* and *y* directions. The black line represents the Fermi energy for an undoped graphene crystal. Close to this Fermi level the energy spectrum is characterized by six double cones where the dispersion relation (energy versus momentum, $\hbar k$) is linear. This corresponds to massless excitations.

As a consequence, at low energies, even neglecting the true spin, the electrons can be described by an equation that is formally equivalent to the massless Dirac equation. Hence, the electrons and holes are called Dirac fermions and the six corners are called the Dirac points [144]. This pseudo-relativistic description is restricted to the chiral limit, i.e., to vanishing rest mass M_0, which leads to interesting additional features [144, 147]:

$$v_F \vec{\sigma} \cdot \nabla \psi(\vec{r}) = E \psi(\vec{r}).$$

``

Here $v_F \sim 10^6$ m/s (.003 c) is the Fermi velocity in graphene, which replaces the velocity of light in the Dirac theory; $\boldsymbol{\sigma}$ is the vector of the Pauli matrices, $\psi(\mathbf{r})$ is the two-component wave function of the electrons, and E is their energy [142].

The equation describing the electrons' linear dispersion relation is

$$E = \hbar v_F \sqrt{k_x^2 + k_y^2};$$

where the wavevector k is measured from the Dirac points (the zero of energy is chosen here to coincide with the Dirac points). The equation uses a pseudospin matrix formula that describes two sublattices of the honeycomb lattice [146].

Single-atom wave propagation. Electron waves in graphene propagate within a single-atom layer, making them sensitive to the proximity of other materials such as high-κ dielectrics, superconductors and ferromagnetics.

Electron transport. Experimental results from transport measurements show that graphene has a remarkably high electron mobility at room temperature, with reported values in excess of 15,000 $\text{cm}^2\cdot\text{V}^{-1}\cdot\text{s}^{-1}$. Additionally, the symmetry of the experimentally measured conductance indicates that hole and electron mobilities should be nearly the same [143]. The mobility is nearly independent of temperature between 10K and 100K [148, 149, 150], which implies that the dominant scattering mechanism is defect scattering. Scattering by the acoustic phonons of graphene intrinsically limits room temperature mobility to 200,000 $\text{cm}^2\cdot\text{V}^{-1}\cdot\text{s}^{-1}$ at a carrier density of 10^{12} cm^{-2} [150, 151], which was later demonstrated and is 10 times greater than copper.

The corresponding resistivity of the graphene sheet would be 10^{-6} Ω·cm. This is less than the resistivity of silver, the lowest known at room temperature. However, for room temperature graphene on SiO_2 substrates, scattering of electrons by optical phonons of the substrate is a larger effect than scattering by graphene’s own phonons. This limits mobility to 40,000 $\text{cm}^2\cdot\text{V}^{-1}\cdot\text{s}^{-1}$ [150].

The sheet conductivity of a 2D material is given by $\sigma = en\mu$. The mobility is theoretically limited to μ=200,000 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ by acoustic phonons at a carrier density of $n=10^{12}$ cm^{-2}. The 2D sheet resistivity, also called the resistance per square, is then 31 Ω. Our fictional hammock measuring 1m^2 would thus have a resistance of 31 Ω. Using the layer thickness we get a bulk conductivity of $0.96\text{x}10^6$ $\Omega^{-1}\text{cm}^{-1}$ for graphene.

Electrical resistance in 40-nanometer-wide nanoribbons of epitaxial graphene changes in discrete steps. The ribbons' conductance exceeds predictions by a factor of 10. The ribbons can act more like optical waveguides or quantum dots, allowing electrons to flow smoothly along the ribbon edges. In copper, resistance increases in proportion to length as electrons encounter impurities [152].

``

Transport is dominated by two modes. One is ballistic and temperature independent, while the other is thermally activated. Ballistic electrons resemble those in cylindrical carbon nanotubes. At room temperature, resistance increases abruptly at a particular length—the ballistic mode at 16 micrometres and the other at 160 nanometres (1% of the former length).

The ribbons were grown on the edges of three-dimensional structures etched into silicon carbide wafers. When the wafers are heated to approximately 1,000 °C (1,270 K; 1,830 °F), silicon is preferentially driven off along the edges, forming nanoribbons whose structure is determined by the pattern of the three-dimensional surface. The nanoribbons had perfectly smooth edges, annealed by the fabrication process. Electron mobility measurements surpassing one million correspond to a sheet resistance of one ohm per square— two orders of magnitude lower than in two-dimensional graphene.

Graphene electrons can cover micrometer distances without scattering, even at room temperature [142].

Despite zero carrier density near the Dirac points, graphene exhibits a minimum conductivity on the order of $4e^2/h$. The origin of this minimum conductivity is still unclear. However, rippling of the graphene sheet or ionized impurities in the SiO_2 substrate may lead to local puddles of carriers that allow conduction [143]. Several theories suggest that the minimum conductivity should be $4e^2/h$; however, most measurements are of order $4e^2/h$ or greater and depend on impurity concentration [153].

Near zero carrier density graphene exhibits positive photoconductivity and negative photoconductivity at high carrier density. This is governed by the interplay between photoinduced changes of both the Drude weight and the carrier scattering rate.

Graphene doped with various gaseous species (both acceptors and donors) can be returned to an undoped state by gentle heating in vacuum [153, 154]. Even for dopant concentrations in excess of 10^{12} cm^{-2} carrier mobility exhibits no observable change [154]. Graphene doped with potassium in ultra-high vacuum at low temperature can reduce mobility 20-fold [153, 155]. The mobility reduction is reversible on heating the graphene to remove the potassium.

Due to graphene's two dimensions, charge fractionalization (where the apparent charge of individual pseudoparticles in low-dimensional systems is less than a single quantum [156]) is thought to occur. It may therefore be a suitable material for constructing quantum computers [157] using anyonic circuits [158].

Anomalous quantum Hall effect. The quantum Hall effect is a quantum mechanical version of the Hall effect, which is the production of transverse (perpendicular to the main current) conductivity in the presence of a magnetic field. The quantization of the Hall effect σ_{xy} at integer

``

multiples (the "Landau level") of the basic quantity e^2/h (where e is the elementary electric charge and h is Planck's constant) It can usually be observed only in very clean silicon or gallium arsenide solids at temperatures around 3 K and very high magnetic fields.

At low temperatures and high magnetic fields, the exceptional mobility of graphene allows for the observation of the quantum Hall effect for both electrons and holes (Fig. 14.51) [13, 137]. Due to its unique band structure, the graphene quantum hall effect exhibits a subtle difference from the conventional quantum Hall effect in that plateaus occur at half integers of $4e^2/h$ rather than the typical $4e^2/h$.

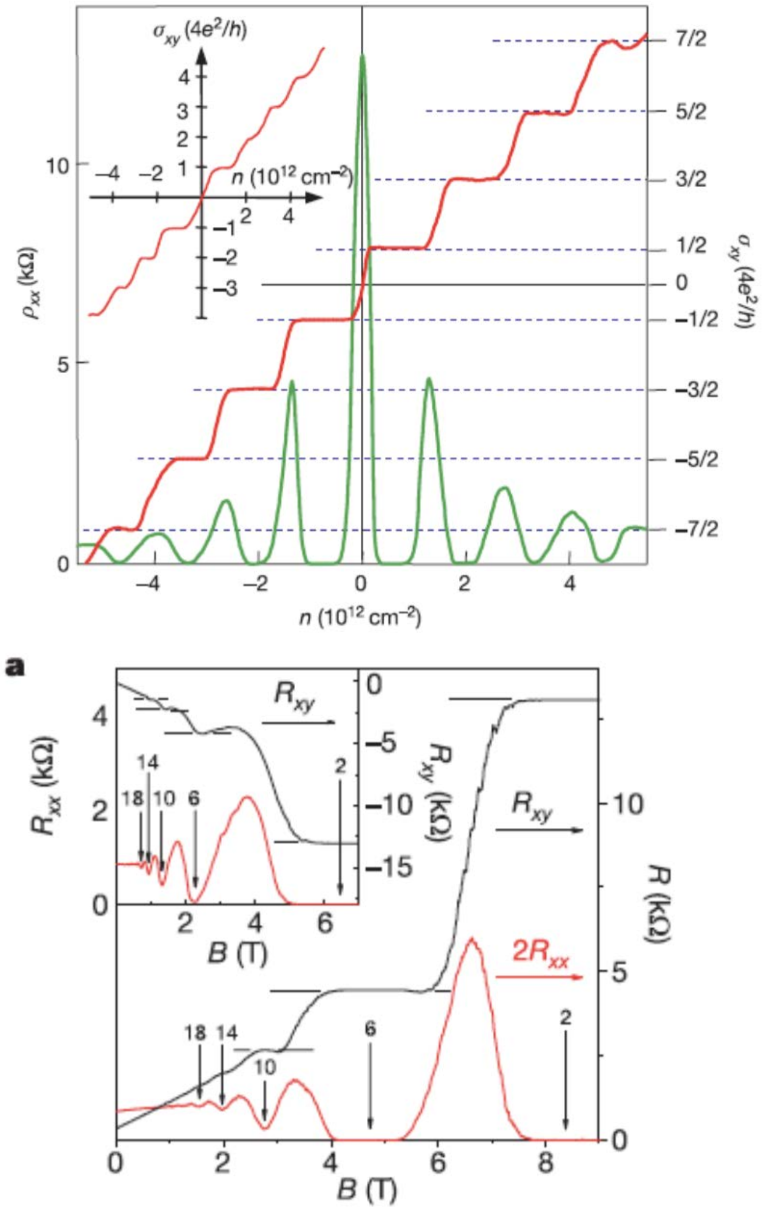


Fig. 14.51. Experimental observation of the unusual quantum Hall effect in graphene. (Top) The Hall conductivity (red) and the longitudinal resistivity (green) as a function of carrier density. The inset shows the Hall conductivity for bi-layer graphene. Note that the spacing between plateaus for graphene is $4e^2/h$, i.e. bigger than for the usual quantum Hall effect and that the steps occur at half

``

integer multiples of this value. For a bilayer of graphene the step height is the same, but the steps occur at integer multiples of $4e^2/h$ but with no step at zero density [148]. (Bottom) The longitudinal and Hall resistance as a function of magnetic flux density for an electron doped sample. The inset shows the same data for a hole doped sample [159].

Graphene shows the quantum Hall effect with respect to conductivity-quantization: the effect is *anomalous* in that the sequence of steps is shifted by 1/2 with respect to the standard sequence and with an additional factor of 4. Graphene's Hall conductivity is,

$$\sigma_{xy} = \pm 4 \cdot (N + 1/2)\mathrm{e}^2/\mathrm{h},$$

where N is the Landau level and the double valley and double spin degeneracies give the factor of 4. These anomalies are present at room temperature, i.e. at roughly 20 °C (293 K) [148].

This behavior is a direct result of graphene's massless Dirac electrons. In a magnetic field, their spectrum has a Landau level with energy precisely at the Dirac point. This level is a consequence of the Atiyah–Singer index theorem and is half-filled in neutral graphene [144], leading to the "+1/2" in the Hall conductivity [160]. Bilayer graphene also shows the quantum Hall effect, but with only one of the two anomalies (i.e. $\sigma_{xy} = \pm 4 \cdot N \cdot \mathrm{e}^2/\mathrm{h}$). In the second anomaly, the first plateau at $N=0$ is absent, indicating that bilayer graphene stays metallic at the neutrality point.

Unlike normal metals, graphene's longitudinal resistance shows maxima rather than minima for integral values of the Landau filling factor in measurements of the Shubnikov–de Haas oscillations, whereby the term *integral* quantum Hall effect. These oscillations show a phase shift of π, known as Berry's phase [148, 143]. Berry's phase arises due to the zero effective carrier mass near the Dirac points [161]. The temperature dependence of the oscillations reveals that the carriers have a non-zero cyclotron mass, despite their zero effective mass [148].

Graphene samples prepared on nickel films, and on both the silicon face and carbon face of silicon carbide, show the anomalous effect directly in electrical measurements [162 – 166]. Graphitic layers on the carbon face of silicon carbide show a clear Dirac spectrum in angle-resolved photoemission experiments, and the effect is observed in cyclotron resonance and tunneling experiments [167].

Strong magnetic fields. In magnetic fields above 10 Teslas or so additional plateaus of the Hall conductivity at $\sigma_{xy} = \nu e^2/h$ with $\nu = 0, \pm 1, \pm 4$ are observed [168]. A plateau at $\nu = 3$ [169] and the fractional quantum Hall effect at $\nu = ⅓$ were also reported [169, 170]. These observations with $\nu = 0, \pm 1, \pm 3, \pm 4$ indicate that the four-fold degeneracy (two valley and two spin degrees of freedom) of the Landau energy levels is partially or completely lifted.

``

14.4.4. Optical Properties of Graphene

Graphene is practically transparent. In the optical region it absorbs only 2.3% of the light (Fig.14.52). This number is in fact given by π α, where α is the fine structure constant that sets the strength of the electromagnetic force. In contrast to low temperature 2D systems based on semiconductors, graphene maintains its 2D properties at room temperature. Graphene also has several other interesting properties, which it shares with carbon nanotubes. It is substantially stronger than steel, very stretchable and can be used as a flexible conductor. Its thermal conductivity is much higher than that of silver.

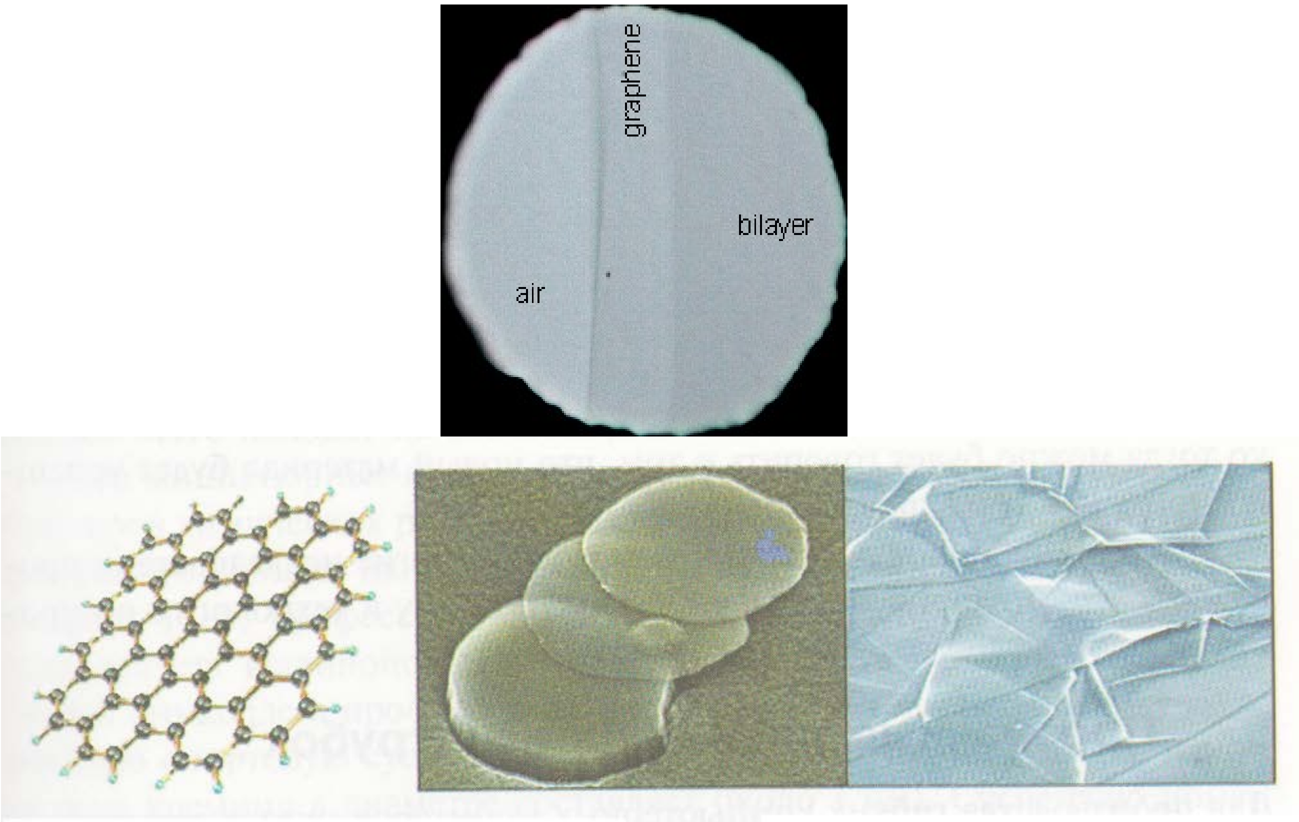


Fig. 14.52. Photograph of graphene in transmitted light. This one-atom-thick crystal can be seen with the naked eye because it absorbs approximately 2.6% of green light [171], and 2.3% of red light [172].

Graphene's unique optical properties produce an unexpectedly high opacity for an atomic monolayer in vacuum, absorbing $\pi\alpha \approx 2.3\%$ of red light, where α is the fine-structure constant [173]. This is a consequence of the unusual low-energy electronic structure of monolayer graphene that features electron and hole conical bands meeting each other at the Dirac point, which is qualitatively different from more common quadratic massive bands [172]. Based on the Slonczewski–Weiss–McClure (SWMcC) band model of graphite, the interatomic distance, hopping

``

value and frequency cancel when optical conductance is calculated using Fresnel equations in the thin-film limit.

Although confirmed experimentally, the measurement is not precise enough to improve on other techniques for determining the fine-structure constant.

Graphene's band gap can be tuned from 0 to 0.25 eV (about 5 micrometre wavelength) by applying voltage to a dual-gate bilayer graphene field-effect transistor (FET) at room temperature [174]. The optical response of graphene nanoribbons is tunable into the terahertz regime by an applied magnetic field [175]. Graphene/graphene oxide systems exhibit electrochromic behavior, allowing tuning of both linear and ultrafast optical properties [176].

A graphene-based Bragg grating (one-dimensional photonic crystal) has been fabricated and demonstrated its capability for excitation of surface electromagnetic waves in the periodic structure by using 633 nm He–Ne laser as the light source [177].

Saturable absorption. Such unique absorption could become saturated when the input optical intensity is above a threshold value. This nonlinear optical behavior is termed saturable absorption and the threshold value is called the saturation fluence. Graphene can be saturated readily under strong excitation over the visible to near-infrared region, due to the universal optical absorption and zero band gap. This has relevance for the mode locking of fiber lasers, where fullband mode locking has been achieved by graphene-based saturable absorber. Due to this special property, graphene has wide application in ultrafast photonics. Moreover, the optical response of graphene/graphene oxide layers can be tuned electrically [176, 178]. Saturable absorption in graphene could occur at the Microwave and Terahertz band, owing to its wideband optical absorption property. The microwave saturable absorption in graphene demonstrates the possibility of graphene microwave and terahertz photonics devices, such as microwave saturable absorber, modulator, polarizer, microwave signal processing and broad-band wireless access networks [179].

Nonlinear Kerr effect. Under more intensive laser illumination, graphene could also possess a nonlinear phase shift due to the optical nonlinear Kerr effect. Based on a typical open and close aperture z-scan measurement, graphene possesses a giant non-linear Kerr coefficient of 10^{-7} $cm^2 \cdot W^{-1}$, almost nine orders of magnitude larger than that of bulk dielectrics [180]. This suggests that graphene may be a powerful nonlinear Kerr medium, with the possibility of observing a variety of nonlinear effects, the most important of which is the soliton [181].

14.4.5. Mechanical Properties of Graphene

Graphite is unique in that the elastic constants in the direction perpendicular are vastly different than the elastic constants along the basal plane. This was known for quite some time and

``

was experimentally measured during the 1960s and 1970s. The experimental mechanical properties of graphene are largely unexplored and the time is ripe to revisit some of the old assumptions about bulk graphite to determine how the elastic constants scale down to the atomic thicknesses. By working with single atomic layers or few atomic layers some of the uncertainties involved in working with large single crystals such as dislocations and defects are avoided.

The carbon–carbon bond length in graphene is about 0.142 nanometers. Graphene sheets stack to form graphite with an interplanar spacing of 0.335 nm. Graphene is the strongest material ever tested [182], with an intrinsic Tensile strength of 130 GPa and a Young's modulus (stiffness) of 1 TPa (150,000,000 psi) [183]. The Nobel announcement illustrated this by saying that a 1 square meter graphene hammock would support a 4 kg cat but would weigh only as much as one of the cat's whiskers, at 0.77 mg (about 0.001% of the weight of 1 m^2 of paper).

Graphene has a breaking strength of 42N/m. Steel has a breaking strength in the range of 250-1200 MPa= 0.25-1.2x10^9 N/m^2. For a hypothetical steel film of the same thickness as graphene (which can be taken to be 3.35Å=3.35x10^{-10} m, *i.e.* the layer thickness in graphite), this would give a 2D breaking strength of 0.084-0.40 N/m. Thus graphene is more than 100 times stronger than the strongest steel. In our 1 m2 hammock tied between two trees you could place a weight of approximately 4 kg before it would break. It should thus be possible to make an almost

The spring constant of suspended graphene sheets has been measured using an atomic force microscope (AFM). Graphene sheets were suspended over SiO_2 cavities where an AFM tip was used to apply a stress to the sheet to test its mechanical properties. Its spring constant was in the range 1–5 N/m and the stiffness was 0.5 TPa, which differs from that of bulk graphite. These intrinsic properties could lead to applications such as NEMS as pressure sensors and resonators [184].

Due to its large surface energy and out of plane ductility, flat graphene sheets are unstable with respect to scrolling, i.e. bending into a cylindrical shape, which is its lower-energy state [185].

As is true of all materials, regions of graphene are subject to thermal and quantum fluctuations in relative displacement. Although the amplitude of these fluctuations is bounded in 3D structures (even in the limit of infinite size), the Mermin–Wagner theorem shows that the amplitude of long-wavelength fluctuations grows logarithmically with the scale of a 2D structure, and would therefore be unbounded in structures of infinite size. Local deformation and elastic strain are negligibly affected by this long-range divergence in relative displacement. It is believed that a sufficiently large 2D structure, in the absence of applied lateral tension, will bend and crumple to form a fluctuating 3D structure. Researchers have observed ripples in suspended layers of grapheme [122], and it has been proposed that the ripples are caused by thermal fluctuations in the material. As a consequence of these dynamical deformations, it is debatable whether graphene is truly a 2D

``

structure [123, 124, 186]. It has recently been shown that these ripples, if amplified through the introduction of vacancy defects, can impart a negative Poisson's ratio into graphene, resulting in the thinnest auxetic material known so far.

Fracture Toughness. The researchers have indicated that despite its strength, graphene is also relatively brittle, with a fracture toughness of ~4 MPa√m. This indicates that imperfect graphene is likely to crack in a brittle manner like ceramic materials, as opposed to many metallic materials which tend to have fracture toughnesses in the range of 15-50 MPa√m. Later it has been announced that graphene showed a greater ability to distribute force from an impact than any known material, 10x that of steel per unit weight. The force was transmitted at 22.2 kilometres per second (13.8 mi/s).

14.4.6. Applications of Graphene

Potential graphene applications include lightweight, thin, flexible, yet durable display screens, electric circuits, and solar cells, as well as various medical, chemical and industrial processes enhanced or enabled by the use of new graphene materials.

Electronics. For integrated circuits, graphene has a high carrier mobility, as well as low noise, allowing it to be used as the channel in a field-effect transistor. Single sheets of graphene are hard to produce and even harder to make on an appropriate substrate [187]. In 2008, the smallest transistor so far, one atom thick, 10 atoms wide was made of graphene [188]. IBM announced that they had fabricated and characterized graphene transistors operating at GHz frequencies. Later an n-type transistor was announced meaning that both n and p-type graphene transistors had been created [189]. A functional graphene integrated circuit was demonstrated– a complementary inverter consisting of one p- and one n-type graphene transistor [190]. However, this inverter suffered from a very low voltage gain. Graphene was epitaxially grown on SiC in a quantity and with quality suitable for mass production of integrated circuits. At high temperatures, the quantum Hall effect could be measured in these samples. IBM built 'processors' using 100 GHz transistors on 2-inch (51 mm) graphene sheets [191]. It had been succeeded in creating the first graphene-based integrated circuit, a broadband radio mixer [192]. The circuit handled frequencies up to 10 GHz. Its performance was unaffected by temperatures up to 127 °C. An 8 transistor 1.28 GHz ring oscillator circuit was described in.

Transistors. Graphene exhibits a pronounced response to perpendicular external electric fields, potentially forming field-effect transistors (FET) (Fig. 14.53). FETs with an on-off ratio of ~30 at room temperature has been obtained. Also an all-graphene planar FET with side gates has

``

been prepared. The devices showed changes of 2% at cryogenic temperatures. The first top-gated FET (on–off ratio of <2) was demonstrated in [193]. Graphene nanoribbons may prove generally capable of replacing silicon as a semiconductor. Hundreds of transistors on a single chip and very high frequency transistors have been produced [194, 195]. A switching effect based on a reversible chemical modification of the graphene layer that gives an on–off ratio of greater than six orders of magnitude has been demonstrated. These reversible switches could potentially be employed in nonvolatile memories [196]. Researchers demonstrated four different types of logic gates, each composed of a single graphene transistor [197]. Practical uses for these circuits are limited by the very small voltage gain they exhibit. Typically, the amplitude of the output signal is about 40 times less than that of the input signal. Moreover, none of these circuits operated at frequencies higher than 25 kHz. At the same time, tight-binding numerical simulations demonstrated that the band-gap induced in graphene bilayer field effect transistors is not sufficiently large for high-performance transistors for digital applications, but can be sufficient for ultra-low voltage applications, when exploiting a tunnel-FET architecture [198, 199]. Transistors with an on/off rate of 100 gigahertz, far exceeding the rates of previous attempts, and exceeding the speed of silicon transistors with an equal gate length has been announced. The 240 nm devices were made with conventional silicon-manufacturing equipment [200]. 3d printing (additive manufacturing) as a method for fabricating graphene devices has been used [201]. Researchers demonstrated graphene's high mobility in a detector that allows broad band frequency selectivity ranging from the THz to IR region (0.76–33 THz) [202]. A separate group created a terahertz-speed transistor with bistable characteristics, which means that the device can spontaneously switch between two electronic states. The device consists of two layers of graphene separated by an insulating layer of boron nitride a few atomic layers thick. Electrons move through this barrier by quantum tunneling. These new transistors exhibit “negative differential conductance,” whereby the same electrical current flows at two different applied voltages [203].

``

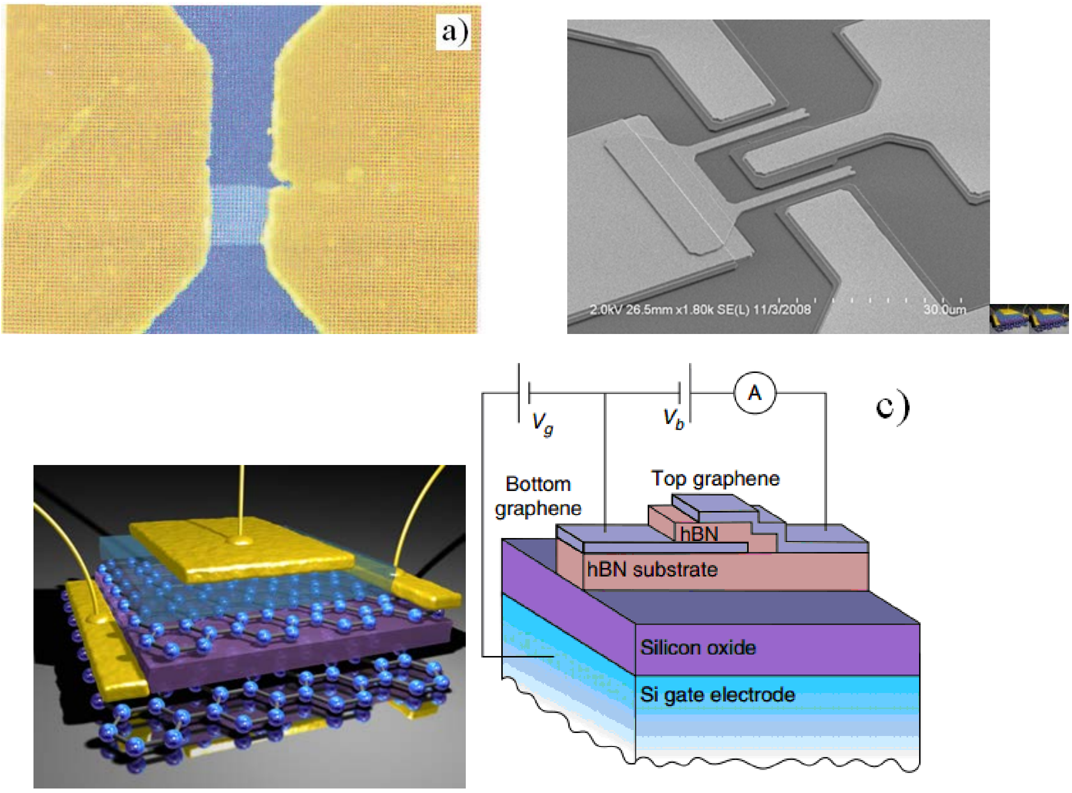


Fig. 14.53. (a) Graphene transistor. (b) Illustration of tunneling transistor based on vertical graphene heterostructures. Tunneling current between two graphene layers can be controlled by gating. (c) Schematic diagram of graphene-BN resonant tunneling transistor.

Graphene does not have an energy band-gap, which presents a hurdle for its applications in digital logic gates. The efforts to induce a band-gap in graphene via quantum confinement or surface functionalization have not resulted in a breakthrough. The negative differential resistance experimentally observed in graphene field-effect transistors of "conventional" design allows for construction of viable non-Boolean computational architectures with the gap-less graphene. The negative differential resistance — observed under certain biasing schemes — is an intrinsic property of graphene resulting from its symmetric band structure. The results present a conceptual change in graphene research and indicate an alternative route for graphene's applications in information processing [204]. Researchers reported the creation of transistors printed on flexible plastic that operate at 25 gigahertz, sufficient for communications circuits and that can be fabricated at scale. The researchers first fabricate the non-graphene-containing structures—the electrodes and gates—on plastic sheets. Separately, they grow large graphene sheets on metal, then peel it off and transfer it to the plastic. Finally, they top the sheet with a waterproof layer. The devices work after being soaked in water, and are flexible enough to be folded.

Capitalizing on graphene’s electrical awesomeness, HRL Laboratories (owned by Boeing and General Motors) built the world’s first functional radio frequency (RF) transistor using graphene in 2008. The tiny device, known as an RF field-effect transistor, can pick up radio frequencies while hardly gobbling any electrical power.

The electronic circuit can be also created based on grephene nanoribbons (Fig. 14. 54).

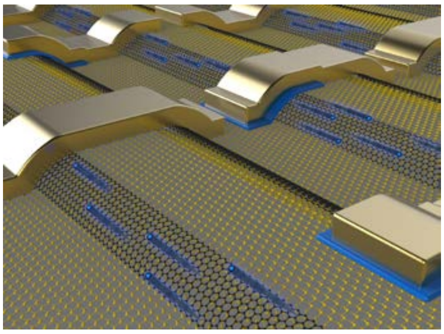


Fig. 14. 54. Conceptual drawing of an electronic circuit composed of interconnected graphene nanoribbons (black atoms) that are epitaxially grown on steps etched in silicon carbide (yellow atoms). Electrons (blue) travel ballistically along the ribbon and then from one ribbon to the next via the metal contacts. Electron flow is modulated by electrostatic gates.

Trilayer graphene. An electric field can change trilayer graphene's crystal structure, transforming its behavior from metal-like to semiconductor-like. A sharp metal scanning tunneling microscopy tip was able to move the domain border between the upper and lower graphene configurations. One side of the material behaves as a metal, while the other side behaves as a semiconductor. Trilayer graphene can be stacked in either Bernal or rhombohedral configurations, which can exist in a single flake. The two domains are separated by a precise boundary at which the middle layer is strained to accommodate the transition from one stacking pattern to the other.

Silicon transistors function as either p-type or n-type semiconductors, whereas graphene could operate as both. This lowers costs and is more versatile. The technique provides the basis for a field-effect transistor. Scalable manufacturing techiques have yet to be developed. In trilayer graphene, the two stacking configurations exhibit very different electronic properties. The region between them consists of a localized strain soliton where the carbon atoms of one graphene layer shift by the carbon–carbon bond distance. The free-energy difference between the two stacking configurations scales quadratically with electric field, favoring rhombohedral stacking as the electric field increases. This ability to control the stacking order opens the way to new devices that combine structural and electrical properties. Graphene-based transistors could be much thinner than modern silicon devices, allowing faster and smaller configurations.

Transparent conducting electrodes. Graphene's high electrical conductivity and high optical transparency make it a candidate for transparent conducting electrodes, required for such

``

applications as touch screens, liquid crystal displays, organic photovoltaic cells, and organic light-emitting diodes. In particular, graphene's mechanical strength and flexibility are advantageous compared to indium tin oxide, which is brittle. Graphene films may be deposited from solution over large areas [205, 206].

Large-area, continuous, transparent and highly conducting few-layered graphene films were produced by chemical vapor deposition and used as anodes for application in photovoltaic devices. A power conversion efficiency (PCE) up to 1.71% was demonstrated, which is 55.2% of the PCE of a control device based on indium tin oxide [206].

Organic light-emitting diodes (OLEDs) with graphene anodes have been demonstrated in [207]. The electronic and optical performance of graphene-based devices are similar to devices made with indium tin oxide.

A carbon-based device called a light-emitting electrochemical cell (LEC) was demonstrated with chemically-derived graphene as the cathode and the conductive polymer PEDOT as the anode [208]. Unlike its predecessors, this device contains only carbon-based electrodes, with no metal.

Energy.

Ethanol distillation. Graphene oxide membranes allow water vapor to pass through, but are impermeable to other liquids and gases [209]. This phenomenon has been used for further distilling of vodka to higher alcohol concentrations, in a room-temperature laboratory, without the application of heat or vacuum as used in traditional distillation methods. Further development and commercialization of such membranes could revolutionize the economics of biofuel production and the alcoholic beverage industry.

Solar cells. Graphene has a unique combination of high electrical conductivity and optical transparency, which make it a candidate for use in solar cells. A single sheet of graphene is a zero-bandgap semiconductor whose charge carriers are delocalized over large areas, implying that carrier scattering does not occur. Because this material only absorbs 2.6% of green light and 2.3% of red light [210], it is a candidate for applications requiring a transparent conductor. Graphene can be assembled into a film electrode with low roughness. However, graphene films produced via solution processing contain lattice defects and grain boundaries that act as recombination centers and decrease the material's electrical conductivity. Thus, these films must be made thicker than one atomic layer to obtain useful sheet resistances. This added resistance can be combated by incorporating conductive filler materials, such as a silica matrix. Reduced graphene film's electrical conductivity can be improved by attaching large aromatic molecules such as pyrene-1-sulfonic acid sodium salt (PyS) and the disodium salt of 3,4,9,10-perylenetetracarboxylic diimide bisbenzenesulfonic acid (PDI). These molecules, under high temperatures, facilitate better π-

``

conjugation of the graphene basal plane. Graphene films have high transparency in the visible and near-infrared regions and are chemically and thermally stable. For graphene to be used in commercial solar cells, large-scale production is required. However, no scalable process for producing graphene is available, including the peeling of pyrolytic graphene or thermal decomposition of silicon carbide.

Graphene's high charge mobilities recommend it for use as a charge collector and transporter in photovoltaics (PV). Using graphene as a photoactive material requires its bandgap to be 1.4–1.9 eV. Single cell efficiencies of nanostructured graphene-based PVs of over 12% were achieved. Organic photovoltaics could be "devices in which semiconducting graphene is used as the photoactive material and metallic graphene is used as the conductive electrodes. Graphene-silicon heterojunction solar cell, where graphene served as a transparent electrode and introduced a built-in electric field near the interface between the graphene and n-type silicon to help collect photo-generated carriers has been prepared. More studies promote this new type of photovoltaic device. The efficiency of 8.6% for a prototype cell consisting of a wafer of silicon coated with a layer of graphene doped with trifluoromethanesulfonyl-amide (TFSA) has been reported. Chemical doping could improve the graphene characteristics and significantly enhance the efficiency of graphene-silicon solar cell to 9.6% [211]. 15.6% percent efficiency using a combination of titanium oxide and graphene as a charge collector and perovskite as a sunlight absorber has been achieved. The device is manufacturable at temperatures under 150°C using solution-based deposition. This lowers production costs and offers the potential using flexible plastics [212].

Large scale production of highly transparent graphene films by chemical vapor deposition was achieved. In this process, ultra-thin graphene sheets are created by first depositing carbon atoms in the form of graphene films on a nickel plate from methane gas. A protective layer of thermoplastic is laid over the graphene layer and the nickel underneath is dissolved in an acid bath. The final step is to attach the plastic-protected graphene to a flexible polymer sheet, which can then be incorporated into an OPV cell. Graphene/polymer sheets range in size up to 150 square centimeters and can be used to create dense arrays of flexible OPV cells. It may eventually be possible to run printing presses covering extensive areas with inexpensive solar cells, much like newspaper presses print newspapers (roll-to-roll).

Silicon generates only one current-driving electron for each photon it absorbs, while graphene can produce multiple electrons. Solar cells made with graphene could offer 60% conversion efficiency – double the widely accepted maximum efficiency of silicon cells.

Fuel cells. Appropriately perforated graphene (and hexagonal boron nitride hBN) can allow protons to pass through it, offering the potential for using graphene monolayers as a barrier that blocks hydrogen atoms but not protons/ionized hydrogen (hydrogen atoms with their electrons stripped off). They could even be used to extract hydrogen gas out of the atmosphere that could power electric generators with ambient air. The membranes are more effective at elevated temperatures and when covered with catalytic nanoparticles such as platinum. Graphene could solve a major problem for fuel cells: fuel crossover that reduces efficiency and durability. At room temperature, proton conductivity with monolayer hBN, outperforms graphene, with resistivity to proton flow of about 10 Ω cm^2 and a low activation energy of about 0.3 electronvolts. At higher temperatures, graphene outperforms with resistivity estimated to fall below 10^{-3} Ω cm^2 above 250 °C.

Sensors

Molecular adsorbtion. Theoretically graphene makes an excellent sensor due to its 2D structure. The fact that its entire volume is exposed to its surrounding environment makes it very efficient to detect adsorbed molecules. However, similar to carbon nanotubes, graphene has no dangling bonds on its surface. Gaseous molecules cannot be readily adsorbed onto graphene surfaces, so intrinsically graphene is insensitive [213]. The sensitivity of graphene chemical gas sensors can be dramatically enhanced by functionalization, for example, coating the film with a thin layer of certain polymers. The thin polymer layer acts like a concentrator that absorbs gaseous molecules. The molecule absorption introduces a local change in electrical resistance of graphene sensors. While this effect occurs in other materials, graphene is superior due to its high electrical conductivity (even when few carriers are present) and low noise, which makes this change in resistance detectable [214].

Piezoelectric effect. Density functional theory simulations predict that depositing certain adatoms on graphene can render it piezoelectrically responsive to an electric field applied in the out-of-plane direction. This type of locally engineered piezoelectricity is similar in magnitude to that of bulk piezoelectric materials and makes graphene a candidate for control and sensing in nanoscale devices [215].

Body motion. Rubber bands infused with graphene("G-bands") can be used as inexpensive body sensors. The bands remain pliable and can be used as a sensor to measure breathing, heart rate, or movement. Light weight sensor suits for vulnerable patients could make it possible to remotely monitor subtle movement. These sensors display 10-fold increases in resistance and work at strains exceeding 800%. Gauge factors of up to 35 were observed. Such sensors can function at vibration frequencies of at least 160 Hz. At 60 Hz, strains of at least 6% at strain rates exceeding 6000 %/s can be monitored.

Medicine

``

Tissue engineering. Graphene has been investigated for tissue engineering. It has been used as a reinforcing agent to improve the mechanical properties of biodegradable polymeric nanocomposites for engineering bone tissue applications [216]. Dispersion of low weight % of graphene (~0.02 wt.%) increased in compressive and flexural mechanical properties of polymeric nanocomposites.

Contrast agents/bioimaging. Functionalized and surfactant dispersed graphene solutions have been designed as blood pool MRI contrast agents [217]. Additionally, iodine and manganese incorporating graphene nanoparticles have served as multimodal MRI-CT contrast agents [218]. Graphene micro- and nano-particles have served as contrast agents for photoacoustic and thermoacoustic tomography [219]. Graphene has also been reported to be efficiently taken up cancerous cells thereby enabling the design of drug delivery agents for cancer therapy [220]. Graphene nanoparticles of various morphologies are non-toxic at low concentrations and do not alter stem cell differentiation suggesting that they may be safe to use for biomedical applications [221].

Polymerase chain reaction. Graphene is reported to have enhanced PCR by increasing the yield of DNA product [222]. Experiments revealed that graphene's thermal conductivity could be the main factor behind this result. Graphene yields DNA product equivalent to positive control with up to 65% reduction in PCR cycles.

Graphene's modifiable chemistry, large surface area, atomic thickness and molecularly gatable structure make antibody-functionalized graphene sheets excellent candidates for mammalian and microbial detection and diagnosis devices [223]. Graphene is so thin water has near-perfect wetting transparency which is an important property particularly in developing bio-sensor applications [224]. This means that a sensors coated in graphene have as much contact with an aqueous system as an uncoated sensor, while it remains protected mechanically from its environment. Integration of graphene (thickness of 0.34 nm) layers as nanoelectrodes into a nanopore [225] can potentially solve a bottleneck for nanopore-based single-molecule DNA sequencing. New elastic composite materials for condoms containing nanomaterials like graphene have been developed. Graphene-based, transparent (across infrared to ultraviolet frequencies), flexible, implantable medical sensor microarrays were announced that allow the viewing of brain tissue hidden by implants. Optical transparency was >90%. Applications demonstrated include optogenetic activation of focal cortical areas, *in vivo* imaging of cortical vasculature via fluorescence microscopy and 3D optical coherence tomography.

Drug delivery. It has been discovered that the sheet of graphene oxide can be transformed into liquid crystal droplets spontaneously – like a polymer - simply by placing the material in a solution and manipulating the pH. The graphene droplets change their structure at the presence of an external magnetic field. This finding opens the door for potential use of carrying drug in the graphene droplets and drug release upon reaching the targeted tissue when the droplets change shape under the magnetic field. Another possible application is in disease detection if graphene is found to change shape at the presence of certain disease markers such as toxins [226].

A graphene ‘flying carpet’ was demonstrated to deliver two anti-cancer drugs sequentially to the lung tumor cells (A549 cell) in a mouse model. Doxorubicin (DOX) is embedded onto the graphene sheet, while the molecules of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) are linked to the nanostructure via short peptide chains. Injected intravenously, the graphene strips with the drug playload preferentially concentrate to the cancer cells due to common blood vessel leakage around the tumor. Receptors on the cancer cell membrane bind TRAIL and cell surface enzymes clip the peptide thus release the drug onto the cell surface. Without the bulky TRAIL, the graphene strips with the embedded DOX are swallowed into the cells. The intracellular acidic environment promotes DOX’s release from graphene. TRAIL on the cell surface triggers the apoptosis while DOX attacks the nucleus. These two drugs work synergistically and were found to be more effective than either drug alone.

14.5. Conclusion

Carbon based nanostructures have interesting and unique properties and mane promising applications. Carbon has the extraordinary chemical flexibility to serve as the sole basic unit of such a variety of compounds. The materials discussed here are testing grounds for important concepts in condensed matter theory; at the same time there is a constant effort to find potentially useful applications.

Fullerene-like structures are known in more complex materials, including biologically important molecules, including the HIV virus. C60 was investigated as a potential drug delivery agent, it may have antiviral activity, and endohedral C_{60} compounds can be possibly used as a spin label (or relaxation agent) in magnetic resonance spectroscopy. Bulk quantities of nanotubes are considered for reinforcement additives in metals and for storage of hydrogen gas. In scanning tunneling and force microscopy the nanotubes may serve as the sharpest possible microscope tip. Single nanotubes were shown to work as single electron transistors, reaching the ultimate limit of nanoscale electronics. Electron emission and light emission from the tips of the nanotubes have been seen at relatively low applied voltages, and the nanotube-based nanoemissive display (NED) technology enables manufacturers to design large flat panel displays that exceed the image quality characteristics of plasma and LCD screens.

``

Possibilities for nanotube-based applications are numerous and will be a step forward in technology. Most of these applications have shown to function in laboratories, but even less applications are sufficiently developed for production. Molecular integrated circuits are still a rather big step away, although single molecular transistors can be made with the use of nanotubes. More accurate placement techniques and smaller electrodes are needed for better device functioning, because large transistor densities are necessary if carbon nanotube transistors want to compete with conventional transistors. Energy storage in carbon nanotubes is even a bit further away from application because storage capacities are not yet satisfactory. Some future applications of carbon nanotubes look very promising. All we need are better production techniques for large amounts of purified nanotubes that have to be found in the near future. In our opinion, nanotubes promises to open up a way to new applications that might be cheaper, lower in weight and have a better efficiency.

Graphene is a unique system in many ways. It is truly 2D, has unusual electronic excitations described in terms of Dirac fermions that move in a curved space, is an interesting mix of a semiconductor zero density of states and a metal gaplessness, and has properties of soft matter. The electrons in graphene seem to be almost insensitive to disorder and electron-electron interactions and have very long mean free paths. Hence, graphene’s properties are different from what is found in usual metals and semiconductors. Graphene has also a robust but flexible structure with unusual phonon modes that do not exist in ordinary 3D solids. In some sense, grapheme brings together issues in quantum gravity and particle physics, and also from soft and hard condensed matter.

Interestingly enough, these properties can be easily modified with the application of electric and magnetic fields, addition of layers, control of its geometry, and chemical doping. Moreover, graphene can be directly and relatively easily probed by various scanning probe techniques from mesoscopic down to atomic scales, because it is not buried inside a 3D structure. This makes graphene one of the most versatile systems in condensed-matter research.

Besides the unusual basic properties, graphene has the potential for a large number of applications from chemical sensors to transistors. Graphene can be chemically and/or structurally modified in order to change its functionality and hence forth its potential applications. Moreover, graphene can be easily obtained from graphite, a material that is abundant on the Earth’s surface. This particular characteristic makes grapheme one of the most readily available materials for basic research since it frees economically challenged research.

The majority of the theoretical and experimental efforts have concentrated on the single layer, perhaps because of its simplicity and the natural attraction that a one atom thick material, which can be produced by simple methods in almost any laboratory, creates. Nevertheless, few-layer graphene is equally interesting and unusual with a technological potential, perhaps larger than

``

the single layer. Indeed, the theoretical understanding and experimental exploration of multilayers is far behind the single layer. This is a fertile and open field of research for the future.

Chemical modification of graphene can lead to entirely new physics. Depending on the nature of chemical dopants and how they are introduced into the graphene lattice adsorption, substitution, or intercalation, there can be many results. Small concentrations of adsorbed alkali metal can be used to change the chemical potential while adsorbed transition elements can lead to strong hybridization effects that affect the electronic structure. In fact, the introduction of d- and f-electron atoms in the grapheme lattice may produce a significant enhancement of the electron-electron interactions. Hence, it is easy to envision a plethora of many-body effects that can be induced by doping and have to be studied in the context of Dirac electrons: Kondo effect, ferromagnetism, antiferromagnetism, and charge- and spin-density waves.

The carbon nanostructures, may have a real impact on our everyday life in the not-too distant future.

``

Chapter 15. NANOWIRES AND WHISKERS

Nanowires are especially attractive for nanoscience studies as well as for nanotechnology applications. Nanowires, compared to other low dimensional systems, have two quantum confined directions while still leaving one unconfined direction for electrical conduction. This allows them to be used in applications which require electrical conduction, rather than tunneling transport. Because of their unique density of electronic states, nanowires in the limit of small diameters are expected to exhibit significantly different optical, electrical, and magnetic properties from their bulk 3-D crystalline counterparts. Increased surface area, very high density of electronic states and joint density of states near the energies of their van Hove singularities, enhanced exciton binding energy, diameter-dependent bandgap, and increased surface scattering for electrons and phonons are just some of the ways in which nanowires differ from their corresponding bulk materials. Yet the sizes of nanowires are typically large enough (> 1 nm in the quantum confined direction) to have local crystal structures closely related to their parent materials, thereby allowing theoretical predictions about their properties to be made on the basis of an extensive literature relevant to their bulk properties. Not only do nanowires exhibit many properties that are similar to, and others that are distinctly different from those of their bulk counterparts, nanowires have the advantage from an applications standpoint in that some of the materials parameters critical for certain properties can be independently controlled in nanowires but not in their bulk counterparts. Certain properties can also be enhanced nonlinearly in small diameter nanowires by exploiting the singular aspects of the 1-D electronic density of states. Furthermore, nanowires have been shown to provide a promising framework for applying the “bottom-up” approach for the design of nanostructures for nanoscience investigations and for potential nanotechnology applications. Driven by: (1) these new research and development opportunities, (2) the smaller and smaller length scales now being used in the semiconductor, opto-electronics, and magnetics industries, and (3) the dramatic development of the biotechnology industry where the action is also at the nanoscale, the nanowire research field has developed with exceptional speed in the last years.

One-dimensional (1D) nanostructures represent the smallest dimension structure that can efficiently transport electrical carriers and can play an important role as both interconnect and functional device elements in integrated nanosystems.

15.1. PREPARATION OF SEMICONDUCTOR NANOWIRES

Nanowires made of inorganic semiconducting materials are of considerable interest due to their unique properties and potential applications [1, 2]. For instance, semiconductor nanowires with lateral dimensions less than 15 nm exhibit band gaps in the ultraviolet (UV) and visible regions of light [3]. In addition, nanowires of single-crystalline inorganic semiconductors show excellent transport properties of electrons and holes [4, 5]. As a result, they offer great promise for electronic and optoelectronic devices with interesting properties. A range of approaches have been developed to generate nanowires of various materials (e.g., Si, Ge, III–V compounds, II–VI compounds, and oxides). The techniques can be classified into two general categories, i.e., ‘‘bottom-up’’ and ‘‘top-down’’ approaches.

15.1.1. ‘‘Bottom-Up’’ Approaches

A bottom-up approach, in which functional electronic structures are assembled from chemically synthesized, well-defined nanoscale building blocks, much like the way nature uses proteins and other macromolecules to construct complex biological systems, represents a flexible alternative to conventional top-down methods [6-8], and moreover, the bottom-up approach has the potential to go far beyond the limits of top-down technology by defining key nanometer scale metrics through synthesis and subsequent assembly - not by lithography.

To enable this bottom-up pathway requires that three key problems, which are at the heart of devices and integration in the broadest terms, be addressed (Fig. 15.1). First, the bottom-up approach requires nanoscale building blocks with precisely controlled and tunable chemical composition, structure, morphology, and size, since these characteristics determine their corresponding electronic and optical properties. To meet this goal necessitates developing methods that enable rational design and predictable synthesis of the building blocks. Second, it is essential to develop and explore the limits of functional devices based on these building blocks. Nanodevices may behave in ways similar to current electronic and optoelectronic devices, although it is also expected that new and potentially revolutionary device concepts will emerge from these building blocks, for example, due to quantum properties. Third and central to the bottom-up concept is the development of device architectures that enable high-density integration with desired function, and the development of hierarchical assembly methods that can organize building blocks into these architectures.

Addressing and overcoming the hurdles in these three major areas of the bottom-up approach could revolutionize fabrication and manufacturing, make a quantum jump in miniaturization, and lead to reduced power consumption and increased speed in next generation electronics and photonics. Moreover, it is very likely that the bottom-up approach could enable entirely new device concepts and new systems. For example, it is possible to combine seamlessly chemically distinct nanoscale building blocks, which could not be integrated together in top-down processing, into the same device architecture and thereby obtain unique function and/or combinations of function in an integrated system. Small and highly perfect building blocks may also lead to quantum electronic or quantum optical devices that enable quantum computing in architecture that has many common features with digital systems.

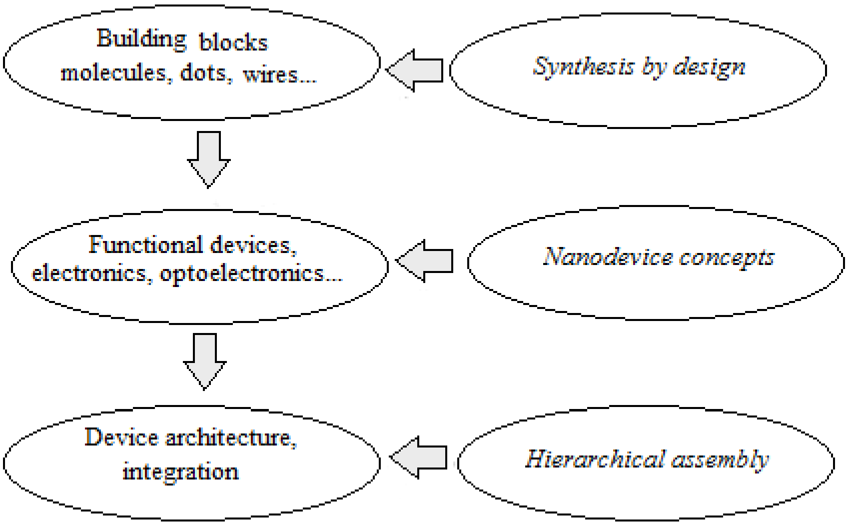


Fig. 15.1. Schematic outlining key challenges (open ellipses) and specific research areas (shaded ellipses) required to enable the bottom-up approach to nanoelectronic systems.

Atoms of semiconductor materials can crystallize and grow into nanowires in both gaseous and liquid media by controlling reaction conditions [9-12]. Among all approaches the vapor-based methods, the vapor–liquid– solid (VLS) process seems to be the most successful and widely used method for generating single crystal nanowires in large quantities (Fig. 15.2) [13]. The obtained results confirm that the VLS process starts with the dissolution of vapors of target materials into nanosized liquid droplets of a metal catalyst (e.g., Au), followed by nucleation and growth of single-crystalline wires. The liquid droplets are generated from metal nanoparticles at elevated temperatures; their size determines the diameter of wires.

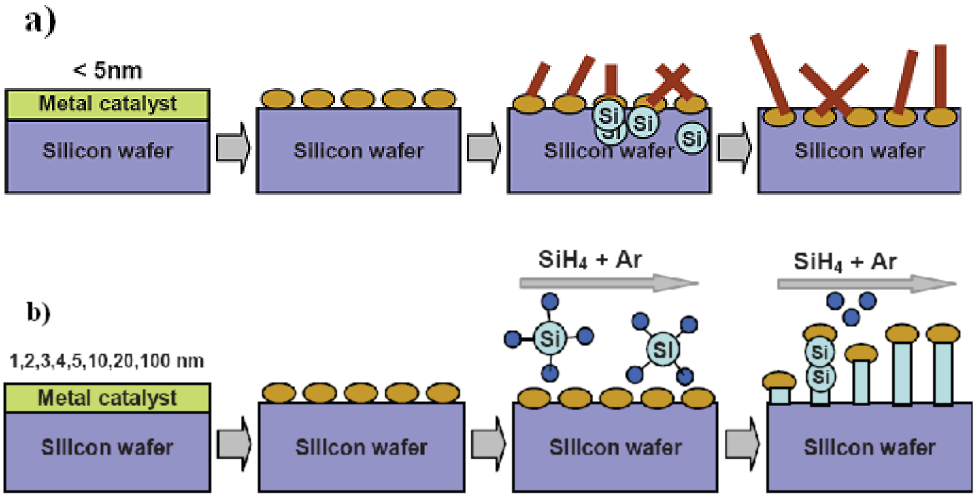


Fig. 15.2. Schematic images of (a) solid-liquid-solis (SLS) and (b) vapor–liquid– solid (VLS) processes.

Chemical Vapor Deposition Process

The vapor-liquid-solid method (VLS) is a mechanism for the growth of one-dimensional structures, such as nanowires, from chemical vapor deposition (CVD) (Fig. 15.3.). The growth of a crystal through direct adsorption of a gas phase on to a solid surface is generally very slow. The VLS mechanism circumvents this by introducing a catalytic liquid alloy phase which can rapidly adsorb a vapor to supersaturation levels, and from which crystal growth can subsequently occur from nucleated seeds at the liquid-solid interface. The physical characteristics of nanowires grown in this manner depend, in a controllable way, upon the size and physical properties of the liquid alloy. The VLS mechanism is typically described in three stages [14]: (*i*) preparation of a liquid alloy droplet upon the substrate from which a wire is to be grown, (*ii*) introduction of the substance to be grown as a vapor, which adsorbs on to the liquid surface, and difluses in to the droplet, and (*iii*) supersaturation and nucleation at the liquid/solid interlace leading to axial crystal growth.

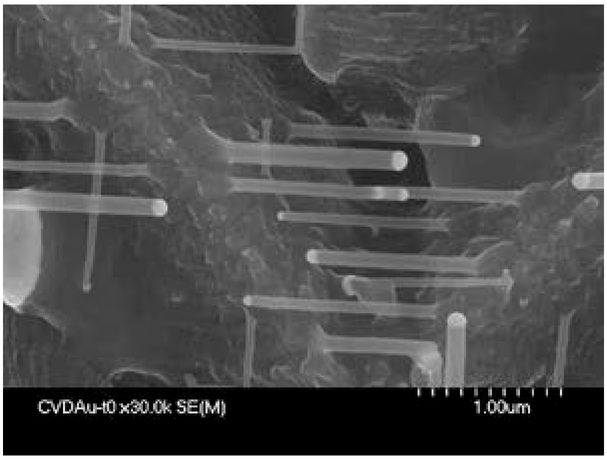


Fig. 15.3. CVD growth of Si nanowires using Au particle catalysts.

The VLS process takes place as follows. A thin (~1-10 nm) Au film is deposited onto a silicon (Si) wafer substrate by sputter deposition or thermal evaporation. The wafer is annealed at temperatures higher than the Au-Si eutectic point, creating Au-Si alloy droplets on the wafer surface (the thicker the Au film, the larger the droplets). Mixing Au with Si greatly reduces the melting temperature of the alloy as compared to the alloy constituents. The melting temperature of the Au:Si alloy reaches a minimum (~363 °C) when the ratio of its constituents is 4:1 Au:Si, also known as the Au:Si eutectic point. Lithography techniques can also be used to controllably manipulate the diameter and position of the droplets (and the resultant nanowires). One-dimensional crystalline nanowires are then grown by a liquid metal-alloy droplet-catalyzed chemical or physical vapor deposition process, which takes place in a vacuum deposition system Au-Si droplets on the surface of the substrate act to lower the activation energy of normal vapor-solid growth. For example, Si can be deposited by means of a $SiCl_4$:H_2 gaseous mixture reaction (CVD), only at temperatures above 800 °C, in normal vapor-solid growth. Moreover, below this temperature almost no Si is deposited on the growth surface. However, Au particles can form Au-Si eutectic droplets at temperatures above 363 °C and adsorb Si from the vapor state (because Au can form a solid-solution with all Si concentrations up to 100%) until reaching a supersaturated state of Si in Au. Furthermore nanosized Au-Si droplets have much lower melting points because the surface area-to-volume ratio is increasing, becoming energetically unfavorable, and nanometer-sized particles act to minimize their surface energy by forming droplets (spheres or half-spheres). Si has a much higher melting point (~1414 °C) than that of the eutectic alloy, therefore Si atoms precipitate out of the supersaturated liquid-alloy droplet at the liquid- alloy/solid- Si interface and the droplet rises from the surface. This process is illustrated in Fig. 15.4 and Fig. 15.5.

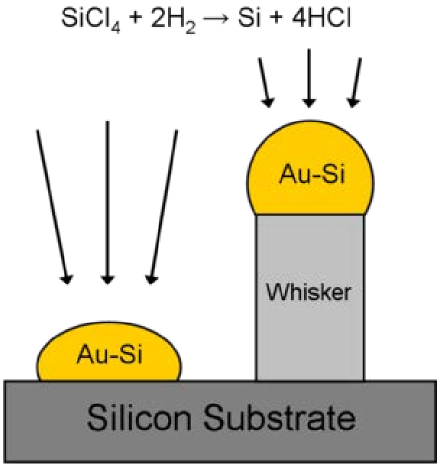


Fig. 15.4. Schematic illustration of Si whisker growth from the reaction SiCl4 and H2 vapor phase. This reaction is catalyzed by gold-silicon droplet deposited on the wafer surface prior to whisker growth.

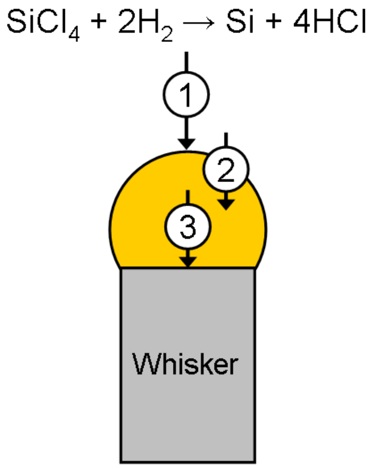


Fig. 15.5. Schematic illustration of metal-alloy catalyzed whisker growth depicting the pathway of source materials through the droplet to the growing interface.

Typical features of the VLS method are (*i*) greatly lowered reaction energy compared to normal vapor-solid growth, (*ii*) wires grow in the areas activated by the metal catalysts and the size and position of the wires are determined by that of the metal catalysts, and (*iii*) this growth mechanism can also produce highly anisotropic nanowire arrays from a variety of materials.

The catalyst must form a liquid solution with the crystalline material to be grown at the nanowire growth temperature. The solid solubility of the catalyzing agent is low in the solid and liquid phases of the substrate material. The equilibrium vapor pressure of the catalyst over the liquid alloy must be small so that the droplet does not vaporize, shrink in volume (and therefore radius), and decrease the radius of the growing wire until, ultimately, growth is terminated. The catalyst must be inert (non-reacting) to the reaction products (during CVD nanowire growth). The vapor-solid, vapor-liquid, and liquid-solid interfacial energies play a key role in the shape of the droplets and therefore must be examined before choosing a suitable catalyst; small contact angles between the droplet and solid are more suitable for large area growth, while large contact angles result in the formation of smaller (decreased radius) whiskers. The solid-liquid interlace must be well-defined crystallographicalfy in order to produce highly directional growth of nanowires. It is also important to point out that the solid-liquid interface cannot, however be completely smooth. Furthermore, if the solid liquid interface was atomically smooth, atoms near the interface trying to attach to the solid would have no place to attach to until a new island nucleates (atoms attach at step ledges), leading to an extremely slow growth process. Therefore, "rough" solid surfaces or surfaces containing a large number of surface atomic steps (ideally 1 atom wide, for large growth rates) are needed for deposited atoms to attach and nanowire growth to proceed [15].

The materials system used, as well as the cleanliness of the vacuum system and therefore the amount of contamination and/or the presence of oxide layers at the droplet and wafer surface during the experiment both greatly influence the absolute magnitude of the forces present at the droplet/surface interface and in turn determine the shape of the droplets. The shape of the droplet, i.e. the contact angle (β_0, see Fig. 15.6) can, be modeled mathematically, however, the actual forces present during growth are extremely difficult to measure experimentally. Nevertheless, the shape of a catalyst particle at the surface of a crystalline substrate is determmed by a balance of the forces of surface tension and the liquid-solid interface tension. The radius of the droplet varies with the contact angle as:

$$R = \frac{r_0}{\sin(\beta_0)},$$

where r_0 is the radius of the contact area and β_0 is defined by a modified Young's equation:

$$\sigma_1 \cos(\beta_0) = \sigma_s - \sigma_{ls} - \frac{\tau}{r_0},$$

It is dependent on the surface (σ_s) and liquid-solid interface (σ_{ls}) tensions, as well as an additional line tension (τ) which comes into effect when the initial radius of the droplet is small (nanosized). As a nanowire begins to grow, its height increases by an amount dh and the radius of the contact area decreases by an amount dr (see Fig. 15.6). As the growth continues, the inclination angle at the base of the nanowires (α, set as zero before whisker growth) increases, as does β_0,:

$$\sigma_1 \cos(\beta_0) = \sigma_s \cos(\alpha) - \sigma_{ls} - \frac{\tau}{r_0}.$$

The line tension therefore greatly influences the catalyst contact area. The most import result from this conclusion is that different line tensions will result in different growth modes. If the line tensions are too large, nanohillock growth will result and thus stop the growth.

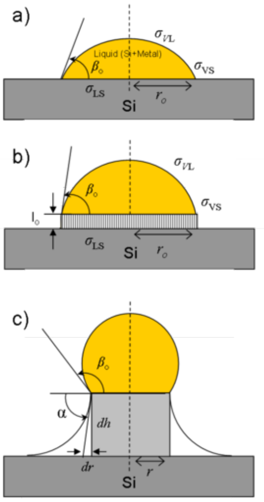


Fig. 15.6. Schematic illustration of metal-alloy catalyzed whisker growth depicting the catalyst droplet formation during the early stages of whisker growth.

The diameter of the nanowire which is grown depends upon the properties of the alloy droplet. The growth of nano-sized wires requires nano-size droplets to be prepared on the substrate. In an equilibrium situation this is not possible as the minimum radius of a metal droplet is given by [16]

$$R_{\min} = \frac{2V_l}{RT\ln(s)}\sigma_{lv},$$

where V_l is the molar volume of the droplet, σ_{lv} the liquid-vapor surface energy, and s is the degree of supersaturation [17] of the vapor. This equations restricts the minimum diameter of the droplet, and of any crystals which can be grown from it, under typically conditions to well above the nanometer level. Several techniques to generate smaller droplets have been developed, including the use of monodispersed nanoparticles spread in low dilution on the substrate, and the laser ablation of a substrate-catalyst mixture so to form a plasma which allows well-separated nanoclusters of the catalyst to form as the systems cools [18].

During VLS whisker growth, the rate at which wickers grow is dependent on the whicker diameter: the larger the whisker diameter, the faster the nanowire grows axially. This is because the supersaturation of the metal- alloy catalyst ($\Delta\mu$) is the main driving force for nanowhisker growth and decreases with decreasing whisker diameter (also known as the Gibbs-Thompson effect):

$$\Delta\mu = \Delta\mu_0 - \frac{4\alpha\Omega}{d}.$$

Again $\Delta\mu$ is the main driving force for nanowhisker growth (the supersaturation of the metal droplet). More specifically, $\Delta\mu_0$ is the difference between the chemical potential of the depositing species (Si in the above example) in the vapor and solid whisker phase. $\Delta\mu$ is the initial difference proceeding whisker growth (when $d \to \infty$), while Ω is the atomic volume of Si and α the specific free energy of the wire surface. Examination of the above equation, indeed reveals that small diameters (<100 nm) exhibit small driving forces for whisker growth while large wire diameters exhibit large driving forces.

Laser-assisted growth

Laser-assisted growth involves the removal of material from metal-containing solid targets by irradiating the surface with high-powered (~100 mJ/pulse) short (10 Hz) laser pulses, usually with wavelengths in the ultraviolet (UV) region of the light spectrum (Fig. 15.7, Fig. 15.8). When such a laser pulse is adsorbed by a solid target, material from the surface region of the target absorbs the laser energy and either (a) evaporates or sublimates from the surface or is (b) converted into a plasma. These particles are easily transferred to the substrate where they can nucleate and grow into nanowires. The laser-assisted growth technique is particularly useful for growing nanowires with high melting temperatures, multiconponent or doped nanowires, as well as nanowires with extremely high crystalline quality. The high intensity of the laser pulse incident at the target allows the deposition of high melting point materials, without having to try to evaporate the material using extremely high temperature resistive or electron bombardment heating. Furthermore, targets can simply be made from a mixture of materials or even a liquid. Finally, the plasma formed during the laser absorption process allows for the deposition of charged particles as well as a catalytic means to lower the activation barrier of reactions between target constituents.

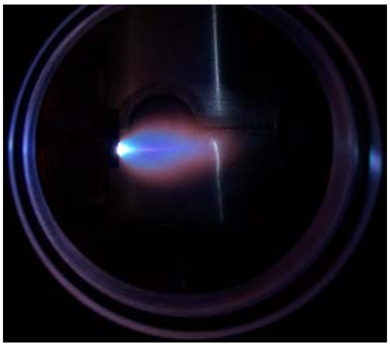


Fig. 15.7. A plasma plume ejected from target during pulse laser deposition.

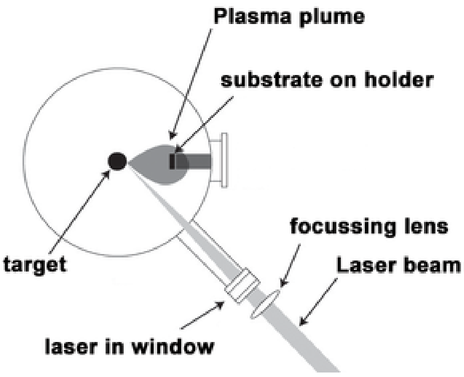


Fig. 15.8. One possible configuration of a PLD deposition chamber.

The vapor, which can be generated by laser ablation [19] or thermal decomposition of organometallic compounds [20], determines the composition of the wires. In a typical setup, target materials (e.g., InP, CdS, and Si) and silicon substrates covered with gold nanoparticles (serving as catalysts) lie at the upstream end and central part of a tube furnace, respectively. The substrate is heated to high temperature of 600–1200 °C while the temperature of target material remains room temperature. A pulsed laser ablates the target to generate a vapor, which transports, through a flow of argon, to the surface of the substrate. This vapor dissolves in the droplets of catalyst until the catalyst is supersaturated with material from the target. At this point, the vapor condenses into nanowires through the VLS process [21]. This method has, in principle, the ability to synthesize nanowires of most semiconductor materials that can be evaporated by tuning the parameters (e.g., power, wavelength and pulse width) of the laser. The size of the catalyst droplets controls the diameter of nanowires. The growth time and ablation conditions control the length of the nanowires [22].

Solution-Liquid-Solid Process

By combing nanoparticles of metals (e.g., In, Ga, or Bi) with low melting points as catalysts with thermal decomposition of organometallic precursors, an analog of VLS process can be extended to liquid phase. This process, so-called solution–liquid–solid process, or SLS, has been used to synthesize nanowires made of III–V compounds [23, 24]. For example, nanowires of GaAs and InP with small diameters (e.g., 3–17 nm) and relatively narrow diameter distributions were synthesized by refluxing 1,3-diisopropylbenzene solutions of corresponding precursors (i.e., $(tBu)_3Ga$ and $As(SiMe_3)_3$ for GaAs, and $[Me_2InP(SiMe_3)_2]_2$ for InP) in the presence of In nanoparticles. The SLS process is mainly limited by the availability of suitable organometallic precursors.

The nanowires synthesized via ‘‘bottom-up’’ approaches always have broad distributions in both lateral and longitudinal dimensions. Furthermore, lengths greater than 100 μm are difficult to achieve. Their surface properties, compositional purity, doping uniformity, and concentration, which all strongly influence their electrical properties, are less well defined than those of the single-crystalline wafers used in the semiconductor industry.

15.1.2. ‘‘Top-Down’’ Approaches

The rapid miniaturization of electronics to the submicron scale has led to remarkable advances in computing power while at the same time reducing costs. These advances have been made possible by many scientific and technological innovations associated with ‘‘top-down’’ manufacturing, in which small features are patterned in bulk semiconductor materials by lithography, deposition, and other processing to form functional devices. This remarkable trend in miniaturization is now universally referred to as Moore’s Law, which dictates that the number of transistors per chip doubles every 18–24 months [25, 26]. The magnitude of what has been accomplished can be recognized by the fact that it is now possible to fabricate a 100 million transistors on a chip of the same size as a single transistor when it was first invented a half century ago. However, as the microelectronic industry advances towards ever smaller devices, it is believed that physical and economic limits of current top-down silicon technology will be soon reached. First, photolithography- based top-down strategies will soon reach a fundamental resolution limit of 70 nm that will limit further reduction in feature sizes using relatively conventional optical sources, although solutions such as extreme ultraviolet lithography are being explored to overcome this limit. Second, the exponentially increasing cost to construct each new generation of fabrication line may limit further miniaturization on the basis of economics alone. These and other limitations of current top-down technology have motivated efforts worldwide to search for new strategies to meet the expected demand for increased computational power as well as for integrating low-cost and flexible computing in unconventional environments in the future [27].

Microwires and nanowires can be fabricated from bulk wafers via ‘‘top-down’’ approaches. These types of nanowires have well-controlled doping levels, dimensions and crystallinity, which, in turn, give them excellent electrical properties for applications in highperformance electronics. The critical step involved in ‘‘top-down’’ approaches is to generate patterned mask lines on the nanometer scale, using simple lithographic techniques such as phase-shift photolithography with soft-phase masks made of poly(dimethylsiloxane) (PDMS) or nanoimprinting [28–30]. For example, Si nanowires have been fabricated from multi-layered wafers (i.e., silicon-on-insulator, or SOI wafers) through the near-field optical lithography [31]. In this case, exposing photoresist on an SOI wafer through an elastomeric phase mask, which forms a conformal contact with the layer of photoresist, generates patterned lines with widths ranging from 50 to 300 nm. The features patterned in the photoresist film can be transferred onto the underlying wafer using reactive ion etching or wet-etching of the top Si layer. After the buried SiO_2 layer is selectively dissolved with HF solution, Si nanowires are released from the mother wafer. The resulting wires can be further thinned through oxidation and selective etching of the outer SiO_2 layer. For many applications, this approach is unsuitable due to the relatively high cost of these kinds of multilayered wafers (e.g., ~$500 for each 6-in. SOI wafer versus ~$5 for each 6-in. Si wafer). Developing convenient methods for generating nanowires from inexpensive wafers is therefore crucial to commercialize electronics.

Anisotropic chemical etching of single-crystalline wafers provides a route to fabricate nanowires from inexpensive wafers with uniform compositions. For example, anisotropic etching of a (110) Si wafer with patterned mask lines in a hot aqueous KOH solution can generate trenches with vertical profiles (see Fig. 15.9A, step *i*) [32]. The top surface and partial side surfaces of each wall can be covered with a thin film of inert materials (e.g., Ti/Au and Pd), which can withstand the etchant of Si, through shadow evaporation (step *iii*) [33]. Isotropic etching in $HF–HNO_3–H_2O$ can cut the walls at the edges of mask films on side surfaces to release free-standing wires (step *iv*). Fig. 15.9B gives a scanning electron microscopy (SEM) image of the sample before Si wires are released. The width of the top surface of each wall is 1.5 μm; this dimension can easily be reduced to the nanometer scale by using phase-shift photolithography [34] or nanoimprinting [35]. Fig. 15.9C shows an SEM image of the vertical walls with width of 100 nm fabricated by phase-shift photolithography. The height of each wire can be easily tuned by controlling the incident angle (α) of the evaporated material. Many other related approaches can generate similar structures. Anisotropic etching of semiconductor wafer that leads to the formation of reverse mesas (i.e., structures with newly formed sidewalls with an acute angle relative to the original surface of wafer) can be exploited to generate nanowires using only a single etching step. Compound semiconductor materials made of group III and V elements are usually crystallized in the form of zinc blende face-centered cubic (f.c.c.) lattice, whose {111} planes are the most stable [36]. The {111} surface terminated by the group III atoms, also called {111}A, is deficient in electron density compared with the {111} surface terminated by the group V atoms, or {111}B. This difference makes the {111}A surface resistant to chemical oxidation and generally difficult to etch. As a result, sidewalls with an acute angle (rather than obtuse angle) relative to top surface can be generated by controlling the orientation of stripes of etching mask. As shown in Fig. 15.10A, anisotropic etching of (100) wafer patterned with parallel mask lines along $(01^{-}1^{-})$ crystallographic direction in appropriate etchant (e.g., H_2O_2–H_3PO_4–H_2O for GaAs and Br_2-methanol for InP) generates reverse mesas with side surfaces terminated by $(1^{-}1^{-}1)$A and $(1^{-}11^{-})$A planes, respectively (step *i*) [37, 38]. With long etching times, the two side walls of each reverse mesa connect to release a wire with triangular cross section (step *ii*). Pure wires with single crystallinity, uniform doping concentrations, and well-defined crystalline surfaces are collected after removal of the etching mask (step *iii*).

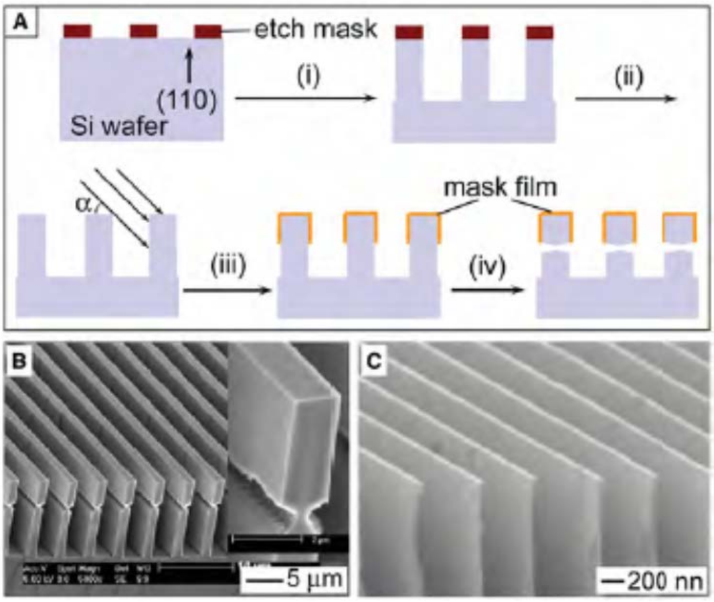


Fig. 15.9. Formation of Si wires from a (110) Si wafer. (A) Steps of the fabrication process: (*i*) anisotropic chemical etching of Si in hot KOH solution using nitride or oxide mask stripes; (*ii*) removal of mask stripes; (*iii*) shadow deposition of new mask material on the top and side surfaces of each vertical wall; (*iv*) isotropic etching to lift off Si wires. (B) SEM image of Si walls with width of ~1.5 μm before lift off. (C) SEM image of Si walls with top surface width of ~100 nm.

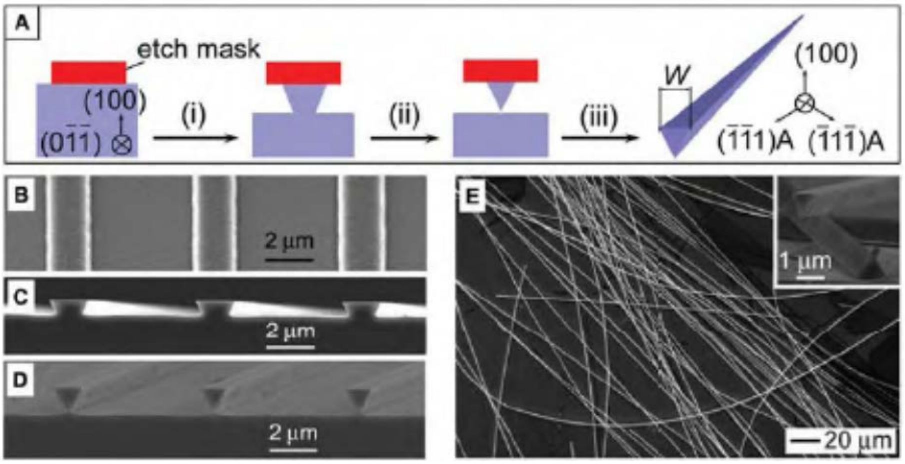


Fig. 15.10. Fabrication of wires by anisotropic chemical etching of a III–V semiconductor wafer with top surface of (100) plane. (A) Schematic illustration of the steps: ((*i*) anisotropic etching using mask stripes oriented along $(01^{-}1^{-})$ direction; (*ii*) continuous etching and release of wire; (*iii*) removal of etch mask and generation of free-standing, clean wire. (B–E) SEM images of the samples formed by etching a GaAs wafer patterned with photoresist (PR) stripes in an aqueous solution consisting of $1H_3PO_4$ (85 wt.%) – $13H_2O_2$ (30 wt.%) – $12H_2O$ (v/v/v) at different etching times. (B) PR stripes on GaAs wafer before etching. (C and D) Cross section of reverse mesas. (E) A random assembly of clean GaAs wires. Etching processes were performed in ice-water bath and ambient environment.

Panels B–E of Fig. 15.10 show a series of SEM images of samples obtained at different etching times. Fig. 15.10B presents an SEM image of 2 μm wide photoresist (Shipley 1805) lines on a (100) GaAs wafer, clearly showing the straight and uniform width along the longitudinal axis. The SEM images shown in Panels C and D display reverse mesas formed at short and long etching times. Comparison of the lateral dimension between reverse mesas and photoresist lines reveals lateral undercutting along with the vertical etching. Fig. 15.10E shows an SEM image of free-standing GaAs wires randomly assembled on a GaAs substrate. The wires exhibit uniform widths (*W*) of 840 nm along their longitudinal axis. The SEM image shown in the inset displays the triangular cross section of each wire, the roughness of the side walls, and the flatness of the top surface. The roughness of the side walls originates mainly from the edge roughness of the mask lines. The formation of curved structures indicates that the wires are flexible. The bendability increases with decreasing lateral dimensions. Wires with widths of ~400 nm can form arc shapes with radii as small as ~20 μm [37]. The excellent flexibility of the nanowires allows them to serve as active components in flexible macroelectronic devices.

It is worth noting that the etching generates undercutting and maintains high anisotropy after the wires lift off the substrate [37]. As a result, the lateral dimensions of the wires can be easily decreased to tens of nanometers by controlling the etching time even when the mask lines have widths in the micrometer scale. Fig. 15.11A–D shows SEM images of individual wires obtained by etching a GaAs wafer patterned with 2 μm wide mask lines. The wires become thinner with increasing etching time, but the triangular cross sections of the wires are preserved. The dependence of wire width (*W*) on etching time is plotted in Fig. 15.11E, which indicates that GaAs wires with widths down to 50 nm can be obtained in this fashion. The minimum width is determined by the roughness on the side walls. The lengths of the GaAs wires can also be easily tuned by controlling the length of mask stripes. Figure 15.11F and G shows SEM images of GaAs rods with average lengths of 6.4 and 15.5 μm, respectively. In principle, wires with lengths as long as tens of centimeters, i.e., the diameter of original wafer can be fabricated. Figure 15.12 gives a composition of several SEM images of long GaAs nanowires with lengths of 1 cm, which indicates that most wires are continuous in the observed regions.

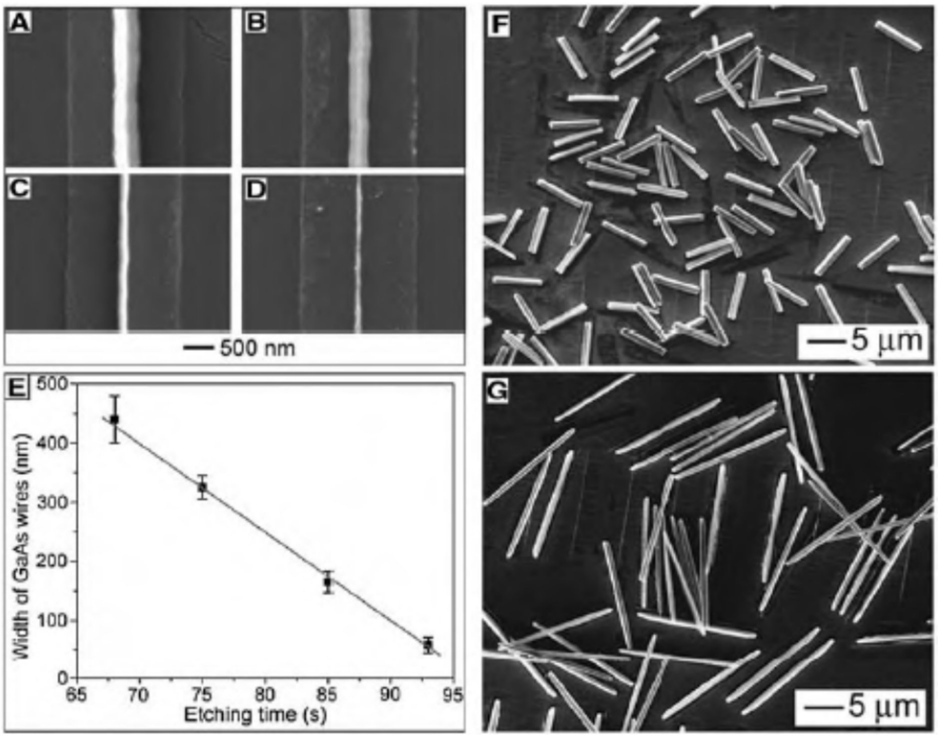


Fig. 15.11. (A–D) SEM images of GaAs wires that were generated by etching a GaAs wafer patterned with 2 μm wide SiO_2 stripes at different times: (A) 68, (B) 75, (C) 85, and (D) 93 sec. (E)

Plot of the dependence of the average width of GaAs wires with etching time. (F and G) SEM images of GaAs rods obtained by etching GaAs wafers patterned with PR stripes of different lengths: (F) 10 and (G) 20 μm. Etching conditions were similar to that of Fig. 15.10.

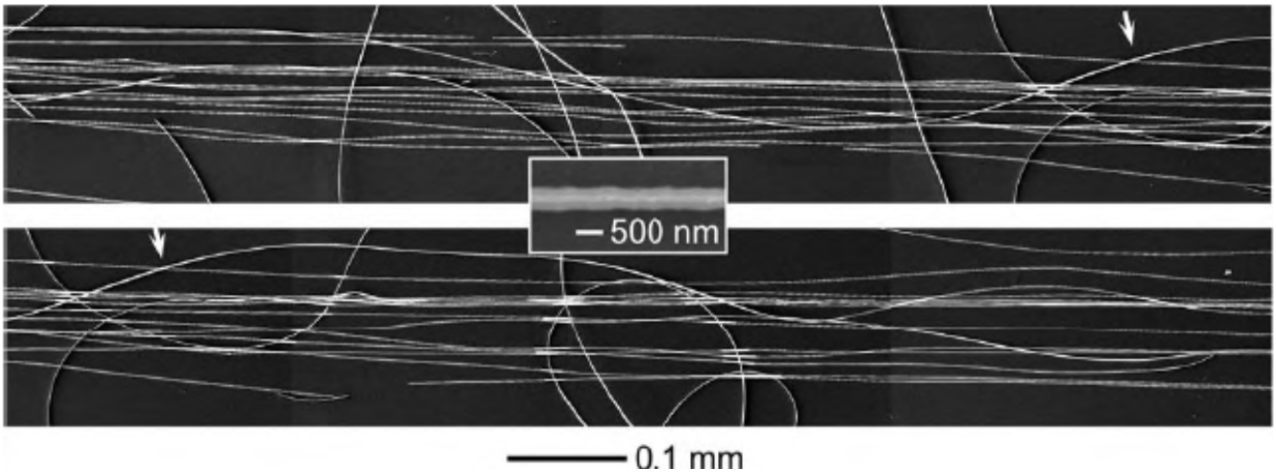


Fig. 15.12. SEM images of long GaAs wires. The arrows indicate the same position of the sample.

The physical parameters of nanowires fabricated via ‘‘top-down’’ approaches exhibit good uniformity (comparable to the high-quality wafer precursors) and enable easy control over the processing. The cost of the wires can be greatly decreased if the resultant mother wafers are polished and reused to generate more wires [37]. A disadvantage of the ‘‘top-down’’ approach is that the composition of the wires is limited to materials (e.g., Si, Ge, GaAs, InP, etc.) that are already available in wafer or thin film forms [37–40].

15.1.3. Nanoscale Building Blocks

Individual molecules [41-45] and quantum dots [46-48], which can be classified as zero-dimensional (0D) structures, have been proposed as building blocks for bottom-up assembly of nanoscale electronics. These 0D structures have been intensively pursued over the past decade since they represent the smallest building blocks with corresponding high potential for massive integration. However, the use of individual molecules or quantum dots in nanoelectronics has been limited by challenges in establishing reliable electrical contacts needed to study their fundamental properties and interconnect them. It has thus been difficult to elucidate and understand the intrinsic properties of individual devices, and moreover, to develop and demonstrate realistic schemes for scalable interconnection and integration of 0D devices into functional architectures.

One-dimensional nanostructures have also been the focus of extensive studies worldwide due to their unique physical properties and potential to revolutionize broad areas of nanotechnology. First, 1D nanostructures represent the smallest dimension structure that can efficiently transport electrical carriers, and thus are ideally suited to the critical and ubiquitous task of moving and routing charges (information) in nanoscale electronics and optoelectronics. Second, 1D nanostructures can also exhibit device function, and thus can be exploited as both the wiring and device elements in architectures for functional nanosystems [6, 49]. In this regard, two material classes, carbon nanotubes (NTs) [49–60] and semiconductor NWs [61–68], have shown particular promise.

Single-walled carbon NTs can exhibit either metallic or semiconducting behavior depending on diameter and helicity [50]. The unique electronic properties of NTs open up the possibility of creating a number of different devices that could have potential in nanoelectronics [49, 51-53]. For example, single-walled NTs have been used to fabricate room-temperature field effect transistors (FETs) [54, 55], diodes [56, 57] and recently, logic circuits [58, 59]. However, the inability to control whether NT building blocks are semiconducting or metallic makes specific device fabrication largely a random event. Hence, moving beyond proof-of-concept single device elements to the integrated arrays required for nanoelectronics poses a serious issue for NT-based approaches. A creative solution to the problem of coexisting metallic and semiconducting NTs involves selective destruction of metallic tubes [60], although such an approach requires extensive top-down lithography and subsequent processing to implement and may not be practical for highly integrated nanoelectronics systems.

Semiconductor NWs [6, 61, 62] represent another important type of nanometer scale wire structure. In contrast to NTs, however, semiconductor NWs can be rationally and predictably synthesized in single crystal form with all key parameters controlled, including chemical composition, diameter and length, and doping/electronic properties [63-65]. Semiconductor NWs thus represent one of best-defined and controlled class of nanoscale building blocks, which correspondingly have enabled a wide range of devices and integration strategies to be pursued. For example, semiconductor NWs have been assembled into nanometer scale FETs [65, 66], *p*–*n* diodes [66, 67], light emitting diodes (LEDs) [66], bipolar junction transistors [67], complementary inverters [67], complex logic gates, and even computational circuits that have been used to carry out basic digital calculations [68]. In contrast to NTs, NW devices can be assembled in a rational and predictable manner because the size, interfacial properties, and electronic properties of the NWs can be precisely controlled during synthesis, and moreover, reliable methods exist for their parallel assembly [69]. In addition, it is possible to combine distinct NW building blocks in ways not possible in conventional electronics and to leverage the knowledge base that exists for the chemical modification of inorganic surfaces [70, 71] to produce semiconductor NW devices that achieve new function and correspondingly could lead to unexpected device and system concepts.

15.2. GENERATION OF ARRAYS OF ALIGNED NANOWIRES

A key motivation underlying research on nanoscale devices is the potential to achieve integration at a level not possible in conventional microelectronics. To achieve this goal in future nanosystems will require the development and implementation of efficient and scalable strategies for assembly of nanoscale building blocks into increasingly complex architectures. First, methods are needed to assemble NWs into highly integrated arrays with controlled orientation and spatial position. Second, approaches must be devised to assemble NWs on multiple length scales and to make interconnects between nano-, micro-, and macroscopic worlds. To address these critical next levels of organization, the significant efforts have been focused on developing complementary strategies for hierarchical assembly of NWs on surfaces.

Hierarchical assembly of nanowires into well-defined architectures is the cornerstone for fabrication of functional devices. In particular, monolayers of aligned nanowires with high density and uniform end-to-end registry are critical to build high-performance thin field transistors (TFTs) for macroelectronics. A number of approaches have been demonstrated to organize nanowires dispersed in solutions into parallel arrays by applying various external forces (e.g., microfluidic shear force, electric field, surface compression, and magnetic field) to overcome Brownian motion [72-77]. In the VLS process, arrays of aligned nanowires can be in situ grown from metal nanoparticles (i.e., catalysts) patterned on substrates with specific crystallographic planes [78-83]. Nanowires fabricated using ‘‘top-down’’ approaches can be designed to maintain the order defined by lithographic process. The resultant wire arrays can be transfer printed onto any desired substrates including plastic sheets [37-40].

Nanowires synthesized through ‘‘bottom-up’’ approaches are often dispersed in solvents and then deposited onto substrates for further applications. Generating organized arrays, which have relatively low entropy, requires external forces to manipulate the nanowires.

15.2.1. Electrical Field Directed Assembly

Applied electric fields (*E*-fields) can be used effectively to attract and align NWs due to their highly anisotropic structures and large polarizabilities (Fig. 15.13) [66]. The interaction between induced wire polarization and electric field aligns the wires parallel to the field direction (Fig. 15.14B) [73, 74]. This underlying idea of *E*-field directed assembly can be readily seen in images of NW solutions aligned between parallel electrodes (Fig. 15.13B), which demonstrate that virtually all of the NWs aligned in parallel along the *E*-field direction. Electric field directed assembly can also be used to position individual NWs at specific positions with controlled directionality. For example, *E*-field assembly of NWs between an array of electrodes (Fig. 15.13C) clearly shows that individual NWs can be positioned to bridge pairs of diametrically opposed electrodes and form a parallel array. In addition, by changing the *E*-field direction with sequential NW solutions, the alignment can be carried out in a layer-by-layer fashion to produce crossed NW junctions (Fig. 15.13D). These results demonstrate clearly that *E*-field directed assembly can be used to align and position individual NWs into parallel and crossed arrays, which correspond to two basic geometries for integration, and thus provide one robust approach for rational and parallel assembly of nanoscale device arrays.

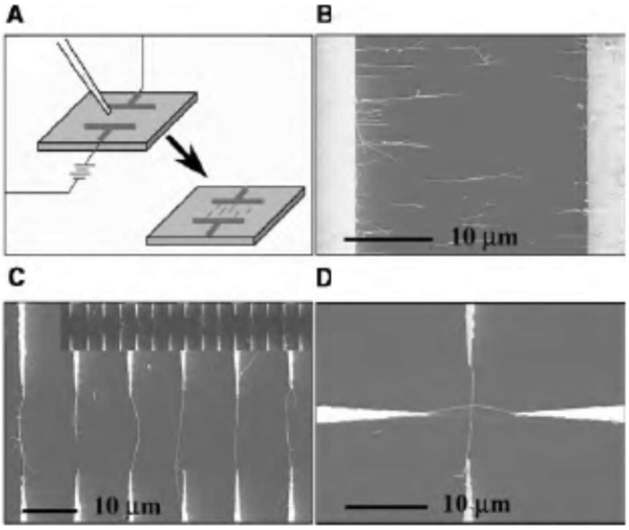


Fig. 15.13. E-field directed assembly of NWs. (A) Schematic view of E-field alignment. (B) Parallel array of NWs aligned between two parallel electrodes. (C) Spatially positioned parallel array of NWs obtained following E-field assembly. The top inset shows 15 pairs of parallel electrodes with individual NWs bridging each diametrically opposed electrode pair. (D) Crossed NW junction obtained using layer-by-layer alignment with the E-field applied in orthogonal directions in the two assembly steps [66].

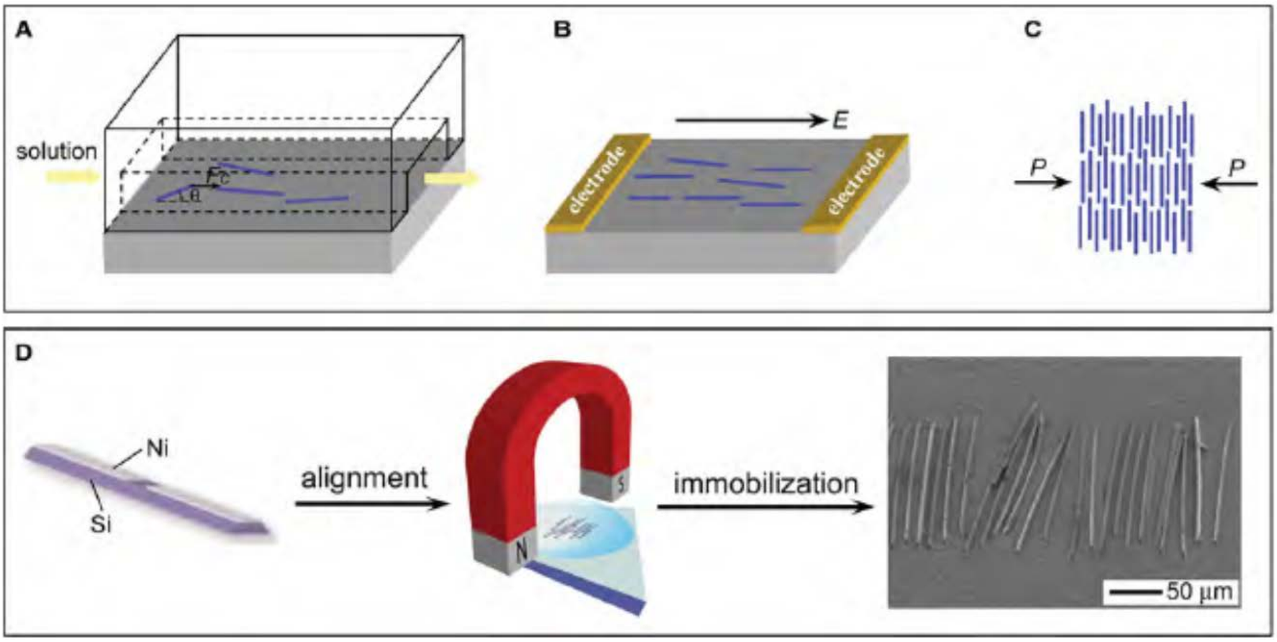


Fig. 15.14. Strategies for assembly of as-synthesized nanowires dispersed in solvents through different external forces: (A) shear flow force formed in microfludic channels; (B) electric field; (C) surface pressure generated through the Langmuir–Blodgett technique; and (D) magnetic field. The wire array demonstrated in (D) consists of Si wires with Ni stripes sitting on the surface of a thin layer of cured polyurethane spin cast on a glass slide.

Electric field directed assembly, which represents the first approach described for assembly of 1D nanostructures, also has limitations, including 1) the need for substantial conventional lithography to pattern microelectrode arrays used to produce aligning fields, 2) the deleterious effect of fringing electric fields at the submicron length scales, 3) it is difficult with this approach to generate uniform films of aligned nanowires with high-density coverage over large areas.

15.2.2. Fluid Flow Directed Assembly

Microchannel fluidic device

To achieve a greater flexibility in rational, parallel assembly of 1D nanostructures into nanosystems, it has been developed a powerful new approach called fluidic flow directed assembly [69]. In this case, a shear flow formed near the surface of substrate provides a force (F_c) to orient the wires parallel to the flow direction before they are immobilized on the substrate (see Fig. 15.14A). The degree of alignment, which is represented by the angle (θ) of nanowires with respect to the flow direction, can be controlled by the flow rate. Meanwhile, the spacing between individual aligned nanowires can be narrowed by increasing the time for deposition. In this method, NWs (or

NTs) can be easily aligned by passing a suspension of NWs through microfluidic channel structures, for example, formed between a poly(dimethylsiloxane) (PDMS) mold [84] and a flat substrate (Fig. 15.15). Parallel and crossed NW arrays can be readily created using single (Fig. 15.15A) and sequential crossed (Fig. 15.15B) flows, respectively, for the assembly process. A disadvantage of this approach is that it is difficult to produce uniform coverage (i.e., the number of tubes per unit area tends to be largest near the place where the nanowire suspension is introduced into the channel).

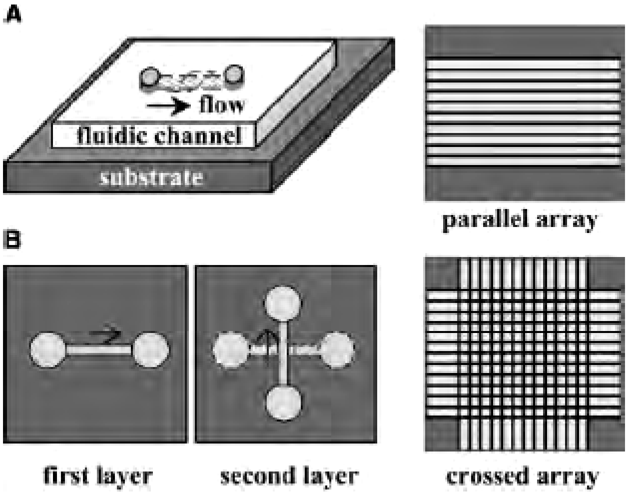


Fig. 15.15. Fluid flow directed assembly NWs. (A) A channel is formed when a trench structure is brought in contact with a flat substrate. Nanowire assembly is carried out by flowing a NW suspension through the channel at a controlled rate and for a set duration. Parallel arrays of NWs are observed in the flow direction on the substrate when the trench structure is removed. (B) Crossed NW arrays can be obtained by changing the flow direction sequentially in a layer-by-layer assembly process [69].

Parallel assembly of nanowire

Images of NWs assembled on substrate surfaces (Fig. 15.16A) within microfluidic flows demonstrate that virtually all of the NWs are aligned along the flow direction. This alignment readily extends over hundreds of micrometers (Fig. 15.16B). Indeed, alignment of the NWs has been found to extend up to millimeter length scales, and is limited only by the size of the fluidic channels used. The alignment of NWs within the channel flow has been explained within the framework of shear flow [85, 86]. Specifically, the channel flow near the substrate surface resembles a shear flow and aligns the NWs in the flow direction before they are immobilized on the substrate. This idea readily provides the intellectual underpinning needed for controlling the degree of alignment and average separation of the NWs. First, higher flow rates produce larger shear forces and will lead to better alignment, and thus the flow rate can be used to control the degree of alignment. Indeed, the width of the NW angular distribution with respect to the flow direction significantly narrows with increasing flow rate (Fig. 15.16C). Studies of the distribution widths measured over a range of flow conditions showed that the width decreased quickly as flow increased to a nearly constant value at ~10mm/sec (where more than 80% of the NWs are aligned within ±5° of the flow direction (Fig. 15.16C, inset).

Second, these studies [69] demonstrated that the average NW density could be controlled by the flow duration (Fig. 15.16D). Experiments carried out at constant flow rate show that the NW density increases systematically with flow duration, and can, upon extended deposition time, produce NW arrays with spacings of the order of 100 nm or less. The relationship between average separation and flow duration also depends on and can be further controlled by the chemical functionality on the NW and substrate surfaces; that is, strong complementary interactions will facilitate deposition from the flow.

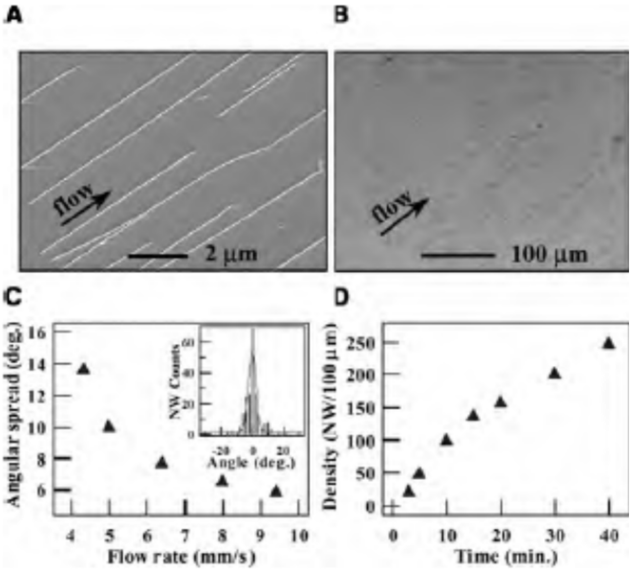


Fig. 15.16. Parallel assembly of NW arrays. (A) SEM image of a parallel array of InP NWs aligned by flow. (B) Optical microscope image of a parallel array of InP NWs aligned over very large area. (C) NW angular spread with respect to the flow direction vs. flow rate. The inset shows histogram of NW angular distribution at a flow rate of 9.40mm=sec. (D) Average density of NWs vs. flow time [69].

Crossed nanowire arrays

The fluidic flow assembly approach can be used to organize NWs into more complex crossed structures, which are critical for building dense nanodevice arrays, using a layer-by-layer deposition process (Fig. 15.15B). The formation of crossed and more complex structures requires that the nanostructure–substrate interaction is sufficiently strong that sequential flow steps do not affect preceding ones: it was find that this condition is readily achieved. For example, alternating the flow in orthogonal directions in a two-step assembly process yields crossbar structures in high yield (Fig. 15.17A,B). These data demonstrate that crossbars extending over 100s of microns on a substrate with only 100s of nanometers separation between individual cross points are obtained through a very straightforward, parallel low cost, and fast process. Fluidic flow directed assembly of multiple crossed NW arrays offers significant advantages over previous efforts. First, it is intrinsically very parallel and scalable with the alignment readily extending over very large length scales. Second, this approach is general for virtually any elongated nanostructure including carbon NTs and DNA molecules. Third, it allows for the directed assembly of geometrically complex structures by simply controlling the angles between flow directions in sequential assembly steps. For example, equilateral triangles (Fig. 15.17C) were easily assembled in a three-layer deposition sequence using 60° angles between the three flow directions. The method of flow alignment thus provides a flexible way to meet the requirements of many device configurations in the future. An important feature of this layer-by-layer assembly scheme is that each NW layer can be independent of the preceding one(s), and thus a variety of homo- and hetero-junction configurations can be obtained at each crossed point by simply changing the composition of the NW suspension used for each flow step. For example, it should be possible to assemble directly and subsequently address individual nanoscale devices using our approach with *n*-type and *p*-type NWs, in which the NWs act as both the wiring and active device elements (see below).

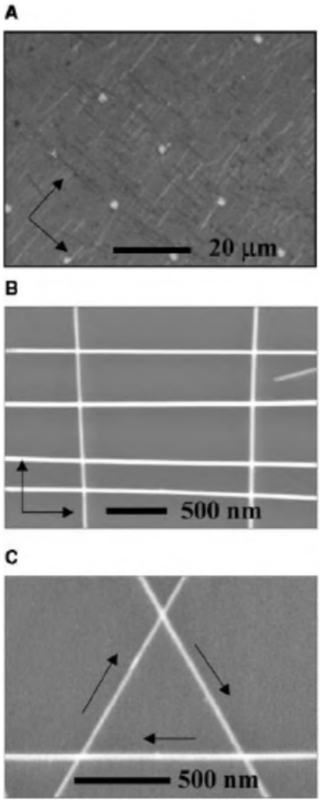


Fig. 15.17. Layer-by-layer assembly of crossed NW arrays. (A, B) Typical optical microscope and SEM images of crossed arrays of InP NWs obtained in a two-step assembly process with orthogonal flow directions for the sequential steps. Arrows indicate the two flow directions. (C) An equilateral triangle of GaP NWs obtained in three-step assembly process, with flow directions highlighted by arrows [69].

Control of periodicity

The above results demonstrate clearly the power of the fluidic assembly approach, although to enable systems organization with greatest control requires in many cases that the spatial position also be defined. To realize this additional constraint on the assembly process, it has been explored complementary chemical interactions between chemically patterned substrates and NWs (Fig. 15.18A). Substrates for alignment are first patterned with two different functional groups, with one of the functional groups designed to have a strong attractive interaction with the NW surface, and then, following flow alignment, regular, parallel NW arrays with lateral periods the same as those of the surface patterns are produced (Fig. 15.18D). These data demonstrate that the NWs are preferentially assembled at positions defined by the chemical pattern, and moreover, show that the periodic patterns can organize the NWs into regular superstructures. In addition, periodic crossed NW arrays can also be envisioned using a substrate with a crossed pattern of chemical functionality.

It is important to recognize that the patterned surface alone does not provide good control of the 1D nanostructure organization. Assembly of NTs [87, 88] and NWs on patterned substrates shows that 1D nanostructures align with bridging and looping structures over the patterned areas and show little directional control. The use of fluidic flows avoids these significant problems and enables controlled assembly in one or more directions. By combining this approach with other surface patterning methods, such as phase separation in diblock copolymers [89] and spontaneous ordering of molecules [90], it should be possible to generate well-ordered NW arrays without the limitations of conventional lithography.

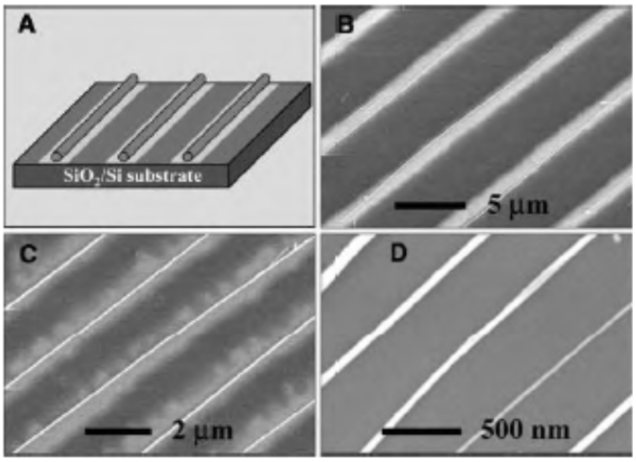


Fig. 15.18. Assembly of periodic NW arrays. (A) Schematic view of NW assembly onto a chemically patterned substrate. (B, C) Parallel arrays of GaP NWs aligned on poly methylmethacrylate (PMMA) patterned surface with 5 and 2 µm separation. (D) Parallel arrays of GaP NWs with 500nm separation obtained with a patterned self-assembled monolayer (SAM) surface [69].

15.2.3. Langmuir–Blodgett (LB) type techniques

Langmuir–Blodgett (LB) type techniques can push nanowires floating on the water–air interface into nematic liquid-crystal structures [75, 76]. In this case, the surfaces of the nanowires are first modified with surfactant molecules that have hydrophobic heads facing the water. The repulsive interaction between these hydrophobic groups and the water molecules lifts the nanowires to the water–air interface. Increasing surface pressure (*P*) forces nanowires to rotate perpendicular to the pressure direction (Fig. 15.14C) in the LB trough. In this manner, a monolayer of aligned nanowires forms on the surface of the water. These wire arrays can be transferred onto other substrates. The advantage of this approach is that it can achieve high density and large area coverage, both of which are critical requirements for fabrication of high-performance macroelectronics. Nevertheless, the surfactant modification on the surfaces of nanowires can affect their electrical properties and thus their performance in applications. The resultant nanowire arrays also have relatively poor end-to-end registry. Finally, the alignment process can be slow, and the transfer mechanism can be inconvenient for many applications.

15.2.4. Magnetic Field Assembly

A different approach exploits magnetism to generate, in a high speed and effective manner, concentrated and aligned arrays [77]. Here, ribbons or wires are coated with a ferromagnetic material such as Ni. An external magnetic field applied to a solution suspension of such elements aligns and concentrates them in controlled ways. As shown in Fig. 15.14D, the interaction between a magnetic field and ferromagnetic Ni coatings induces the ribbons to align parallel to the magnetic field. The aligned ribbons can be fixed such that they maintain their order during the drying process by bonding them to the surface of a photocured layer of polyurethane (PU) film (see a typical SEM image in Fig. 15.5D). This method can precisely position nanostructures by controlling the magnetic field. It is possible to integrate this approach with printing technologies to generate patterns over large areas. In addition, the Ni stripes can serve as electrodes in the further TFT fabrication because Ni can form a silicide with Si to generate ohmic contacts [91].

15.2.5. In Suti Growth of Aligned Nanowires via VLS Process

Nanowires can epitaxially grow vertically to surfaces of appropriate substrates when reaction conditions are precisely controlled. As a result, patterning metal nanoparticle catalysts on a substrate with specific crystallographic orientation can direct the in situ growth of aligned nanowire arrays via the VLS process (see Fig. 15.19A) [78-81]. For example, an array of Si nanowires perpendicular to a (111) Si wafer can grow through catalysis of Au nanoparticles (>20nm) deposited on the surface of Si wafer because the Si wires prefer to grow along the <111> direction [78]. ZnO nanowire arrays can vertically grow on an α-plane (110) sapphire substrate although ZnO nanowires tend to grow along the <001> direction. The epitaxial growth of ZnO nanowires on the (110) plane of sapphire is attributed to the fact that the α-axis of wurzite ZnO and the c-axis of sapphire are related by a factor of four (i.e., $\alpha_{ZnO} = 3.24$A$^{\circ}$ vs. $c_{sapphire} = 12.99$A$^{\circ}$) with mismatching less than 0.08% at room temperature [79]. Locations of nanowire arrays grown in this way can be easily controlled by precisely manipulating the positions of Au nanoparticles through various lithographic techniques including, for example, soft lithography, e-beam lithography, photolithography, and microcontact printing. It is apparent that the density of wire arrays is determined by the density of Au nanoparticles on substrate. Arrays formed using this method are highly dense and well aligned on the mother substrate, but they might be difficult to transfer to plastic sheets while preserving their order. Most recently, it has been reported that the growth orientation of ZnO nanowires on sapphire surface could be altered by controlling the size and spacing of Au nanoparticles [82]. The results indicate it is necessary to have Au nanodroplet catalysts with diameters smaller than 20 nm and spacings larger than 100 nm to achieve horizontal growth in the surface plane of substrate (Fig. 15.19B). Au nanoparticles larger than 20 nm always yield nanowires with offsurface orientations regardless of the interparticle spacing. Aligned ZnO nanowires with diameters less than 10 nm have been grown on the surface of α-plane (112^{-}0) sapphire substrates through VLS process combined with prepatterning of thin Au lines of very narrow widths. These kinds of wire arrays are suitable for transfer onto plastic substrates for TFTs, but it is difficult to increase the wire density due to relatively large spacing between Au nanoparticles.

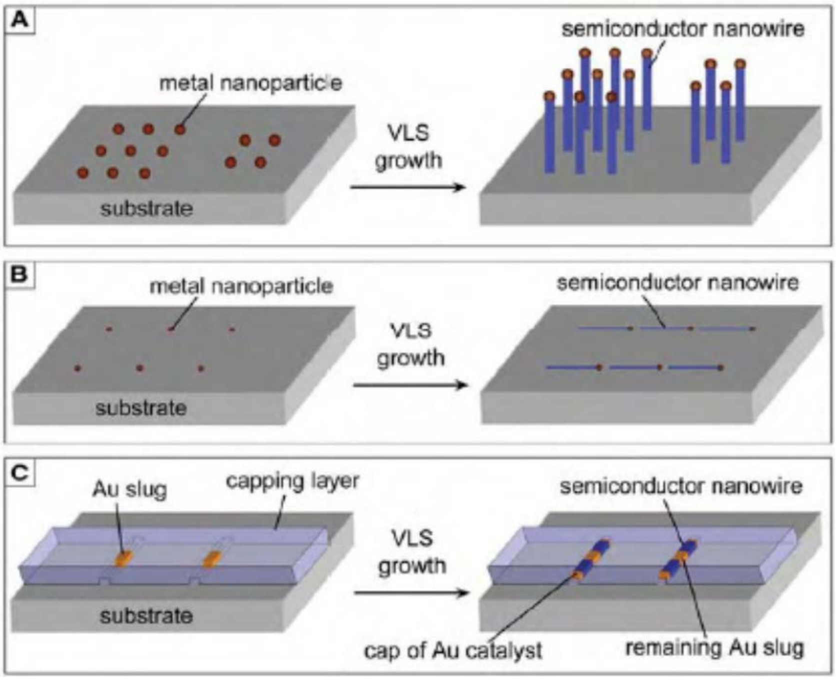


Fig. 15.19. In situ growth of nanowire arrays on substrates via the VLS process: (A) wire array perpendicular to surface of substrate; (B) wire array in the plane of substrate; (C) wire array grown on substrate in nanochannels.

The ‘‘grow-in-place’’ approach demonstrates the ability to produce self-assembled, crystalline Si nanowires by combining the VLS growth mechanism with lithographically fabricated nanochannels as growth templates [83]. As shown in Fig. 15.19C, Au lines that have heights and widths on the nanometer scale are first generated on an insulating substrate through e-beam lithography, deposition of metal and lift-off. A capping layer (e.g., silicon nitride) is then deposited over the substrate, followed by photolithographic patterning and selective dry etching. The Au lines under the capping layer are shortened to slugs with lengths of several micrometers by wet etching. In this configuration, the remaining Au in the center of channels serves as catalyst for the growth of semiconductor nanowires (i.e., Si), while the nanochannels act as templates to confine the dimensions of nanowires. When the Au slug is longer than 10 μm, the silicon cannot saturate the whole slug, but instead locally saturates the ends. The nanodroplet of Au–Si alloy at each end induces the growth of one wire by continuous transfer of $SiH_4 - H_2$ gas. As a result, two Si nanowires separated by the remaining Au slug form in the nanochannel. These two wires join together to form a continuous wire when the Au slug is short enough (e.g., < 2 μm). The resultant wires have well-controlled size, shape, orientation, and positioning. On the other hand, the use of e-beam lithography makes it expensive to generate wire arrays with large area.

15.2.6. Dry Transfer Printing of Wire Arrays

Nanowires fabricated via ‘‘top-down’’ approaches can inherit the order of the lithographic patterns used to produce them. Ordered nanowires produced in this manner can be transfer printed onto any desired substrates, such as plastic sheets. For example, Fig. 15.20A summarizes the steps for generating and transferring nanowire arrays of GaAs to a plastic substrate, i.e., poly(ethylene terephthalate) (PET) sheet [37]. If patterned SiO_2 lines (etch masks) surrounded by bulk SiO_2 film are defined on the surface of a (100) wafer, both ends of each GaAs wire generated from anisotropic chemical etching connect to the mother wafer. This connection confines the wires and preserves the spatial orientation defined by the pattern of SiO_2. GaAs wire arrays can be transfer printed to plastic sheets with retention of their orientation and relative position using a polydimethylsiloxane (PDMS) stamp. In the transfer printing process, the PDMS stamp is slightly oxidized to generate a hydrophilic surface by exposing it to a weak oxygen plasma or ozone. Contacting the oxidized PDMS stamp with SiO_2-covered GaAs wafer generates strong chemical bonding due to the formation of covalent siloxane (Si–O–Si) linkages between PDMS and SiO_2 via a condensation reaction. This bonding is strong enough to break the crystalline connections at the ends of wires. Peeling the PDMS stamp away from the GaAs substrate transfers all of the wires to the stamp. The remaining GaAs wafer after this transfer step can be polished to regenerate a flat surface for fabrication of more wires. The combination of wafer polishing with wire fabrication has the capability to generate a huge number of GaAs wires from a single piece of wafer, thus significantly decreasing the cost per wire.

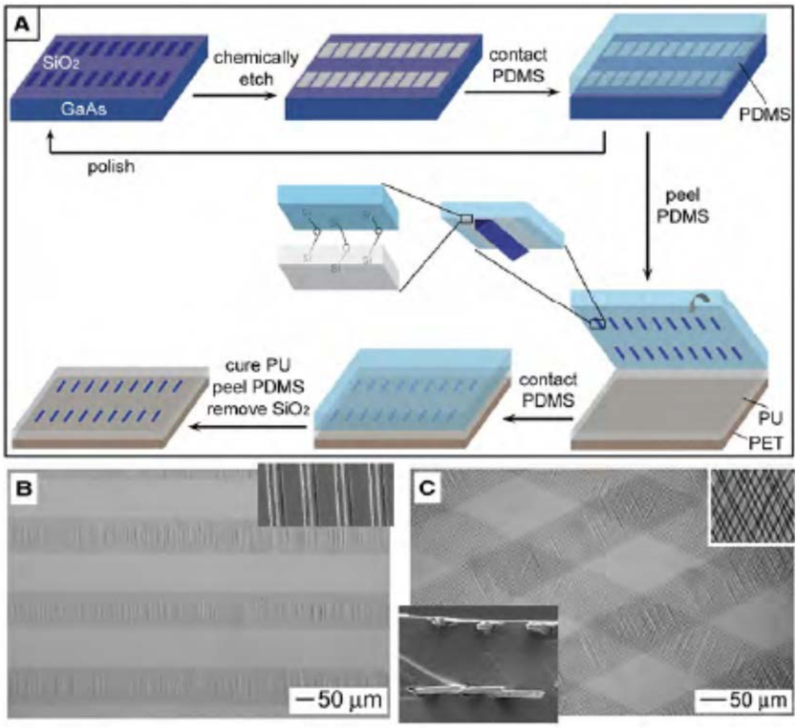


Fig. 15.20. Transfer printing wire arrays formed via ‘‘top-down’’ approaches: (A) schematic illustration of steps for transfer printing GaAs wire arrays generated by anisotropic etching; (B and C) optical microscopic images of (B) single and (C) double layers of GaAs wire arrays transfer printed on PET substrates.

In the next step, a photocurable polymer, such as polyurethane (PU), is used as a glue to bind the GaAs wires to a polyethylene terephthalate (PET) substrate. Placing the PDMS stamp (with GaAs wires) against a PET sheet coated with a thin layer of PU (thickness between one and tens of micrometers, determined by spin speed) forms conformal contact by pushing the surfaces of the wires into the liquid PU. Illuminating the sample with a UV lamp crosslinks and solidifies the PU layer, resulting in the formation of a strong bond between the cured PU and the GaAs wires and the underlying PET substrate. Peeling off the PDMS stamp and dissolving the SiO_2 stripes leave the GaAs wire arrays embedded in the matrix of cured PU with order and crystallographic orientation similar to those of the wires prior to transfer. The top flat surfaces of the wires are facing up, and can be exploited for subsequent device fabrication steps.

Figure 15.20B gives an optical image of a PU/PET substrate with GaAs wire arrays, indicating that all of GaAs wires were transfer printed onto the PET substrate with preservation of order and spatial orientation. The inset shows an SEM image taken from the same sample, clearly showing that the GaAs wires were well embedded in the cured PU. The flat top surface of each wire is exposed for further applications. The transfer printing process can be repeated to generate multiple layers of GaAs wire arrays on the same PET substrate. Fig. 15.20C presents an optical image of a sample with two layers of wire arrays formed by rotating the second layer with ~45° relative to the first layer. The bottom left inset is an SEM image of the cross section of this sample, indicating that the two individual layers of GaAs wire arrays are well separated by the cured PU. These PU layers can electrically insulate circuits or devices built on different layers of GaAs wire arrays. This ‘‘dry transfer printing’’ process combined with the ‘‘top-down’’ fabrication approach is promising for applications related to high-performance macroelectronics. It controls well the crystallographic orientation of transferred wires, with the top (100) surfaces facing up to provide an extremely flat surface (similar to that of original wafer) for device fabrication. The use of SiO_2 stripes as etch mask and buffer layer for the transfer prevents the (100) surfaces of the GaAs wires from becoming contaminated by organics used in the processing. In addition, conformal contact between PDMS stamp and liquid PU generates GaAs wires conformally bound to, and embedded in cured PU, which prevents the wires from moving, especially when the plastic substrates are bent or twisted.

Circuits for macroelectronics are fabricated on substrates with sizes much larger than semiconductor wafers. Active semiconductor films usually occupy only 10–20% area of the whole substrate, and the remaining regions are used to build interconnection between individual device units (e.g., transistors). Therefore, wire arrays generated on a mother wafer can be selectively transferred from certain regions rather than the whole surface, thus further decreasing the cost of fabrication. Fig. 15.21A shows a diagram for printing nanowire arrays generated from a single wafer over large areas on a plastic substrate. This process uses a PDMS stamp with patterned posts. Placing the stamp against the wafer and peeling it off picks up only the wires that contact with the surface of extrusive posts. The printing process can transfer patches of wire arrays onto plastic substrates. Repeating the steps over the remaining wires on the mother wafer can yield organized arrays of wires over large areas on a plastic substrate. Fig. 15.21B gives an optical image of patches consisting of GaAs wire arrays that were transfer printed on a PU+PET substrate. The image with relatively high magnification (as shown in Fig. 15.21C) shows that the wires of each patch are well aligned and have good end-to-end registry. This transfer printing process can be extended to wire arrays of other materials (e.g., InP, Si, etc.) fabricated via ‘‘top-down’’ approaches [37, 39, 92]. For instance, Fig. 15.21D shows a 5” × 5” PET sheet with patterned patches of Si ribbons.

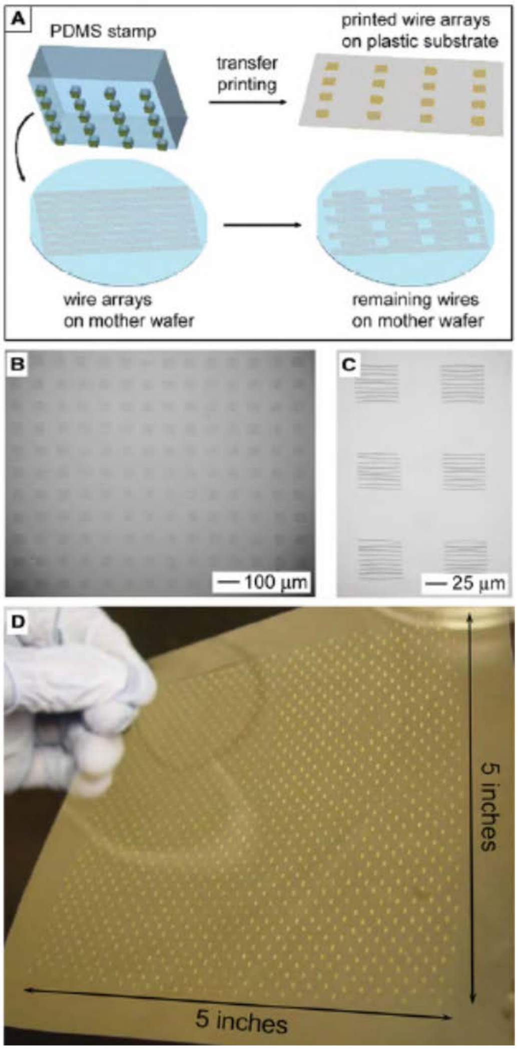


Fig. 15.21. Transfer printing patches of wire arrays on plastic substrates using PDMS stamps with patterned posts: (A) schematic illustration of this process; (B and C) optical images of patches of arrays of 50 μm long GaAs wires; (D) a large-area sample with transfer printed Si ribbons on a PET substrate.

15.2.7. Future Directions

The above data demonstrate clearly ordering of NW structures over multiple length scales - organization of nanometer diameter wires with 100 nm to micrometer scale separations over millimeter scale areas. This hierarchical order can readily bridge the microscopic and macroscopic worlds, although eventual device formation still requires conventional lithography to define metallic contact electrodes to the NWs. It is possible to eliminate this step by direct assembly of NWs onto predefined metallic electrode arrays [74] that have been deposited by either via conventional or unconventional (e.g., microcontact printing or imprinting methods) lithography techniques. Lastly, it will also be beneficial to develop further these approaches in the future to enable more complex assembly, such as the selective parallel assembly of two or more electronically distinct NWs and/or NWs with different lengths.

15.3. CHARACTERIZATION AND PHYSICAL PROPERTIES OF NANOWIRES

The discovery and investigation of nanostructures were stimulated by advances in various characterization and microscopy techniques that enable materials characterization to take place at smaller and smaller length scales, reaching down to individual atoms. For applications, characterization of the nanowire structural properties is especially important so that a reproducible relationship between their desired functionality and their geometrical and structural characteristics can be established. Due to the enhanced surface-to-volume ratio in nanowires, their properties may depend sensitively on their surface condition and geometrical configuration. Even nanowires made of the same material may possess dissimilar properties due to differences in their crystal phase, crystalline size, surface conditions, and aspect ratios, which depend on the synthesis methods and conditions used in their preparation.

15.3.1. Structural Characterization

Structural and geometric factors play an important role in determining the various attributes of nanowires, such as their electrical, optical and magnetic properties. Various novel tools, therefore, have been developed and employed to obtain this important structural information at the nanoscale. At the micron scale, optical techniques are extensively used for imaging structural features. Since the sizes of nanowires are usually comparable to or, in most cases, much smaller than the wavelength of visible light, traditional optical microscopy techniques are usually limited in characterizing the morphology and surface features of nanowires. Electron microscopy techniques, therefore, play a more dominant role at the nanoscale. Since electrons interact more strongly than photons, electron microscopy is particularly sensitive relative to X-rays for the analysis of tiny samples.

Let’s review and give examples of how scanning electron microscopy, transmission electron microscopy, scanning probe spectroscopies, and diffraction techniques are used to characterize the structure of nanowires. To provide the necessary basis for developing reliable structure-property relations, multiple characterization tools to the same samples are applied.

Scanning Electron Microscopy (SEM)

SEM usually produces images down to length scales of ~10 nm and provides valuable information regarding the structural arrangement, spatial distribution, wire density, and geometrical features of the nanowires. Examples of SEM micrographs shown in Figs. 15.22 and 15.23 indicate that structural features at the 10 nm to 10μm length scales can be probed, providing information on the size, size distribution, shapes, spatial distributions, density, nanowire alignment, filling factors, granularity etc.

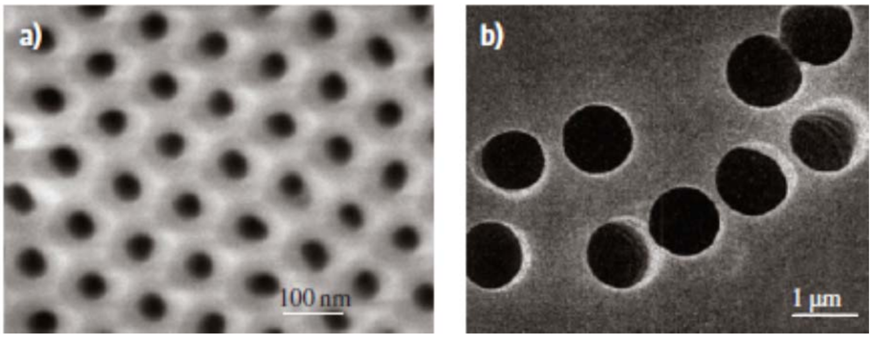


Fig. 15.22. (a) SEM images of the top surfaces of porous anodic alumina templates anodized with an average pore diameter of 44 nm [93]. (b) SEM image of the particle track-etched polycarbonate membrane, with a pore diameter of 1μm [94].

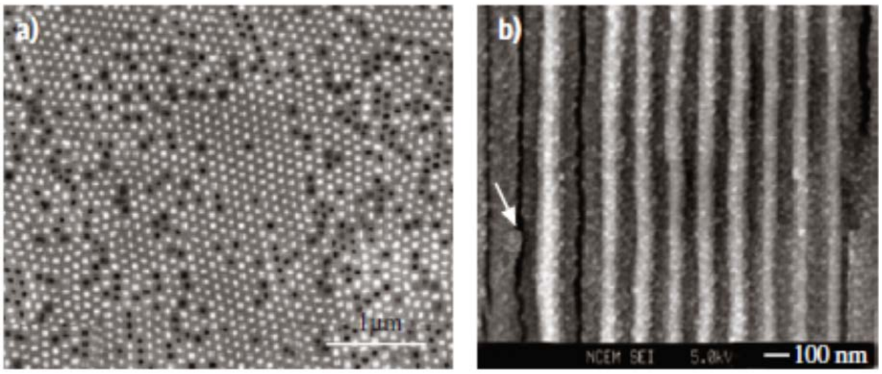


Fig. 15.23. (a) SEM image of a Bi$_2$Te$_3$ nanowire array in cross section showing a relatively high pore filling factor. (b) SEM image of a Bi$_2$Te$_3$ nanowire array composite along the wire axis [95].

As another example, Fig. 15.24a shows an SEM image of ZnO nanowire arrays grown on a sapphire substrate [96], which provides evidence for the nonuniform spatial distribution of the nanowires on the substrate. This distribution was attained by patterning the catalyst film to define high density growth regions and nanowire-free regions. Figure 15.24b, showing a higher magnification of the same system, indicates that these ZnO nanowires grow perpendicular to the substrate, are well aligned with approximately equal wire lengths, and have wire diameters in the range of $20 < d_W < 150$ nm. The SEM micrograph in Fig. 15.24c provides further information about the surface of the nanowires, showing it to be well-faceted and forming a hexagonal crosssection, indicative of nanowire growth along the <0001> direction. The uniformity of the nanowire size, its alignment perpendicular to the substrate, and its uniform growth direction, as suggested by the SEM data, are linked to the good epitaxial interface between the (0001) plane of the ZnO nanowire and the (110) plane of the sapphire substrate. (The crystal structures of ZnO and sapphire are essentially incommensurate, with the exception that the *a* axis of ZnO and the *c* axis of sapphire are related almost exactly by a factor of 4, with a mismatch of less than 0.08% at room temperature [96].) The well-faceted nature of these nanowires has important implications for their lasing action. Figure 15.25 shows an SEM image of GaN nanowires synthesized by a laser-assisted catalytic growth method [97], indicating a random spatial orientation of the nanowire axes and a wide diameter distribution for these nanowires, in contrast to the ZnO wires in Fig. 15.24 and to arrays of well-aligned nanowires prepared by template-assisted growth (see Fig. 15.23).

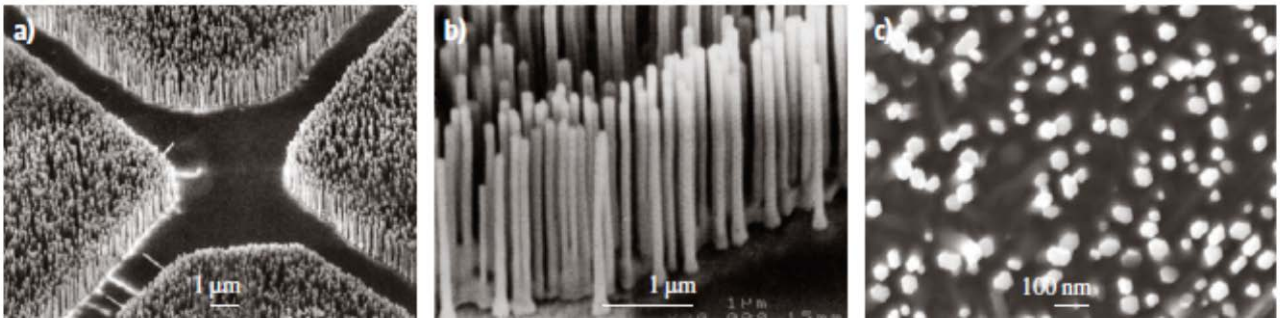


Fig. 15.24. SEM images of ZnO nanowire arrays grown on a sapphire substrate, where (a) shows patterned growth, (b) shows a higher resolution image of the parallel alignment of the nanowires, and (c) shows the faceted side-walls and the hexagonal cross section of the nanowires. For nanowire growth, the sapphire substrates were coated with a 1.0 to 3.5 nm thick patterned layer of Au as the catalyst, using a TEM grid as the shadow mask. These nanowires have been used for nanowire laser applications [96].

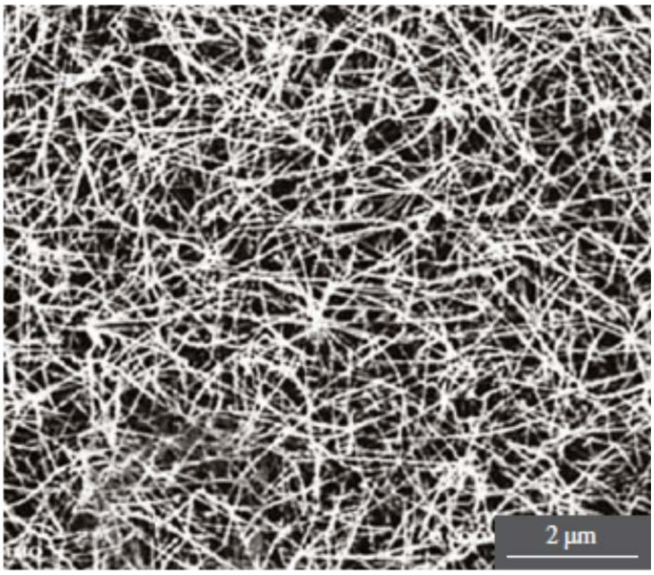


Fig. 15.25. SEM image of GaN nanowires in a mat arrangement synthesized by laser-assisted catalytic growth. The nanowires have diameters and lengths on the order of 10 nm and 10μm, respectively [97].

Transmission Electron Microscopy (TEM)

TEM and high resolution transmission electron microscopy (HRTEM) are powerful imaging tools to study nanowires at the atomic scale, and they usually provide more detailed geometrical features than are seen in SEM images. TEM studies also yield information regarding the crystal structure, crystal quality, grain size, and crystal orientation of the nanowire axis. When operating in the diffraction mode, selected area electron diffraction (SAED) patterns can be made to determine the crystal structure of nanowires. As an example, the TEM images in Fig. 15.26 show four different morphologies for Si nanowires prepared by the laser ablation of a Si target [98]: (a) spring-shaped; (b) fishbone-shaped (indicated by solid arrow) and frog-egg-shaped (indicated by the hollow arrow), (c) pearl-shaped, while (d) shows the poly-sites of nanowire nucleation. The crystal quality of nanowires is revealed from high resolution TEM images with atomic resolution, along with selected area electron diffraction (SAED) patterns. For example, Fig. 15.27 shows a TEM image of one of the GaN nanowires from Fig. 15.25, indicating single crystallinity and showing (100) lattice planes, thus demonstrating the growth direction of the nanowire. This information is supplemented by the corresponding electron diffraction pattern in the upper right.

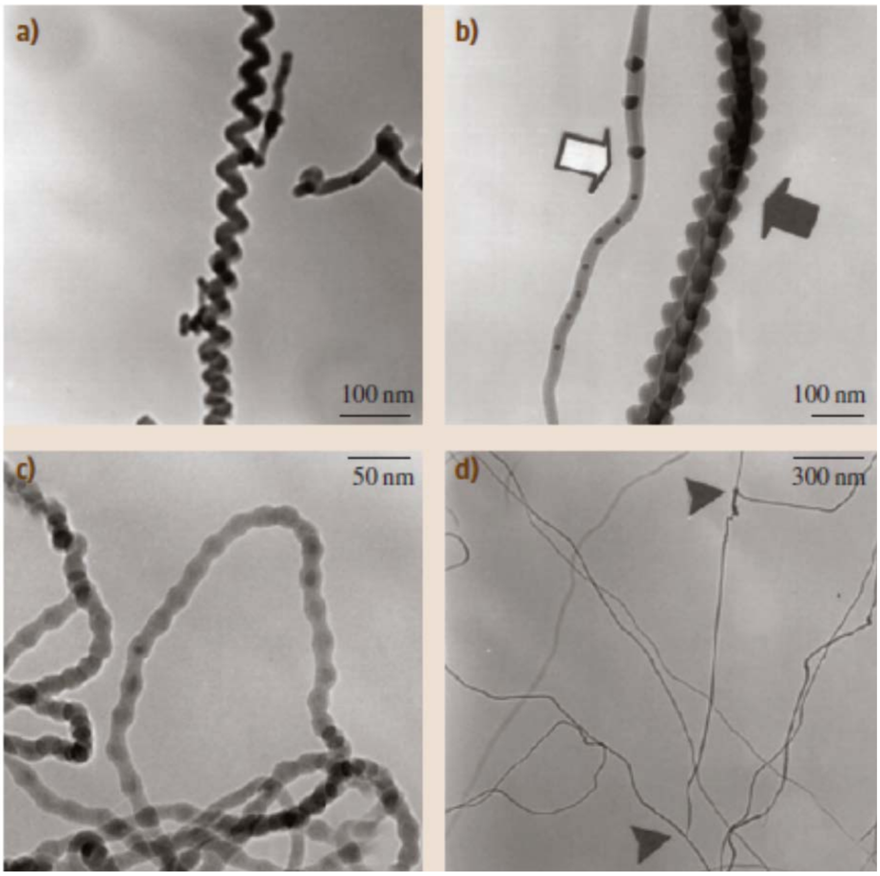


Fig. 15.26. TEM morphologies of four special forms of Si nanowires synthesized by the laser ablation of a Si powder target. (a) A springshaped Si nanowire; (b) fishboneshaped (indicated by a *solid arrow*) and frog-egg-shaped (indicated by a *hollow arrow*) Si nanowires; and (c) pearl-shaped nanowires, while (d) shows poly-sites for the nucleation of silicon nanowires (indicated by *arrows*) [98].

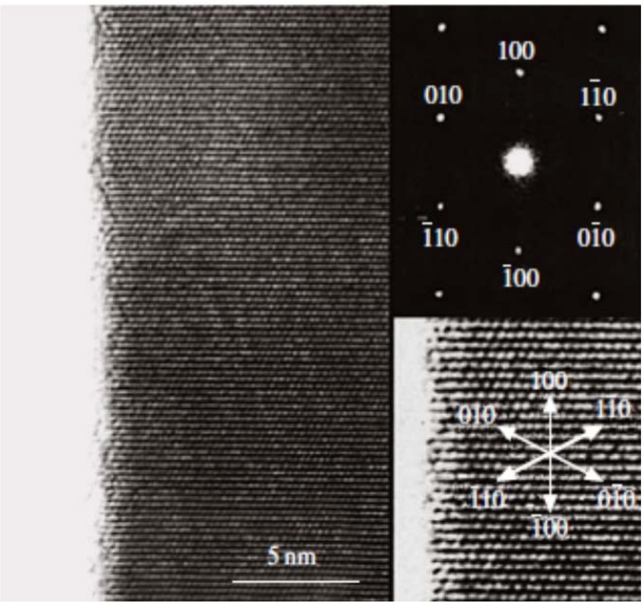


Fig. 15.27. Lattice resolved high resolution TEM image of one GaN nanowire (*left*) showing that (100) lattice planes are visible perpendicular to the wire axis. The electron diffraction pattern (*top*

right) was recorded along the <001> zone axis. A lattice-resolved TEM image (*lower right*) highlights the continuity of the lattice up to the nanowire edge, where a thin native oxide layer is found. The directions of various crystallographic planes are indicated in the *lower right* figure [97].

The high resolution of the TEM also allows for the investigation of the surface structure of the nanowires. In many cases the nanowires are sheathed with a native oxide layer or an amorphous oxide layer that forms during the growth process.

This can be seen in Fig. 15.28b for silicon nanowires and in Fig. 15.29 for germanium nanowires [99], showing a mass-thickness contrast TEM image and a selected-area electron diffraction pattern of a Ge nanowire. The main TEM image shows that these Ge nanowires possess an amorphous GeO_2 sheath with a crystalline Ge core oriented in the [211] direction.

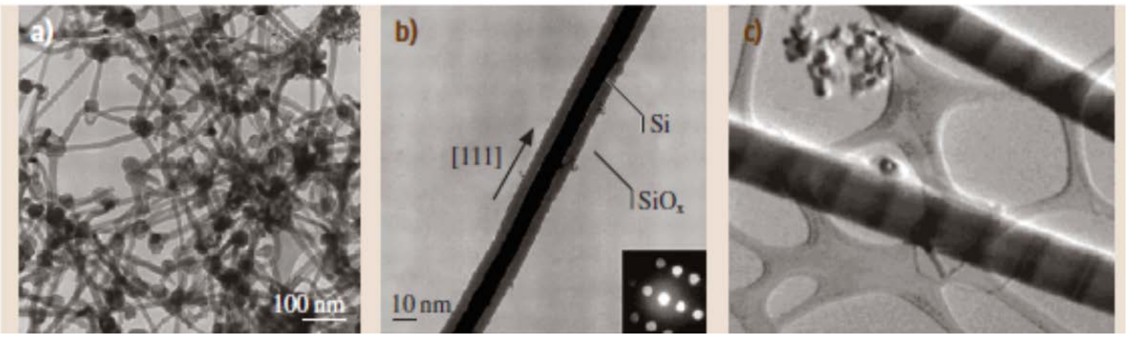


Fig. 15.28. (a) TEM images of Si nanowires produced after laser ablating a Si0.9Fe0.1 target. The dark spheres with a slightly larger diameter than the wires are solidified catalyst clusters [100]. (b) Diffraction contrast TEM image of a Si nanowire. The crystalline Si core appears darker than the amorphous oxide surface layer. The inset shows the convergent beam electron diffraction pattern recorded perpendicular to the wire axis, confirming the nanowire crystallinity [100]. (c) STEM image of Si/Si1−*x*Ge*x* superlattice nanowires in the bright field mode. The scale bar is 500 nm [101].

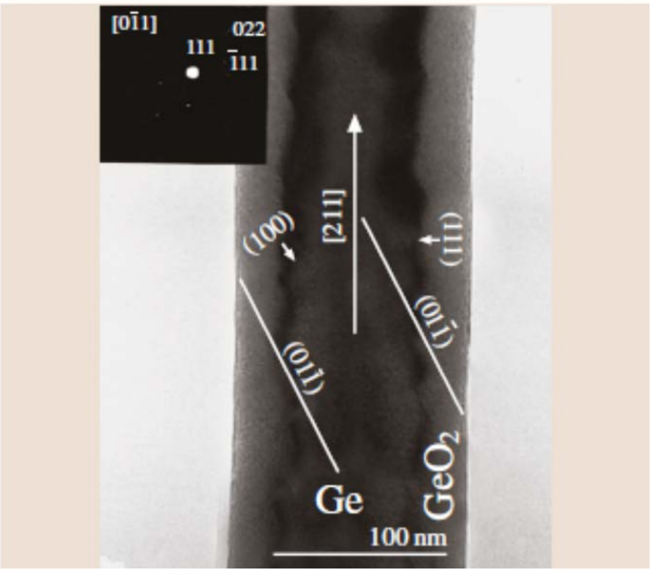


Fig. 15.29. A mass-thickness contrast TEM image of a Ge nanowire taken along the [01$^-$1] zone axis and a selected area electron diffraction pattern (*upper left inset*) [99]. The Ge nanowires were synthesized by laser ablation of a mixture of Ge and GeO_2 powder. The core of the Ge nanowire is crystalline, while the surface GeO_2 is amorphous.

Dynamical processes of the surface layer of nanowires can be studied by using an in situ environmental TEM chamber, which allows TEM observations to be made while different gases are introduced or as the sample is heat treated at various temperatures, as illustrated in Fig. 15.30. The figure shows high resolution TEM images of a Bi nanowire with an oxide coating and the effect of a dynamic oxide removal process carried out within the environmental chamber of the TEM [102]. The amorphous bismuth-oxide layer coating the nanowire (Fig. 15.30a) is removed by exposure to hydrogen gas within the environmental chamber of the TEM, as indicated in Fig. 15.30b.

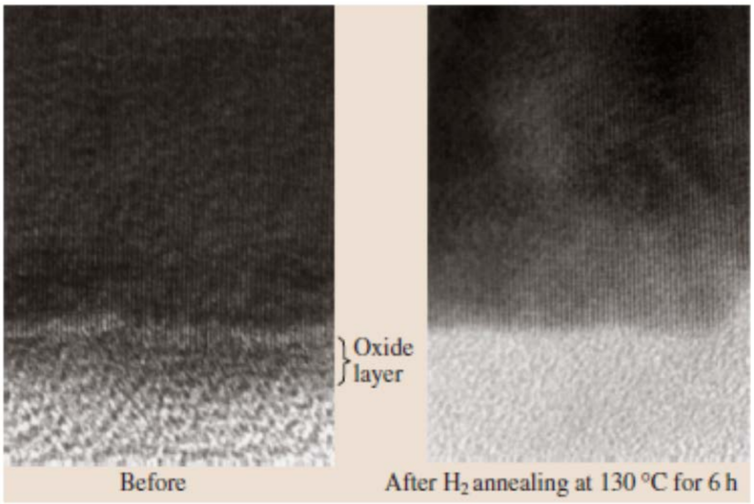


Fig. 15.30. High resolution transmission electron microscope (HRTEM) image of a Bi nanowire (*left*) before and (*right*) after annealing in hydrogen gas at 130 ◦C for 6 hours within the environmental chamber of the HRTEM instrument to remove the oxide surface layer [102].

By coupling the powerful imaging capabilities of TEM with other characterization tools, such as an electron energy loss spectrometer (EELS) or an energy dispersive X-ray spectrometer (EDS) within the TEM instrument, additional properties of the nanowires can be probed with high spatial resolution. With the EELS technique, the energy and momentum of the incident and scattered electrons are measured in an inelastic electron scattering process to provide information on the energy and momentum of the excitations in the nanowire sample. Fig. 15.31 shows the dependence on nanowire diameter of the electron energy loss spectra of Bi nanowires. The spectra were taken from the center of the nanowire, and the shift in the energy of the peak position (Fig. 15.31) indicates the effect of the nanowire diameter on the plasmon frequency in the nanowires. The results show changes in the electronic structure of Bi nanowires as the wire diameter decreases [103]. Such changes in electronic structure as a function of nanowire diameter are also observed in their transport and optical properties and are related to quantum confinement effects.

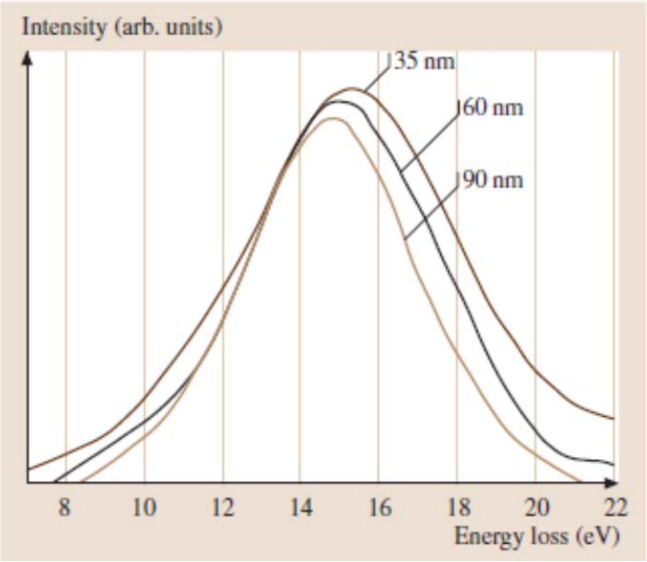


Fig. 15.31. Electron energy loss spectra (EELS) taken from the center of bismuth nanowires with diameters of 35, 60, and 90 nm. The shift in the volume plasmon peaks is due to the wire diameter effects on the electronic structure [103].

EDS measures the energy and intensity distribution of X-rays generated by the impact of the electron beam on the surface of the sample. The elemental composition within the probed area can be determined to a high degree of precision. The technique was particularly useful for the compositional characterization of superlattice nanowires [101] and core-sheath nanowires [104].

Scanning Tunneling Probes

Several scanning probe techniques [105], such as scanning tunneling microscopy (STM), electric field gradient microscopy (EFM) [106], magnetic field microscopy (MFM) [107], and scanning thermal microscopy (SThM) [108], combined with atomic force microscopy (AFM), have been employed to study the structural, electronic, magnetic, and thermal properties of nanowires. A scanning tunneling microscope can be employed to reveal both topographical structural information, such as that illustrated in Fig. 15.32, as well as information on the local electronic density of states of a nanowire, when used in the STS (scanning tunneling spectroscopy) mode.

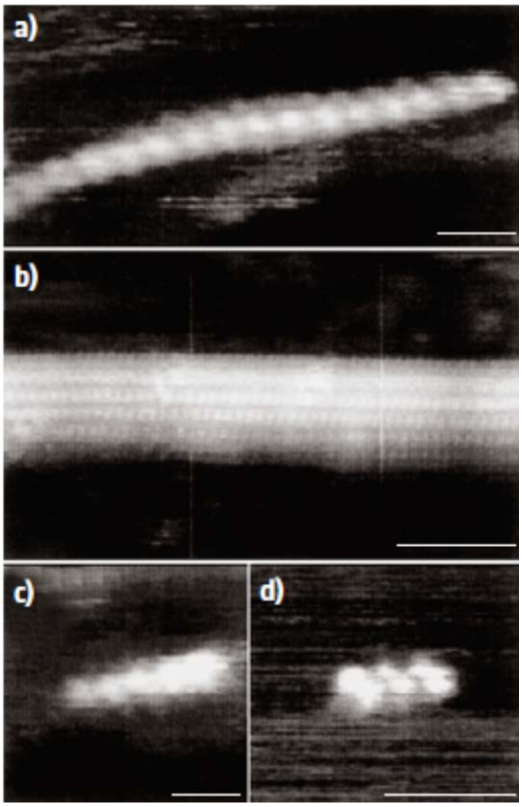


Fig. 15.32. STM height images, obtained in the constant current mode, of MoSe chains deposited on an Au(111) substrate. (a) A single chain image, and (b) a MoSe wire bundle. (c) and (d) Images of MoSe wire fragments containing 5 and 3 unit cells, respectively [105]. The scale bars are all 1 nm.

Figure 15.32 shows STM height images (taken in the constant current STM mode) of MoSe molecular wires deposited from a methanol or acetonitrile solution of $Li_2Mo_6Se_6$ on to Au substrates. The STM image of a single MoSe wire (Fig. 15.32a) exhibits a 0.45 nm lattice repeat distance in a MoSe molecular wire. When both STM and STS measurements are made on the same sample, the electronic and structural properties can be correlated, for example, as in the joint STM/STS studies on Si nanowires [109], showing alternating segments of a single nanowire identified with growth along [110] and [112] directions, and different I–V characteristics measured for the [110] segments as compared with the [112] segments.

Magnetic field microscopy (MFM) has been employed to study magnetic polarization of magnetic nanowires embedded in an insulating template, such as an anodic alumina template. For example, Fig. 15.33a shows the topographic image of an anodic alumina template filled with Ni nanowires, and Fig. 15.33b demonstrates the corresponding magnetic polarization of each nanowire in the template. This micrograph shows that a magnetic field microscopy probe can distinguish between spin-up and spin-down nanowires in the nanowire array, thereby providing a method for measuring interwire magnetic dipolar interactions [107].

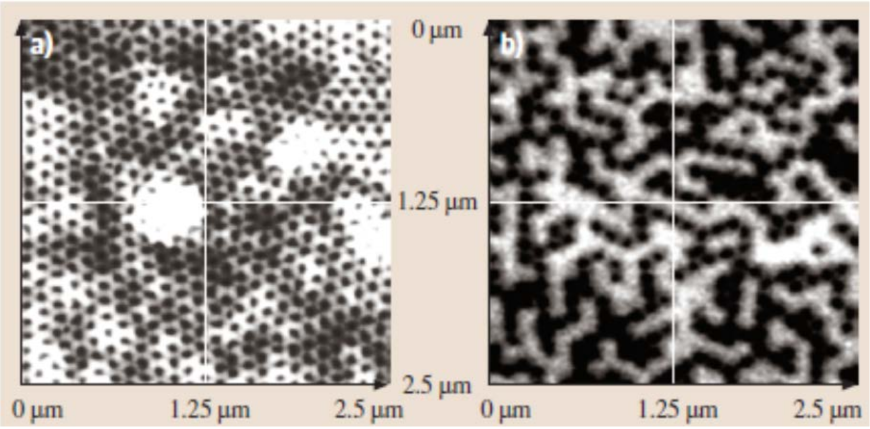


Fig. 15.33. (a) Topographic image of a highly ordered porous alumina template with a period of 100 nm filled with 35 nm diameter nickel nanowires. (b) The corresponding MFM (magnetic force microscope) image of the nano-magnet array, showing that the pillars are magnetized alternately "up" (*white*) and "down" (*black*) [107].

X-Ray Analysis

Other characterization techniques commonly used to study the crystal structure and chemical composition of nanowires include X-ray diffraction and X-ray energy dispersion analysis (EDAX). The peak positions in the X-ray diffraction pattern can be used to determine the chemical composition and the crystal phase structure of the nanowires. For example, Fig. 15.34 shows that Bi nanowires have the same crystal structure and lattice constants as bulk bismuth.

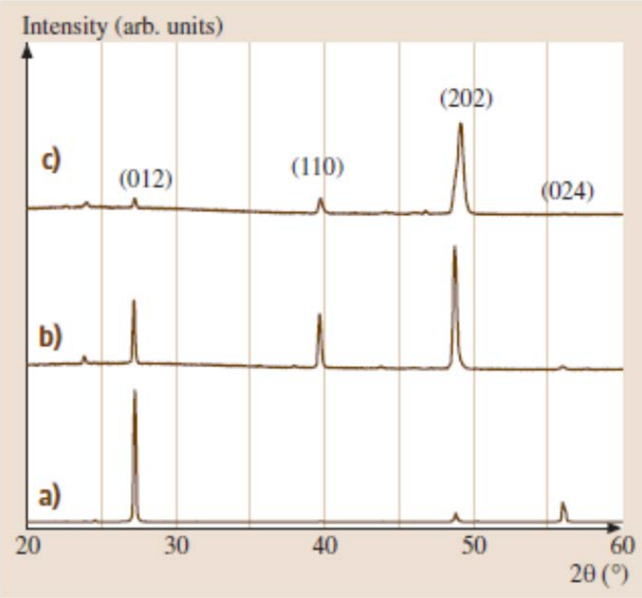


Fig. 15.34. XRD patterns of bismuth/anodic alumina nano-composites with average bismuth wire diameters of (a) 40 nm, (b) 52 nm, and (c) 95 nm [110]. The Miller indices corresponding to the lattice planes of bulk Bi are indicated above the individual peaks. The majority of the Bi nanowires are oriented along the [101^{-}1] and [011^{-}2] directions for $d_W \geq 60$ nm and $d_W \leq 50$ nm, respectively [106, 110]. The existence of more than one dominant orientation in the 52-nm Bi nanowires is attributed to the transitional behavior of *intermediate*-diameter nanowires as the preferential growth orientation is shifted from [101^{-}1] to [011^{-}2] with decreasing d_W.

Both the X-ray diffraction pattern (XRD) for an array of aligned Bi nanowires (Fig. 15.34) and the SAED pattern for individual Bi nanowires [106] suggest that the nanowires have a common axis of crystal orientation. As another example of an XRD pattern for an array of aligned nanowires, Fig. 15.35 shows the X-ray iffraction pattern of the ZnO nanowires displayed in Fig. 15.24. Only (00*l*) diffraction peaks are observed for these aligned ZnO nanowires, indicating that their preferred growth direction is (001) along the wire axis. A similar preferred growth orientation was also observed for Bi nanowires produced by high pressure injection, while the vapor phase and low pressure injection filling techniques give preferred alignment along lower symmetry axes (see Fig. 15.34). EDAXhas been used to determine the chemical composition, stoichiometry of compound nanowires, or the impurity content in the nanowires. But the results from EDAX analysis should be interpreted carefully to avoid systematic errors.

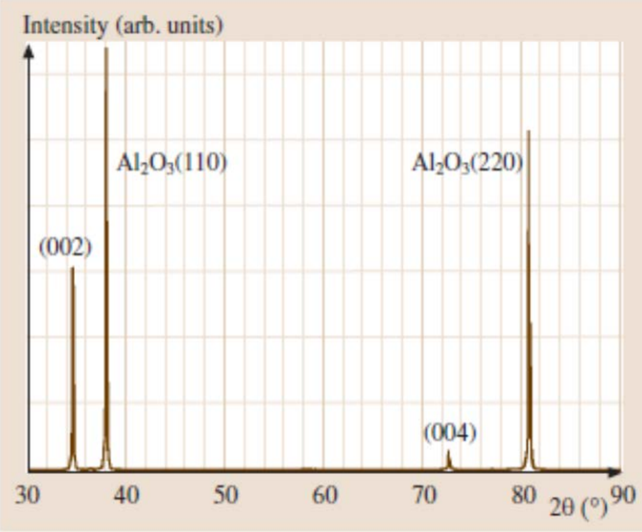


Fig. 15.35. X-ray diffraction pattern of aligned ZnO nanowires (see Fig. 15.24) grown on a sapphire substrate. Only [00*l*] diffraction peaks are observed for the nanowires, owing to their well-oriented growth orientation. Strong diffraction peaks for the sapphire substrate are also found [96].

15.3.2. Transport Properties

The study of nanowire electrical transport properties is important for nanowire characterization, electronic device applications, and the investigation of unusual transport phenomena arising from one-dimensional quantum effects. Important factors that determine the transport properties of nanowires include the wire diameter (important for both classical and quantum size effects), material composition, surface conditions, crystal quality, and the crystallographic orientation along the wire axis, which is important for materials with anisotropic materials parameters, such as the effective mass tensor, the Fermi surface, or the carrier mobility. Electronic transport phenomena in low-dimensional systems can be roughly divided into two categories: ballistic and diffusive transport. Ballistic transport phenomena occur when electrons travel across the nanowire without any scattering. In this case, the conduction is mainly determined by the contacts between the nanowire and the external circuit, and the conductance is quantized into an integral number of universal conductance units $G_0 = 2e^2/h$ [111, 112]. Ballistic transport phenomena are usually observed in very short quantum wires, such as those produced by using mechanically controlled break junctions (MCBJ) [113, 114] where the electron mean free path is much longer than the wire length, and the conduction is a pure quantum phenomenon. To observe ballistic transport, the thermal energy must also obey the relation $k_BT << \varepsilon_j - \varepsilon_{j-1}$, where $\varepsilon_j - \varepsilon_{j-1}$ is the energy separation between subband levels j and $j-1$. On the other hand, for nanowires with lengths much larger than the carrier mean free path, the electrons (or holes) undergo numerous scattering events when they travel along the wire. In this case, the transport is in the diffusive regime, and the conduction is dominated by carrier scattering within the wires due to phonons (lattice vibrations), boundary scattering, lattice and other structural defects, and impurity atoms.

Conductance Quantization in Metallic Nanowires

The ballistic transport of 1-D systems has been extensively studied since the discovery of quantized conductance in 1-D systems in 1988 [111, 112]. The phenomena of conductance quantization occur when the diameter of the nanowire is comparable to the electron Fermi wavelength, which is on the order of 0.5nm for most metals [115]. Most conductance quantization experiments up to the present were performed by joining and separating two metal electrodes. As the two metal electrodes are slowly separated, a nano-contact is formed before it breaks completely (see Fig. 15.36a), and conductance in integral multiple values of G_0 is observed through these nano-contacts. Fig. 15.36b shows the conductance histogram built with 18,000 contact breakage curves between two gold electrodes at room temperature [116], with the electrode separation up to ~ 1.8 nm. The conductance quantization behavior is found to be independent of the contact material and has been observed in various metals, such as Au [116], Ag, Na, Cu [117], and Hg [118]. For semimetals such as Bi, conductance quantization has also been observed for an electrode separation of as long as 100 nm at 4K because of the long Fermi wavelength (~26 nm) [115], indicating that the conductance quantization may be due to the existence of well-defined quantum states localized at a constriction instead of resulting from the atom rearrangement as the electrodes separate. Since the conductance quantization is observed only in breaking contacts, or for very narrow and very short nanowires, most nanowires of practical interest (possessing lengths of several microns) lie in the diffusive transport regime, where the carrier scattering is significant and should be considered.

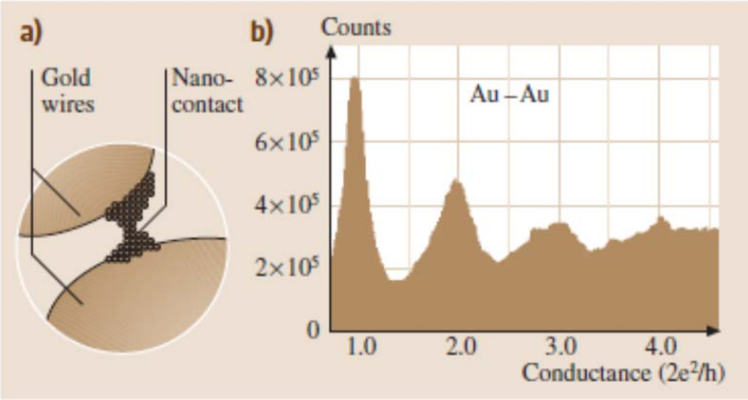


Fig. 15.36. (a) Schematic representation of the last stages of the contact breakage process [116]. (b) Histogram of conductance values built with 18,000 gold contact breakage experiments in air at

room temperature, showing conductance peaks at integral values of G_0. In this experiment the gold electrodes approach and separate at 89,000 A°/s [116].

I–V Characterization of Semiconducting Nanowires.

The electronic transport behavior of nanowires may be categorized based on the relative magnitude of three length scales: carrier mean free path l_W, the de Broglie wavelength of electrons λ_e, and the wire diameter d_W. For wire diameters much larger than the carrier mean free path ($d_W >> l_W$), the nanowires exhibit transport properties similar to bulk materials, which are independent of the wire diameter, since the scattering due to the wire boundary is negligible compared to other scattering mechanisms. For wire diameters comparable or smaller than the carrier mean free path ($d_W \sim l_W$ or $d_W < l_W$) but still much larger than the de Broglie wavelength of the electrons ($d_W >> \lambda_e$), the transport in nanowires is in the classical finite size regime, in which the band structure of the nanowire is still similar to that of bulk while the scattering events at the wire boundary alter their transport behavior. For wire diameters comparable to the electronic wavelength $d_W \sim \lambda_e$, the electronic density of states is altered dramatically and quantum subbands are formed due to the quantum confinement effect at the wire boundary. In this regime, the transport properties are further influenced by the change in the band structure. Transport properties for nanowires in the classical finite size and quantum size regimes, therefore, are highly diameter-dependent.

Researchers have investigated the transport properties of various semiconducting nanowires and have demonstrated their potential for diverse electronic devices, such as for *p-n* diodes [119, 120], field effect transistors [119], memory cells, and switches [121]. The nanowires studied so far in this context have usually been made from conventional semiconducting materials, such as group IV and III-V compound semiconductors via the VLS growth method, and their nanowire properties have been compared to their well-established bulk properties. Interestingly, the physical principles for describing bulk semiconductor devices also hold for devices based on these semiconducting nanowires with wire diameters of tens of nanometers. For example, Fig. 15.37 shows the current–voltage (I–V) behavior of a 4-by-1 crossed *p*-Si/*n*-GaN junction array at room temperature [119]. The long horizontal wire in the figure is a *p*-Si nanowire (10–25 nm in diameter) and the four short vertical wires are *n*-GaN nanowires (10–30 nm in diameter). Each of the four nanoscale cross points independently forms a *p-n* junction with current rectification behavior, as shown by the I–Vcurves in Fig. 15.37, and the junction behavior (e.g., the turn-on voltage) can be controlled by varying the oxide coating on these nanowires [119].

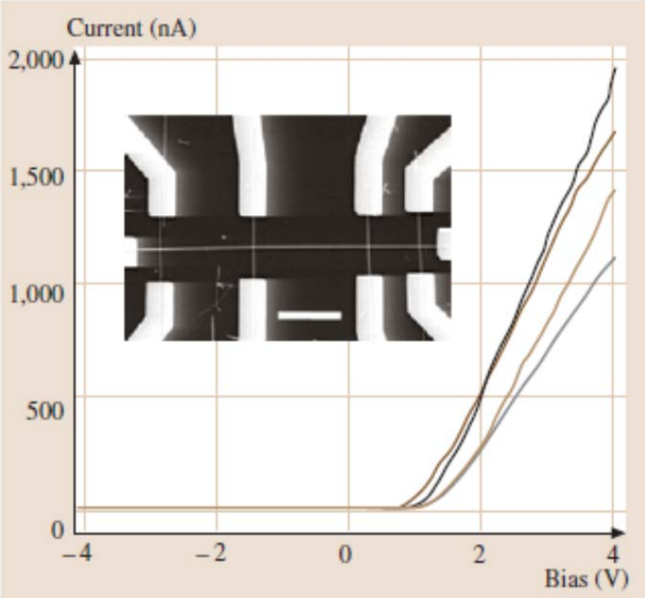


Fig. 15.37. I–V behavior for a 4(p) by 1(n) crossed p-Si/ n-GaN junction array shown schematically in the inset. The four curves represent the I–V response for each of the four junctions, showing similar current rectifying characteristics in each case. The length scale bar between the two middle junctions is 2μm [119]. The p-Si and n-GaN nanowires are 10–25 nm and 10–30 nm in diameter, respectively

The nanowire junction diodes with a high turn-on voltage (~ 5V) by increasing the oxide thickness at the junctions have been demonstrated. The high turn-on voltage enables the use of the junction in a nanoscale FET as shown in Fig. 15.38 [119], where I–V data for a p-Si nanowire are presented, for which the *n*-GaN nanowire with a thick oxide coating is used as a nanogate. By varying the nano-gate voltage, the conductance of the *p*-Si nanowire can be changed by more than a factor of 10^5 (lower curve in the inset), whereas the conductance changes by only a factor of ten when a global back-gate is used (top curve in the inset of Fig. 15.38). This behavior may be due to the thin gate dielectric between the crossed nanowires and the better control of the local carrier density through a nano-gate. Based on the gate-dependent I–V data of these *p*-Si nanowires, it is found that the mobility for holes in the *p*-Si nanowires may be higher than that for bulk p-Si, although further investigation is required for a complete understanding.

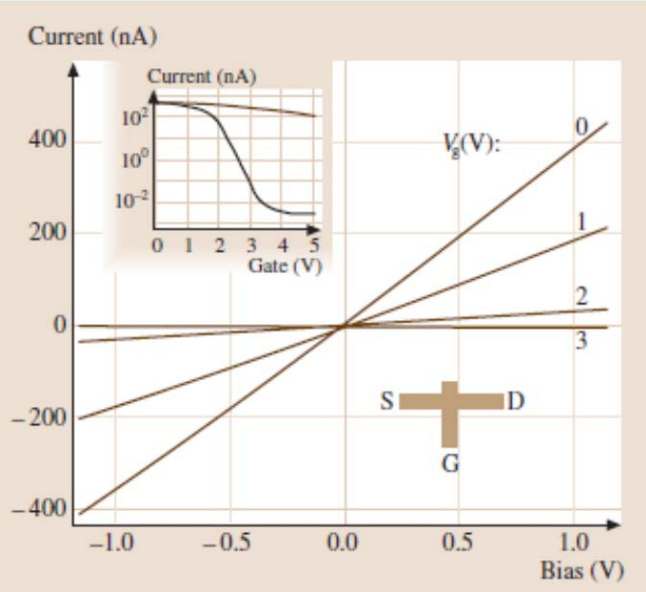


Fig. 15.38. Gate-dependent I–V characteristics of a crossed nanowire field-effect transistor (FET). The n-GaN nanowire is used as the nano-gate, with the gate voltage indicated (0, 1, 2, and 3 V). The inset shows the current vs. V_{gate} for a nanowire gate (*lower curve*) and for a global back-gate (*top curve*) when the bias voltage is set at 1V [119].

Because of the enhanced surface-to-volume ratio of nanowires, their transport behavior may be modified by changing their surface conditions. For example, researchers have found that by coating *n*-InP nanowires with a layer of redox molecules such as cobalt phthalocyanine, the conductance of the InP nanowires may change by orders of magnitude by altering the charge state of the redox molecules to provide bistable nanoscale switches [121]. The resistance (or conductance) of some nanowires (e.g., Pd nanowires) is also very sensitive to the presence of certain gases (e.g., H_2) [122, 123], and this property may be utilized for sensor applications to provide improved sensitivity compared to conventional sensors based on bulk material.

Although it remains unclear how the size may influence the transport properties and device performance of semiconducting nanowires, many of the larger diameter semiconducting nanowires are expected to be described by classical physics, since their quantization energies $\hbar^2/(2m_e d^2_W)$ are usually smaller than the thermal energy k_BT. By comparing the quantization energy with the thermal energy, the critical wire diameter below which quantum confinement effects become significant is estimated to be 1 nm for Si nanowires at room temperature, which is much smaller than the size of many of the semiconducting nanowires that have been investigated so far. By using material systems with much smaller effective carrier masses m_e (such as bismuth), the critical diameter for which such quantum effects can be observed is increased, thereby facilitating the study of quantum confinement effects. It is for this reason that the bismuth nanowire system has been studied so extensively. Furthermore, since the crystal structure and lattice constants of bismuth nanowires are the same as for 3-D crystalline bismuth, it is possible to carry out detailed model calculations to guide and to interpret transport and optical experiments on bismuth nanowires. For these reasons, bismuth can be considered as a model system for studying 1-D effects in nanowires.

Temperature-Dependent Resistance Measurements.

Although nanowires with electronic properties similar to their bulk counterparts are promising for constructing nano-devices based on well-established knowledge of their bulk counterparts, it is expected that quantum size effects in nanowires will likely be utilized to generate new phenomena absent in bulk materials and thus provide enhanced performance and novel functionality for certain applications. In this context, the transport properties of bismuth (Bi) nanowires have been extensively studied, both theoretically [124] and experimentally [110, 125-130] because of their promise for enhanced thermoelectric performance. Transport studies of ferromagnetic nanowire arrays, such as Ni or Fe, have also received much attention because of their potential for high-density magnetic storage applications.

The very small electron effective mass components and the long carrier mean free paths in Bi facilitate the study of quantum size effects in the transport properties of nanowires. Quantum size effects are expected to become significant in bismuth nanowires with diameters smaller than 50 nm [124], and the fabrication of crystalline nanowires of this diameter range is relatively easy.

Figure 15.39a shows the T dependence of the resistance $R(T)$ for Bi nanowires ($7 \leq d_W <$ 200 nm) synthesized by vapor deposition and pressure injection [125], illustrating the quantum effects in their temperature-dependent resistance. In Fig. 15.39a, the $R(T)$ behavior of Bi nanowires is dramatically different from that of bulk Bi and is highly sensitive to the wire diameter. The $R(T)$ curves in Fig. 15.39a show a nonmonotonic trend for larger-diameter (70 and 200 nm) nanowires, although $R(T)$ becomes monotonic with T for small-diameter (≤ 48 nm) nanowires. This dramatic change in the behavior of $R(T)$ as a function of d_W is attributed to a unique semimetal-semiconductor transition phenomena in Bi [131], induced by quantum size effects. Bi is a semimetal in bulk form, in which the T-point valence band overlaps with the L-point conduction band by 38 meV at 77K. As the wire diameter decreases, the lowest conduction subband increases in energy, and the highest valence subband decreases in energy. Model calculations predict that the band overlap should vanish in Bi nanowires (with their wire axes along the trigonal direction) at a wire diameter ~ 50 nm [124].

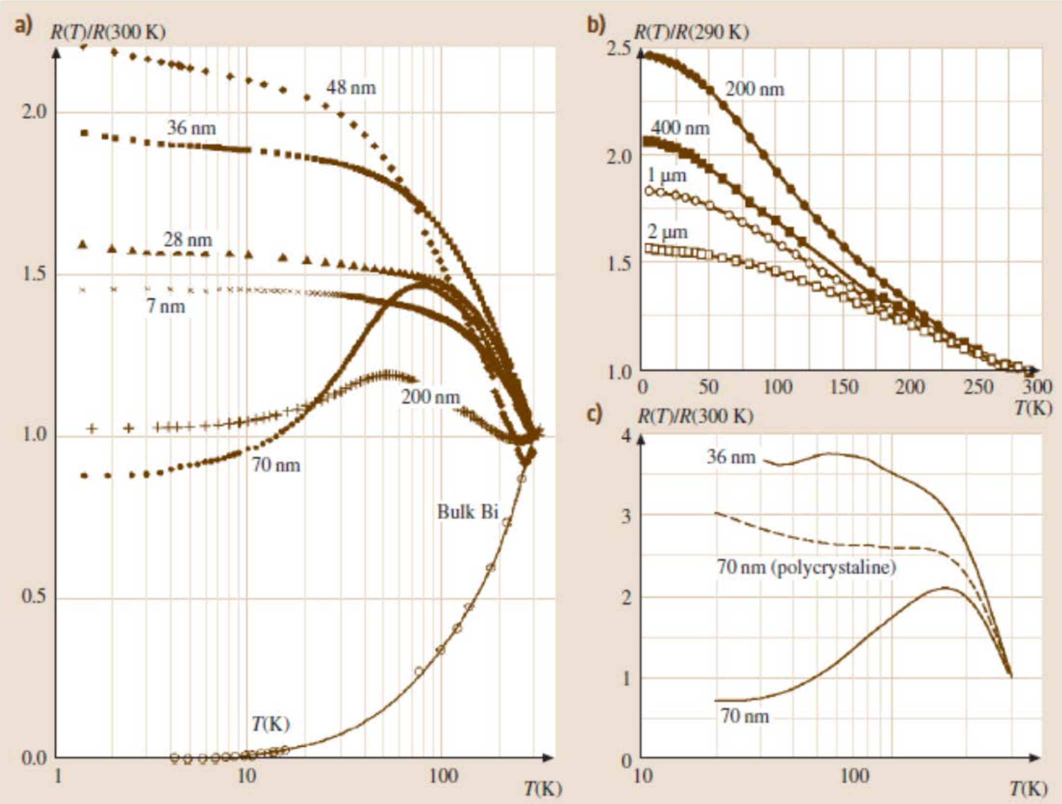


Fig. 15.39. (a) Measured temperature dependence of the resistance *R(T)* normalized to the room temperature (300 K) resistance for bismuth nanowire arrays of various wire diameters d_W [125]. (b) *R(T)/R(*290 K*)* for bismuth wires of larger d_W and lower mobility [126]. (c) Calculated *R(T)/R(*300 K*)* of 36-nm and 70-nm bismuth nanowires. The *dashed curve* refers to a 70-nm polycrystalline wire with increased boundary scattering [131].

The resistance of Bi nanowires is determined by two competing factors: the carrier density that increases with *T*, and the carrier mobility that decreases with *T*. The non-monotonic *R(T)* for large-diameter Bi nanowires is due to a smaller carrier concentration variation at low temperature (≤ 100 K) in semimetals, so that the electrical resistance is dominated by the mobility factor in this temperature range. Based on the semiclassical transport model and the established band structure of Bi nanowires, the calculated *R(T)/R(*300 K) for 36-nm and 70-nm Bi nanowires is shown by the solid curves in Fig. 15.39c to illustrate different *R(T)* trends for semiconducting and semimetallic nanowires, respectively [110]. The curves in Fig. 15.39c exhibit trends consistent with experimental results. The condition for the semimetal–semiconductor transition in Bi nanowires can be experimentally determined, as shown by the measured resistance ratio *R*(10 K)/*R*(100 K) of Bi nanowires as a function of wire diameter [132] in Fig. 15.40. The maximum in the resistance ratio *R*(10 K)/*R*(100 K) at $d_W \sim 48$ nm indicates the wire diameter for the transition of Bi nanowires from a semimetallic phase to a semiconducting phase. The semimetal–semiconductor transition and the semiconducting phase in Bi nanowires are examples of new transport phenomena, resulting from lowdimensionality and absent in the bulk 3-D phase, which further increase the possible benefits from the properties of nanowires for desired applications.

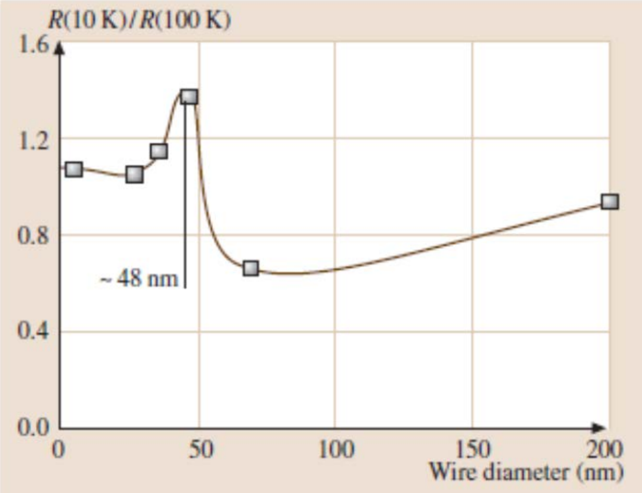


Fig. 15.40. Measured resistance ratio $R(10\text{K})/R(100\text{K})$ of Bi nanowires as a function of diameter. The peak indicates the transition from a semimetallic phase to a semiconducting phase as the wire diameter decreases [133].

It should be noted that good crystal quality is essential for observing the quantum size effect in nanowires, as shown by the $R(T)$ plots in Fig. 15.39a. For example, Fig. 15.39b shows the normalized $R(T)$ measurements of Bi nanowires with larger diameters (200 nm–2µm) prepared by electrochemical deposition [126], and these nanowires possess monotonic $R(T)$ behaviors, quite different from those of the corresponding nanowire diameters shown in Fig. 15.39a. The absence of the resistance maximum in Fig. 15.39b is due to the lower crystalline quality for nanowires prepared by electrochemical deposition, which tends to produce polycrystalline nanowires with much lower carrier mobility. This monotonic $R(T)$ for semimetallic Bi nanowires at a higher defect level is also confirmed by theoretical calculations, as shown by the dashed curve in Fig. 15.39c for 70-nm wires with increased grain boundary scattering [131].

The theoretical model developed for Bi nanowires not only provides good agreement with experimental results, but it also plays an essential role in understanding the influence of the quantum size effect, the boundary scattering, and the crystal quality on their electrical properties. The transport model has also been generalized to predict the transport properties of Te-doped Bi nanowires [110], Sb nanowires [134], and BiSb alloy nanowires [135], and good agreement between experiment and theory has also been obtained for these cases. While the electronic density of states may be significantly altered due to quantum confinement effects, various scattering mechanisms related to the transport properties of nanowires can be accounted for by Matthiessen's rule.

For nanowires with diameters comparable to the phase-breaking length, their transport properties may be further influenced by localization effects. It has been predicted that in disordered systems, the extended electronic wave functions become localized near defect sites, resulting in the trapping of carriers and giving rise to a different transport behavior. Localization effects are also expected to be more pronounced as dimensionality and sample size are reduced. Localization effects on the transport properties of nanowire systems have been studied on Bi nanowires [136] and, more recently, on Zn nanowires [137]. Figure 15.41 shows the measured $R(T)/R(300\text{K})$ of Zn nanowires fabricated by vapor deposition in porous silica or alumina [137]. While 15 nm Zn nanowires exhibit an $R(T)$ behavior with a $T1$ dependence as expected for a metallic wire, the $R(T)$ of 9 nm and 4 nm Zn nanowires exhibits a temperature dependence of $T^{-1/2}$ at low temperature, consistent with 1-D localization theory. Thus, due to this localization effect, the use of nanowires with very small diameters for transport applications may be limited.

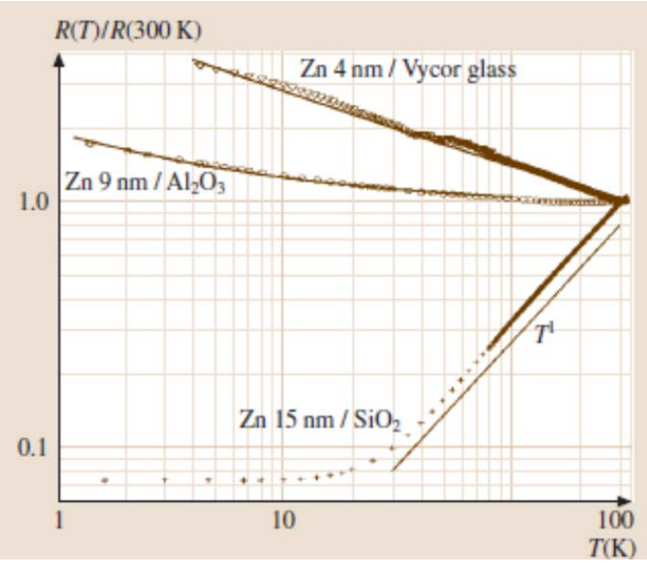


Fig. 15.41. Temperature dependence of the resistance of Zn nanowires synthesized by vapor deposition in various porous templates [137]. The data are given as points, the *full line* are fits to a T^1 law for 15 nm diameter Zn nanowires in an SiO_2 template, denoted by Zn/SiO_2. Fits to a combined T^1 and $T^{-1/2}$ law were made for the smaller nanowire diameter composite denoted by 9 nm Zn/Al_2O_3 and 4 nm Zn/Vycor glass

Magnetoresistance.

Magnetoresistance (MR) measurements provide an informative technique for characterizing nanowires because these measurements yield a great deal of information about the electron scattering with wire boundaries, the effects of doping and annealing on scattering, and localization effects in the nanowires [129]. For example, at low fields the MR data show a quadratic dependence on the B field from which carrier mobility estimates can be made (see Fig. 15.42 at low B field).

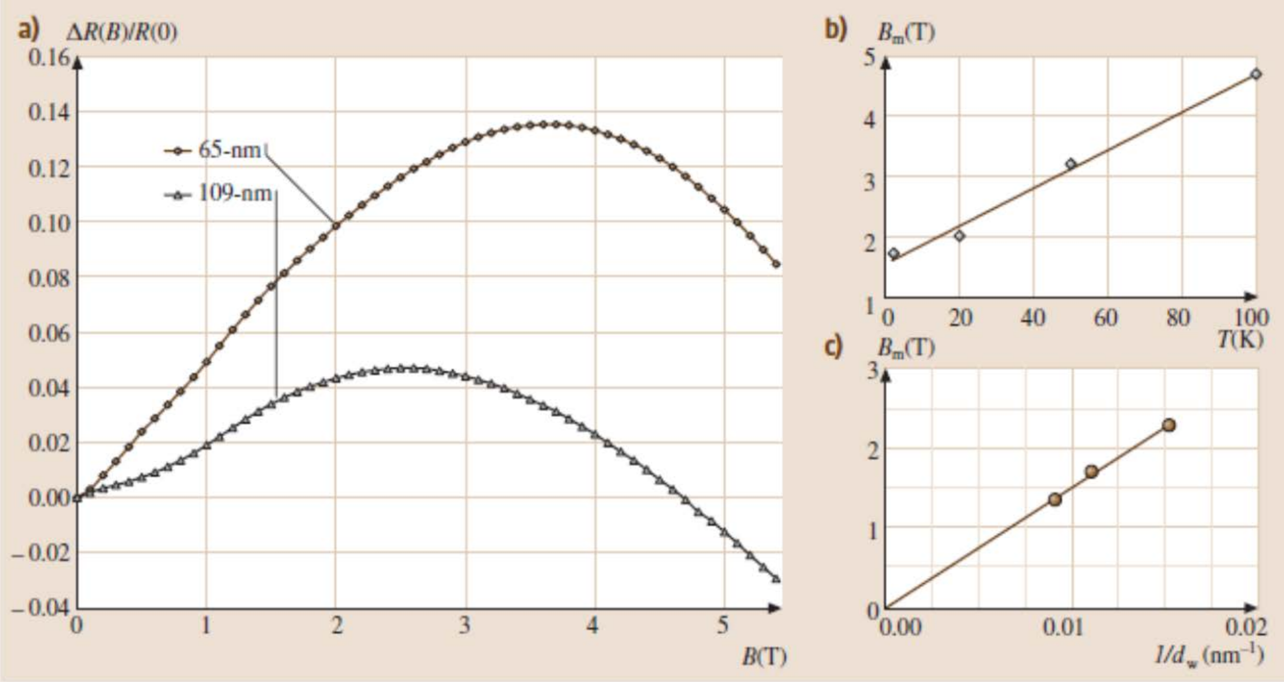


Fig. 15.42. (a) Longitudinal magnetoresistance, $\Delta R(B)/R(0)$, at 2K as a function of B for Bi nanowire arrays with diameters 65 and 109 nm before thermal annealing. (b) The peak position B_m as a function of temperature for the 109 nm diameter Bi nanowire array after thermal annealing. (c) The peak position B_m of the longitudinal MR (after thermal annealing) at 2K as a function of $1/d_W$, the reciprocal of the nanowire diameter [138].

Figure 15.42 shows the longitudinal magnetoresistance ($\boldsymbol{B}$ parallel to the wire axis) for 65 nm and 109 nm Bi nanowire samples (before thermal annealing) at 2K. The MR maxima in Fig. 15.42a are due to the classical size effect, where the wire boundary scattering is reduced as the cyclotron radius becomes smaller than the wire radius in the high field limit, resulting in a decrease in the resistivity. This behavior is typical for the longitudinal MR of Bi nanowires in the diameter range of 45 nm to 200 nm [125, 128, 129, 138], and the peak position B_m moves to lower B field values as the wire diameter increases, as shown in Fig. 15.42c [138], in which B_m varies linearly with $1/d_W$. The condition for the occurrance of B_m is approximately given by $B_m \sim 2c\hbar k_F/ed_W$ where k_F is the wave vector at the Fermi energy. The peak position, B_m, is found to increase linearly with temperature in the range of 2 to 100K, as shown in Fig. 15.42b [138]. As T is increased, phonon scattering becomes increasingly important, and therefore a higher magnetic field is required to reduce the resistivity associated with boundary scattering sufficiently to change the sign of the MR.

Likewise, increasing the grain boundary scattering is also expected to increase the value of B_m at a given T and wire diameter.

The presence of the peak in the longitudinal MR of nanowires requires a high crystal quality with long carrier mean free paths along the nanowire axis, so that most scattering events occur at the wire boundary instead of at a grain boundary, at impurity sites, or at defect sites within the nanowire. The MR of 400-nm Bi nanowires synthesized by electrochemical deposition has been investigated [139], and no peak in the longitudinal MR is observed. The absence of a magnetoresistance peak may be attributed to a higher defect level in the nanowires produced electrochemically and to a large wire diameter, much longer than the carrier mean free path. The negative MR observed for the Bi nanowire arrays above B_m (see Fig. 15.42) shows that wire boundary scattering is a dominant scattering process for the longitudinal magnetoresistance, thereby establishing that the mean free path is larger than the wire diameter and that a ballistic transport behavior is indeed observed in the high field regime.

In addition to the longitudinal magnetoresistance measurements, transverse magnetoresistance measurements ($\boldsymbol{B}$ perpendicular to the wire axis) have also been performed on Bi nanowires array samples [125, 129, 138], where a monotonically increasing B^2 dependence over the entire range $0 \le B \le 5.5$ T is found for all Bi nanowires studied thus far. This is as expected, since the wire boundary scattering cannot be reduced by a magnetic field perpendicular to the wire axis. The transverse magnetoresistance is also found always to be larger than the longitudinal magnetoresistance in nanowire arrays.

By applying a magnetic field to nanowires at very low temperatures (≤ 5K), one can induce a transition from a 1-D confined system at low magnetic fields to a 3-D confined system as the field strength increases, as shown in Fig. 15.43 for the longitudinal MR of Bi nanowire arrays of various nanowire diameters (28–70 nm) for $T <$ 5K [129]. In these curves, a subtle step-like feature is seen at low magnetic fields, which is found to depend only on the wire diameter, and is independent of temperature, the orientation of the magnetic field, and even on the nanowire material (e.g., in Sb nanowires [134]). The lack of a dependence of the magnetic field at which the step appears on temperature, field orientation, and material type indicates that the phenomenon is related to the magnetic field length, $L_H = (\hbar/eB)^{1/2}$. The characteristic length L_H is the spatial extent of the wave function of electrons in the lowest Landau level, and L_H is independent of the carrier effective masses. Setting $L_H(B_c)$ equal to the diameter d_W of the nanowire defines a critical magnetic field strength, B_c, below which the wave function is confined by the nanowire boundary (the 1-D regime), and above which the wave function is confined by the magnetic field (the 3-D regime). The physical basis for this phenomenon is associated with confinement of a single magnetic flux quantum within the nanowire cross section [129]. This phenomenon, though independent of temperature, is observed for $T \leq 5$K, since the phase breaking length has to be larger than the wire diameter. This calculated field strength, B_c, indicated in Fig. 15.43 by vertical lines for the appropriate nanowire diameters, provides a good fit to the step-like features in these MR curves.

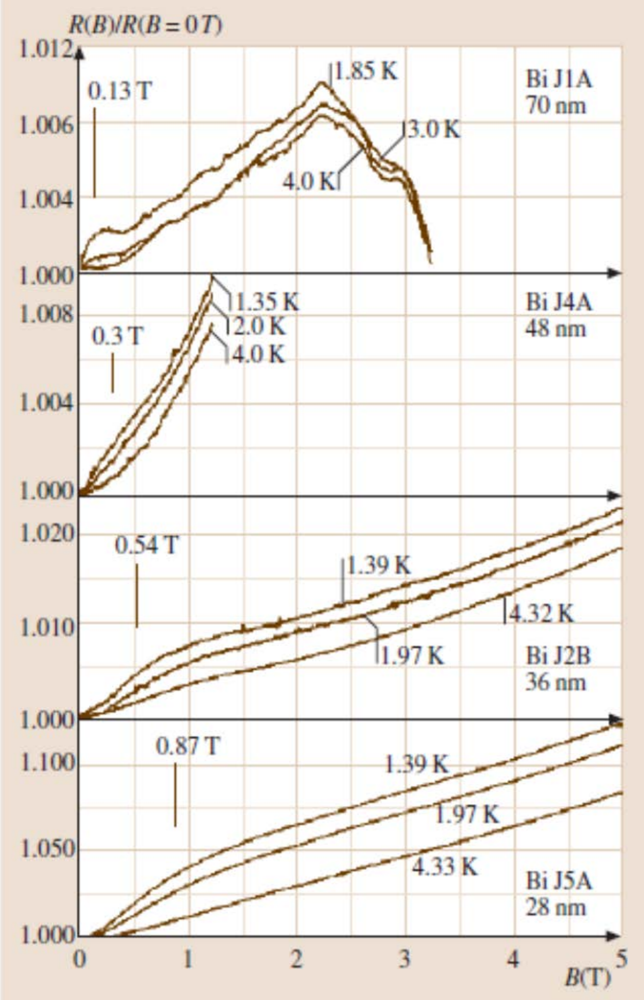


Fig. 15.43. Longitudinal magnetoresistance as a function of magnetic field for Bi nanowires of the diameters indicated. The vertical bars indicate the critical magnetic field B_c at which the magnetic length equals the nanowire diameter [129].

The Shubnikov–de Haas (SdH) quantum oscillatory effect, which results from the passage of the quantized Landau levels through the Fermi energy as the field strength varies, should, in principle, provide the most direct measurement of the Fermi energy and carrier density. For example, it has been demonstrated that SdH oscillations can be observed in Bi nanowire samples with diameters down to 200 nm [140] and have demonstrated that Te doping can be used to raise the Fermi energy in Bi nanowires. Such information on the Fermi energy is important because, for certain applications based on nanowires, it is necessary to place the Fermi energy near a subband edge where the density of states has a sharp feature. But due to the unusual 1-D geometry for nanowires, other characterization techniques commonly used in bulk materials to determine the Fermi energy and the carrier concentration (e.g., the Hall measurement) cannot be applied for nanowire systems. The observation of the SdH oscillatory effect requires very high crystal quality samples that allow carriers to execute a complete cyclotron orbit in the nanowire before they are scattered. For small nanowire diameters, large magnetic fields are required to produce cyclotron radii smaller than the wire radius. For some nanowire systems, all Landau levels may have passed through the Fermi level at such high field strength, and in such a case, no oscillations can be observed. The localization effect may also prevent the observation of SdH oscillations for very small diameter (≤ 10 nm) nanowires. Observing SdH oscillations in highly doped samples (as may be required for certain applications) may be difficult because impurity scattering reduces the mean free path, requiring high ***B*** fields to satisfy the requirement that carriers complete a cyclotron orbit prior to scattering. Although SdH oscillations provide the most direct method of measuring the Fermi energy and carrier density of nanowire samples, this technique may not work, however, for smaller diameter nanowires, or for heavily doped nanowires.

Thermoelectric Properties.

Nanowires are predicted to be promising for thermoelectric applications [124, 141], due to their novel band structure compared to their bulk counterparts and the expected reduction in thermal conductivity associated with enhanced boundary scattering. Due to the sharp density of states at the 1-D subband edges (where the van Hove singularities occur), nanowires are expected to exhibit enhanced Seebeck coefficients compared to their bulk counterparts. Since the Seebeck coefficient measurement is independent of the number of nanowires contributing to the signal, the measurements on nanowire arrays of uniform wire diameter are, in principle, as informative as single-wire measurements. The major challenge in measuring the Seebeck coefficient of nanowires lies in the design of tiny temperature probes to determine accurately the temperature difference across the nanowire. Figure 15.44a shows the schematic experimental setup for the Seebeck coefficient measurement of nanowire arrays [142], where two thermocouples are placed on both faces of a nanowire array and a heater is attached on one face of the array to generate a temperature gradient along the nanowire axis. Ideally the size of the thermocouples should be much smaller than the thickness of the nanowire array template (i. e., the nanowire length) to minimize error. However, due to the thinness of most templates (≤ 50μm) and the large size of commercially available thermocouples (~ 12μm), the measured Seebeck coefficient values are usually underestimated.

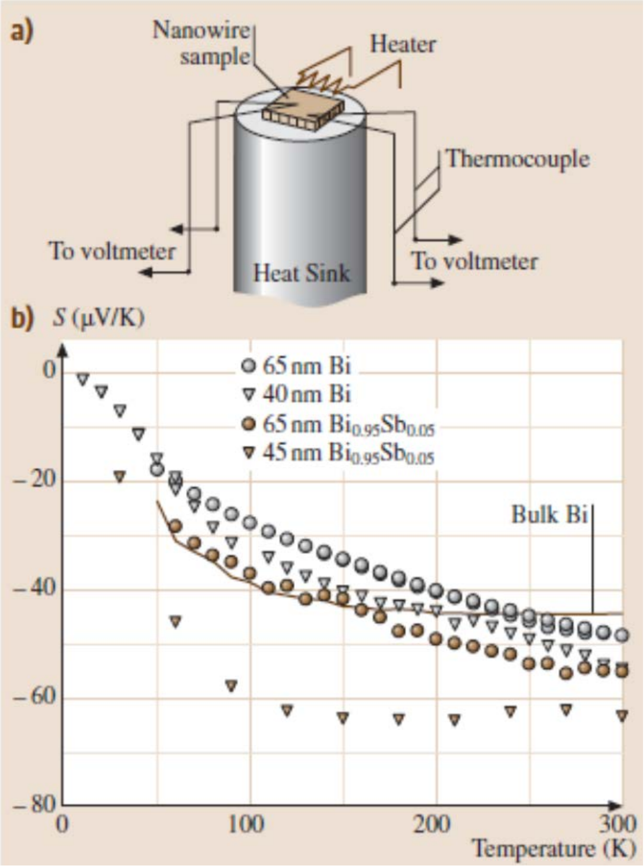


Fig. 15.44. (a) Experimental setup for the measurement of the Seebeck coefficient in nanowire arrays [142]. (b) Measured Seebeck coefficient as a function of temperature for Bi and $Bi_{0.95}Sb_{0.05}$ nanowires with different diameters. The *solid curve* denotes the Seebeck coefficient for bulk Bi [131].

The thermoelectric properties of Bi nanowire systems have been investigated extensively because of their potential as good thermoelectric materials. Figure 15.44b shows the measured Seebeck coefficient *S(T)* as a function of temperature for nanowire arrays with diameters of 40 and 65 nm and different isoelectronic Sb alloy concentrations [131]; *S(T)* results for bulk Bi are shown (*solid curve*) for comparison. Enhancement in the thermopower is observed in Fig. 15.44b as the wire diameter decreases and as the Sb content increases, which is attributed to the semimetal–semiconductor transition induced by quantum confinement and to Sb alloying effects in $Bi_{1-x}Sb_x$ nanowires. It has been observed a substantial increase in the thermopower of Bi nanowires as the wire diameter further decreases, as shown in Fig. 15.45a for 15 nm Bi/silica and 9 nm Bi/alumina nanocomposites [137]. The enhancement is due to the sharp density of states near the Fermi energy in a 1-D system. Although the samples in Fig. 15.45a also possess very high electrical resistance (~ GΩ), the results for the 9 nm Bi/alumina samples show that the Seebeck coefficient can be enhanced by almost 1,000 times relative to bulk material. But for Bi nanowires with very small diameters (~ 4 nm), the localization effect becomes dominant, which compromises the thermopower enhancement. Therefore, for Bi nanowires, the optimal wire diameter range for the largest thermopower enhancement is found to be between 4 to 15 nm [137].

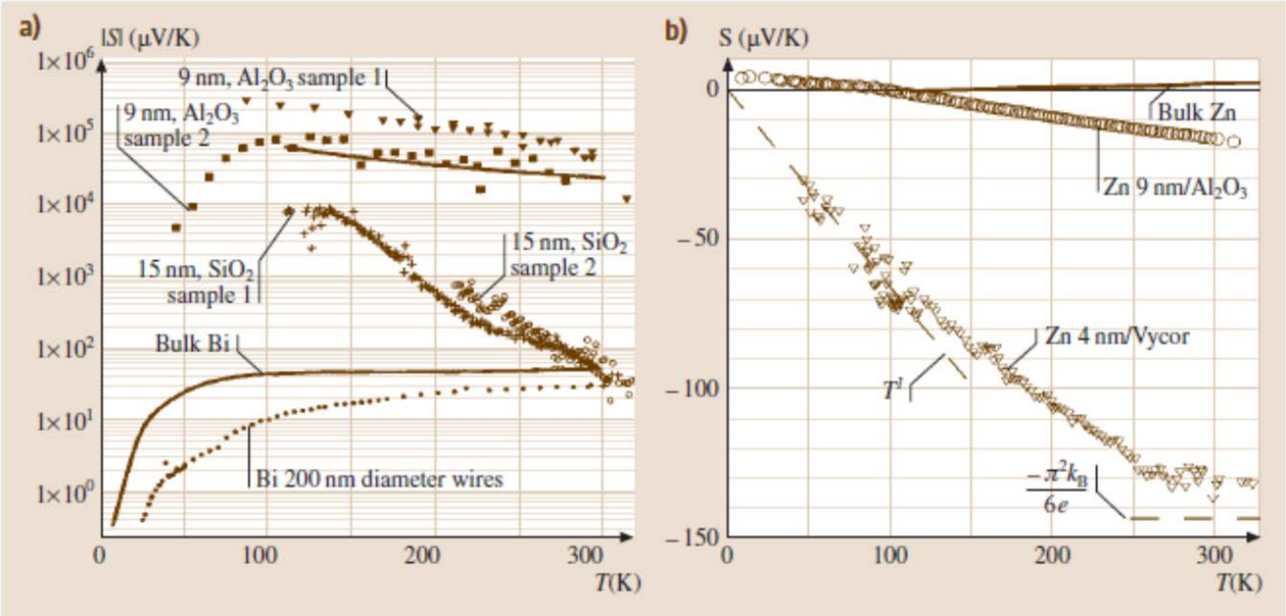


Fig. 15.45. (a) Absolute value of the Seebeck coefficient of two 15 nm Bi/silica samples, and two 9 nm Bi/alumina nanocomposite samples, in comparison to bulk Bi and 200 nm Bi nanowires in the pores of alumina templates [137]. The *full line* on top of the figure is a fit to a T^{-1} law. The Seebeck coefficient of the 9 nm Bi/alumina composite is positive; the rest are negative. (b) The Seebeck coefficient of 9 nm Zn/Al_2O_3 and 4 nm Zn/Vycor glass nanocomposite samples in comparison to bulk Zn [137].

The effect of the nanowire diameter on the thermopower of nanowires has also been observed in Zn nanowires [137]. Figure 15.45b shows the Seebeck coefficient of 9 nm Zn/alumina and 4 nm Zn/Vycor glass nanocomposites, exhibiting enhanced thermopower as the wire diameter decreases. It is found that while 9 nm Zn nanowires still exhibit metallic behavior, the thermopower of 4 nm Zn nanowires shows a different temperature dependence, which may be due to the 1-D localization effect, although further investigation is required for a definitive identification of the conduction mechanism in such small nanowires.

Quantum Wire Superlattices

The studies on superlattice nanowires, which possess a periodic modulation in their materials composition along the wire axis, have attracted much attention recently because of their promise in such applications as thermoelectrics [143], nanobarcodes [101], nanolasers [144], one-dimensional waveguides, and resonant tunneling diodes [145, 146]. Figure 15.46a shows a schematic structure of a superlattice nanowire consisting of interlaced quantum dots of two different materials, as denoted by A and B. Various techniques have been developed to synthesize superlattice nanowire structures with different interface conditions.

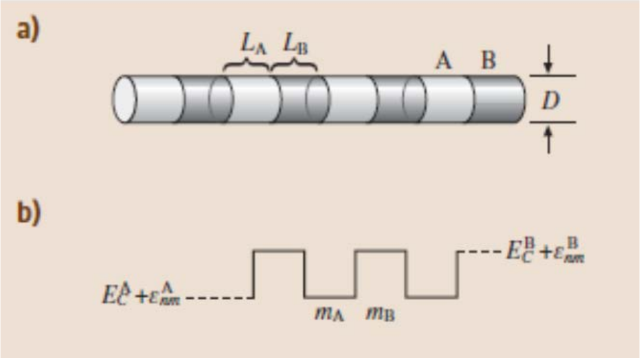


Fig. 15.46. (a) Schematic diagram of superlattice (segmented) nanowires consisting of interlaced nanodots A and B of the indicated length and wire diameter. (b) Schematic potential profile of the subbands in the superlattice nanowire [143].

In this superlattice (SL) nanowire structure, the electronic transport along the wire axis is made possible by the tunneling between adjacent quantum dots, while the uniqueness of each quantum dot and its 0-D characteristic behavior is maintained by the energy difference of the conduction or valence bands between quantum dots of different materials (see Fig. 15.46b), which provides some amount of quantum confinement. Recently it has been have observed interesting nonlinear I–V characteristics with a negative differential resistance in one-dimensional heterogeneous structures made of InAs and InP, where InP serves as the potential barrier [145, 146]. The nonlinear I–V behavior is associated with the double barrier resonant tunneling in one-dimensional structures, demonstrating the capability of transport phenomena in superlattice nanowires via tunneling and the possibility of controlling the electronic band structure of the SL nanowires by carefully selecting the constituent materials. This kind of new structure is especially attractive for thermoelectric applications, because the interfaces between the nanodots can reduce the lattice thermal conductivity by blocking the phonon conduction along the wire axis, while electrical conduction may be sustained and even benefit from the unusual electronic band structures due to the periodic potential perturbation. For example, Fig. 15.47 shows the calculated dimensionless thermoelectric figure of merit $ZT = S^2\sigma T/\kappa$ where κ is the total thermal conductivity (including both the lattice and electronic contribution) of 10-nm diameter PbS/PbSe superlattice nanowires as a function of the segment length. A higher thermoelectric performance than for $PbSe_{0.5}S_{0.5}$ alloy nanowires can be achieved for a 10 nmdiameter superlattice nanowire with segment lengths ≤ 7 nm. But the localization effect, which may become important for very short segment lengths, may jeopardize this enhancement in the *ZT* of superlattice nanowires [133].

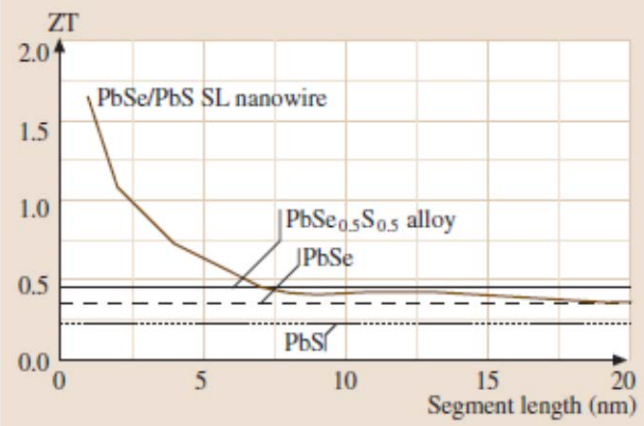


Fig. 15.47. Optimal *ZT* calculated as a function of segment length for 10-nm diameter PbSe/PbS nanowires at 77 K, where "optimal" refers to the placement of the Fermi level to optimize *ZT*. The optimal *ZT* for 10-nm diameter PbSe, PbS, and PbSe0.5S0.5 nanowires are 0.33, 0.22, and 0.48, respectively [133].

Thermal Conductivity of Nanowires

Experimental measurements of the temperature dependence of the thermal conductivity $\kappa(T)$ of individual suspended nanowires have been carried out to study the dependence of $\kappa(T)$ on the wire diameter. In this context, measurements have been made on nanowires down to only 22 nm in diameter [147]. Such measurements are very challenging and are nowpossible because of technological developments in fabricating and using nanometer size thermal scanning probes [108, 148, 149]. The experiments show that the thermal conductivity of small homogeneous nanowires may be more than one order of magnitude smaller than in the bulk, arising mainly from strong boundary scattering effects [150]. Phonon confinement effects may eventually become important at still smaller diameter nanowires. Measurements on mats of nanowires (see, for example, Fig. 15.25) do not generally give reliable results because the contact thermal resistance between adjacent nanowires tends to be high, which is in part due to the thin surface oxide coating that most nanowires have. This surface oxide coating may also be important for thermal conductivity measurements on individual suspended nanowires because of the relative importance of phonon scattering at the lateral walls of the nanowire.

The most extensive experimental thermal conductivity measurements have been done on Si nanowires [147] where $\kappa(T)$ measurements have been made on nanowires in the diameter range of $22 \leq d_W \leq 115$ nm. The results show a large decrease in the peak of $\kappa(T)$ associated with umklapp processes as d_W decreases, indicating a growing importance of boundary scattering and a corresponding decreasing importance of phonon–phonon scattering. At the smallest wire diameter of 22 nm, a linear $\kappa(T)$ dependence is found experimentally, consistent with a linear T dependence of the specific heat for a 1-D system and a temperature independent mean free path and velocity of sound.

Model calculations for $\kappa(T)$ based on a radiative heat transfer model have been carried out for Si nanowires [151]. These results show that the predicted $\kappa(T)$ behavior for Si nanowires is similar to that observed experimentally in the range of $37 \leq d_W \leq 115$ nm regarding both the functional form of $\kappa(T)$ and the magnitude of the relative decrease in the maximum thermal conductivity κ_{max} as a function of d_W. But the model calculations predict a substantially larger magnitude for $\kappa(T)$ (by 50% or more) than is observed experimentally. Furthermore, the model calculations (see Fig. 15.48) do not reproduce the experimentally observed linear T dependence for the 22 nm nanowires, but rather predict a 3-D behavior for both the density of states and the specific heat in 22 nm nanowires [151, 152].

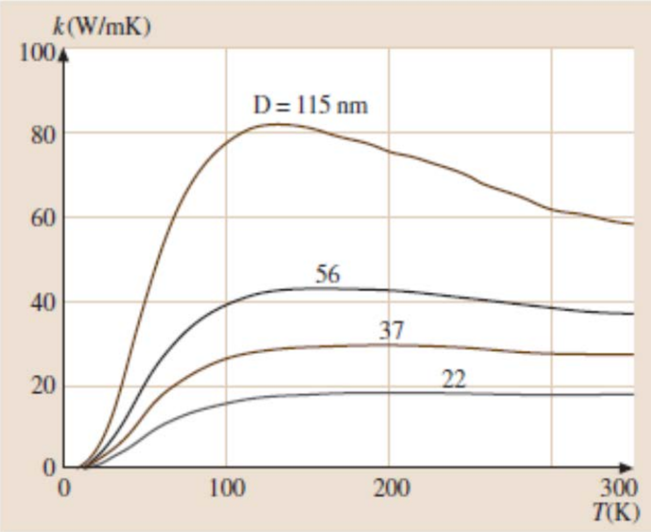


Fig. 15.48. Predicted thermal conductivity of Si nanowires of various diameters [151].

Thermal conductance measurements on GaAs nanowires below 6 K show a power law dependence, but the T dependence becomes somewhat less pronounced below ~ 2.5 K [148]. This deviation from the power lawtemperature dependence led to a more detailed study of the quantum limit for the thermal conductance. To carry out these more detailed experiments, a mesoscopic phonon resonator and waveguide device were constructed that included four~ 200 nm wide and 85 nm thick silicon nitride nanowire-like nano-constrictions (see Fig. 15.49a) to establish the quantized thermal conductance limit of $g_0 = \pi^2 k^2_B T/3h$ (see Fig. 15.49b) for ballistic phonon transport [153, 154]. For temperatures above 0.8K, the thermal conductance in Fig. 15.49b follows a T^3 law, but as

T is further reduced, a transition to a linear T dependence is observed, consistent with a phonon mean free path of ~ 1 μm, and a thermal conductance value approaching $16g_0$, corresponding to four massless phonon modes per channel and four channels in their phonon waveguide structure (see Fig. 15.49a). Ballistic phonon transport occurs when the thermal phonon wavelength (380 nm for the experimental structure) is somewhat greater than the width of the phonon waveguide at its constriction.

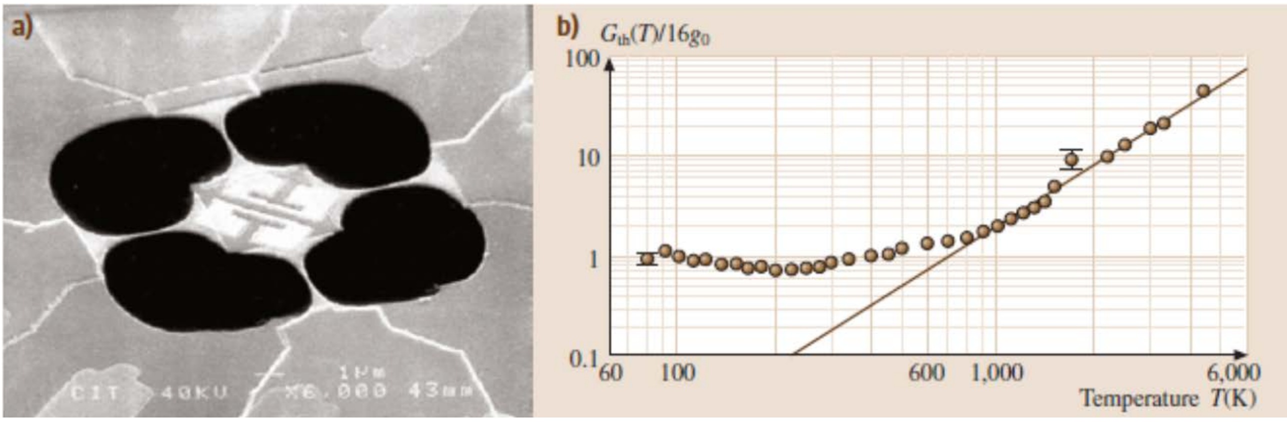


Fig. 15.49. (a) Suspended mesoscopic phonon device used to measure ballistic phonon transport. The device consists of an 4×4 μm "phonon cavity" (*center*) connected to four Si_3N_4 membranes, 60 nm thick and less than 200 nm wide. The two bright "C" shaped objects on the phonon cavity are thin film heating and sensing Cr/Au resistors, whereas the *dark regions* are empty space. (b) Log–log plot of the temperature dependence of the thermal conductance G_0 of the structure in (a) normalized to $16g_0$ (see text) [154].

15.3.3. Optical Properties

Optical methods provide an easy and sensitive tool for measuring the electronic structure of nanowires since optical measurements require minimal sample preparation (for example, contacts are not required) and the measurements are sensitive to quantum effects. Optical spectra of 1-D systems, such as carbon nanotubes, often show intense features at specific energies near singularities in the joint density of states formed under strong quantum confinement conditions. A variety of optical techniques have shown that the properties of nanowires are different from those of their bulk counterparts.

Although optical properties have been shown to provide an extremely important tool for the characterization of nanowires, the interpretation of these measurements is not always straightforward. The wavelength of light used to probe the sample is usually smaller than the wire length but larger than the wire diameter. Hence, the probe light used in an optical measurement cannot be focused solely onto the wire, and the wire and the substrate on which the wire rests (or host material, if the wires are embedded in a template) are simultaneously probed. For measurements, such as photo-luminescence (PL), if the substrate does not luminescence or absorb in the frequency range of the measurements, PL directly measures the luminescence of the nanowires and the substrate can be ignored. In reflection and transmission measurements, however, even a nonabsorbing substrate can modify the measured spectra of nanowires.

The Dielectric Function

Let's review the use of effective medium theory as a method to handle the optical properties of nanowires whose diameters are typically smaller than the wavelength of light, noting that observable optical properties of materials can be related to the complex dielectric function [155, 156]. Effective medium theories [157, 158] can be applied to model the nanowire and substrate as one continuous composite with a single complex dielectric function $(\varepsilon_1+i\varepsilon_2)$, where the real and imaginary parts of the dielectric function ε_1 and ε_2 are related to the index of refraction (n) and the absorption coefficient (K) by the relation $\varepsilon_1+i\varepsilon_2 = (n+iK)^2$. Since photons at visible or infrared wavelengths "see" a dielectric function for the composite nanowire array/substrate system that is different from that of the nanowire itself, the optical transmission and reflection are different from what they would be if the light were focused only on the nanowire. One commonly observed consequence of effective medium theory is the shift in the plasma frequency in accordance with the percentage of nanowire material contained in the composite [159]. The plasma resonance occurs when $\varepsilon_1(\omega)$ becomes zero, and the plasma frequency of the nanowire composite will shift to lower (higher) energies when the magnitude of the dielectric function of the host materials is larger (smaller) than that of the nanowire.

Although reflection and transmission measurements probe both the nanowire and the substrate, the optical properties of the nanowires can be independently determined. One technique for separating the dielectric function of the nanowires from the host is to use an effective medium theory in reverse. Since the dielectric function of the host material is often known, and since the dielectric function of the composite material can be measured by the standard method of using reflection and transmission measurements in combination with either the Kramer–Kronig relations or Maxwell's equations, the complex dielectric function of the nanowires can be deduced. This approach has been used successfully, for example, in determining the frequency dependence of the real and imaginary parts of the dielectric function $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ for a parallel array of bismuth nanowires filling the pores of an alumina template [160].

Optical Properties Characteristic of Nanowires

A wide range of optical techniques are available for the characterization of nanowires to distinguish their properties from those of their parent bulk materials. Some differences in properties are geometric, such as the small diameter size and the large length-to-diameter ratio (also called the aspect ratio), while other differences focus on quantum confinement issues.

Probably the most basic optical technique is to measure the reflection and/or transmission of a nanowire to determine the frequency-dependent real and imaginary parts of the dielectric function. This technique has been used, for example, to study the band gap and its temperature dependence in gallium nitride nanowires in the 10–50 nm range in comparison to bulk values [161]. The plasma frequency, free carrier density, and donor impurity concentration as a function of temperature were also determined from the infrared spectra, which is especially useful for nanowire research since Hall effect measurements cannot be made on nanowires.

Photo-luminescence (PL) or fluorescence spectroscopy is a common method to study nanowires. Emission techniques probe the nanowires directly, and the effect of the host material does not have to be considered. This characterization method has been used to study many properties of nanowires, such as the optical gap behavior, oxygen vacancies in ZnO nanowires [162], strain in Si nanowires [163], and quantum confinement effects in InP nanowires [144]. Figure 15.50 shows the photo-luminescence of InP nanowires as a function of wire diameter, thereby providing direct information on the effective band gap. As the wire diameter of an InP nanowire is decreased so that it becomes smaller than the bulk exciton diameter of 19 nm, quantum confinement effects set in, and the band gap is increased. This results in an increase in the PL peak energy because of the stronger electron–hole Coulomb binding energy within the quantum-confined nanowires as the wire radius gets smaller than the effective Bohr radius for the exciton for bulk InP. The smaller the effective mass, the larger the quantum confinement effects are. When the shift in the peak energy as a function of nanowire diameter (Fig. 15.50a) is analyzed using an effective mass model, the reduced effective mass of the exciton is deduced to be $0.052m_0$, which agrees quite well with the literature value of $0.065m_0$ for bulk InP. Although the line widths of the PL peak for the small diameter nanowires (10 nm) are smaller at low temperature (7 K), the observation of strong quantum confinement and band gap tunability effects at room temperature are significant for photonics applications of nanowires.

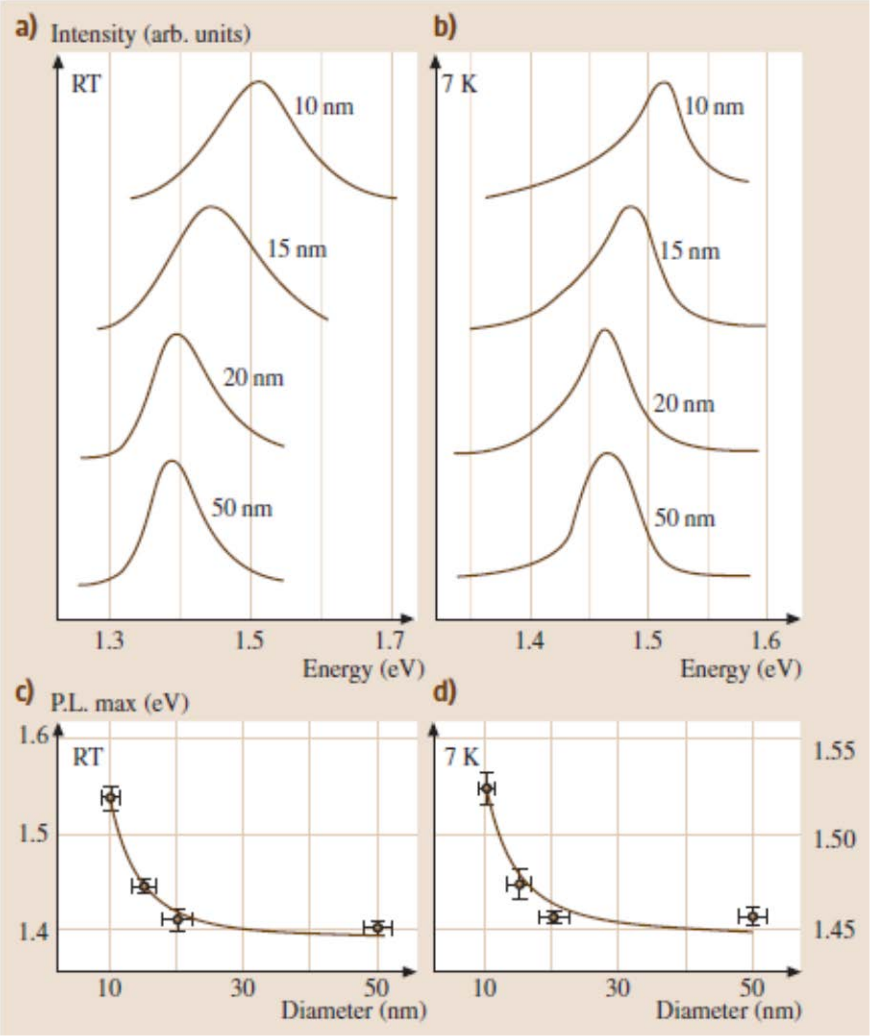


Fig. 15.50. Photo-luminescence of InP nanowires of varying diameters at 7K ((b) and (d)) and room temperature ((a) and (c)), showing quantum confinement effects of the exciton for wire diameters less than 20 nm [144].

The resolution of photo-luminescence (PL) optical imaging of a nanowire is, in general, limited by the wavelength of light. But when a sample is placed very close to the detector, the light is not given a chance to diffract, and so samples much smaller than the wavelength of light can be resolved. This technique is known as near-field scanning optical microscopy (NSOM) and has been used successfully [164] to image nanowires. For example, Fig. 15.51 shows the topographical (a) and PL (b) NSOM images of a single ZnO nanowire.

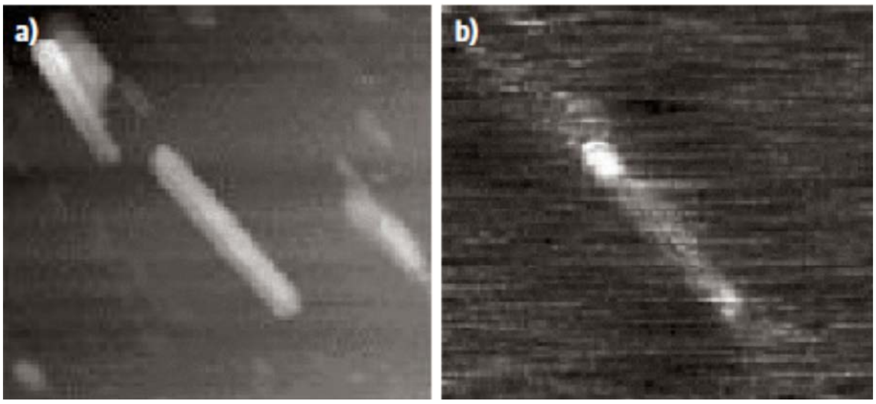


Fig. 15.51. (a) Topographical and (b) photoluminescence (PL) nearfield scanning optical microscopy (NSOM) images of a single ZnO nanowire waveguide [164].

Magneto-optics can be used to measure the electronic band structure of nanowires. For example, magneto-optics in conjunction with photo-conductance has been proposed [165] as a tool to determine band parameters for nanowires, such as the Fermi energy, electron effective masses, and the number of subbands to be considered [165]. Since different nanowire subbands have different electrical transmission properties, the electrical conductivity changes when light is used to excite electrons to higher subbands, thereby providing a method for studying the electronic structure of nanowires optically. Magneto-optics can also be used to study the magnetic properties of nanowires in relation to bulk properties [166, 167]. For example, the surface magneto-optical Kerr effect has been used to measure the dependence of the magnetic ordering temperature of Fe-Co alloy nanowires on the relative concentration of Fe and Co [167], and it was used to find that, unlike the case of bulk Fe-Co alloys, cobalt in nanowires inhibits magnetic ordering.

Nonlinear optical properties of nanowires have received particular attention since the nonlinear behavior is often enhanced over that in bulk materials and since the nonlinear effects can be utilized for many applications. One such study measured the second harmonic generation (SHG) and third harmonic generation (THG) in a single nanowire by using near-field optical microscopy [168]. ZnO nanowires were shown to have strong SHG and THG effects that are highly polarization-sensitive, and this polarization sensitivity can be explained on the basis of optical and geometrical considerations. Some components of the second harmonic polarization tensor are found to be enhanced in nanowires while others are suppressed as the wire diameter is decreased, and such effects could be of interest for device applications. The authors also showed that the second-order nonlinearities are mostly wavelength independent for $\lambda < 400$ nm, which is in the transparent regime for ZnO, below the onset of band gap absorption; this observation is also of interest for device applications.

Reflectivity and transmission measurements have also been used to study the effects of quantum confinement and surface effects on the low energy indirect transition in bismuth nanowires [169]. An intense and sharp absorption peak in bismuth nanowires, which is not observed in bulk bismuth has been investigated in [169]. The energy position E_p of this strong absorption peak increases with decreasing diameter. But the rate of increase in energy with decreasing diameter $|\partial E_p/\partial d_W|$ is an order of magnitude less than that predicted for either a direct interband transition or for intersubband transitions in bismuth nanowires. On the other hand, the magnitude of $|\partial E_p/\partial d_W|$ agrees well with that predicted for an indirect *L*-point valence to *T*-point valence band transition (see Fig. 15.52). Since both the initial and final states for the indirect *L*-*T* point valence band transition downshift in energy as the wire diameter d_W is decreased, the shift in the absorption peak results from a *difference* between the effective masses and not from the actual value of either of the masses. Hence the diameter dependence of the absorption peak energy is an order of magnitude less for a valence to valence band indirect transition than for a direct interband *L*-point transition.

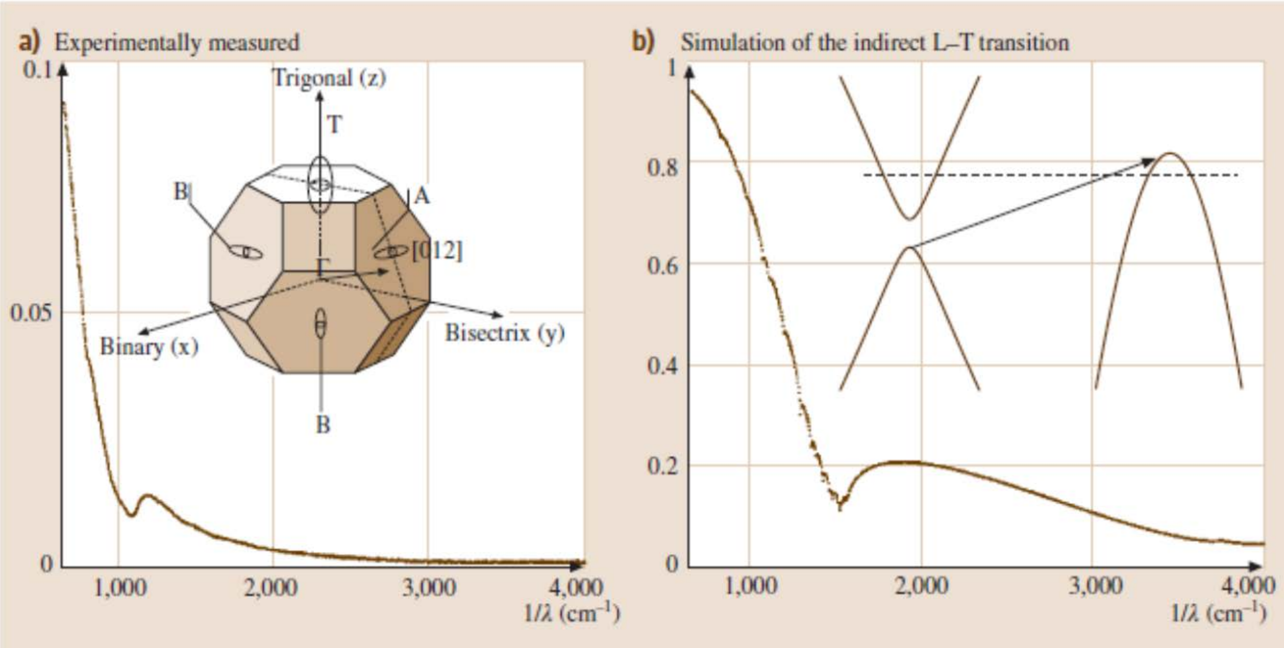


Fig. 15.52. (a) The measured optical transmission spectra as a function of wavenumber ($1/\lambda$) of a ~ 45 nm diameter bismuth nanowire array. (b) The simulated optical transmission spectrum resulting from an indirect transition of an *L* point electron to a *T* point valence subband state. The insert in (a) shows the bismuth Brillouin zone, and the location of the *T*-point hole and of the three *L*-point electron pockets, including the nondegenerate A, and the doubly degenerate B pockets. The insert in (b) shows the indirect *L* to *T* point electronic transition induced by a photon with an energy equal to the energy difference between the initial and final states minus the phonon energy (about 100 cm^{-1}) needed to satisfy conservation of energy in a Stokes process [169].

Furthermore, the band-tracking effect for the indirect transition gives rise to a large value for the joint density of states, thus accounting for the high intensity of this feature. The enhanced coupling of this indirect transition to an applied optical field arises through the gradient of the dielectric function, which is large at the bismuth-air or bismuth-alumina interfaces. It should be noted that, in contrast to the surface effect for bulk samples, the whole nanowire contributes to the optical absorption due to the spatial variation in the dielectric function, since the penetration depth is larger than or comparable to the wire diameter. In addition, the intensity can be quite significant because abundant initial state electrons, final state holes, and appropriate phonons exist for making an indirect *L*-*T* point valence band transition at room temperature. Interestingly, the polarization dependence of this absorption peak is such that the strong absorption is present when the electric field is perpendicular to the wire axis but is absent when the electric field is parallel to the wire axis; this is contrary to a traditional polarizer, such as a carbon nanotube in which the optical *E* field is polarized by the nanotube itself to be aligned along the carbon nanotube axis. The observed polarization dependence for bismuth nanowires is consistent with a surface-induced effect that increases the coupling between the *L*-point and *T*-point bands throughout the full volume of the nanowire. Figure 15.52 shows the experimentally observed transmission spectrum in bismuth nanowires of ~ 45 nm diameter (a), and for comparison, the simulated optical transmission from an indirect transition in bismuth nanowires of ~ 45 nm diameter is also shown in (b). The indirect *L* to *T* point valence band transition mechanism [169] is also consistent with observations of the effect on the optical spectra of a decrease in the nanowire diameter and of n-type doping of bismuth nanowires with Te.

Phonon Confinement Effects

Phonons in nanowires are spatially confined by the nanowire cross-sectional area, crystalline boundaries and surface disorder. These finite size effects give rise to phonon confinement, causing an uncertainty in the phonon wave vector, which typically gives rise to a frequency shift and a lineshape broadening. Since zone center phonons tend to correspond to maxima in the phonon dispersion curves, the inclusion of contributions from a broader range of phonon wave vectors results in both a downshift in frequency and an asymmetric broadening of the Raman line that developes a low frequency tail. These phonon confinement effects have been theoretically predicted [170, 171] and experimentally observed in GaN [172], as shown in Fig. 15.53 for GaN nanowires with diameters in the range 10–50 nm. The application of these theoretical models indicates that broadening effects should be noticeable as the wire diameter in GaN nanowires decreases to ~ 20 nm. When the wire diameter further decreases to ~ 10 nm, the frequency downshift and asymmetric

Raman line broadening effects should become observable in the Raman spectra for the GaN nanowires but are not found in the corresponding spectra for bulk GaN.

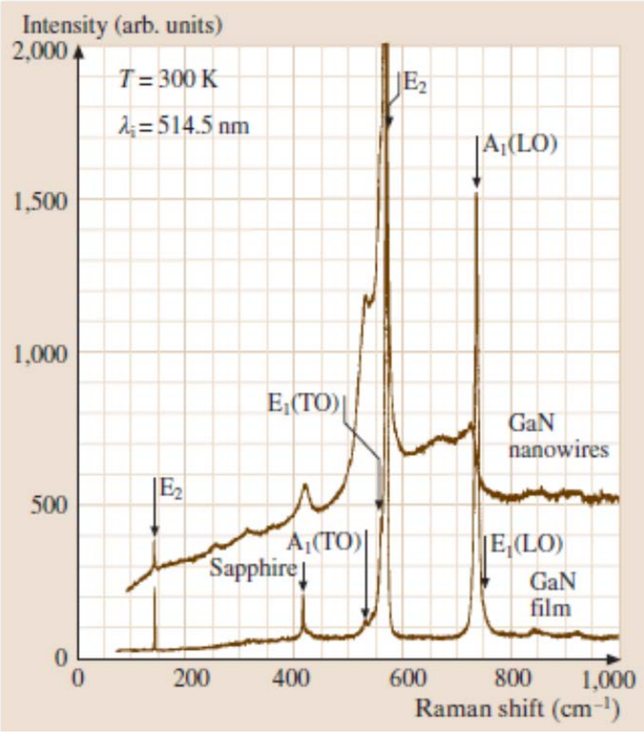


Fig. 15.53. Room temperature Raman-scattering spectra of GaN nanowires and of a 5μm thick GaN epilayer film with green (514.5 nm) laser excitation. The Raman-scattering response was obtained by dividing the measured spectra by the Bose–Einstein thermal factor [172].

The experimental spectra in Fig. 15.53 show the four $A_1+ E_1+2E_2$ modes expected from symmetry considerations for bulk GaN crystals. Two types of quantum confinement effects are observed. The first type is the observation of the downshift and asymmetric broadening effects discussed above. Observations of such downshifts and asymmetric broadening have also been recently reported in 7 nm diameter Si nanowires [173]. A second type of confinement effect found in Fig. 15.53 for GaN nanowires is the appearance of additional Raman features not found in the corresponding bulk spectra and associated with combination modes and a zone boundary mode. Resonant enhancement effects were also observed for the A_1(LO) phonon at 728 cm^{-1} (see Fig. 15.53) at higher laser excitation energies [172].

15.4. NANOWIRE ELECTRONIC DEVICE ELEMENTS

The availability of a wide range of NW materials with controlled chemical composition, physical size, and electronic properties opens up many exciting opportunities ranging from fundamental studies of the role of dimensionality on physical properties to a range of potential applications in areas such as nanoscale electronics and optoelectronics. Let's consider the fabrication and electrical transport properties of basic nanoscale device elements fashioned from these NW building blocks, including 1) single NW-FETs, 2) crossed NW *p–n* diodes, intra-NW *p–n* diode, 3) bipolar transistors, 4) crossed NW-FETs, and 5) NW quantum interference devices [174].

15.4.1. Nanowire Field Effect Transistors

Device structure and underlying principles

The basic FET structure fabricated from single semiconducting NWs is illustrated in Fig. 15.54. The FET is supported on an oxidized silicon substrate with the underlying conducting silicon used as a global back gate electrode to vary the electrostatic potential of the NW. In a typical NW-FET device (Fig. 15.54, inset), two metal contacts, which correspond to source and drain electrodes, are defined by electron beam lithography followed by evaporation of suitable metal contacts. Current (I) vs. source–drain voltage (V_{sd}) and I vs. gate voltage (V_g) is then recorded for a NW-FET to characterize its electrical properties.

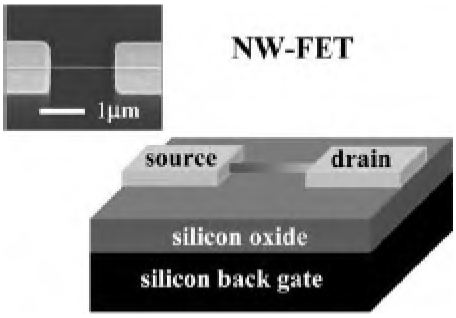


Fig. 15.54. Schematic of a NW-FET. Inset: SEM image of a NW-FET; two metal electrodes, which correspond to source and drain, are visible at the left and right sides of the image.

Variation of V_g during characterization of the NWFET enables important qualitative and quantitative properties to be elucidated. For example, changes in V_g produce variations in the electrostatic potential of the NW, and hence change the carrier concentration and conductance of the NW. As shown in Fig. 15.55, *p*- and *n*-type semiconductor NWs, which are contacted at both ends to metal electrodes, respond in opposite ways to the applied gate. When a positive V_g is applied, the bands are lowered, which depletes the holes and suppresses conductivity in *p*-NWs, but leads to an accumulation of electrons and an enhancement in conductivity in *n*-NWs. Conversely, a negative V_g will increase the conductivity of *p*-type NWs and decrease the conductivity of the *n*-type NWs.

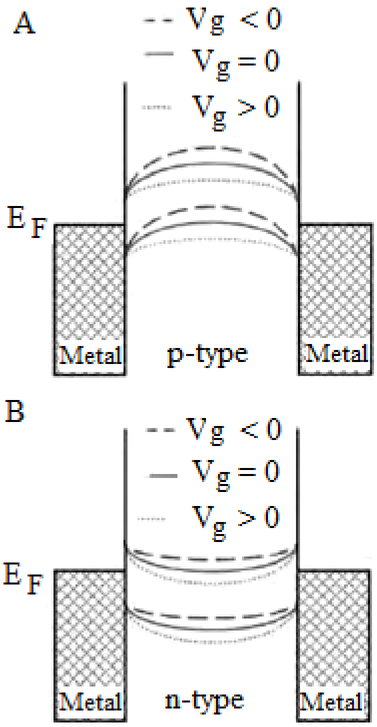


Fig. 15.55. (A,B) Band diagrams illustrating the underlying principle for *p*- and *n*-channel NW-FETs. When a positive voltage is applied, the bands are lowered, which depletes the holes in *p*-NWs and suppresses conductivity, but leads to an accumulation of electrons in *n*-NWs and enhances the conductivity. Conversely, a negative gate voltage will raise the bands and increase the conductivity of *p*-type NWs and decrease the conductivity of the n-type NWs.

p-Channel nanowire transistors

Typical *I* vs. V_{sd} data obtained from a single boron doped Si NW-FET at different V_{gs} are shown in Fig. 15.56. The two-terminal I–V_{sd} curves are linear, which indicates that the metal electrodes make ohmic contacts to the NW, and moreover, the gate response demonstrates that the

NW is *p*-type; that is, the conductance of the *p*-Si NW decreases (increases) with increasingly positive (negative) V_g. The transfer characteristics, $I–V_g$, of *p*-Si NW devices (Fig. 15.56, inset) exhibit behavior typical of *p*-channel metal-oxide–semiconductor FETs (MOSFETs) [175, 176]. Significantly, the conductance modulation of the *p*-Si NW-FET exceeds 10^3, where the V_g required for switching (–10 to 10 V) could be reduced significantly by reducing the thick (600 nm) oxide dielectric layer in these back-gated devices (see below).

Gate-dependent measurements have also been used to estimate the hole concentration in *p*-channel NW-FETs. The total NW charge can be expressed as $Q = C \times V_{th}$, where C is the NW capacitance V_{th} the threshold gate voltage required to deplete completely the NW. The capacitance is given by $C \approx 2\pi\varepsilon\varepsilon_0 L/\ln(2h/r)$, where ε is the effective gate oxide dielectric constant, h is the thickness of the SiO_2 layer on the substrate, L is the NW length, and r is the NW radius. The hole density, $n_h = Q/(e\ \pi r^2 L)$, is estimated to be $\sim 10^{18} cm^{-3}$ for the device shown in Fig. 15.56. In addition, it is possible to estimate the carrier mobility of the NW-FETs from the transconductance $dI/dV_g = \mu(C=L^2)\ V_{sd}$, where μ is the carrier mobility. Plots of dI / dVg vs. V_{sd} are linear for Si NWs, as expected for this model, yield hole mobilities in the range of 50–800 cm^2/V sec. Significantly, the *p*-Si NW-FET mobilities are comparable to or larger than the best *p*-Si planar devices, 100–300 cm^2/V sec, at comparable hole densities ($p \sim 10^{17}–10^{18}\ cm^{-3}$) [175].

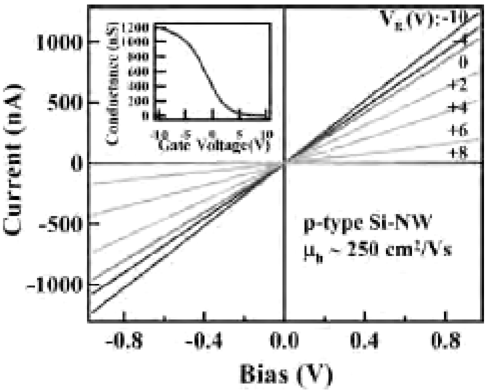


Fig. 15.56. Current vs. voltage for a *p*-type Si NW-FET. The numbers inside the plot indicate the corresponding gate voltages (V_g). The inset shows current vs. V_g for V_{sd} of 1 V.

n-Channel nanowire transistors

It is also possible to assemble *n*-channel NW-FETs in a similar way from *n*-type NWs. For example, gate-dependent $I–V_{sd}$ data recorded from an InP NW-FET exhibits increased conductance for positive V_g and decreased conductance for negative V_g (Fig. 15.57), as expected for an *n*-channel device. The *n*-InP NWFET transfer characteristics (I-V_g) show that the current increases rapidly from below 1 nA at V_g = –2V to above 400 nA at V_g = +2V (Fig. 15.57, inset), and tends to saturate at higher voltages, which can be attributed to contact resistance and other factors. Nevertheless, the conductance changes up three orders of magnitude for only a few volts' change in the gate voltage in these unoptimized devices.

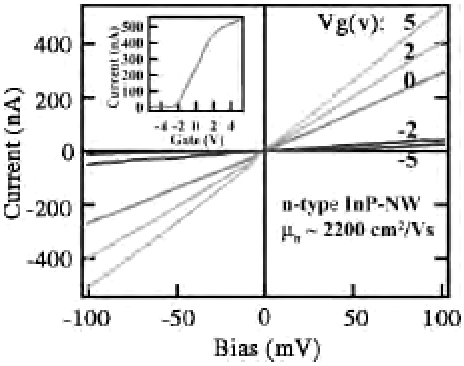


Fig. 15.57. Current vs. voltage for an n-type InP NW-FET. The numbers inside the plot indicate the corresponding gate voltages (V_g). The inset shows current vs. V_g for V_{sd} of 0.1 V.

The electron concentration and mobility in the *n*-channel NW-FETs have been estimated as described above for *p*-channel devices. For the *n*-InP NW-FET shown in Fig. 15.57, the electron mobility is 2200 cm^2/V sec for an electron concentration of ~10^{18} cm^{-3}. Studies of a number of different devices yields mobility values from 400 to 3000 cm^2/V sec, which is comparable to or larger than bulk InP, 1000–2000 cm^2/V sec, at similar carrier concentrations [175]. These mobilities are believed to represent a lower limit in our NW materials since the contact resistance and surface depletion have not been included. Significantly, surface passivation studies suggest that substantially higher carrier mobilities are possible in the NW-FETs. Taken together, these results suggest that the NW-FETs could be essential elements in high performance (e.g., high gain, high speed and low power) nanoelectronics and photonics.

15.4.2. Crossed Nanowire *p–n* Diodes

The availability of well-defined *n*- and *p*-type NW building blocks opens up the possibility of creating complex functional devices by forming junctions between two or more wires. To explore this exciting opportunity, we have studied the transport behavior of *n–n*, *p–p*, and *p–n* junctions formed by crossing two *n*-type, two *p*-type, and one *n*-type and one *p*-type NW, respectively (Fig. 15.58A) [66]. Significantly, the types of junctions studied in an experiment are reproducible since we can select the specific type of NW used at each of the two stages of device assembly.

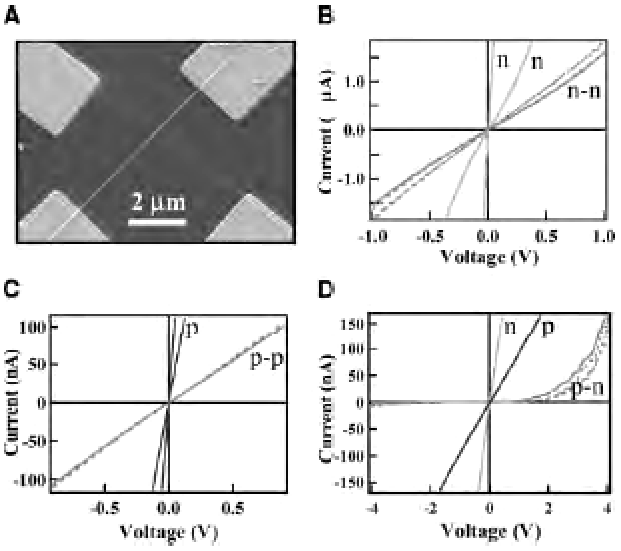


Fig. 15.58. Crossed NW junctions. (A) SEM image of a typical crossed InP NW device with four metal electrodes contacted to each of the four arms. (B–D) I–V behavior of *n–n*, *p–p*, and *p–n* junctions, respectively. The I–V behavior of individual *n*- and *p*-NWs in the junctions is indicated by ‘‘*n*’’ and ‘‘*p*’’, respectively. The I–V behavior across the junctions is designated by ‘‘*n–n*’’, ‘‘*p–p*,’’ and ‘‘*p–n*.’’ The solid lines represent transport behavior across one pair of adjacent arms, and the dashed lines represent that of the other three pairs of adjacent arms [66].

First, I–V data recorded on the individual NWs in *n–n* and *p–p* crossed junctions show linear or nearly linear I–V behavior (Fig. 15.58B,C), indicating that the metal electrodes used in the experiments make ohmic or nearly ohmic contact to the NWs. This point is important since it shows that the NW–metal contacts will not make nonlinear contributions to the I–V measurements across the nanoscale junctions. In general, transport measurements recorded across the *n–n* and *p–p* junctions show linear or nearly linear behavior. These results indicate that interface oxide between individual NWs does not produce a significant tunneling barrier since a tunneling barrier would lead to highly nonlinear I–V behavior. In addition, the I–V curves recorded through each pairs of adjacent arms show similar current levels, which are smaller than that of the individual NWs themselves, demonstrating that the junction dominates the transport behavior. Taken as a whole, these data show that individual NWs can make good electrical contact with each other, despite the small contact area (10^{-12}–10^{-10} cm^2) and simple method of junction fabrication.

Initial studies designed to probe the utility of this new approach for creating functional devices were focused on *p–n* junctions from crossed *p*- and *n*-type NWs. These junctions can be made reproducibly by sequential deposition of dilute solutions of *n*- and *p*-type NWs with intermediate drying. Typical *I–V* behavior of a crossed InP NW *p–n* junction is shown in Fig. 15.58.

The linear *I*–*V* of the individual *n*- and *p*-type NWs components indicates ohmic contact between the NWs and metal electrodes, while transport across the *p*–*n* junction shows clear current rectification; that is, little current flows in reverse bias, while there is a sharp current onset in forward bias. Significantly, this behavior is similar to conventional semiconductor *p*–*n* junctions. In a standard *p*–*n* junction, rectification arises from the potential barrier formed at the interface between *p*- and *n*-type materials [176]. In the case of our crossed NW *p*–*n* junctions, this picture is probably modified due to the presence of some interface oxide (Fig. 15.59), although a thin oxide will not change substantially the overall *I*–*V* response.

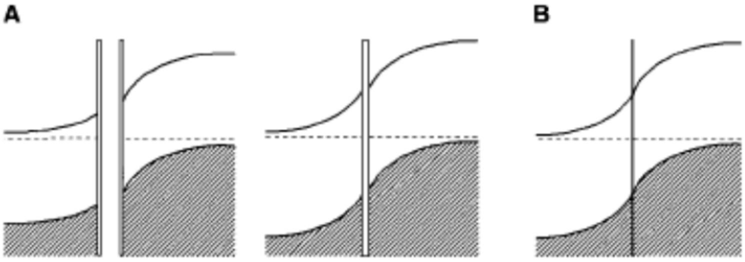


Fig. 15.59. Band diagrams of the crossed NW *p*–*n* junction. (A) Band bending occurs prior to for *p*–*n* contact due to surface Fermi level pinning; the bands bend further when the *p*- and *n*-type NWs come into contact to achieve equilibrium; (B) band bending in a clean *p*–*n* junction structure.

The assignment of the observed rectification to the *p*–*n* junction formed at the crossing point between *p* and *n*-type InP NWs was further supported by several other pieces of evidence. First, the linear or nearly linear *I*–*V* behavior of individual *p*- and *n*-type NWs shows that ohmic contacts were been made between the NWs and metal electrodes, and thus exclude the possibility that rectification arises from metal– semiconductor Schottky diodes [176]. Second, the *I*–*V* behavior of the junction determined through each pair of adjacent electrodes (Fig. 15.58D) exhibit similar rectification and current level, which is also much smaller than the current through individual NWs, demonstrating that the junction dominates the *I*–*V* behavior. Third, four-terminal measurements in which current is passed through two adjacent electrodes while the junction voltage drop is measured across the two remaining electrodes exhibit similar *I*–*V* and rectification with only a slightly smaller voltage drop (0.1–0.2 V) compared to two-terminal measurements at the same current level. Fourth, measurements made on over 20 independent InP crossed *p*–*n* junctions showed similar rectification in the *I*–*V* data. Lastly, the formation of crossed NW *p*–*n* junctions is not by any means restricted to InP NWs, and is general to the wide range of materials. For example, *p*–*n* junctions have been assembled from *p*-Si/*n*-Si [67], *p*-Si/*n*-GaN [68], *p*-Si/*n*-InP, *p*-Si/*n*-CdS, and *p*-Si/*n*-CdSe NWs

[177], and transport measurements have demonstrated that all of these crossed *p*–*n* junctions show consistent current rectification behavior.

15.4.3. Bipolar Junction Transistors

Since *p*–*n* junctions represent a basic element in many functional electronic devices, including amplifiers and switches, we have explored the possibility of assembling such devices at nanometer scale using the welldefined *p*- and *n*-type NW materials. As an example, integrated bipolar transistors [67], which are active devices capable of current gain, have been assembled from three distinct types of Si NWs in the form of two crossed junctions (Fig. 15.60A). A conventional bipolar transistor requires three distinct material types. For example, in n^+–*p*–*n* structure, a highly doped n^+ layer is used as an emitter (E), a *p*-type layer for the base (B), and an *n*-type layer for the collector (C) [176]. Significantly, this n^+–*p*–*n* basic structure can be easily assembled with Si NWs since NWs with controlled doping type and doping concentration are available (Fig. 15.60C, inset) [65].

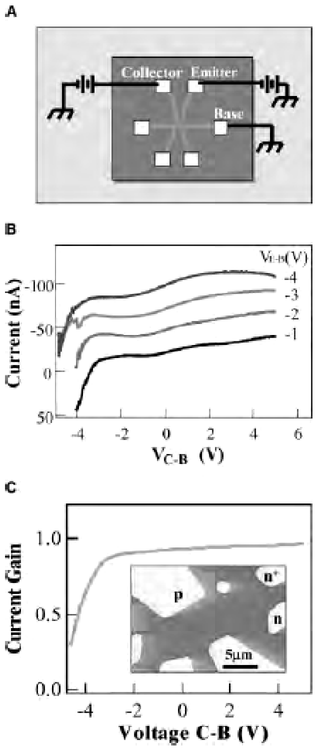


Fig. 15.60. Bipolar junction transistors. (A) Schematic illustrating the common base configuration of an n^+–*p*–*n* bipolar transistor built from crossed Si NWs. (B) Collector current vs. collector–base voltage recorded on an n^+–*p*–*n* transistor with emitter and collector Si NWs 15 μm apart. The numbers inside the plot indicate the corresponding emitter–base voltages. (C) The common base current gain vs. collector– base voltage. Inset: Typical SEM image of Si NW bipolar transistors [67].

To characterize the electrical behavior of assembled NW bipolar transistors, the individual Si NW building blocks and *p*–n^+ and *p*–*n* junctions were first tested, and found to exhibit ohmic or nearly ohmic metal contacts and rectification, respectively. The bipolar transistor characteristics were then assessed from measurements of the collector current as a function of C–B voltage (Fig. 15.60B), while the n^+ Si NW emitter was biased at different values. In general, the collector current is relatively constant (vs. C–B voltage) in the region from 0 to 6 V, which corresponds to the collector in reverse bias with only a very small leakage current, and this current value increases as the emitter forward bias/injected current is increased. The large collector current in reverse bias demonstrates these simple Si NW-based bipolar transistors exhibit behavior similar to that found in standard planar devices, and moreover, can exhibit very good current gain. The common base current gain, which is defined as the ratio of the collector current to emitter current (Fig. 15.60C), and the common emitter current gain, which is defined as the ratio of the collector current to base current, were found to be 0.94 and 16, respectively. The relatively large current gain observed in these simple devices suggests several important points. First, the efficiency of electron injection from emitter to base must be quite high, and can be attributed to the controlled NW doping that yields the desired n^+–*p* E–B junction. Second, large current gains have been achieved in devices with large (e.g., 15 μm) base widths. This fact suggests that the mobility of injected electrons can be quite high in the Si NWs and is consistent with the direct mobility studies described above. These observations also indicate clear directions for improving the Si NW bipolar transistors. For example, it will be interesting to study the current gain as a function of base width, because it is easily possible to assemble structures with separations of the n^+ and *n*-NWs of the order of 100 nm or less.

15.4.4. Crossed Nanowire Field Effect Transistors

A major motivation underlying research on nanoscale devices is to achieve integration at densities higher than possible with current technologies. The NW-FETs discussed above represent nanoscale analogs to conventional MOSFETs and have been very useful for testing basic device behavior (e.g., doping type and carrier mobility). However, the basic device structure of the NW-FETs and similar NT FETs [54, 55] requires lithography to define metallic source–drain electrodes, and use either a global back gate (i.e., the doped silicon substrate) or lithography to define a more local gate. These design and fabrication features pose serious problems for integration. First, lithographically defined metal electrodes (i.e., source, drain, and gate) will limit integration to a level similar to that of conventional silicon technology. Moreover, the use of global back gate electrodes eliminates the possibility of independently addressing individual devices, and thus is incompatible with integration in most architectures. Direct assembly of highly integrated functional electronic circuits based on NWs requires the development of new device concepts that are amenable to scalable integration. To this end, it was recently developed a novel crossed NW-based FET [68]. A crossed NW-FET (cNW-FET) is assembled from two NWs where one or both have an oxide coating that serves as the gate dielectric (Fig. 15.61A,B). This approach is quite flexible since nano-FETs can be readily assembled with *p*- or *n*-type active channel NWs and the gate NW can also be *p*- or *n*-type independent of the channel. For example, using *n*-type GaN crossed NW as the gate for a *p*-type Si NW, a *p*-channel cNW-FET is formed with both a nanoscale channel and a nanoscale gate. Typical I–V_{sd} data recorded for different NW V_gs resemble the characteristics of a conventional depletion mode *p*-channel FET device (Fig. 15.61C). Notably, the conductance of the Si NW responds very sensitively to the voltage applied to *n*-NW gate, and can be changed by more than five orders of magnitude with a 1–2V variation in the NW gate (Fig. 15.61D). In contrast, the conductance changed less than a factor of 10 for this same device when similar gate voltages were applied to the global back gate (Fig. 15.61D). The improved sensitivity can be attributed to the intrinsically thin gate dielectric between the two NWs. In addition, there is no leakage current from the *n*-NW gate when the cNW-FET is operated at low source– drain bias in the deletion mode since the crossed *p*–*n* junction is always reverse biased, and in this regard the device is similar to a junction FETs (JFETs) [176].

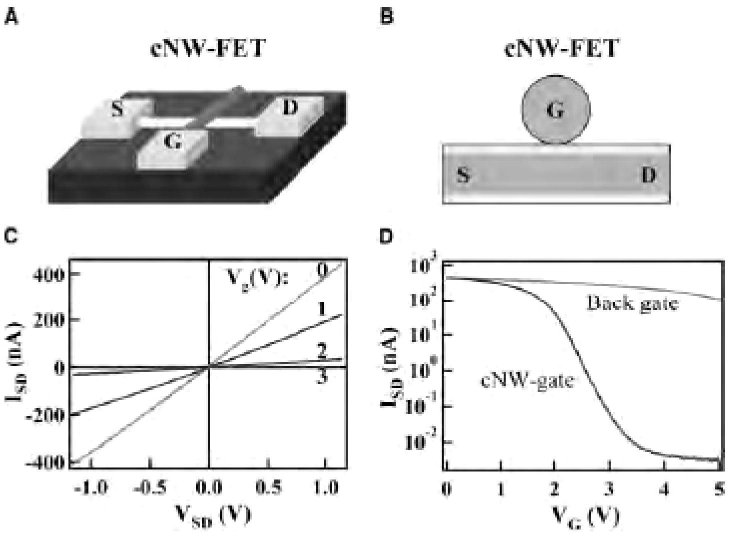


Fig. 15.61. Crossed NW-FET (cNW-FET). (A) Schematics illustrating the cNW-FET concept. A nano-FET with both nanoscale conducting channel and nanoscale gate is obtained with one NW used as the gate for the other NW in a crossed configuration. (B) Schematic showing the critical device dimensions of the cNW-FET. Three intrinsic nanometer scale metrics are naturally defined

by the structure. (C) Gate-dependent I–V_{sd} characteristics of a crossed NW-FET. The NW gate voltage for each I–V curve is indicated (0, 1, 2, and 3V). (D) The curves showing I vs. V_g for n-NW and global back (light gray) gates for V_{sd} of 1V [68].

The cNW-FET represents an important new transistor concept for nanoelectronics. With this concept, three critical nanometer scale device metrics are naturally defined in assembled circuits without lithography: 1) a nanoscale channel width determined by the diameter of the active NW; 2) a nanoscale channel length defined by the crossed NW gate diameter; and 3) a nanoscale gate dielectric thickness determined by the NW surface oxide. Significantly, these distinct nanometer scale metrics are determined and can be controlled with near atomic precision during NW synthesis and subsequent assembly, and should enable higher gain, higher speed, and lower power dissipation devices than possible by conventional approaches. Moreover, the cNW-FET concept can be readily integrated in a parallel manner without lithography, thus enabling one to envision a straightforward way to nanometer scale integrated electronics of the future.

15.4.5. Summary and Future Directions

Overall, in this section, we have shown that semiconductor NW materials can be doped *p*- and *n*-type in a precisely controlled way, and these electronically well-defined NW materials thus enable the assembly of a broad range of nanoscale electronic and optoelectronics devices, including FET, crossed NW *p*–*n* diodes, intra-NW *p*–*n* diode, bipolar transistors, cNW-FETs. Together, these devices represent a set of critical elements for the bottom-up assembly of more complex electronic circuits. To exploit best, this tool box of elements for building highly integrated electronic circuits will require several issues to be addressed, including: 1) production of device elements in high yield, 2) development of rational methods for assembly of these elements into integrated device arrays, and 3) development of system architectures that best utilize these device element function(s) and assembly capabilities.

The availability of electronically well-defined NW building blocks, which have enabled to demonstrate a series of nanoscale electronic devices, opens many exciting opportunities in nanoscale science and technology. For example, the nanoscale electronic devices may enable us to construct functional electronic circuits with integration density far beyond current technologies. The wide range of electronic applications described above are based primarily upon diffusive motion of electrons as in conventional devices—their uniqueness arises from the small size scale and our ability to obtain unique function from combinations of materials that would be incompatible with conventional fabrication processes. In addition, other effects, such as the quantum mechanical or wave-like nature of carriers, will need to be accounted for sufficiently small structures [48, 178], and at the same time could open up entirely new possibilities for electronic manipulation and devices [179].

The intrinsically nanometer scale devices assembled from NWs or NTs represent ideal systems to probe quantum phenomena and exploit devices exhibiting such phenomena for fundamentally new applications such as quantum computation [180, 181]. The availability of a wide range of NW materials opens up considerable opportunity for exploring these exciting directions since fundamental properties, such as the Fermi wavelength and number of conductance channels, can be systematically varied through choice of NW material (e.g., InAs vs. Si) and diameter, respectively.

As an example of the potential of NWs to enable behavior not readily accessible with conventional planar devices, we consider recent studies of InAs NWs (Fig. 15.62A). Significantly, temperature dependent three-terminal transport studies show that the conductance of individual InAs NWs remain essentially unchanged over the temperature range from 300K down to 1.6K (Fig. 15.62B). The temperature independent transport suggests strongly that electrons are not inelastically scattered by either impurities or phonons within the NWs, that is, that the electron transport is ballistic. This important conclusion is supported by several other pieces of data, including 1) most of the InAs NWs exhibit conductance values ranging from one quantum conductance ($G_0 = 2e^2/h$) to $8G_0$ and 2) the conductance shows little or no dependence on NW length. It has been also obtained similar results for clean Si NWs, and thus believes that the ballistic transport phenomena can be considered as a general property to exploit in NW-based nanoelectronics.

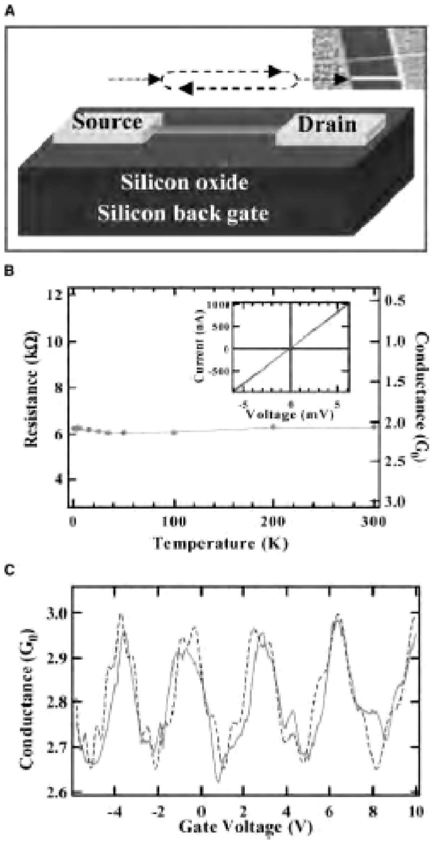


Fig. 15.62. Low temperature studies of InAs NWs. (A) Schematic showing an InAs NW device. Electron waves inside the NWs are partially reflected at the two metallic contacts, and the forward and backward reflected waves quantum mechanically interfere with each other. (B) Conductance vs. temperature of an InAs NW device. (C) Conductance vs. gate voltage (V_{sd} = 0.5 mV) at 1.6 K. The dashed line corresponds to a theoretical fit based on multichannel Landauer formalism.

Transport studies also suggest that InAs and Si NWs can exhibit coherent phenomena. Specifically, low-temperature gate voltage-dependent conductance measurements exhibit reproducible, quasi-periodic conductance oscillations (Fig. 15.62C). These oscillations can be interpreted as quantum interference between electron waves reflected by two metallic contacts, much like optical waves interfere within a Fabry–Perot cavity. In our NW electron cavities, the gate-voltage modulates the Fermi level, and hence changes the Fermi wavelength and the interference of the electron waves inside the NW. The observed electron–wave interference has been quantitatively analyzed using a multichannel Landauer formalism [178, 182], where the overall conductance is determined from two 6 × 6 scattering matrices representing the NW–metal contacts and one 6 × 6 transmission matrix inside the NWs [181]. Significantly, the calculated conductance behavior accounts for both major and minor oscillatory features in the data, and show that these different features arise from the different sub-band modes of relevance to the InAs NW cavity.

The unique transport properties of these InAs and similar NWs could lead to a number of exciting opportunities for devices and systems integration. For example, ballistic transport will be important to advance interconnects and highly integrated nanosystems since it eliminates power dissipation within the NW. Ballistic and coherent transport should also enable spin-dependent transport and spin-based devices [183]. Moreover, the ability of NWs to exhibit coherence over large distances—that is, transmit quantum mechanical information—should also be important in the emerging field of quantum computing [180]. In a quantum computer, the usual digital ‘‘bits’’ are replaced by quantum bits (qubits), and these qubits can be manipulated in very different ways by forming superpositions or entangled states, and thereby used to solve computing tasks inaccessible to digital systems. We believe that the availability of high quality NW materials opens up a new pathway for assembling quantum devices and potentially exploiting the potential of quantum computing.

15.5. SUMMARY AND PERSPECTIVES

Two different paradigms have been demonstrated to have capability for the fabrication of electronic devices based on semiconductor nanowires. In the first strategy, nanowires synthesized via ‘‘bottom-up’’ approaches form aligned arrays through either postassembly or in situ catalyst-positioning growth (i.e., nanowires grow from patterned catalyst nanoparticles into aligned arrays). Chemically synthetic ‘‘bottom-up’’ approaches can produce nanowires with a wide range of materials by controlling reaction conditions and precursors. The versatility of compositions provides a flexibility to choose suitable nanowires to meet specific applications. However, the moderate quality of nanowires (in terms of dimensional uniformity, purity, dopant concentration, doping uniformity, and surface crystallography) and assembled nanowire film (in terms of density, orientation, and end-to-end registry) might make it difficult to build large arrays of high-performance devices with them. The current lack of effective approaches for assembling high-quality nanowire thin films represents the main challenge for successful technology applications of these systems. In the second strategy, nanowires fabricated from high-quality, single-crystal, bulk wafers via ‘‘top-down’’ approaches adopt the organization and orientation defined by the pattern of mask. The resultant nanowire arrays can be transferred onto substrates through the ‘‘dry transfer printing’’ technique. A drawback of this approach is that it is limited to materials that currently exist in bulk wafer or thin film form.

There is a very bright future for these electronically well-defined NW building blocks both in fundamental science and nanotechnologies. Specifically, we believe that only the very edge of a broad range of ideas has been touched, and that there remain many fascinating fundamental problems in these NW systems, such as the role of finite size and coherent states. Well-defined semiconductor NWs also represent exciting systems to probe fundamental questions about localization or delocalization of electrical carriers and optical excitons in 1D. Further investigations will be not only critical to understanding fundamental issues in 1D systems, but also central to improved device characteristics and possibly conceptually and fundamentally new type of devices, such as single photon emitters and detectors, which could be critical for future quantum communication and computation. Continuing efforts will be required to develop even better control of NW synthesis and more and more sophisticated assembly approaches that can vary device functionality over multilength scales. Lastly, developing new device concepts and integrated architectures will be increasingly essential as we move closer to ultrahigh density integrated nanosystems.

Ch 16 Photonic Crystals

Nature provides a large variety of shapes and colors in living creatures. It makes use of many approaches to produce its colorful kaleidoscope, the most common methods being coloration by pigments, dyes, or structure. Although pigments and dyes owe their color to the absorption of light, structural color has its origin in geometric structures capable of manipulating the diffraction and/or reflection of light through periodically arranged photonic nanostructures. Usually, this form of color generation is more efficient. Insects, unlike plants, widely use this principle. Photonic crystals (PhCs) can be considered as periodic arrangements of regularly shaped materials with different dielectric constants. Periodicity can vary from one-dimensional (1D) to three-dimensional (3D). PhCs have distinct wavelengths of reflection that are governed by the distance between the layers or spheres, and this causes their specific color. If the periodicity of the crystal is changed, for example, by a (bio)chemical stimulus, the wavelength of maximum reflectance also will change. Hence, the effect offers a convenient tool for sensing, in particular if the effect can be made specific for the stimulus. Such effects also are even amenable to visual readout by unskilled operators. Moreover, methods are likely to be cheap, and handling is fairly easy. Some of these features make them more attractive than other kinds of stimuli-responsive materials. It therefore does not come as a surprise that such color changes have become the basis for a variety of physical, chemical, and biological sensors.

2. Fundamentals of Photonic Crystals

It is well known that properties of free electrons and electrons inside the crystal solids are strongly different. This difference, mainly caused by scattering processes of electrons on the regular crystal potential (Fig.1a). It is very tempting try to make something similar to this for photons. Indeed, as photon as electron are the quantum objects characterized by wave-particle duality. Namely, the both theses objects are characterized by a wavelength (for example, the characteristic electron wavelength in semiconductors is ~ 10 nm, and wavelength of photon in the visible region is ~500 nm), from one side. The both these objects are charactrised by a mass and then by a pulse, from other side. The question arises: if the bond electron exists, why does the similar state not exist for the photon?

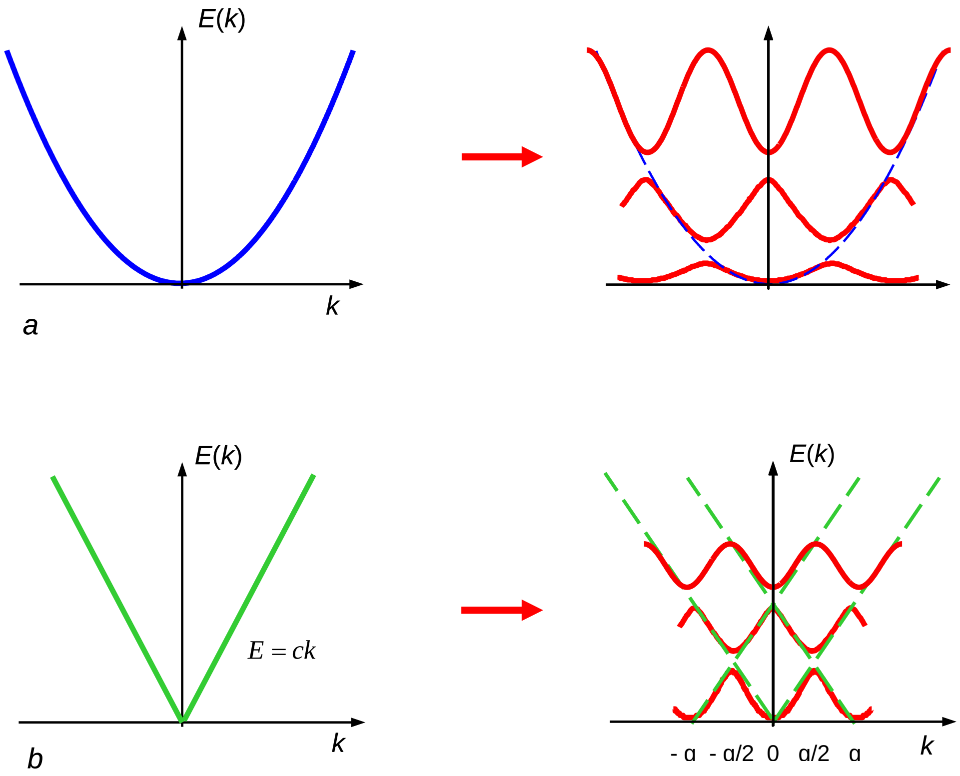


Fig.1. Formation of band structure by scattering of electrons (a) and photons (b) on the crystal potential

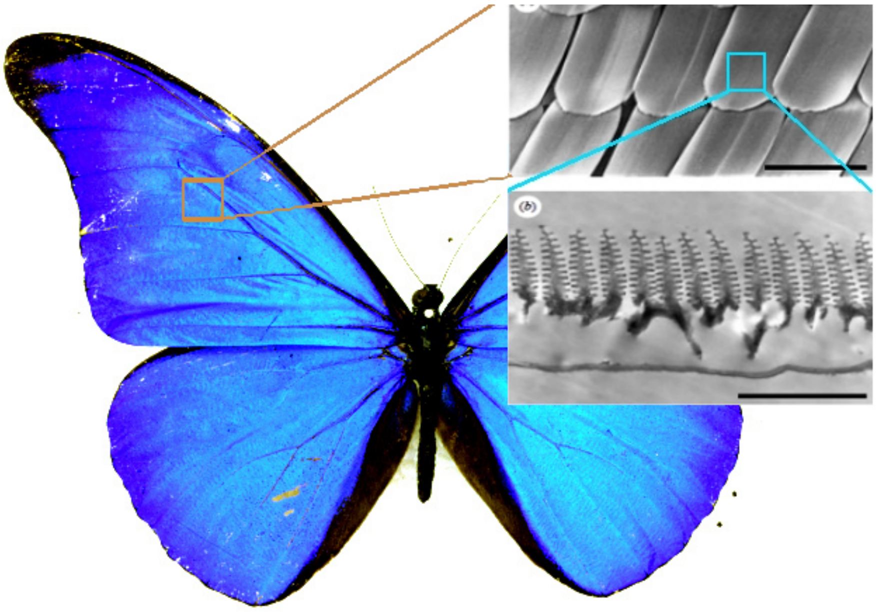


Fig 2. a) Photography of the wings of Morpho rhetenor. Transmission electron microscopy image of the cross-section of its wing scales (insert)

Indeed, it could be a possible to make some regular structure o scatterers at which the photons could form something similar “band photons” (Fig.1b). If such structures can exist they should be named as photonic crystals. The idea to use the inhomogeneous structures which consist if the particles or other inhomo-geneities instead atoms. If dielectric constants of the constituent media are different enough, Bragg scattering off the dielectric interfaces can produce many of the same phenomena for photons as the atomic potential does for electrons. As a result, a photonic crystal could be designed to possess a complete photonic bandgap – stop band. The stop band is a range of frequencies for which light is forbidden to exist within the interior of the crystal.

Indeed, in the absence of external currents and sources, Maxwell's equations can be reduced to the form

$$\left[\nabla\times\frac{1}{\varepsilon(\mathbf{r})}\nabla\times\right]\mathbf{H}(\mathbf{r})=\frac{\omega^{2}}{c^{2}}\mathbf{H}(\mathbf{r}) \ , \qquad (1)$$

with $\mathbf{H}(\mathbf{r})$ magnetic component of electromagnetic field, ω frequency and c light velocity. Because dielectric constant $\varepsilon(\mathbf{r})$ is the parameter of the equation, the solutions $\mathbf{H}(\mathbf{r})$ and ω are determined completely by the strength and symmetry of $\varepsilon(\mathbf{r})$. The dielectric constant $\varepsilon(\mathbf{r})$ in photonic crystal is perfectly periodic. Then, the solutions will be characterized by a wave vector **k** and a band index n. Similarly to crystal solid state theory, the region of all allowed wave vectors lie in Brilloin zone and the set of all solutions is termed a band structure.

Then, one can point out some simirarities between crystal solids and photonic crystals. First, periodic modulation of the refractive index in a photonic crystal forms a lattice similar to atomic lattice of solid-state. Second, behavior of photons in a photonic crystal is similar to electron behavior in an atomic lattice. Third, the existence of band gap as a result of lattice periodicity. Forth, the determination of the eigenfunctions in a photonic crystal is like to calculation of the particle wave functions in the theory of solid state physics.

At the same time, there exist some differences between solid state and photonic crystals. Obviously, the main difference is the particle energy distribution. Electrons obey the Fermi-Dirac distribution while photons obey Boze-Einstein distribution. The electrons in the solids are affected by intracrystalline field which leads to the necessity of taking it into account. Because the structure of intracrystalline field is unknown and can not be measured directly, one needs to use some model approximations such as pseudopotential method. Photons are affected by artificially constructed structures the field of which can be relatively simple calculated.

Nature uses photonic crystals (PhCs) to a wide extent. Animals such as certain butterflies and beetles use them for mimicry and survival techniques, for example, to distract predators. The periodicity of the nanostructures ranges from one- to three-dimensional and can be found in a large variety of living species of higher order. Butterflies (such as Morpho rhetenor) are the organisms most commonly referred to in order to illustrate this effect (Fig.2). Such periodic structures can not only exhibit many hues of color, but can also change on application of an external stimulus, which leads to an alternation of color.

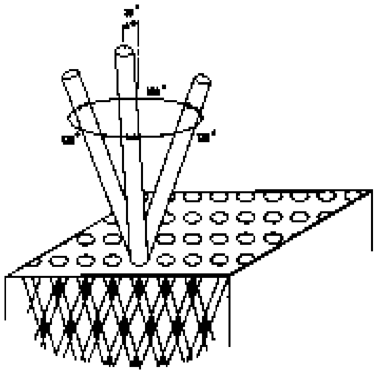


Fig.3. The structure having the photonic crystal properties – yablonovit.

The idea of use the 'crystal' properties of regular structure in optics was expressed by Yablonovitch and John. Both Yablonovitch and John suggested that structures with periodic variations in dielectric constant could influence the nature of photonic modes in an a material. Yablonovitch approach as an aim was to control the radiative properties of such structures. John aim was to provide the effect of photon localization by introducing structures with random refractive index variation Yablonovitch and co-workers were amongst the first to suggest constructing *artificial* three dimensional periodic structures to manipulate the proagation of light (Fig.3). The structure he designed, `Yablonovite' was built on a lengthscale of millimetres and shown to have a complete photonic band gap, that is to prohibit the propagation of microwaves in all directions for certain wavelengths which values are about of lattice constant of the regular structure. These regular structures with periodic variations in dielectric constant named as photonic crystals can continued progress in different branches of optics.

The characteristic properties of photonic crystals is a arising of band gap (the energy band in which the photon states can not exist). The photonic energy structure in 3D photon crystal characterizing by the FCC symmetry is shown in Fig.4. We can see that frequencies in the range . Similarly to solid state the photonic band structure is mainly defined by photonic crystal structure symmetry. Because of the periodic structure of photonic crystals is artificially designed and can be made with prediction, the wide range of practical applications arise.

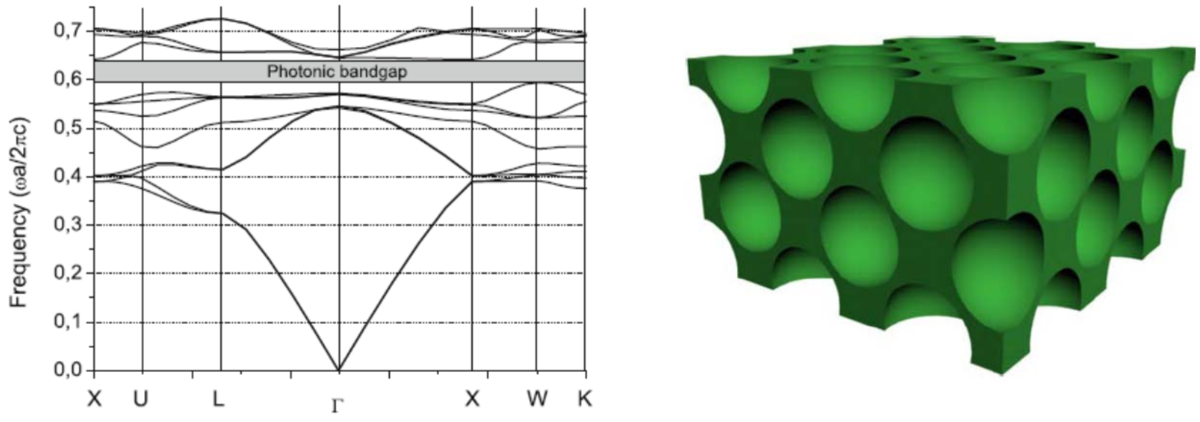


Fig.4. The band structure of the photonic crystal (left) with inverted FCC lattice shown right

The optoelectronic industry uses the optical fiber lines, where the usual frequencies are in the range around 1.5 μm. It means that the lattice constant of a photonic crystal must be of the order of 0.5 μm. Although this dimension is about 10^3 times larger than the lattice constant of solid state crystal, it is still over very small range (micrometers). This is the reason to use the photonic crystal systems in the optoelectronics.

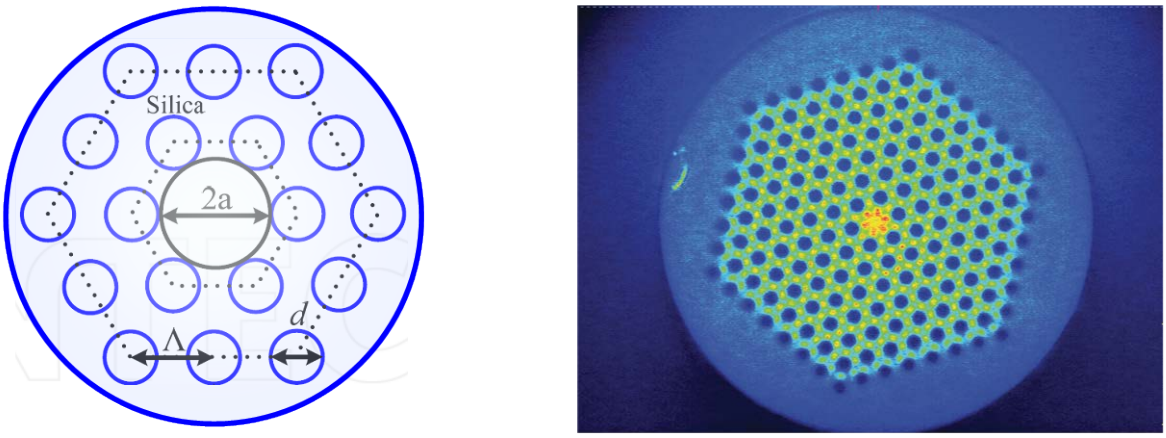


Fig.5. The schematic cross-section of the photonic crystal fiber (left) near-field image of the cross section of photonic crystal fiber (right).

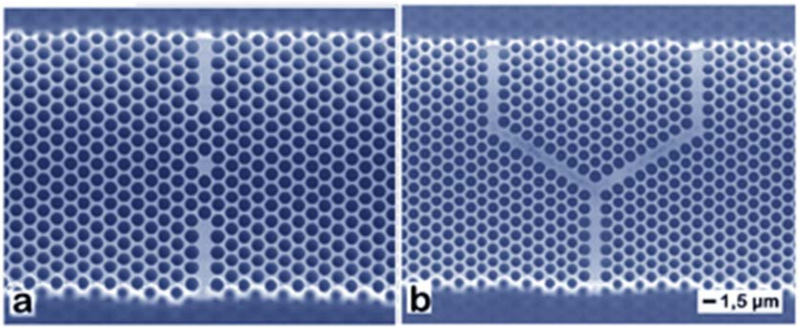


Fig.6. The 2D photonic crystal structures with defect:
micricavity in waveguide (a), and splitter (b)

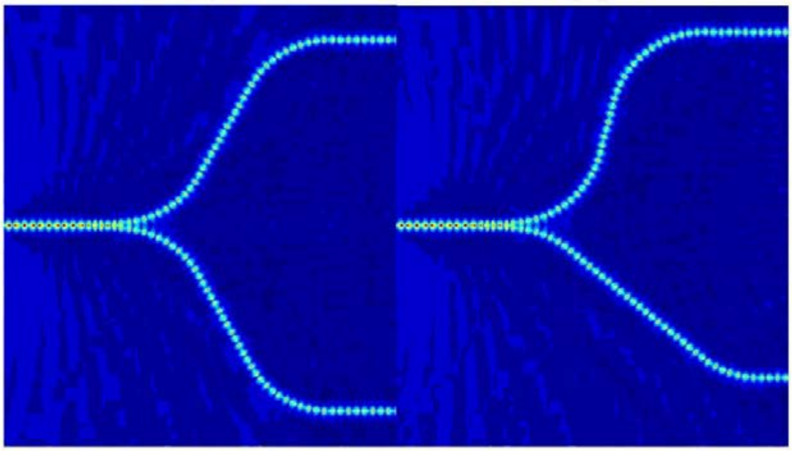


Fig.7. The results of calculations of near-field image of scanning near-field optical microscopy of photonic crystal splitters (symmetric- left) and nonsymmetric (right).

One of usage of photonic crystal structure is photonic crystal fibers, proposed by Russel, in 2003. A pure silica core optical fibers with tiny air holes embedded in the host silica matrix running along the propagation axis, have boosted the fiber optic research due to their remarkable modal properties such as provide single-mode operation for very short operating wavelengths, remain single-mode for large scale fibers. This allows us to achieve high birefringence and controllable dispersion characteristics which cannot be chieved with conventional optical fibers. The photonic crystal fiber consists of a triangular lattice air holes where the core is defined by missing air hole (Fig.5). The core diameter is $2a = \Lambda - d/2$, with Λ the air hole pitch (the distance between the centers of neighboring air holes. The background material is regular silica with a cladding refractive index n=1.45.

The other branch of usage of photonic crystals is optical insulators. Utilization of the photonic crystals as an optical insulator, as a rule, is reduced to the utilization of the photonic crystal possibility to localize the radiation inside the defect of the periodic structure. At that, the radiation wavelength must lie inside the photonic band gap. Main devices which can be developed on the basis of such a photonic crystal property are microcavities, waveguides and sharp waveguide bends, splitters, couplers etc (see, for example, Fig.6). Main function of microresonators is based on the possibility of the photonic crustals to localize the radiation inside the defect area of the periodic structure. In fact, the defect can be represented by the shift, variation of the parameters or missing of some elements or group of elements. These structures can be, for example, used in the photoelectronic devices in the input-output systems. Indeed these systems could have the dimensions about few μm dimensions. The surface plasmons can be splitted in the 2D photonic crystal splitters successfully as it is shown in Fig.7 where the near field images presented

We can see that the both symmetric and nonsymmetric types of photonic crystal splitters can rather efficiently conduct the surface plasmons in 2D structures. Fig.8 demonstartes the experimental results of surface plasmon splitting by 2D photonic crystal. One can see that similarly to the theoretical calculations the experimental scanning near-field image suggests that the surface plasmon splitting is rather efficient and can be used in the optoelectronic devices.
An unusual optical properties of special type of photonic crystals can be observed when the each of the elements of photonic crystal has a special structure. Namely, it could be realized an artificially structured composite which would show to exhibit a negative index of refraction over a band of microwave frequencies. Experimental results revealing new phenomena and potential applications for these artificial materials exhibit exotic properties of this materials main of characteristics of which is a negative index of refraction. The properties such as artificial magnetism, negative

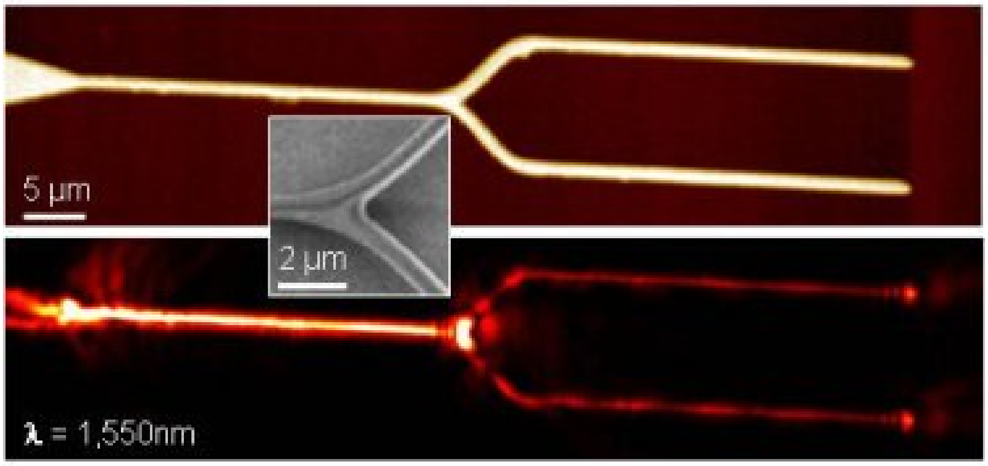


Fig.8. The surface plasmon splitter based on the 2D photonic crystal. The scanning tunnel microscopy image of the structure is shown above, the scanning near-field optical microscopy image is shown below.

ermittivity, and negative permeability have been observed in fabricated metamaterial composites. The main idea of metamaterial was formulated by Veselago. To understand Veselago's idea one should to write Maxwell's equation determined 1D electromagnetic wave propagation within a medium characterized by dielectric constant ε and magnetic permeability μ (cf. Eq.(1))

$$\left[\frac{\partial^2}{\partial x^2}-\varepsilon\mu\frac{\partial^2}{\partial t^2}\right]\mathbf{E}(x,t)=0\,. \tag{2}$$

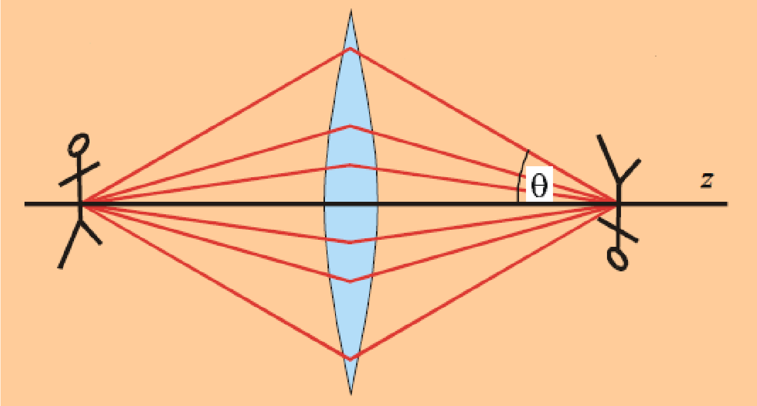


Fig.9. Formation of image by thin lens. Conven-tional lens need a wide aperture for good resolution but even so are limited in resolution by the wavelength employed.

This equation has a solution

$$\mathbf{E}(x,t)=\mathbf{E}_0 e^{i(nkx-\omega t)}\ , \tag{3}$$

where refractive index $n=\sqrt{\varepsilon\mu}$, k is a wave vector and ω is a frequency. In this equation ε and μ enter as a product and it would not appear to matter whether the signs of ε and μ were both positive or were both negative. Propagating solutions exist in the material whether ε and μ are both positive or are both negative. So what, if anything, is the difference between positive and negative materials?

Note, we in principle could write $n=-\sqrt{\varepsilon\mu}$. The important step is that the square root of either ε or μ alone must have a positive imaginary part. This is necessary for a passive material. This briefly stated argument shows why the material Veselago pondered years ago is so unique: the index of refraction is negative. A negative refractive index implies that the phase of a wave advancing through the medium will be negative rather than positive. As Veselago pointed out, this fundamental reversal of wave propagation contains important implications for nearly all electromagnetic phenomena. Many of the exotic effects of negative index have been or are currently being pursued by researchers. But perhaps the most immediately accessible phenomenon from an experimental or computational point-of-view is the reversal of wave refraction. In part henoted that a negative index focusing lens would need to be concave rather than convex. For thin lens, geometrical optics valid for either positive or negative index gives the result that the focal length is related to the radius of curvature of the lens by $f=R/(n-1)$. The denominator in the focal length formula implies an inherent distinction between positive and negative index lenses, based on the fact that an $n=\pm1$ material does not refract electromagnetic fields while an $n=-1$ material does. To make a conventional lens with the best possible resolution a wide aperture is sought. Each ray emanating from an object (see, Fig.9) characterizes by wave vector components along the axis of the lens $k_z=k_0\cos\theta$ and perpendicular to the axis $k_x=k_0\sin\theta$. The former component is responsible for transporting the light from object to image and the latter represents a Fourier component of the image for resolution: the larger we can make k_x , the better. Naturally the best that can be achieved is k_0 and hence the limit to resolution of

$$\Delta\approx\pi/k_0\equiv\lambda/2\ , \tag{4}$$

with λ wavelength. This restriction is a grate problem in many areas of optics. In contrast to the image, there is no limit to the electro-magnetic details contained in the

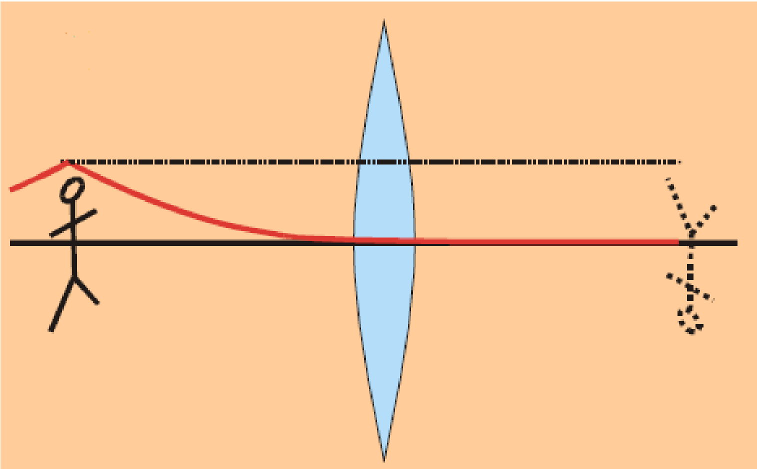


Fig.10. The missing components of the image are contained in the near field which decays exponentially and makes negligible contribution to the image.

object but unfortunately not all of them make it across the lens to the image. The problem lies with the *z*-component of the wave vector which can be written as $k_z = \left(k_0^2 - k_x^2\right)^{1/2}$. Evidently for large values of k_x , corresponding to fine details in the object, k_z is imaginary and the waves acquire an evanescent ones. By the time they reach the image they have negligible amplitude as it is shown in Fig.10. This is near-field regime, when the field does not transfer the energy. Unlike in the far-field regime the ray propagates and transmits the energy. If by any specific way, one could amplify the near fields, one could in principle recoup their contribution, but the amplification would have to be of just the right amount and possibly very strong for the most localized components. This is a tall order but by a remarkable chance the new negative slab lens achieves this feat. This situation is shown in Fig.11 where rays contributing to the image for the negative slab, when $\varepsilon \to -1$, $\mu \to -1$, and $n \to -1$.

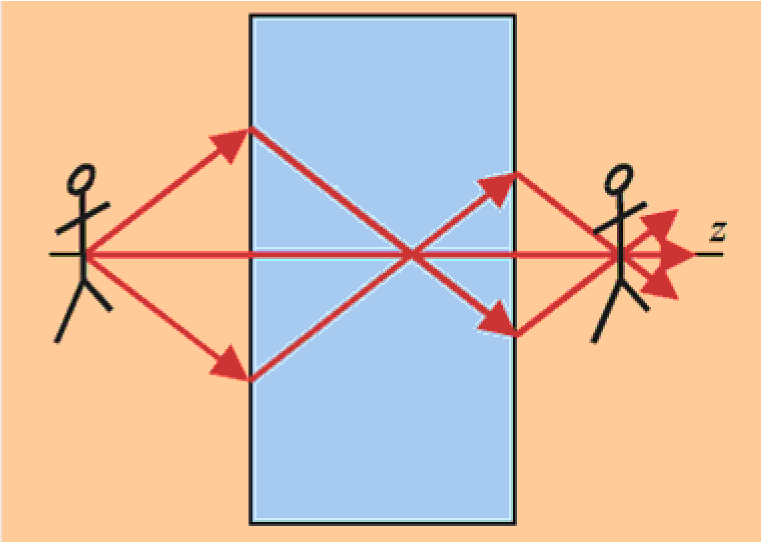
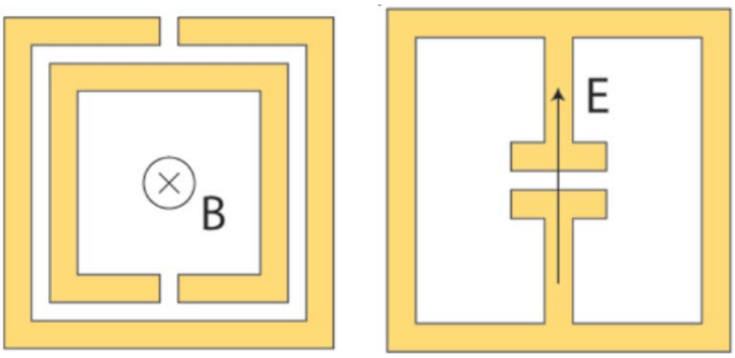


Fig.11. The image formation in the slab when $\varepsilon \to -1$, $\mu \to -1$, and $n \to -1$



(a) (b)

Fig.12. Elements used as split ring resonator providing the response to the magnetic component of the electro-magnetic field (a); the artificial medium used for electric response to the electric component of elecgtromagnetic field

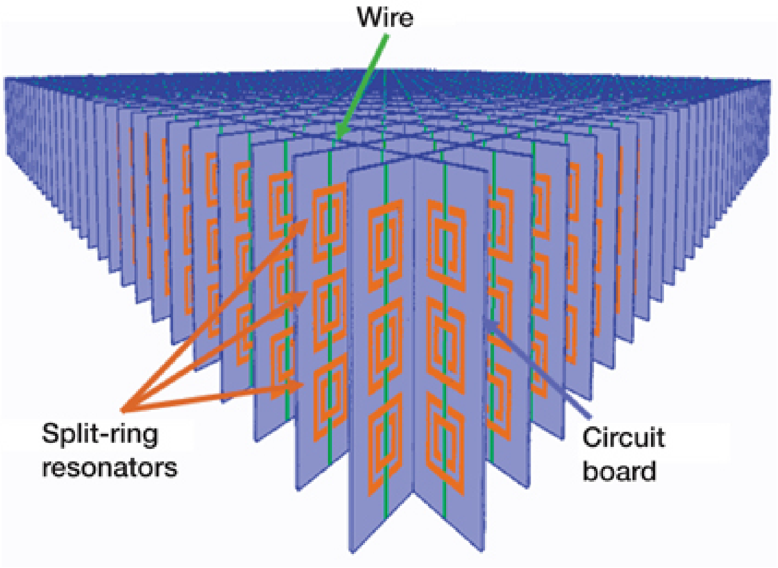


Fig.13. An metamaterial formed by split ring resonators and wires deposited on opposite sides of a circuit board.

To construct the similar "super lens" the idea of periodic photonic structures is fruitful. Namely, photonic crystals are periodic dielectric or metallic structures capable of achieving negative phase velocity and thus negative refractive index. However, these structures are not easily described by bulk parameters such as ε and μ. Rather, we are concerned here with those artificial structures that can be viewed as homogeneous, described by values of ε and μ. The desired material consists of an array of subwavelength elements, designed independently to respond preferentially to the electric or magnetic component of an electromagnetic wave. The unit of such photonic crystal should consist of the split ring resonator (Fig.12 a) which provides the response to the magnetic component of the electromagnetic field and element for electric response (Fig.12b).

The artificial material having the properties of metamaterial is shown in Fig.13. We can see that the both properties of elements provided the response to the magnetic and electric component of the electro-magnetic field are presented in this structure.

Then, we can see that the new artificial materials named as photonic crystals have the new unusual properties which can used in different branches of modern sciences and practical applications.

Ch17 Superconductivity

The phenomenon of superconductivity when the electrical resistance of mercury completely vanishes at low temperatures, was first discovered by the Dutch physicist Kamerlingh Onnes, who was the first to liquefy helium[1]. The view of Kamerlingh Onnes's laboratory is shown in Fig.1. In 1911 Kamerlingh Onnes with colleagues discovered the phenomenon of superconductivity while studying the resistance of mercury at low temperatures (Fig.2). The mercury was chosen because very pure samples could easily be prepared by distillation.

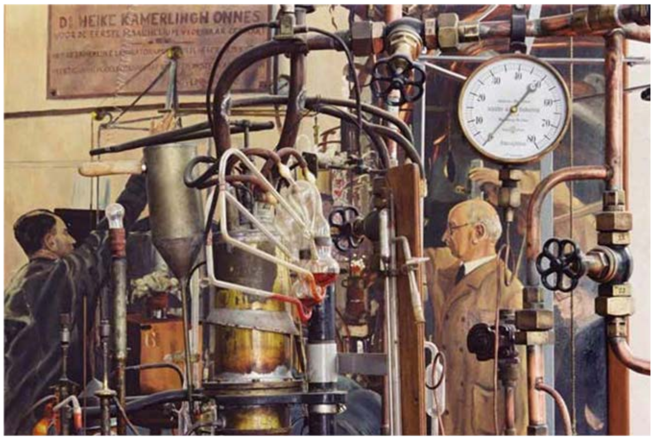


Fig.1. Low temperature laboratory, leading by Kamerlingh Onnes, where the liquid helium was firstly obtained.

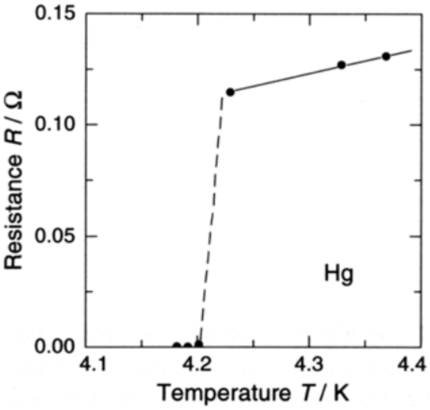


Fig.2 The resistance of mercury measured by Kamerlingh Onnes.

Tabl.1

element	T_C , K
Al	1.19
Hg	4.15
Nb	9.2

---

[1] Mercury boils at 4.2 K at standard pressure.

Pb	7.2
La	4.8
Sn	3.7

The term “superconductivity” refers to a complex of phenomena which are usually found to occur together. It occurs in a wide variety of metals and alloys, below a certain temperature called as transition temperature or critical temperature T_c which currently ranges up to ~ 150 K. Of the various constituent phenomena, the two which are conceptually most important are macroscopic diamagnetism and perfect conductivity. For more than ten years Kamerlingh Onnes's laboratory was unique in which can liquefy helium. Then, the next superconductor discovered in Leiden Kamerlingh Onnes's laboratory was Sn (1913). In 1923 liquid helium was obtained in Berlin and Toronto. Soon after first measurement of critical temperature in Hg and Sn many other elemental metals were found to exhibit zero resistance when their temperatures were lowered below T_C. Some values of critical temperature for the metal elements are shown in Tabl.1. We can see that critical temperatures of metal elements are no exceeded 10 K. Scientists from Leiden laboratory become measuring the influence of magnetic field on the superconductivity. It was established that metal can transmits to normal state at temperature lowered below T_C under action magnetic field. The magnetic field, at which the transition from superconducting state to normal state is observed, was named critical field. The studies of magnetic field acting to the superconductors resulted in two very important discovering. In 1933, Walter Meissner and Robert Ochsenfeld discovered a magnetic phenomenon that showed that superconductors are not just perfect conductors. Namely, they established that superconductors are macroscopic diamagnetic. Comparing behaviors of superconductor and ideal conductor under external magnetic field acting is schematically shown in Figure 3. Imagine that both the ideal conductor and superconductor are above their critical temperature. That is, they both are in a normal conducting state and have electrical resistance. When a magnetic field, **B**, is acted to the both samples, the field penetrates into the both materials. Both samples are then cooled so that the ideal conductor now has zero resistance. It is found that the superconductor expels the magnetic field from inside it, while the ideal conductor maintains its interior field. The phenomena of magnetic expelling from inside superconductor is named Meissner effect. The ideal diamagnetism was discovered by Meissner and Ochsenfeld for rods made of lead or tin. This expulsion effect, similar to the property $R = 0$, can be nicely demonstrated using the “levitated magnet”. In order to show the property a perfect conductivity, in Fig.4 one has lowered the permanent magnet towards the superconducting lead bowl, in this way generating permanent currents by induction. To demonstrate the Meissner-Ochsenfeld effect, one places the permanent magnet into the lead bowl at $T > Tc$ (Fig.4, left) and then cool down further. The field expulsion appears at the superconducting transition: the magnet is repelled from the diamagnetic supercon-ductor, and it is raised up to the equilibrium height (Fig.4,from left to right).

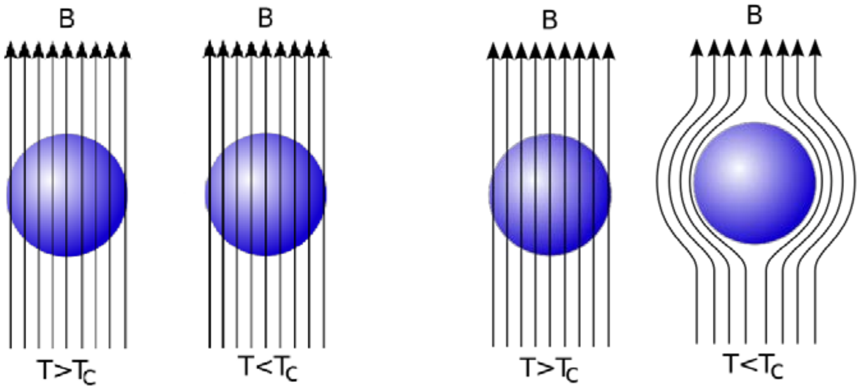


Fig.3

Penetration of external magnetic field **B** inside an ideal conductor (left) and superconductor (right)

The next discovery under magnetic field action was established of two types superconductors I and II. The idea of two different types of superconductor behavior under magnetic field is connected with two types of superconductor was formulated by Shubnikov and his colleagues in 30-th years of XX century. What did Shubnikov establish? When the magnetic field on the surface of the superconductor exceeds the

Fig.4 “Levitated magnet” for demonstrating the Meissner-Ochsenfeld effect in the presence of an applied magnetic field. *Left*: starting position at $T > Tc$. *Middle* & *Right*: equilibrium position at $T < Tc$.

critical value, it starts to penetrate the superconductor, which then phase separates. The superconducting phase now coexists with the normal phase which contains the magnetic field in the bulk of the material. Now, a surface separating these two phases could be expected to exist. As a matter of fact, it does in some materials, but in others, this surface becomes unstable, and the normal phase disperses itself finely into the superconducting phase in the form of vortices containing discrete quanta of the magnetic flux. The materials in which the phase interface between the superconducting and the normal state is stable are designated as type I, while the others are type II[2] (Fig.5). Type II materials enter into a mixed state when the magnetic field begins to penetrate the bulk of the superconductor. Magnetic vortex lines start to nucleate at the surface and drift into the bulk of the material when the magnetic field on the surface becomes stronger than the lower critical field. Magnetic vortex lines continue to accumulate inside the material until they become so

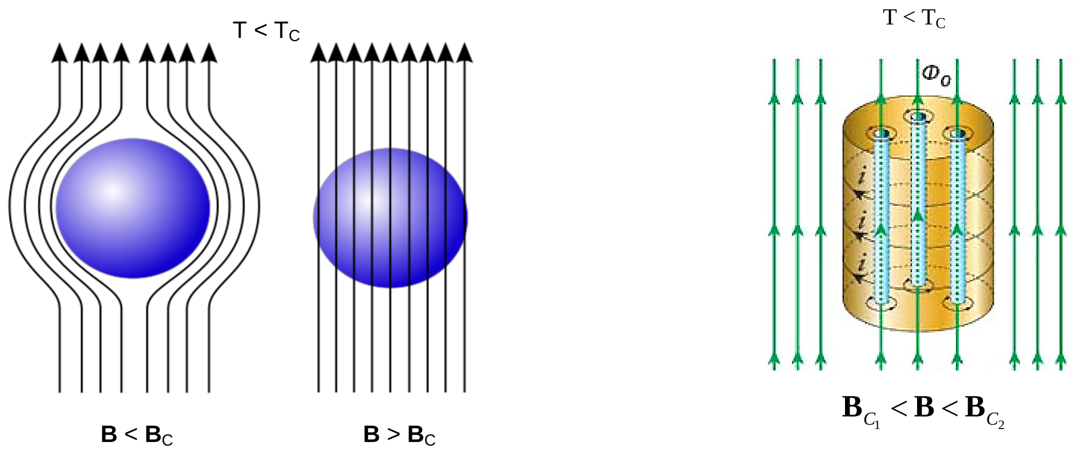


Fig.5. Superconductors I and II types behavior under magnetic field action. I type superconductor (left) expels the external magnetic field from inside it when magnetic field **B** does not exceed the critical value **B**$_C$. Increasing of magnetic field up to value **B** > **B**$_C$ leads to break of superconducting state. II type superconductor expels the external magnetic field from inside it when magnetic field **B** does not exceed the lower critical value $\mathbf{B}_{C_1}$. When external magnetic field lies between lower and upper critical values $\mathbf{B}_{C_1} < \mathbf{B} < \mathbf{B}_{C_2}$, the applied field is able to partially penetrate the superconductor, so the Meissner effect is incomplete. The superconductor is in Shubnikov phase (or mixed state). When external field $\mathbf{B} > \mathbf{B}_{C_2}$, the uperconductivity completely vanishes

---

[2] All the high temperature superconductors that were discovered since 1986 are type II materials.

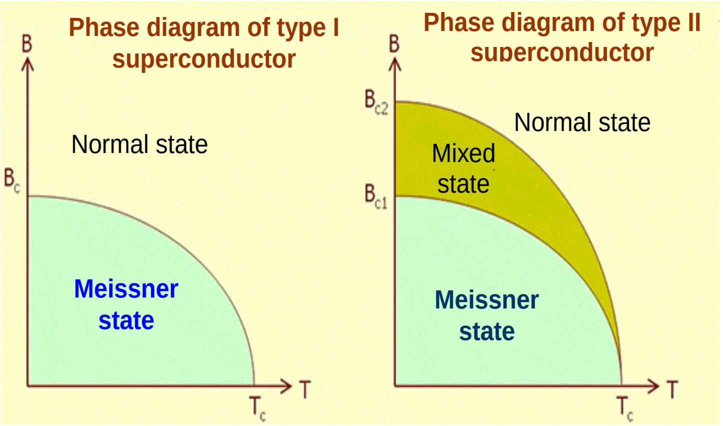


Fig.6. Phase diagrams of superconductors I and II types

dense that they destroy superconductivity in the material altogether. This happens at the upper critical field. The lower critical field is lower, and the upper critical field is higher than the thermodynamic critical field, which is the critical field expected on the basis of the difference in free energy between the superconducting

and normal states of the material. As a result, one can present the phase diagrams of superconductors I and II types in Fig.6. Then, the first type (type-I) superconductors expels the magnetic field up to a maximum value B_c, the critical field. For larger fields, superconductivity breaks down, and the sample assumes the normal

conducting state. Here the critical field depends on the temperature and reaches zero value at the transition temperature T_c. Mercury or lead are examples of a type-I superconductor. The second type (type-II) superconductors demonstrates ideal diamagnetism for magnetic fields smaller than the lower critical magnetic field $\mathbf{B}_{C_1}$. Superconductivity completely vani-shes for magnetic fields larger than the upper critical magnetic field $\mathbf{B}_{C_2}$, which often is much larger than $\mathbf{B}_{C_1}$. Under the external magnetic field $\mathbf{B}_{C_1} < \mathbf{B} < \mathbf{B}_{C_2}$ the mixed state (Shubnikov phase) occurs. Both critical fields reach zero at T_c. This behavior is found in many alloys, but also in the high-temperature superconductors. In the latter, $\mathbf{B}_{C_2}$ can reach even values larger than 100 T.

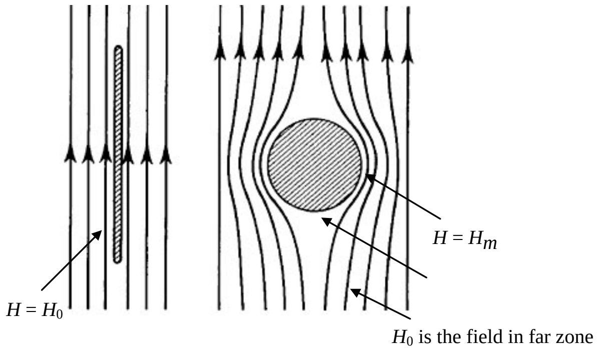


Fig.8. To explanation of intermediate state in I type superconductor.

The penetration of magnetic field inside the superconductor can be observed as in the sample of type II superconductor as type I superconductor. There are mixed state of type II superconductor and intermediate state of I superconductor. What is the intermediate state of I superconductor? Consider the sample of type I superconductor shaped as thin film (Fig.7). One can see, that due to geometry of the sample superconductor divides on the normal and superconducting domains. The reason of this phenomena one can understand considering the acting of external magnetic field to the samples shaped as thin needle and sphere. In Fig.8 the acting the external field H_0 to the samples is shown. We can see that magnetic field is $H = H_0$ at the vicinity of the sample shaped as a needle. But different values of local magnetic field are observed at different domains at the sample shaped as a sphere. It means that in certain domain near the sample the local magnetic field can exceed the maximum value at which superconductivity vanishes. At these domains the penetration of magnetic field inside the sample can be realized. The connection between the external field in far zone H_0 and H_m can be found

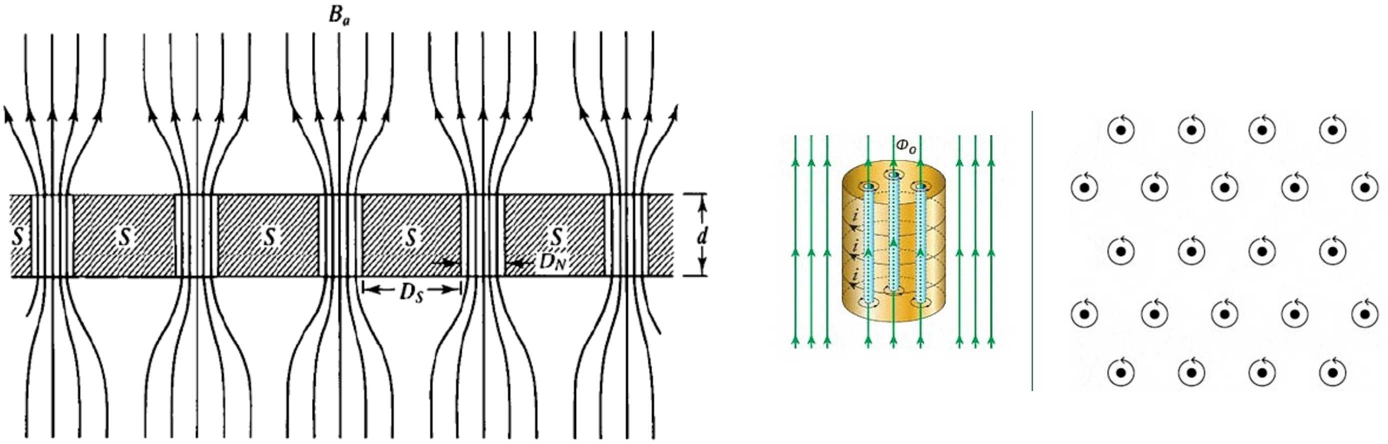


Fig.7. Penetrating the magnetic field into the sample of I type superconductor - the intermediate state of superconductor (left). Penetration the magnetic field into the sample II type superconductor, where the vertexes with core of normal state arise – the mixed state of superconductor (right)

via expression

$$H_m = \frac{H_0}{1-n} \quad , \tag{1}$$

where parameter n is characterized by shaping of the superconductor sample. For example, if the sample is shaped as a cylinder, axis of which is oriented parallel to the field direction, parameter n equals to zero. When cylinder axis is oriented perpendicularly to the field direction, $n = 1/2$. Spherical sample of I type superconductor is characterized by $n = 1/3$. The flat plate oriented perpendicularly to the field direction is characterized by $n = 1$.

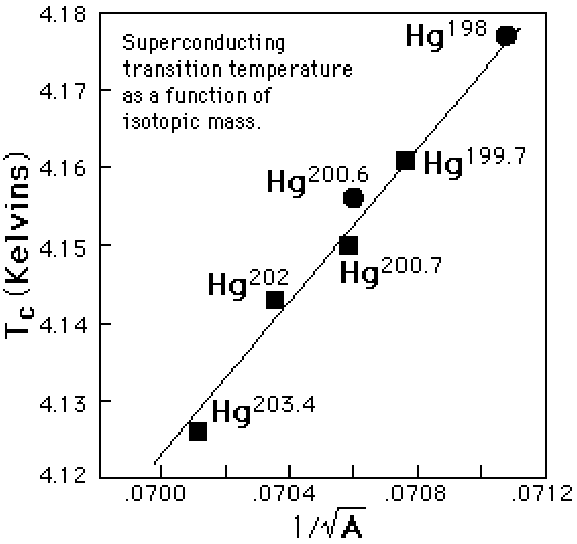


Fig.9. The dependence of transition temperature T_C of mercury on its isotopic mass.

In 50-th years of XX century the isotopic effect was discovered. The effect reflects the connection between the transition temperature to superconducting phase and isotopic mass of the metal (see, Fig. 9). In part, the $T_C = const/\sqrt{M}$ connection was established for most superconductors. Most thermodynamic properties of a superconductor are found to vary as $\exp\left(-\Delta / k_B T\right)$, indicating the existence of a gap, or energy interval with no allowed eigenenergies, in the energy spectrum.

All these experimental facts are needed to be explained.

Consider now the thermodynamic properties of superconductors. To make this consider the I-type superconducting cylinder along which the external magnetic field $\mathbf{H}_0$ is applied (Fig.10). When the field H_0 is less than critical field H_C , the Meissner effect observes. It means that inside the cylinder $\mathbf{B} = 0$ and magnetic momentum of unit of volume of the cylinder $M = -\mathbf{H}_0 / 4\pi$. The external source of the field operates the work when the field changes as $d\mathbf{H_0}$. Then one can write

$$-\mathbf{M} d\mathbf{H}_0 = \mathbf{H}_0 d\mathbf{H}_0 / 4\pi \ . \tag{2}$$

When the field arises from zero up to certain value H_0, the source operates the work

$$A = -\int_0^{H_0} \mathbf{M}\, d\mathbf{H}_0 = H_0^2/8\pi \ . \tag{3}$$

The free energy of the superconductor under the magnetic field action is written as

$$F_{sH} = F_{s0} + H_0^2/8\pi \ . \tag{4}$$

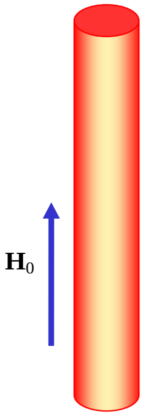


Fig.8. To consideration of thermodynamical properties of superconductor

When the magnetic field increases up to critical value, the sample transmits to normal state. Then,

$$F_n = F_{s0} + H_C^2/8\pi \ . \tag{5}$$

If one subtracts Eq.(4) from Eq.(5), one obtains very important expression

$$F_n - F_{s0} = H_C^2/8\pi \ , \tag{6}$$

which manifests that the critical field of bulk superconductor is a measure of energetic efficiency of superconducting state. Write now the first principle of thermodynamics

$$\delta Q = \delta R + \delta U \ . \tag{7}$$

The term in left part of Eq.(7) means the increment of thermal energy density. The first term in right part of Eq.(6) is the increment of thermal energy density, and second term is the increment of the internal energy. Because of the free energy density is

$$F = U - TS \ , \tag{8}$$

the variation of it can be written as

$$\delta F = \delta U - T\delta S - S\delta T \ . \tag{9}$$

Using that the considered process is reversible one, or $\delta Q = T\delta S$, we obtain from Eq.(7)

$$\delta U = T\delta S - \delta R \,. \tag{10}$$

Then, one can rewrites Eq.(9) in the form

$$\delta F = -\delta R - S\delta T \,. \tag{11}$$

This equation in the case of adiabatic process gives us

$$S = -\left(\frac{\partial F}{\partial T}\right)_R \,. \tag{12}$$

Which with taking into account Eq.(6) immediately leads to

$$S_s - S_n = \frac{H_C}{4\pi}\left(\frac{\partial H_C}{\partial T}\right)_R \,. \tag{13}$$

As it follows from Nernst heat theorem, at T = 0 the entropy of each body S = 0. It means that

$$\left(\frac{\partial H_C}{\partial T}\right)_{T=0} = 0 \,. \tag{14}$$

Then, the curve $H_C(T)$ characterises by zero derivative at T = 0. From the other side, the behavior of $H_C(T)$ established experimentally, obeys the formula

$$H_C(T) = H_C(0)\left[1 - \left(\frac{T}{T_C}\right)^2\right] \,. \tag{15}$$

We can see, that $\mathrm{H_C}(T)$ is the monotonically decreasing quantity. It means that

$$\frac{\partial H_C}{\partial T} < 0 \,. \tag{16}$$

Using this expression, one can se from Eq.(13), that

$$S_s < S_n . \tag{17}$$

This inequality means that superconducting state is more ordered state than normal state. Then, transition from normal state to superconducting state is the phase *transition*.

Because of $H_C = 0$ at $T = T_C$, then $S_s = S_n$ at $T = T_C$. It means that the transition at $T = T_C$ occurs without absorption or latent heat release. Then, the transition from normal metal state to superconducting state at $T = T_C$ is the ***second order phase transition***. On the other side, at $T < T_C$ the transition from superconducting state to normal state occurs under magnetic field action, when $S_C < S_n$. Then, the transition over magnetic field occurs with latent heat absorption. It means that all transitions under magnetic field action in superconductors are the ***first order phase transitions***.

The electrodynamic properties of superconducting can be explained in the frame of phenomenological approach developed by Londons. Londons supposed that two kinds of current densities form the total current inside superconductor: normal $\mathbf{j}_\mathrm{n}$ caused by normal electrons and $\mathbf{j}_\mathrm{S}$ caused by superconducting electrons

$$\mathbf{j} = \mathbf{j}_n + \mathbf{j}_s . \tag{18}$$

One supposes that $\dot{\mathbf{D}} = 0$ and $\mu = 1$. Then, taking into account that $\mathbf{j}_n << \mathbf{j}_s$ one writes Maxwell equations for superconducting electrons

$$\begin{gathered} div\mathbf{E} = 0, \quad div\mathbf{B} = 0, \\ rot\mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}, \quad rot\mathbf{B} = \frac{4\pi}{c}\mathbf{j}_s \ . \end{gathered} \tag{19}$$

These equations were added by London’s constitutive equations:

Acceleration equation describes the current increasing when electric field acts to superconductor (this equation describes an infinite conductance of superconductor)

$$\Lambda \frac{\partial \mathbf{j}_s}{\partial t} = \mathbf{E}, \tag{20}$$

and

$$\Lambda rot\mathbf{j}_s = -\frac{1}{c}\mathbf{B}. \tag{21}$$

Here the designation

$$\Lambda = \frac{m}{n_s e^2} \tag{22}$$

was introduced, with m mass of superconducting electrons, n_s its concentration and e its charge. Origin of Eq.(20) is motion equation of superconducting electrons under action of electric field **E**

$$m\frac{d\mathbf{v}_s}{dt} = e\mathbf{E}. \tag{23}$$

Multiplying the both sides of this equation by concentration of superconducting electrons and taking into account that $\mathbf{j}_s = n_s e\mathbf{v}_s$ one obtains

$$\mathbf{E} = \left(\frac{m}{n_s e^2}\right)\frac{d\mathbf{j}_s}{dt} \tag{24}$$

the first London's equation (22).

The 2-nd London’s equation one can obtain from the next circumstances: Let $\mathcal{F}_{s0}$ is superconductor free energy without current and magnetic field action. The density of kinetic energy of superconducting current

$$W_{kin} = n_s\frac{mv_s^2}{2} = \frac{m}{2n_s e^2}j_s^2. \tag{25}$$

Then, using Maxwell’s equation $\mathrm{rot}\mathbf{H} = \left(4\pi/c\right)\mathbf{j}_s$, one obtains

$$W_{kin} = \frac{\lambda^2}{8\pi}\left(\mathrm{rot}\mathbf{H}\right)^2, \tag{26}$$

with

$$\lambda^2 = \frac{mc^2}{4\pi n_s e^2}. \tag{27}$$

The local density of magnetic energy inside superconductor is $H^2/2\pi$. Total free energy of the superconductor can be written as

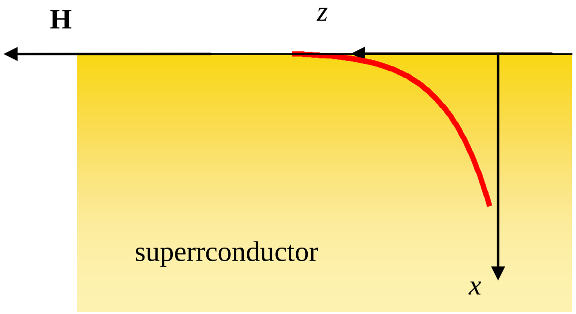
$$\mathcal{F}_{sH}=\mathcal{F}_{s0}+\frac{1}{8\pi}\int d^3r\left[\mathbf{H}^2+\lambda^2\left(\mathrm{rot}\mathbf{H}\right)^2\right]. \tag{28}$$

The variation equation $\delta\mathcal{F}_{sH}=0$ gives the magnetic field function at which $\mathcal{F}_{sH}$ reaches of minimum. Because of $\mathbf{a}\mathrm{rot}\mathbf{b}=\mathbf{b}\mathrm{rot}\mathbf{a}-\mathrm{div}\left[\mathbf{ab}\right]$, one can write the variation

$$\delta\mathcal{F}_{sH}=\int d^3r\left[\mathbf{H}+\lambda^2\left(\mathrm{rotrot}\mathbf{H}\right)\right]\delta\mathbf{H}-\int d^3r\mathrm{div}\left[\mathrm{rot}\mathbf{H},\delta\mathbf{H}\right]. \tag{29}$$

The integral of last term of right part of Eq.(29) can be rewritten with Gauss theorem $\oint_S d\mathbf{S}\left[\mathrm{rot}\mathbf{H},\delta\mathbf{H}\right]=0$, where integration is over superconductor surface. Magnetic field at the surface is an external field which is given quantity. Then $\delta\mathbf{H}=0$. As a result, one obtains

$$\int d^3r\left[\mathbf{H}+\lambda^2\left(\mathrm{rotrot}\mathbf{H}\right)\right]\delta\mathbf{H}=0. \tag{30}$$



Taking into account that field variation is an arbitrary, one obtains from this equation the 2-nd London's equation

$$\mathbf{H}+\lambda^2\left(\mathrm{rotrot}\mathbf{H}\right)=0. \tag{31}$$

To clarify the physical sense of parameter λ, one should consider the penetration of external magnetic field inside superconductor. Let magnetic field directed along O*Z* axis has a value $H_z(0)$ at the surface of superconductor and, then is $\mathbf{H}=\left(0,0,H_z(x)\right)$. From second London's equation one obtains

$$\frac{\partial^2 H_z}{\partial x^2}=\frac{1}{\lambda^2}H_z. \tag{32}$$

Solution of this equation is written in the form

Fig.8. The penetration of magnetic field inside superconductor

$$H_z(x) = H_z(0)e^{-x/\lambda} . \tag{33}$$

This solution, obviously means that the magnetic field penetrates into the superconductor to characteristic depth λ. This penetration depth is called as London's penetration depth. Because λ depends on superconducting electrons concentration n_s which depends on temperature, the London's penetration depth depends on temperature. Then, at $T = 0$, $\lambda = \lambda_{\min}$. At $T = T_C$, $\lambda = \infty$. Note, characteristic value of λ_L is about $\lambda_{charact} = 10^{-6}$ cm.

V.Ginzburg and L.Landau developed the phenomenological theory of superconductivity (GL theory) based on the London's idea of two kind of curriers. The main point of GL approach is the existence of an “order parameter” (“macroscopic wave function”), which has the general nature of a Schrödinger wave function – in particular, is complex scalar.

Let us consider the circumstances lied in the base of GL theory.

The transition from normal to superconducting state is characterized by order parameter. The macroscopic wave function can be chosen as order parameter. Namely,

$$\Psi(\mathbf{r}) \ \rightarrow \ \begin{cases} \neq 0, & T < T_C \\ = 0, & T \geq T_C \end{cases} \tag{34}$$

Let the wave function is normalized as

$$\left|\Psi(\mathbf{r})\right|^2 = n_s / 2 , \tag{35}$$

where n_s is concentration of superconducting electrons. The, we cam write the free energy of superconductor when temperature is about critical temperature ($T \approx T_C, \quad T < T_C$)

$$F_S(T \approx T_C) = F_N + \alpha\left|\Psi(\mathbf{r})\right|^2 + \frac{\beta}{2}\left|\Psi(\mathbf{r})\right|^4 , \tag{36}$$

with α and β the constants which can be found from minimize of free energy. Indeed, the main state of superconductor corresponds to minimum of free energy

$$\frac{dF_S(0)}{d\left|\Psi_S(\mathbf{r})\right|^2} = 0 \ , \tag{37}$$

where $F(0)$ is a free energy when $H = 0$. This equation gives us

$$\alpha + \beta\left|\Psi_S(\mathbf{r})\right|^2 = 0 \ , \tag{38}$$

from which one obtains the value of order parameter, corresponding to superconducting state

$$\left|\Psi_S(\mathbf{r})\right|^2 = -\frac{\alpha}{\beta} \ . \tag{39}$$

This expression, in part, shows us that coefficients α and β have a different signs. Substitution of the order parameter [Eq.(39)] into free energy [Eq.(36)] gives us

$$F_S(0) - F_N = -\frac{\alpha^2}{\beta} + \frac{1}{2}\frac{\alpha^2}{\beta} = -\frac{1}{2}\frac{\alpha^2}{\beta} \ . \tag{40}$$

Taking into account that

$$F_S(0) - F_N = -\frac{\mathbf{H}_C^2}{8\pi} \, , \tag{41}$$

one obtains

$$\frac{\alpha^2}{2\beta} = \frac{\mathbf{H}_C^2}{8\pi} \ . \tag{42}$$

Because of

$$\Psi(\mathbf{r}) \ \rightarrow \ \begin{cases} \neq 0, & T < T_C \ , \\ = 0, & T \geq T_C \ , \end{cases} \tag{43}$$

we can suppose that

$$\alpha(T_C) = 0 \, , \tag{44}$$

which allows us to write

$$\alpha(T) \sim \left(T - T_C\right)^a . \tag{45}$$

Taking into account that when $T \to T_C$ ($T > T_C$), parameter α should be added, then

$$\alpha(T) \sim \left(T - T_C\right). \tag{46}$$

To stable superconducting state, parameter β have to be positive. Indeed, when $T < T_C$, $\alpha < 0$. Because of $|\Psi_s(\mathbf{r})|^2 = -\alpha / \beta$ and $|\Psi_s(\mathbf{r})|^2 > 0$ this expression can be valid only when $\beta > 0$. As a result, one can write

$$F = \alpha_0 (T - T_C)\eta^2 + (\beta / 2)\eta^4 \ . \tag{47}$$

The GL theory can be generalized for inhomogeneous systems. The initial point of the theory is Gibbs free energy in the form

$$G_S(\mathbf{H}) = G_N + \int \left\{ \alpha\left|\Psi\right|^2 + \frac{1}{2}\beta\left|\Psi\right|^4 + \frac{1}{4m_f}\left|\left(-i\hbar\nabla - \frac{2e}{c}\mathbf{A}\right)\Psi\right|^2 + \frac{\left(rot\mathbf{A}\right)^2}{8\pi} - \frac{\left(rot\mathbf{A}\right)\mathbf{H}_0}{4\pi} \right\} dV . \tag{48}$$

The main state of the system is found by varying of the energy

$$\delta_{\Psi^*} G_S(\mathbf{H}) = 0 . \tag{49}$$

As a result, one obtains the first GL equation

$$\alpha\Psi + \beta|\Psi|^2\, \Psi + \frac{1}{4m_f}\left(i\hbar\nabla + \frac{2e}{c}\mathbf{A}\right)^2 \Psi = 0 , \tag{50}$$

with the boundary condition

$$\left(i\hbar\nabla\Psi + \frac{2e}{c}\mathbf{A}\Psi\right)\mathbf{n} = 0 . \tag{51}$$

Here **n** is an orth of normal to the sample surface. The second equation of GL theory can be obtained by varying of Gibbs free energy over vector potential **A**

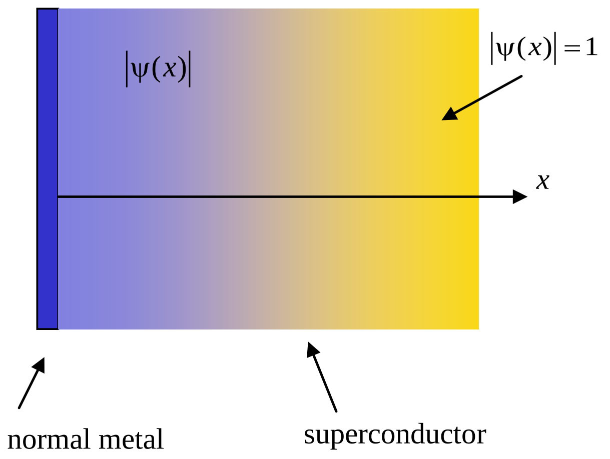
$$\delta_{\mathbf{A}} G_S(\mathbf{H}) = 0 . \tag{52}$$

The variation gives us

$$\frac{i\hbar e}{2m_f c}\left(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*\right) + \left(\frac{2e^2}{m_f c^2}\right)|\Psi|^2 \mathbf{A} + \frac{1}{c}\mathbf{j}_s = 0 . \tag{53}$$

Eqs.(50), (51) and (53) usually are written in the standard form

$$\xi_S^2\left(i\nabla + \frac{2\pi}{\Phi_0}\mathbf{A}\right)^2 \psi - \psi + |\psi|^2 \psi = 0 , \tag{54}$$



$$\left(i\nabla + \frac{2\pi}{\Phi_0}\mathbf{A}\right)\mathbf{n}\psi = 0 , \tag{55}$$

$$rotrot\mathbf{A} = -i\frac{\Phi_0}{4\pi\lambda_L^2}\left(\psi^*\nabla\psi - \psi\nabla\psi^*\right) - \frac{|\psi|^2}{\lambda_L^2}\mathbf{A} . \tag{56}$$

Here the next designations were introduced

$$\Phi_0 = \frac{\pi\hbar c}{e} \tag{57}$$

quant of the flux;

$$\xi_S^2 = \frac{\hbar^2}{4m_f|\alpha|} \tag{58}$$

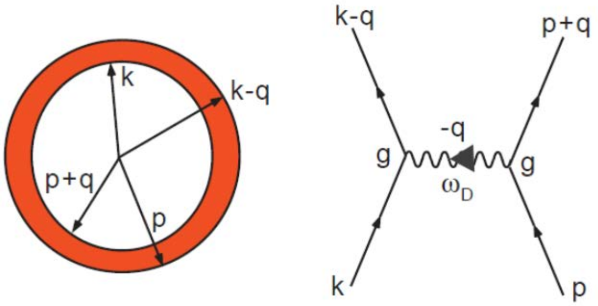
the length of coherency;

$$\lambda_L^2 = \frac{m_f c^2}{4\pi e^2 n_S} \tag{59}$$

penetration length.

To clarify the physical sense of the length of coherency, let us consider contact between the normal metal and superconductor (Fig.9). Suppose that the normal metal film is rather thin. It means that

Fig.9. Contact between the superconductor and thin normal metal film



$$\psi(0) \approx \psi(x \to \infty) = 1. \qquad (60)$$

Write now GL equation for order parameter (wave function of superconducting electrons) ψ

$$-\xi_s^2 \frac{d^2\psi}{dx^2} - \psi + \psi^3 = 0 \ . \qquad (61)$$

Taking into account Eq.(60), we can suppose that at vicinity of interface "normal metal-superconductor" we can write

$$\psi(x) = 1 - \varepsilon(x) \ , \ \ |\varepsilon(x)| << 1. \qquad (62)$$

Then, one obtains equation for small parameter ε(*x*)

$$\xi_s^2 \frac{d^2\varepsilon(x)}{dx^2} - 1 + \varepsilon(x) + [1 - \varepsilon(x)]^3 = \xi_s^2 \frac{d^2\varepsilon(x)}{dx^2} - 1 + \varepsilon(x) + 1 - 3\varepsilon(x) = 0, \qquad (63)$$

or

$$\xi_s^2 \frac{d^2\varepsilon(x)}{dx^2} - 2\varepsilon(x) = 0 \ . \qquad (64)$$

This equation has a solution

$$\varepsilon(x) = C \cdot e^{x/a}, \qquad (65)$$

where parameter *a* is found from characteristic equation $\xi_s^2 / a^2 - 2 = 0$, $a = \pm \xi_s / \sqrt{2}$. Because of $\varepsilon(x \to \infty) = 0$ and $\varepsilon(x = 0) = \varepsilon(0)$, we can write solution [Eq.(65)] in the form

Fig.10. Electron scattered by phonon exchange in the states of the shell of thickness ω_D about Fermi sphere

$$\varepsilon(x) = \varepsilon(0)\exp\left[-\sqrt{2}\,\frac{x}{\xi_s}\right]. \tag{66}$$

This solution sows us that the length ξ_s characterizes the distance at which parameter $\psi(x)$ is inhomogeneous. Other words, parameter ξ_s is the length at which wave function of supeconducting electrons $\psi(x)$ is coherent.

Let us rewrite the expression for penetration depth in the form $\lambda_L^2 = \frac{m_f c^2}{4\pi e^2 n_s} = \frac{m_f c^2}{8\pi e^2 |\Psi|^2} = \frac{m_f c^2 \beta}{8\pi e^2 |\alpha|}$.

Taking into account Eq.(58) we ca write

$$\begin{cases} \lambda_L^2 = \dfrac{m_f c^2 \beta}{8\pi e^2 |\alpha|}, \\ \xi_S^2 = \dfrac{h^2}{4m_f |\alpha|}. \end{cases} \tag{67}$$

Then, using the fact that dependence of parameter α on temperature is described by Eq.(46), one obtains

$$\begin{cases} \lambda_L : \dfrac{1}{\sqrt{T - T_C}}, \\ \xi_S : \dfrac{1}{\sqrt{T - T_C}}. \end{cases} \tag{68}$$

It means that penetration depth and length of coherency leads to infinity when $T \to T_C$ as $(T - T_C)^{-1/2}$.

The knowledge of superconductivity in the 1950's allows to suppose that superconducting current is provided with electrons pair and it was already suspected that the electron phonon interaction was responsible. Leon Cooper considered interaction between of two electrons which states are in the vicinity of Fermi sphere. Cooper showed that the noninteracting electrons are unstable towards the addition of a single pair of electrons with attractive interactions. Earlier Frölich showed that the electron-phonon interaction can be origin of the attractive potential. According to Cooper, the electron-phonon interaction is strongest for those electrons with single-particle energies $\in_{\mathbf{k}}$ within the shell of thickness $\omega_{\mathbf{q}}$ at Fermi sphere (Fig.10). To

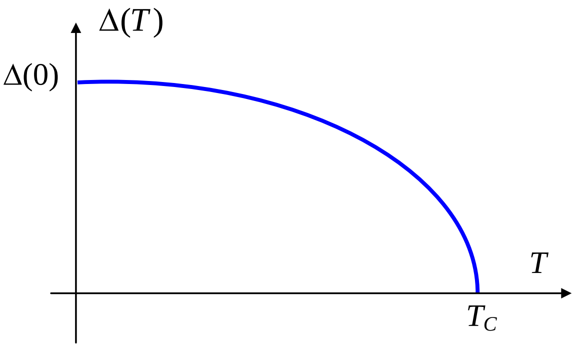


Fig. 12. The dependence of order parameter on temperature

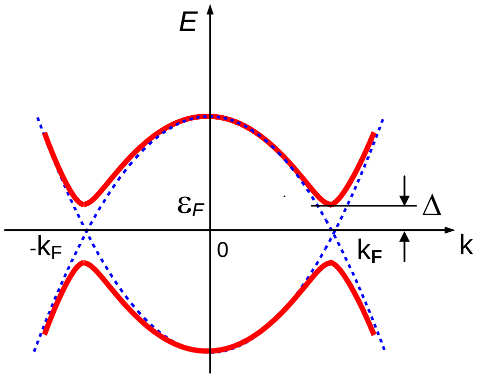
understand why the electron-phonon interaction leads to electron pairs formation, Cooper supposed that effective electron-electron interaction can be presented in the form

$$V_{kk'} = \begin{cases} -V \ , & \in_k, \in_{k'} < \omega_C \\ 0 \ , & \text{otherwise} \end{cases} \tag{69}$$

As a result, Cooper obtained the bound state equation in the form

$$u_k = \frac{V \sum_{k'} u_{k'}}{2E_k - E} \ , \tag{70}$$

where summation is only over the states k such that $E_F < E_k < E_F + \omega_C$. This supposition allows us to eliminate u_k by the summing both sides of equation (70) $\sum_k \ldots$.



This yields

$$\frac{1}{V} = \sum_k \frac{1}{2E_k - E} \approx \frac{1}{2} N_0 \int_{E_F}^{E_F+\omega_C} d\xi \frac{1}{2\xi - E} = \frac{1}{2} N_0 \ln \frac{2E_F + 2\omega_C - E}{2E_F - E} \ , \tag{71}$$

where N_0 is a density of states at the Fermi surface. Then, for a weak coupling ($N_0 V << 1$), one obtins the pair binding energy

$$\Delta_C \approx \sim 2\omega_C e^{-2/N_0 V} \ . \tag{72}$$

The energy Δ_C is an energy deficit when two electrons create the Cooper pair. Other words, there is the energy of pair destruction.

Fig.11. The dispersion curves for elementary excitations of superconductor in the BCS model.

Bardeen, Cooper and Schrieffer developed the microscopic theory of superconductivity. This theory, in part gives the dispersion relation for elementary excitations of fermion-type in the form

$$\in(\mathbf{k}) = \left(\eta_{\mathbf{k}}^2 + \Delta^2\right)^{1/2} , \tag{73}$$

with Δ energy gap, and $\eta_{\mathbf{k}} = E(\mathbf{k}) - E_F$. The dispersion curves have a view shown in Fig.11. Analyzing the dispersion relation of superconducting state, one can suppose that the energy gap (order parameter) should be depended on temperature. Indeed, more destroyed pairs – more elementary excitations arise in superconductor which leads to decreasing energy gap. At least one can write

$$\begin{cases} \Delta(T=0) = \Delta, & \text{maximum value of the energy gap} \\ \Delta(T=T_C) = 0, & \text{at } T = T_C \text{ the energy gap vanishes} \end{cases} \tag{74}$$

BCS theory gives the equation for $\Delta(T)$

$$\frac{1}{N_0 V} = \int_0^{\hbar\omega_{ph}} \frac{d\nu}{\sqrt{\nu^2 + |\Delta(T)|^2}} \,\mathrm{th}\, \frac{\sqrt{\nu^2 + |\Delta(T)|^2}}{2T} \quad . \tag{75}$$

The dependence of the order parameter on temperature according to Eq.(75) is shown in Fig.12. As we can see, order parameter satisfies Eq.(74). Supposing in Eq.(75) $\Delta(T \to T_C) \to 0$, one obtains from Eq.(75) the equation for T_C

$$\frac{1}{N_0 V} = \int_0^{\hbar\omega_{ph}} \frac{d\nu}{\nu} \,\mathrm{th}\, \frac{\nu}{2T_C} \quad . \tag{76}$$

The solution of this equation is written in the form

$$T_C = 1.14\hbar\omega_{ph} \exp\left(-1/N_0 V\right) \quad . \tag{77}$$

This equation, in part, reflects the fact that the exchange by phonons between electrons is the main factor of superconductivity arising.

Consider the superconductor ring acted by external magnetic field shown in Fig.13. Due to an induction the permanent current is generated inside the ring. It is known, that the magnetic flux can be written as the product of the self-inductance L of the ring and the current I flowing along the ring

$$\Phi = L \cdot I\,. \tag{78}$$

The magnetic flux through the simple metallic ring could take any arbitrary value. Unlike the magnetic flux through a superconducting ring can take up only discrete values

$$\Phi = n\Phi_0\,, \tag{79}$$

where the flux quantum is written via the Plank constant h and charge of two electrons – $2e$ which reflects that the superconducting current is formed by the pairs of electrons (Cooper pairs)

$$\Phi_0 = \frac{h}{2e}\,. \tag{80}$$

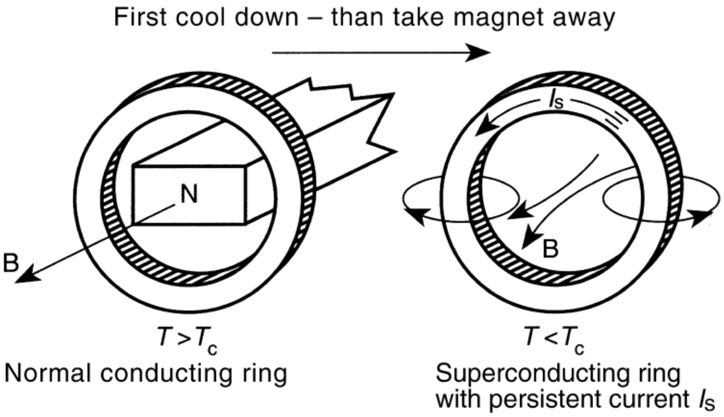


Fig.13. the flux quantization in the superconducting ring

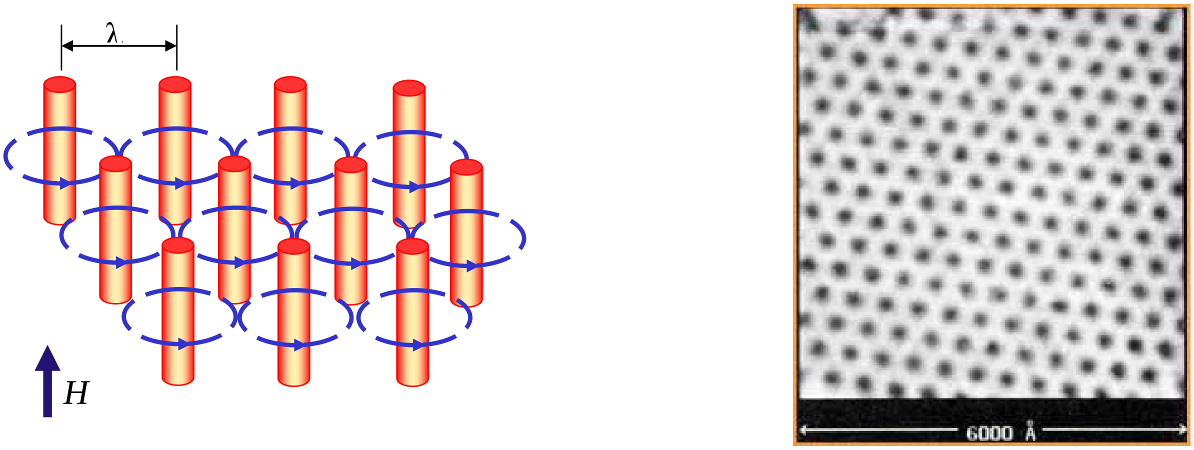


Fig.14. The Abrikosov vertexes in the superconductor II type. Sketch (left); The STM image of abrikosov vetrtexes lattice in $NbSe_2$ at T = 1.8 K under magnetic field H = 1 T action (H.F. Hess et al., Phys. Rev. Lett. 62, 214 (1989))

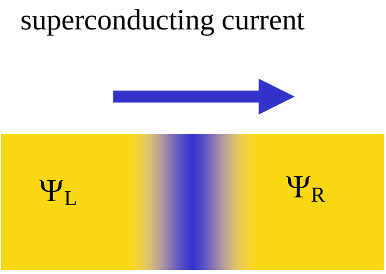


Fig.15. Stationary Josephson effect.

Note the mixed state of superconductor forms with Abrikosov vertexes, the flux quantum located at each of the vertex. The magnetic field flux quantization is the reflection of quantum phenomena in superconductivity.

B.Josephson in 1962 published the paper in which the prediction of two new effects was made. The effects have to be observed in the tunnel transitions between the superconductors. First effect consists in arising the current between the dielectric film separating two superconductor (Fig.15). This is the

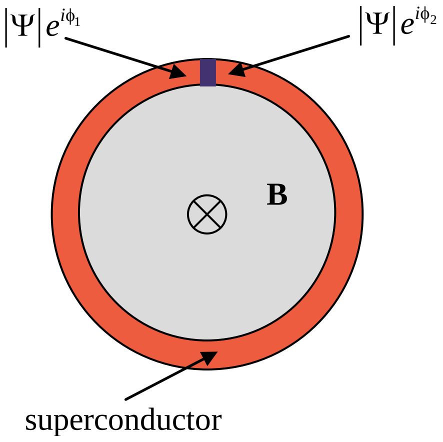


Fig.17. Josephson junction under magnetic field action

stationary Josephson effect. The current defines by different of the wave functions phases in the superconductors $\varphi_J = \vartheta_R - \vartheta_L$

$$I_s = I_c \sin \varphi_J \, . \tag{81}$$

Note that normal current does not arise under the such conditions. In the case when the supercurrent becomes large that any critical value, the tunnel transition radiates the electromagnetic radiation.

Consider the superconducting ring with the Josephson junction (Fig.16). Let the external magnetic field acts to the ring. As a result, the superconducting current arises in the ring. The current satisfies the expression $\int_C \mathbf{j}_S d\mathbf{l} = 0$ which immediately gives

$$\int_C \left(\nabla_{\mathbf{r}} \phi - \frac{2e}{\hbar c} \mathbf{A} \right) d\mathbf{l} = \phi_2 - \phi_1 - \frac{2e}{\hbar c} \Phi \quad , \tag{82}$$

where $\mathbf{\Phi} = \int_C \mathbf{A} d\mathbf{l} = \int \mathbf{B} d\mathbf{S}$ is the magnetic field flux through the ring.

Then, we obtain

$$\phi_2 - \phi_1 = 2\pi \frac{\Phi}{\Phi_0} \, . \tag{83}$$

From the other side, according Josephson effect [Eq.(81)] the current arises between the both sides of Josephson junction. Then, one can obtain

$$I_S = I_C \sin\left(2\pi \Phi / \Phi_0 \right) . \tag{84}$$

Because Φ_0 is a small value, the very small changes of magnetic field acting to the Josephson junction can cause the rather large changes of the current. This fact lies in the base of the superconducting quantum interference device – SQUID. SQUID consists of two Josephson junctions in the superconducting ring (Fig.18). The current inside the SQUID depends on the magnetic flux by expression

$$I = I_C \left[\sin \gamma_{12} + \sin\left(\gamma_{12} + \Phi \right) \right], \tag{85}$$

with

$$\gamma_{ij} = \frac{\Phi_0}{2} \left(\phi_2 - \phi_1 \right) + \int_1^2 dl \cdot A \, . \tag{86}$$

Eq.(85) means that the current will oscillate with the flux changing. As a result, SQUID can measure the magnetic field with very large sensitivity. Megnetometers based on SQUID can achieve the sensitivity ~10^{-11} Gs (note, magnetic field of Earth is about 10^{-1} Gs). SQUID is the most precise device for measurement of magnetic fields. It is used in different branches of modern technologies since earthquake prediction to medical diagnostics.

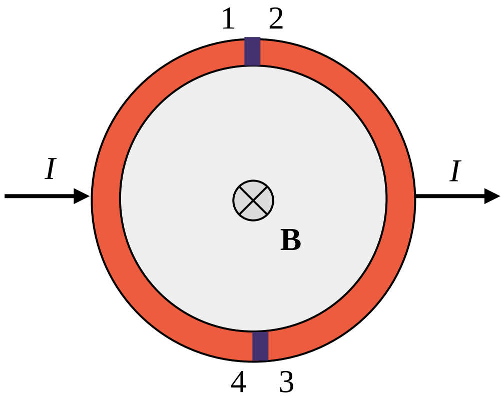


Fig.18. Two Josephson junctions being superconducting quantum interference device

Chapter 18. Porous silicon

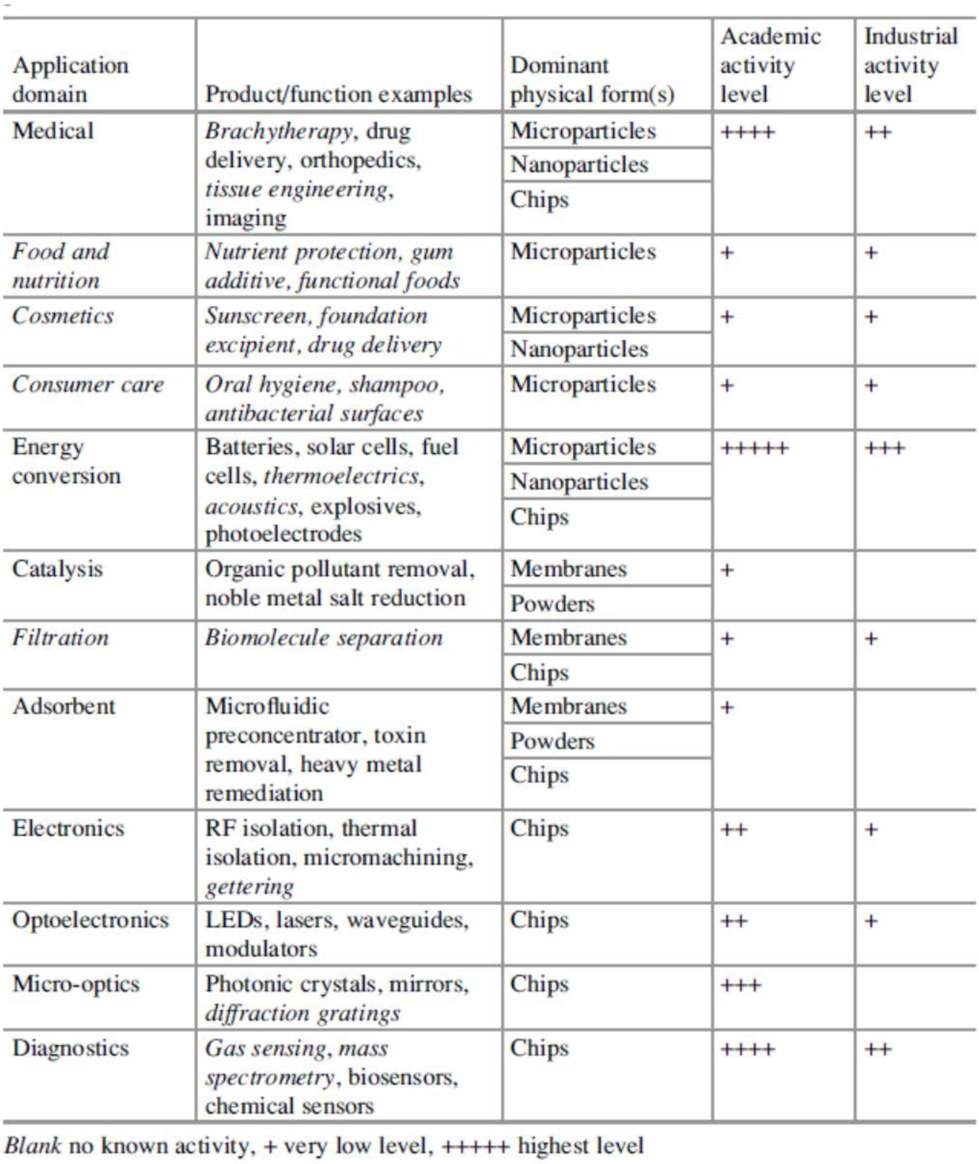
18.1 Introduction

Porous silicon was accidentally discovered by the Uhlirs in the mid 1950s. They were trying to develop an electrochemical method to machine silicon wafers for use in microelectronic circuits. Under the appropriate electrochemical conditions, the silicon wafer did not dissolve uniformly as expected, but instead free holes appeared, propagating primarily in the < 100 > direction in the wafer. Since this did not provide the smooth polish desired, the curious result was reported in a Bell labs technical note , and then the material was more or less forgotten. In the 1970s and 1980s a significant level of interest arose because the high surface area of porous silicon was found to be useful as a model of the crystalline silicon surface in spectroscopic studies, as a precursor to generate thick oxide layers on silicon, and as a dielectric layer in capacitance - based chemical sensors. Interest in porous silicon, and in particular in its nanostructure, exploded in the early 1990s when Ulrich Goesele at Duke University identified quantum confinement effects in the absorption spectrum of porous silicon, and almost simultaneously Leigh Canham at the Defense Research Agency in England reported efficient, bright red – orange photoluminescence from the material. The quantum confinement effects arise when the pores become extensive enough to overlap with each other, generating nanometer - scale silicon filaments. As expected from the quantum confinement relationship, the red to green color of photoluminescence occurs at energies that are significantly larger than the bandgap energy of bulk silicon (1.1 eV, in the near - infrared).

Porous silicon has been fabricated by both “top-down” techniques from solid silicon and “bottom-up” routes from silicon atoms and silicon-based molecules. Over the last 50 years, electrochemical etching has been the most investigated approach for chip-based applications and has been utilized to create highly directional mesoporosity and macroporosity. Chemical conversion of porous or solid silica is now receiving increasing attention for applications that require inexpensive mesoporous silicon in powder form. Very few techniques are currently available for creating wholly microporous silicon with pore size below 2 nm.

With the discovery of efficient visible light emission from porous silicon came a flood of work focused on creating silicon - based optoelectronic switches, displays, and lasers. Problems with the material are chemical and mechanical stability, and its disappointingly low electroluminescence efficiency led to a waning of interest by the mid 1990s. In the same time period, the unique features of the material – large surface area, controllable pore sizes, convenient surface chemistry, and compatibility with conventional silicon microfabrication technologies – inspired research into applications far outside optoelectronics. Many of the fundamental chemical stability problems have been overcome as the chemistry of the material has matured, and various biomedical sensor, optics, and electronics applications have emerged Scientific interest in porous silicon has grown significantly over the last 25 years fundamentally because it is a form of silicon that has both highly tunable and remarkable properties. These properties can be dramatically different from those of solid silicon and have enabled opportunities to arise in diverse fields that started in electronics but now include microsystems, optoelectronics, optics, acoustics, energy conversion, diagnostics, nutrition, medical therapy etc. Table 18.1 provides more detail for each application domain, providing examples of porous silicon specific products or functions and a qualitative indication of the level of industrial and research activity.

Table.18.1 Application of PS.



Application domain	Product/function examples	Dominant physical form(s)	Academic activity level	Industrial activity level
Medical	*Brachytherapy*, drug delivery, orthopedics, *tissue engineering*, imaging	Microparticles Nanoparticles Chips	++++	++
Food and nutrition	*Nutrient protection, gum additive, functional foods*	Microparticles	+	+
Cosmetics	*Sunscreen, foundation excipient, drug delivery*	Microparticles Nanoparticles	+	+
Consumer care	*Oral hygiene, shampoo, antibacterial surfaces*	Microparticles	+	+
Energy conversion	Batteries, solar cells, fuel cells, *thermoelectrics*, *acoustics*, explosives, photoelectrodes	Microparticles Nanoparticles Chips	+++++	+++
Catalysis	Organic pollutant removal, noble metal salt reduction	Membranes Powders	+	
Filtration	*Biomolecule separation*	Membranes Chips	+	+
Adsorbent	Microfluidic preconcentrator, toxin removal, heavy metal remediation	Membranes Powders Chips	+	
Electronics	RF isolation, thermal isolation, micromachining, *gettering*	Chips	++	+
Optoelectronics	LEDs, lasers, waveguides, modulators	Chips	++	+
Micro-optics	Photonic crystals, mirrors, *diffraction gratings*	Chips	+++	
Diagnostics	*Gas sensing, mass spectrometry*, biosensors, chemical sensors	Chips	++++	++

Blank no known activity, + very low level, +++++ highest level

18.2 .Electrochemical formation

18.2.1 Cross-sectional view of a single-tank anodization cell

Usualy, PS is formed by an electrochemical etching of Si in an HF solution. Following an electrochemical reaction occurring at the Si surface a partial dissolution of Si settles in. Let us concentrate on the various factors which rule this process.

HF is sold in an aqueous solution with up to 50% of HF. Thus, the first attempts to form PS were performed using only HF diluted in deionized and ultra-pure water. Due to the hydrophobic character of the clean Si surface, absolute ethanol is usually added to the aqueous solution to increase the wettability of the PS surface. In fact, ethanoic solutions infiltrate the pores, while purely aqueous HF solutions do not. This is very important for the lateral homogeneity and the uniformity of the PS layer in depth. In addition, during the reaction there is hydrogen evolution. Bubbles form and stick on the Si surface in pure aqueous solutions, whereas they are promptly removed if ethanol (or some other surfactant) is present. For the same reason, a careful design of the anodization cell is necessary in order to promote hydrogen bubble removal. Moreover, it has been found that lateral inhomogeneity and surface roughness can be reduced, increasing electrolyte viscosity, either by diminishing the temperature or introducing glycerol to the composition of the HF solution.

The dissolution is obtained either controlling the anodic current or the potential. Generally, it is preferable to work with constant current, because it allows a better control of porosity, thickness and reproducibility of the PS layer. The simplest electrochemical cell is a Teflon beaker . The Si wafer acts as the anode and the cathode is generally made of platinum, or other HF-resistant and conductive material. The cell body is usually made of an highly acid resistant polymer such as Teflon. Using this cell PS is formed all over the wafer surface exposed to HF, including the cleaved edges. The advantages of this cell geometry are the simplicity of equipment and the ability to anodize silicon-on-insulator structures. The drawback is the inhomogeneity in porosity and thickness of PS layers, mainly due to a potential drop. In fact, there is a difference in potential between the top and the bottom, leading to different values of local current density.

The second type of anodization cell (1) is shown in Fig. 18.1b. In this cell, the Si wafer (2) is placed on a metal disk (6) and sealed through an O-ring (4), so that only the front side of the sample is exposed to the electrolyte (5). When an Si wafer with high resistivity (i.e. more than a few Ohm/cm) is used, a high dose implantation on the back surface of the wafer is required to improve the electrical contact between the wafer and the metal disk. This step is crucial to get later homogeneity in the PS layer. This cell is the most widely used because it leads to uniform PS layers, allows an easy control of both porosity and thickness, and it is suitable for front side illumination of the sample during the attach.

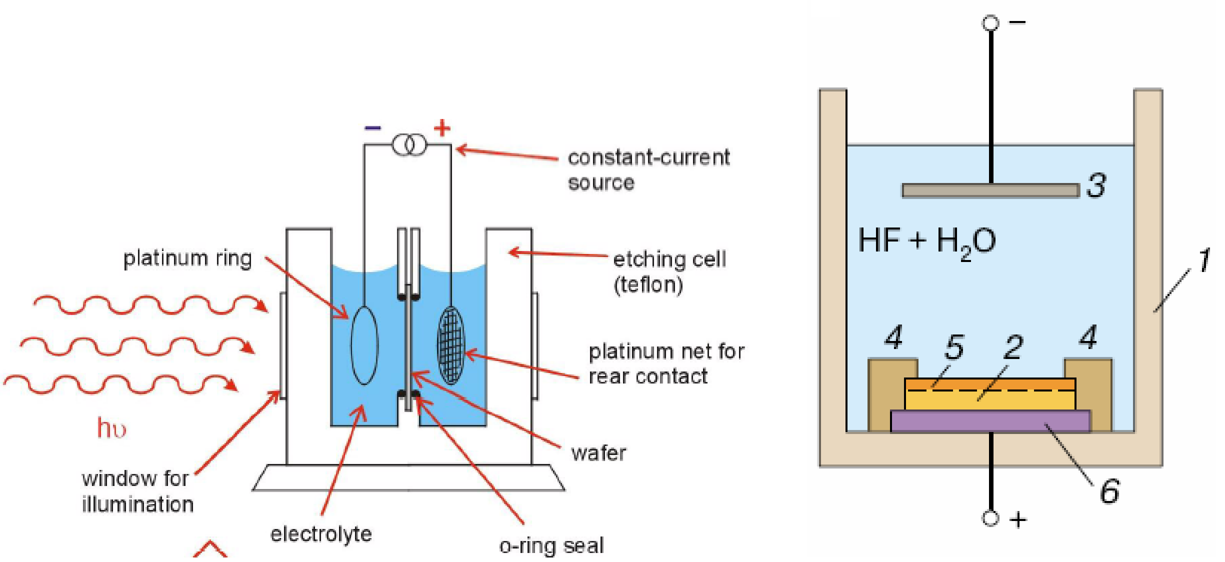


Fig. 18.1 Cross-sectional view of a double-tank anodization cell and single-tank anodization cell

The third type of anodization cell is a double-tank geometry with an electrolytic back-side contact. This cell (Fig. 18.1a) consists of two half-cells in which Pt electrodes are immersed and the Si wafer is used to separate the two half-cells. HF solution, circulated by chemical pumps to remove the gas bubbles and avoid the decrease in the local concentration of HF, is used both to etch the front side and as a back contact. A better uniformity is obtained using symmetrical and large Pt plates as the cathode and the anode. The current flows from one electrode to the other through the Si wafer. While stirring of the solution has shown no major effects on the quality of the PS layer, the HF circulation during the etch helps to achieve a good depth uniformity and it is also preferable for security reasons. The use of closed loop pumps to circulate the electrolyte is possible also for the other geometries. The back-side of the Si wafer acts as a secondary cathode where proton reduction takes place leading to hydrogen evolution, while the front side of the wafer acts as a secondary anode, where PS is formed. Since the back-side contact is made electrolytically, no metalization is required, but a high-dose implantation is still necessary for high resistive wafers. The uniformity of the layers obtained under these conditions is comparable to that with a single-tank cell. Most of the problems encountered with the solid back contact in highly resistive samples are greatly reduced. If illumination is required, the material used in the cell should be Plexiglas, which is transparent and HF resistant (up to 15% HF). With this cell, both front and back-side illumination is possible.

18.2.2. Current-voltage characteristics

When a potential is applied to Si in aqueous solution, a measurable external current flows through the system. However, for any current to pass the Si/electrolyte interface, it must first change from electronic to ionic current. This means that a specific chemical redox reaction must occur at the Si interface. Application of a potential then induces a precise chemical reaction, the nature of which is fundamental to the formation of PS. Fig. 18.2 shows the ``typical" I-V curves for n- and p-type doped Si in aqueous HF. The I-V curves show some similarities to the normal Schottky diode behavior expected for a semiconductor/electrolyte interface, but there are some important differences. For instance, while the sign of majority carriers changes between n- and p-type, the chemical reactions at the interface remain the same. Moreover, the reverse-bias dark currents are at least three orders of magnitude higher than normal Schottky diode expectations. Another anomaly is the open circuit potential for n- and p-type doped Si, which is not consistent with the difference between bulk Fermi levels.

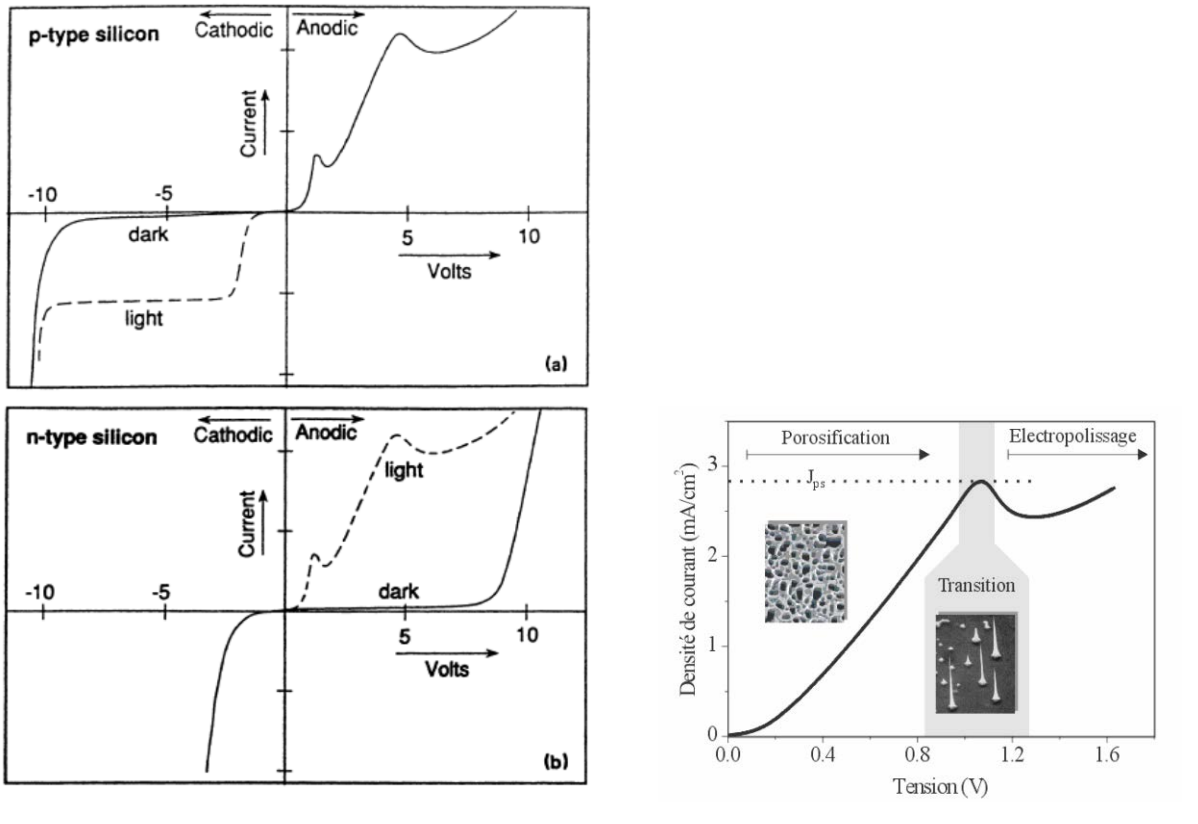


Fig. 18.2. Typical I-V curves for (a) p-type and (b) n-type silicon, (c) shows the I-V curve at low positive potential

Under cathodic polarization, for both n- and p-type materials, Si is stable. The only important cathodic reaction is the reduction of water at the Si/HF interface, with formation of hydrogen gas. This usually occurs only at high cathodic overpotentials, or using Schottky diode terminology, at reverse breakdown. Under anodic polarization Si dissolves. At high anodic overpotentials the Si surface electropolishes and the surface retains a smooth and planar morphology. In contrast, with low anodic overpotentials, the surface morphology is dominated by a vast labyrinth of channels that penetrate deep into the bulk of the Si. PS is formed. Pore formation occurs only in the initial rising part of the I-V curve for a potential value below the potential of the small sharp peak (see Fig. 18.2). The current peak is called the electropolishing peak. The quantitative values of the I-V curves, as well as the values corresponding to the electropolishing peak, depend on etching parameters and wafer doping. For n-type substrates, this typical I-V behavior is observed only under illumination because hole supply is needed.

18.2.3. Dissolution chemistries

The exact dissolution chemistries of Si are still in question, and different mechanisms have been proposed. However, it is generally accepted that holes are required for both electropolishing and pores formation. During pore formation two hydrogen atoms evolve for every Si atom dissolved. The hydrogen evolution diminishes approaching the electropolishing regime and disappears during electropolishing. Current efficiencies are about two electrons per dissolved Si atom during pore formation, and about four electrons in the electropolishing regime. The global anodic semireactions can be written during pore formation as

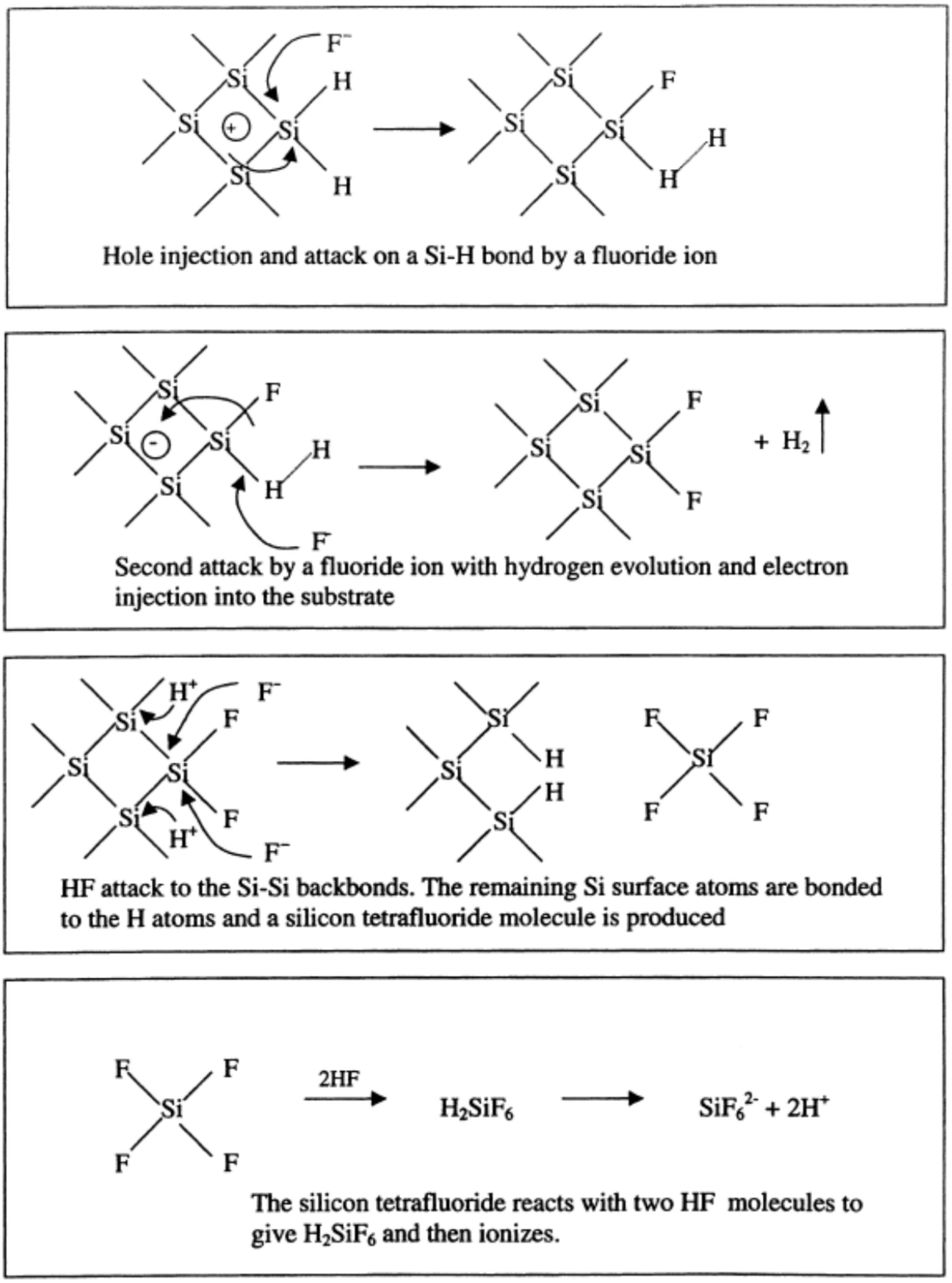
$$\mathrm{Si} + 6\mathrm{HF} \rightarrow \mathrm{H_2SiF_6} + \mathrm{H_2} + 2\mathrm{H^+} + 2e^-$$

\

and during electropolishing as

$$\mathrm{Si} + 6\mathrm{HF} \rightarrow \mathrm{H_2SiF_6} + 4\mathrm{H^+} + 4e^-$$

The final and stable product for Si in HF is in any case H_2SiF_6 or some of its ionized forms. This means that during pore formation only two of the four available Si electrons participate in an interfacial charge transfer, while the remaining two undergo a corrosive hydrogen liberation. In contrast, during electropolishing, all four Si electrons are electrochemically active. Lehmann and Gosele have proposed a dissolution mechanism which is so far the most accepted (Fig. 18.3). It is based on a surface bound oxidization scheme, with hole capture, and subsequent electron injection, which leads to the divalent Si oxidization state.



Hole injection and attack on a Si-H bond by a fluoride ion

Second attack by a fluoride ion with hydrogen evolution and electron injection into the substrate

HF attack to the Si-Si backbonds. The remaining Si surface atoms are bonded to the H atoms and a silicon tetrafluoride molecule is produced

The silicon tetrafluoride reacts with two HF molecules to give H_2SiF_6 and then ionizes.

Fig. 18.3. Silicon dissolution scheme proposed by Lehmann and Gosele.

According to Fig. 18.3, the Si hydride bonds passivate the Si surface unless a hole is available. This hypothesis is also supported by the experimental observation that hydrogen gas continues to evolve from the porous layer after the release of the applied potential for a considerable time. In addition, various spectroscopic techniques have confirmed the presence of Si-H surface bonds during PS formation.

18.2.4 Pore formation

While it is generally accepted that pore initiation occurs at surface defects or irregularities, different models have been proposed to explain pore formation in PS. Some basic requirements have to be fulfilled for electrochemical pore formation to occur:

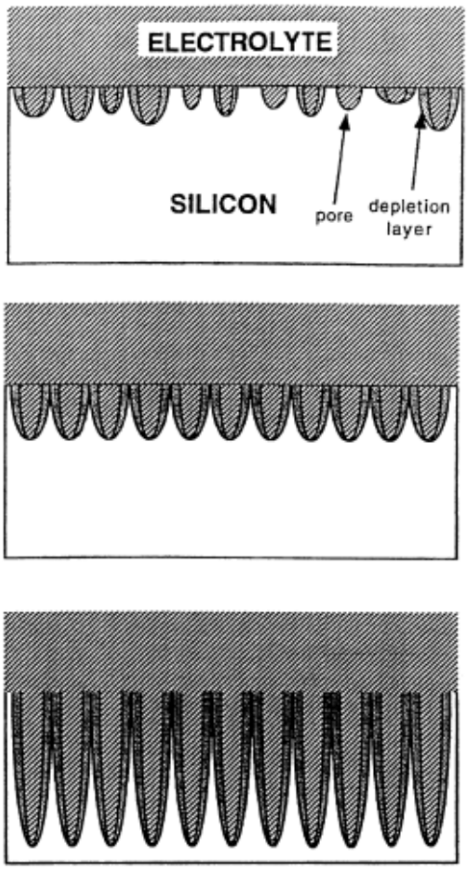


Fig. 18.4. Mechanism of pore formation in PS. Random pore initiation on the Si surface (top panel), formation of depletion layers and directional growth of pores (middle panel) and dissolution advance only at the pore tips (bottom panel).

1. Holes must be supplied by bulk Si, and be available at the surface.
2. While the pore walls have to be passivated, the pore tips have to be active in the dissolution reaction. Consequently, a surface which is depleted of holes is passivated to electrochemical attach, which means that (i) the electrochemical etching is self-limiting and (ii) hole depletion

occurs only when every hole that reaches the surface reacts immediately. The chemical reaction is not limited by mass transfer in the electrolyte.

3. The current density should be lower than the electropolishing critical value. For current densities above such a value, the reaction is under ionic mass transfer control, which leads to a surface charged of holes and to a smoothing of Si surface (electropolishing). The behavior at high current densities turns out to be useful to produce PS free-standing layers. Raising the current density above the critical value at the end of the anodization process results in a detachment of the PS film from the Si substrates.

In the low current density regime, where PS forms, some considerations apply:

-A surface region depleted in mobile carriers is formed at the Si/electrolyte interface (Fig. 18.4). This region is highly resistive (comparable to intrinsic Si). The thickness of the depleted region depends on the doping density. It is several mm thick for lightly n-type doped Si. It is thin for highly n- or ptype doped Si, and it does not exist for lightly to moderately p-type doped Si (fig.18.5).

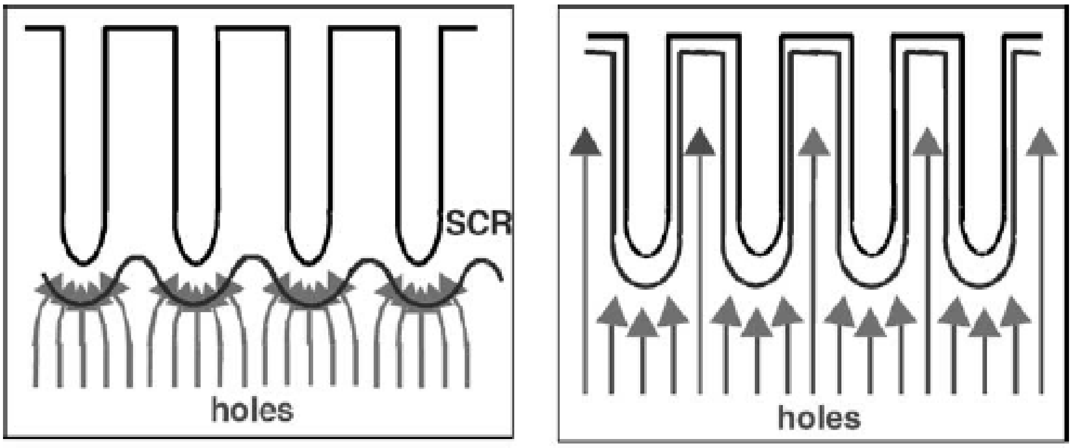


Fig.18.5 Influence of space charge region in interface electrolyte/Si on PS formation

-The size of the pores is related both to the depletion layer width and to the mechanism of charge transfer (fig.18.5) .

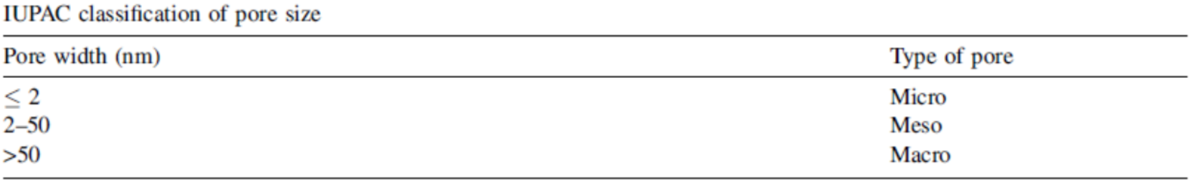
-In highly doped substrates charge transfer is dominated by tunneling of the carriers, and the pore size reflects the width of the depletion region, being typically around 10 nm.

-In lightly n-type doped Si anodized in the dark, generation of carriers occurs at breakdown. The pore dimensions are about 10-100 nm (mesopores), regardless of doping density. Under illumination the pore size is dependent on doping density and anodization conditions, with diameters in the range 0.1-20 mm (macropores).

-A hole depletion is expected in any case if the dimensions of the nanocrystals are about few nanometers, independent of the substrate type and doping. In this size region, quantum confinement is effective and the Si band gap is increased. A hole needs to overcome an energy barrier to enter this region. This is highly improbable. The quantum confinement is responsible for pore diameters below 2 nm, denoted as micropores. Micropores can be found on every type of PS samples, but only in moderately and lightly p-type doped substrates pure microPS exists.

-Both mechanisms coexist during PS formation, resulting in a superposition of micro and meso (or macro) structures (see Table 18.2 for a classification), whose average size and distribution depend on substrate and anodization condition.

Tabl.18.2 clasification of of pore size.



IUPAC classification of pore size

Pore width (nm)	Type of pore
≤ 2	Micro
2–50	Meso
>50	Macro

18.2.5 Effect of anodization conditions

All the properties of PS, such as porosity, thickness, pore diameter and microstructure, depend on anodization conditions. These conditions include HF concentration, current density, wafer type andresistivity, anodization duration, illumination (n-type mainly), temperature, ambient humidity and drying conditions (see Table 18.3). Porosity is defined as the fraction of void within the PS layer and can be easily determined by weight measurements. The wafer is weighted before anodization (m_1), just after anodization (m_2), and after a rapid dissolution of the whole porous layer in a 3% KOH solution (m_3). The porosity is given by the following equation:

$$P(\%) = \frac{(m_1 - m_2)}{(m_1 - m_3)}.$$

Guessing the Si density ρ, one can also get the PS layer thickness d

$$d = \frac{m_1 - m_3}{\rho S},$$

where S is the etched surface.

Tabl.18.3

Effect of anodization parameters on PS formation

An increase of . . . yields a	Porosity	Etching rate	Critical current
HF concentration	Decreases (see text)	Decreases	Increases
Current density	Increases	Increases	–
Anodization time	Increases	Almost constant	–
Temperature	–	–	Increases
Wafer doping (p-type)	Decreases	Increases	Increases
Wafer doping (n-type)	Increases	Increases	–

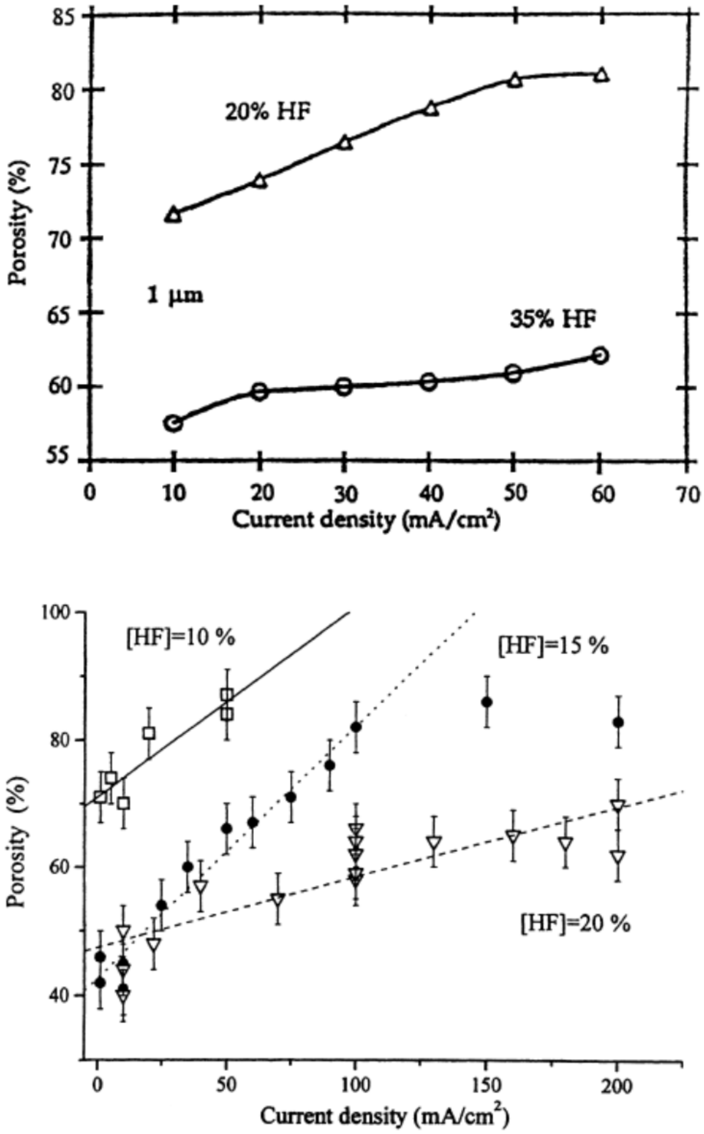


Fig. 18.6. Porosity as a function of current densities for different HF concentrations. Highly doped (bottom panel) and lightly doped (top panel) p-type silicon substrate.

For p-type doped substrates, and for a given HF concentration the porosity increases with increasing current density. For fixed current density, the porosity decreases with HF concentration (see Fig. 18.6). With fixed HF concentration and current density, the porosity increases with thickness and porosity.

The porosity gradients in depth occur. This happens because of the extra chemical dissolution of PS layer in HF. The thicker the layer, the longer the anodization time, and the longer the residence of Si in the HF reach solutions, the higher the mass of chemically dissolved PS. This effect is much more important for lightly doped Si, while it is almost negligible for heavily doped Si, because of the lower specific surface area. For heavily n-type doped Si, the porosity as a function of current density is quite different from the corresponding curves obtained for p-type doped substrates. The porosity exhibits a sharp minimum around 20mA/cm^2 (see Fig. 18.7). For higher current densities the behavior is similar to the p-type doped substrates, but for lower current densities the porosity increases sharply. This large increase in porosity is not explained simply by chemical dissolution (with given thickness, the lower the current density, The longer the anodization time), but it is due to a difference in microstructure. In n-type doped Si, the layers obtained at low current density have a finer structure and are, therefore, more luminescent.

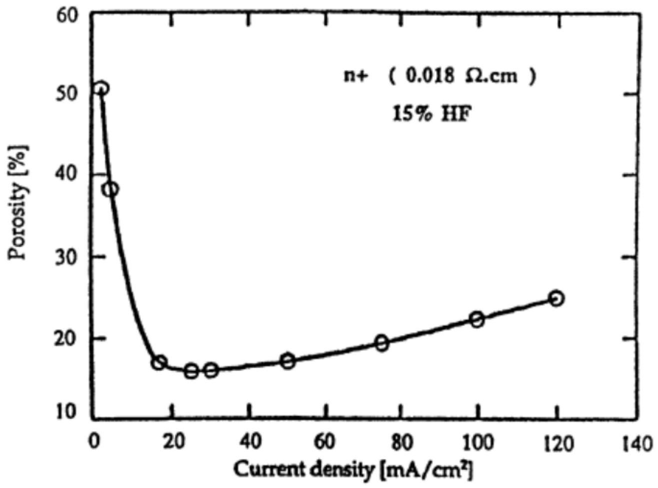


Fig. 18. 7 Porosity as a function of current densities for highly doped n-type silicon substrate.

To obtain higher porosities, with samples emitting in the green and blue regions, a two-step process has been proposed. The PS layers are prepared by standard anodization which is followed by illumination under open-circuit conditions in the same electrochemical cell for 0-15 min. The formation of PS is selective with respect to the doping of the substrate. Heavily doped regions are etched faster than low doped regions, in the dark, n-type doped regions embedded into p-type doped regions are not attacked, controlled doping profiles result in controlled PS formation. The dependence of the PS formation on the doping of the Si wafer has been also exploited for impurity profiling

18.3. Morphology of the porous silicon.

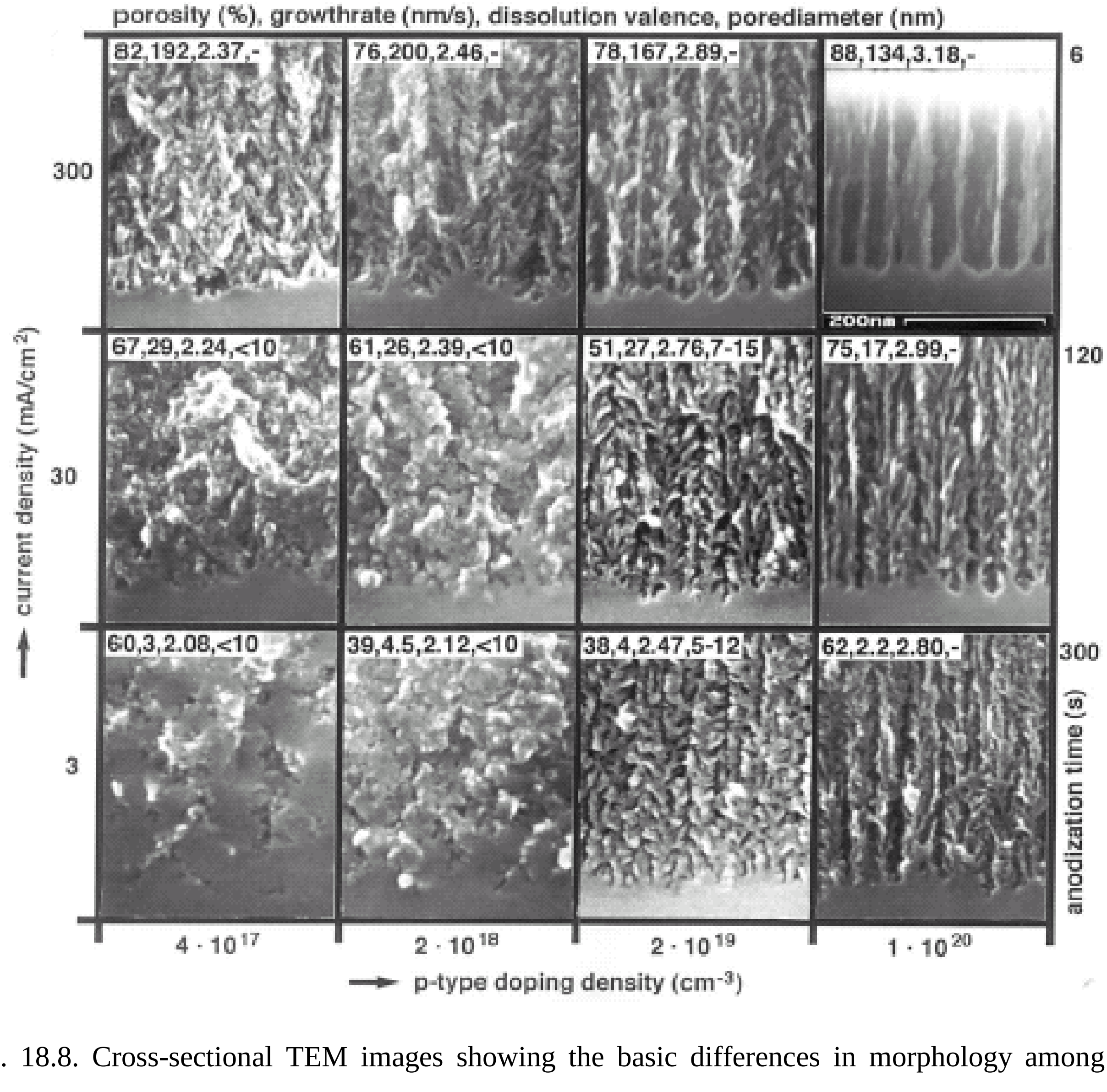


Fig. 18.8. Cross-sectional TEM images showing the basic differences in morphology among different types of samples

Porosity is a macroscopic parameter which helps in discussing trends but which does not give microscopic information on the morphology of the PS layers. Information on pore sizes and shapes are not easy to obtain, and different experimental techniques have been used, depending on the sizes of the pores. A better knowledge of the PS structure can be achieved if the pore shape, size and topological distribution are determined. However, no complete understanding of the mechanisms that determine PS morphology still exists, because of the large number of parameters involved (fig.18.8).

Nevertheless, some general trends can be derived for different types of starting Si substrates. Fig. 9 shows four cross-sectional TEM images of PS samples with different starting substrates. The difference in morphologies is evident. For p-type doped Si both pore size and inter-pore spacing are very small, typically between 1 and 5 nm, and the pore network looks very homogeneous and interconnected. As the dopant concentration increases, pore sizes and inter-pore spacing increase, while the specific surface area decreases. The structure becomes anisotropic, with long voids running perpendicular to the surface, very evident in highly p-type doped Si (p.), as shown in Fig. 18.8.

For n type doped Si the situation is more complicated. Generally, pores in n-type doped Si are much larger than in p-type doped Si, and pore size and inter-pore spacing decreases with increasing dopant concentrations. Lightly doped n-type substrates anodized in the dark have low porosity (1-10%) with pores in the micrometer range. Under illumination higher values of porosity can be achieved, and mesopores are formed together with macropores. The final structure depends strongly on anodization conditions, especially on light intensity and current density. While highly n- and p-type doped Si show similar structures, in n-type doped Si pores form a randomly directed filamentary net and tend to ``pipe" forming large straight channels approaching electropolishing regime. Both filaments and channels propagate only in the 100 direction, probably because the (1 0 0) planes present the most sterically favored geometry for the chemical attach.

18.3.2 Macroporous silicon formation

Macropores formation in p-type doped Si is more difficult. First reports claimed that high resistivity wafers and peculiar electrolytes (e.g. water-free acetonitrile) were needed. Later it was discovered that many electrolytes are suitable and that the substrate doping was not critical. The structure of the etched pores are not as good as with n-type doping, but the typical sizes are smaller. Pores elongate preferentially along the (1 1 1) or the (3 1 1) directions, and the use of O-free etchant yields better results (fig.18.9).

The etching of n-silicon at transient regime (near peak, fig18.2) allows to obtain tube, pilar of wall –type structures (fig.18.10)

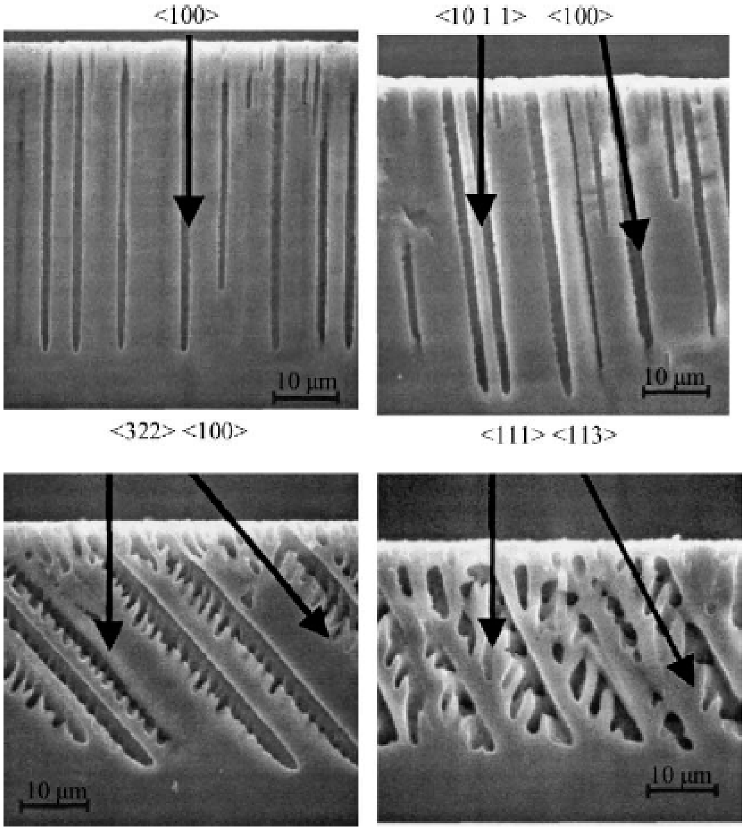
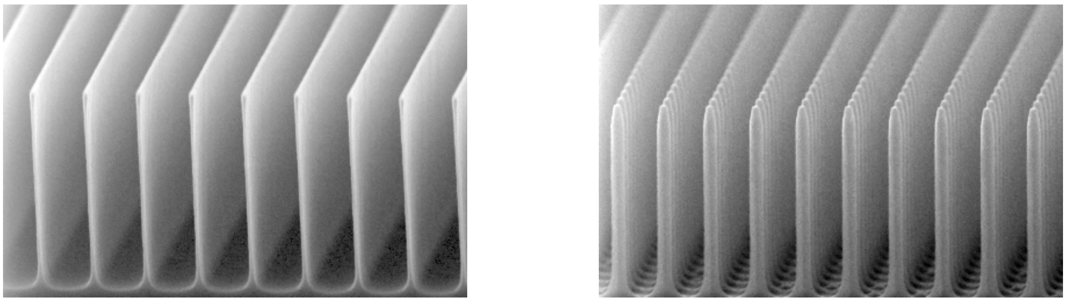
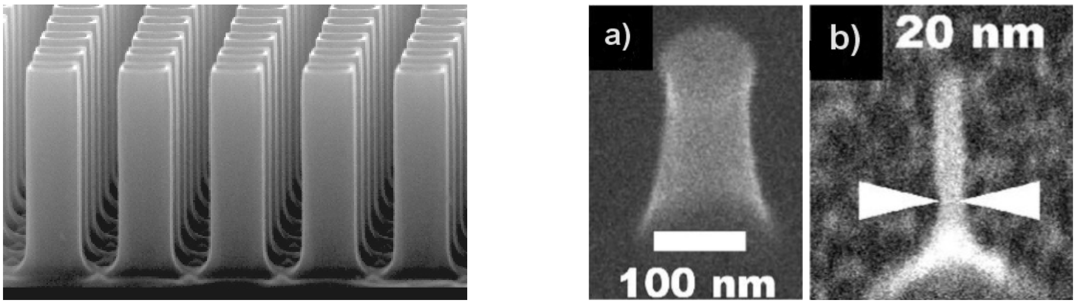


Fig. 18.9. Macropores in n-silicon of different orientations



Réseau de Murs
Période 1.6 µm ; Largeur 110 nm ; hauteur 7 µm

Réseau de Piliers
Période 1.6µm ; Largeur 450 nm ; hauteur 7 µm



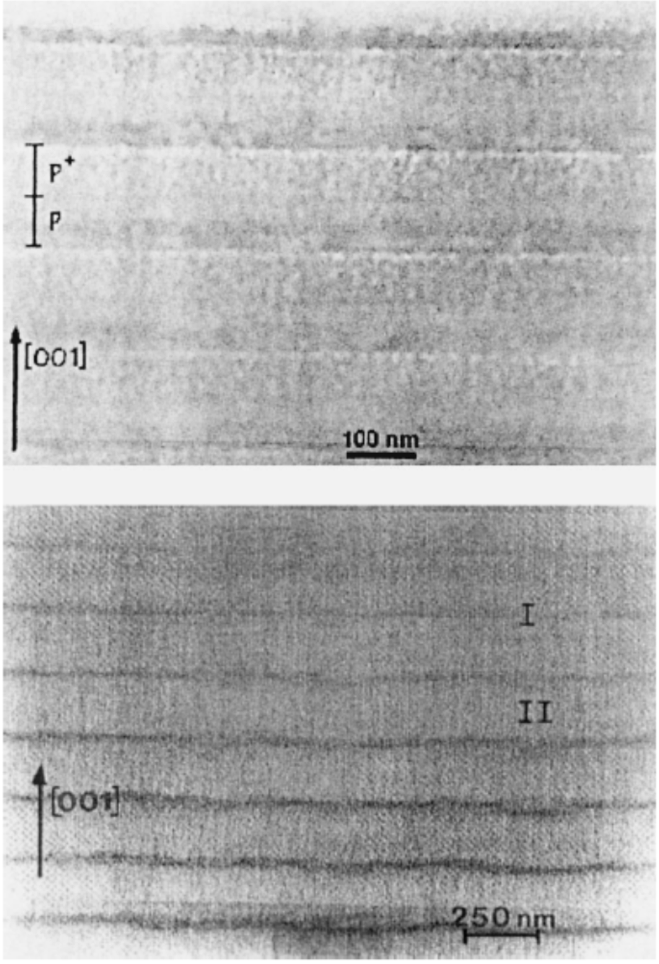
Réseau de Tubes
Période 3.2 µm ; Murs 220 nm ; hauteur 7 µm

a) Pilier obtenu par gravure plasma puis b) aminci par gravure électrochimique

Fig.18.10 Morphology of PS obtained in transient mode.

18.3.3. Porosity multilayers

The possibility of forming multilayer structure by using PS of different porosities relies on the basic characteristics of the etching process. The etching is self-limiting and occurs only in correspondence with the pore tips. That is, the already etched structure is not affected by further electrochemical etching of the wafer. For the formation of different layers the influence of the etch parameters is crucial. There are basically two types of PS multilayers, classified bythe way the porosity is changed from one layer to another. In the first type of multilayers the current density is changed during anodization, whereas in the second type the change in porosity is determined by changing the depth in doping level of the substrate . This last approach produces very sharp interfaces, as shown in Fig.18.11(a) , but it turned out to be less convenient than the current variation approach, because it requires epitaxial growth of the Si substrate. For this reason, the first type of PS multilayers are by far more common in the literature. With the first method, it is possible to vary the porosity, and therefore, the refractive index only at the etch front by changing the current density during the anodization process. In this way an arbitrary current versus time profile is transferred in a porosity versus depth profile. HF concentration and substrate doping have to be carefully chosen in order to obtain the maximum variations of the refractive index. Much larger variations of the refractive index are possible for heavily p-type doped substrates, both for the large value of critical current density and the wide range of porosities achievable with a given HF concentration. Moreover, the etch rate is higher and the inner surface is lower for p.-type doped substrates, and, hence, the additional chemical dissolution due to the permanence in HF is almost negligible. However, the refractive index and etching rate for a single layer are modified by the presence of the multilayer structure, and values lower than those determined for thick layers obtained with the same current density have been systematically observed for both the refractive index and the etching rate. To the best of our knowledge, no studies of this kind have been reported for n-type doped substrates.



\

Fig. 18.11Cross-section TEM images of PS multilayers. Upper image: sample obtained on a periodically doped substrate.

18.4. Physical properties

18.4.1 TEM and AFM analysis

Focusing the attention on luminescent material, TEM specimens for porosities as high as 80-85% has been prepared by direct cleavage or scraping. The TEM images of Fig. 18.12 show narrow undulating Si columns with diameters less than 3 nm. The crystalline nature of the Si nanostructures is clear. PS layers which have been heavily oxidized show the presence of dispersed Si nanocrystals with dimensions of a few nanometers.

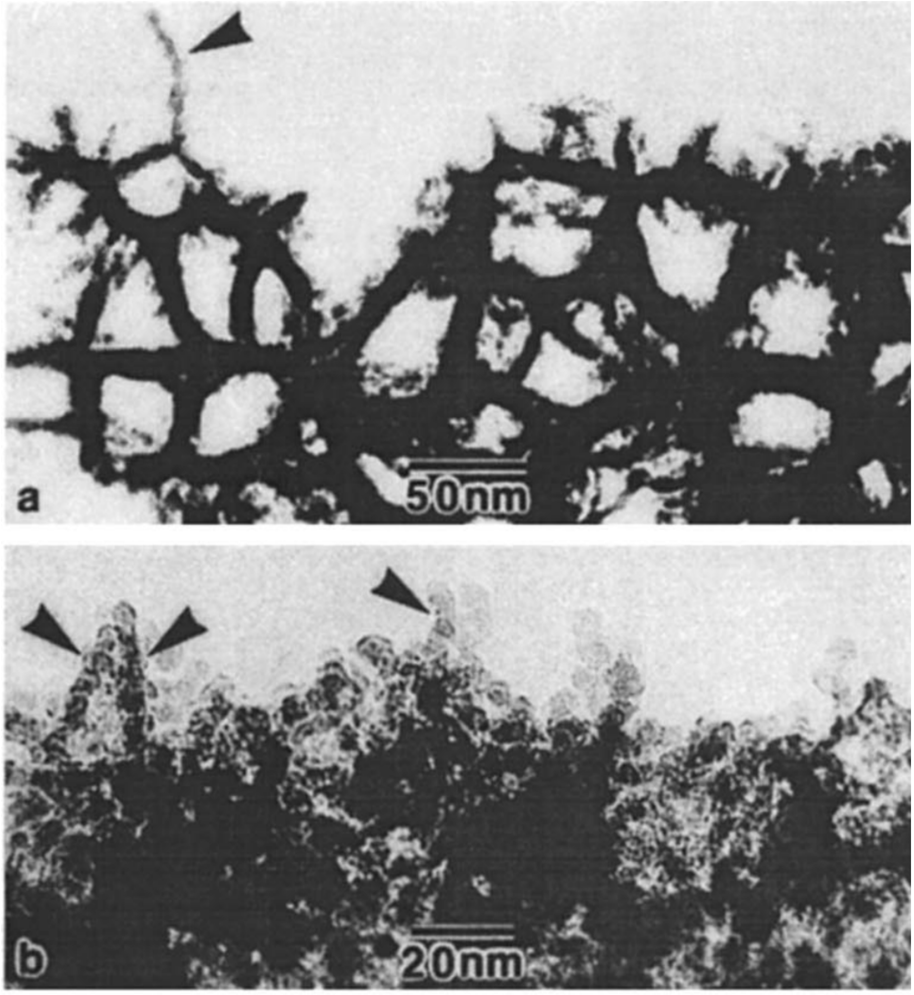


Fig. 18.12. TEM images (bright-field, under focus) of thin, high porosity Si layers: (a) non-luminescent samples; (b) luminescent samples (nanometer scale, columnar Si structures arrowed).

In addition to electron microscopy studies a variety of scanning probe microscopy investigations can be found in the literature. Atomic force microscope (AFM) studies focus entirely on the nanoscale characterization of PS films. In most studies AFM is applied together with other optical or morphological characterization techniques. AFM proves the presence of silicon features with diameters of about 10 nm (Fig. 18.13). The presence of sub 5 nm large topography features in images of PS has been also claimed. Due to the geometry of commonly used tips, AFM does not allow to image deep pores. As a result AFM does only allow imaging of the top end of the pores.

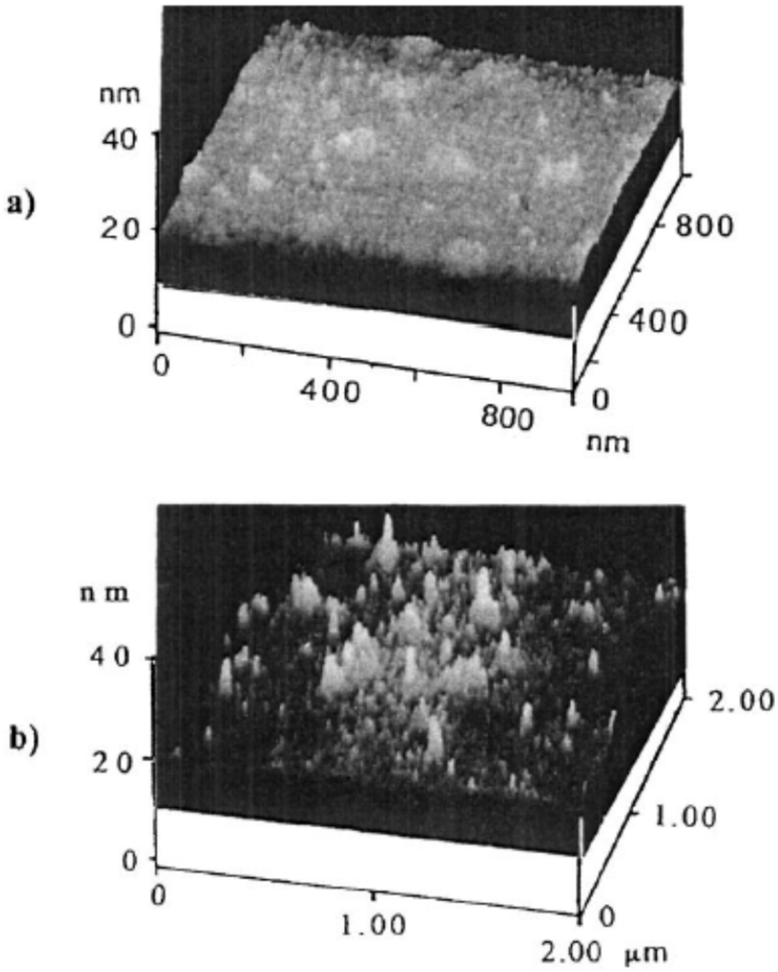


Fig. 18.13. Three-dimensional AFM image of the as-anodized porous silicon surface structure formed on p-type (5-8 Ohm, (1 0 0) Si), I= 10mA/cm2 for: (a) t = 8 s and (b) t= 30 min.

18.4.2. Raman spectroscopy

Among the other optical investigation methods, Raman spectroscopy, and especially micro-Raman spectroscopy, can give indirect information on the microstructure of PS. While in bulk Si interaction of optical phonons with incident photons is limited to the center of the Brillouin zone, the presence of nanocrystals in PS relaxes the k-selection rule, and the associated Raman peak broadens (see Fig. 18.14). The Raman spectra are analyzed in terms of specific contributions from amorphous and nanocrystalline Si. A gaussian line centered at 470 cm^{-1} with a width of 55 cm^{-1} is used as representative of the amorphous contribution. The Si nanocrystals contribution is described by using a phenomenological model based on the spatial confinement of phonons:

$$I(\omega) = \int_0^\infty I_L(\omega)P(L)\,\mathrm{d}L.$$

I(ω) is the Raman spectrum, I_L(ω) the first order Raman scattering from phonons confined within nanocrystals of size L, and P(L) is the nanocrystals size distribution, i.e. a gaussian distribution with mean L_0 and dispersion s. The results of a fitting procedure is shown in Fig. 26 where the fitting parameters are L_0 = 6,5 nm; s=2,2 nm and the relative weight of the amorphous to the crystalline part is 10% for both the spectra.

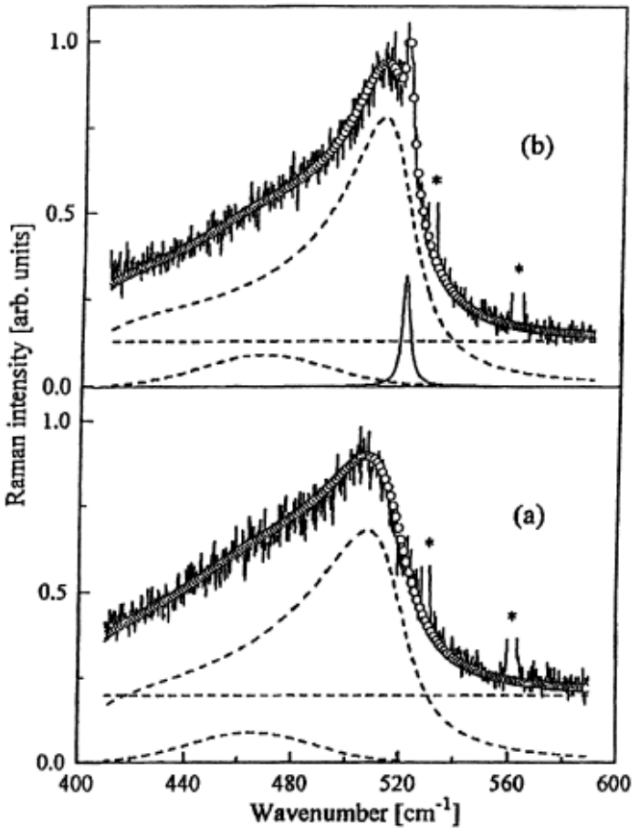


Fig. 18.14. Room temperature Raman spectra of a 15 mm thick and 74% porosity sample. (a) Spectrum measured near the interface air/PS. (b) Spectrum measured near the interface PS/bulk silicon. The continuous lines represent the experimental data, the open circles are the lineshape fit which considers the contribution from nanocrystalline Si, amorphous Si and underlying bulk Si. The different contributions are shown as dashed lines

18.4.3. Chemical composition and internal surface of as-prepared PS

The internal surface of a PS layer is very large. Values as high as 1000m2=cm3 can be measured (see Fig. 18.15). Such a large surface contains an enormous quantity of impurities coming from the electrolyte used for electrochemical etching and from the ambient air. Even the vessels in which the samples are stored could contribute to the sample contamination. It is necessary to know the chemical composition of PS because optical and electrical properties depend on impurity content and surface passivation.

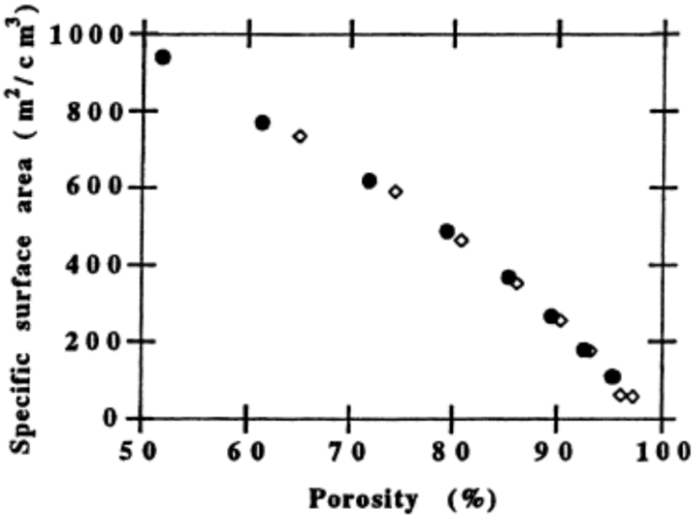


Fig. 18.15. Specific surface area as a function of PS porosity. Data from PS layer starting porosity of 51 (dots) and 65% (rhombuses). Layer thickness . 1 mm.

The original impurity which is always found in PS layers is hydrogen. Infrared absorption (IR) experiments have shown the presence of SiH_x groups (x = 1; 2; 3) on the internal PS surface during the etching process. A typical IR spectra of a fresh and oxidized PS samples are shown in Fig. 18.16, 18.17, while IR frequencies commonly observed in PS and their attributions are reported in Table 18.4. After formation and drying, the SiHx groups are still present on the inner surface for weeks and even for months, as it as been demonstrated using IR absorption and nuclear magnetic resonance. Hydrogen desorption occurs during annealing, and it has been shown that hydrogen desorbs from SiH_3 groups between 300 and 400°C, while desorption from SiH_2 occurs at 400°C and from SiH around 500°C. The atomic ratio H/Si in PS has been determined using secondary ion mass spectrometry (SIMS) and elastic recoil detection analysis (ERDA). In freshly anodized samples it is as high as 0.1±0.6, depending on the porosity and surface area of the samples. This result means that the surface of freshly etched PS is almost totally covered by SiHx groups. Fluorine comes only from residual electrolyte in the pores

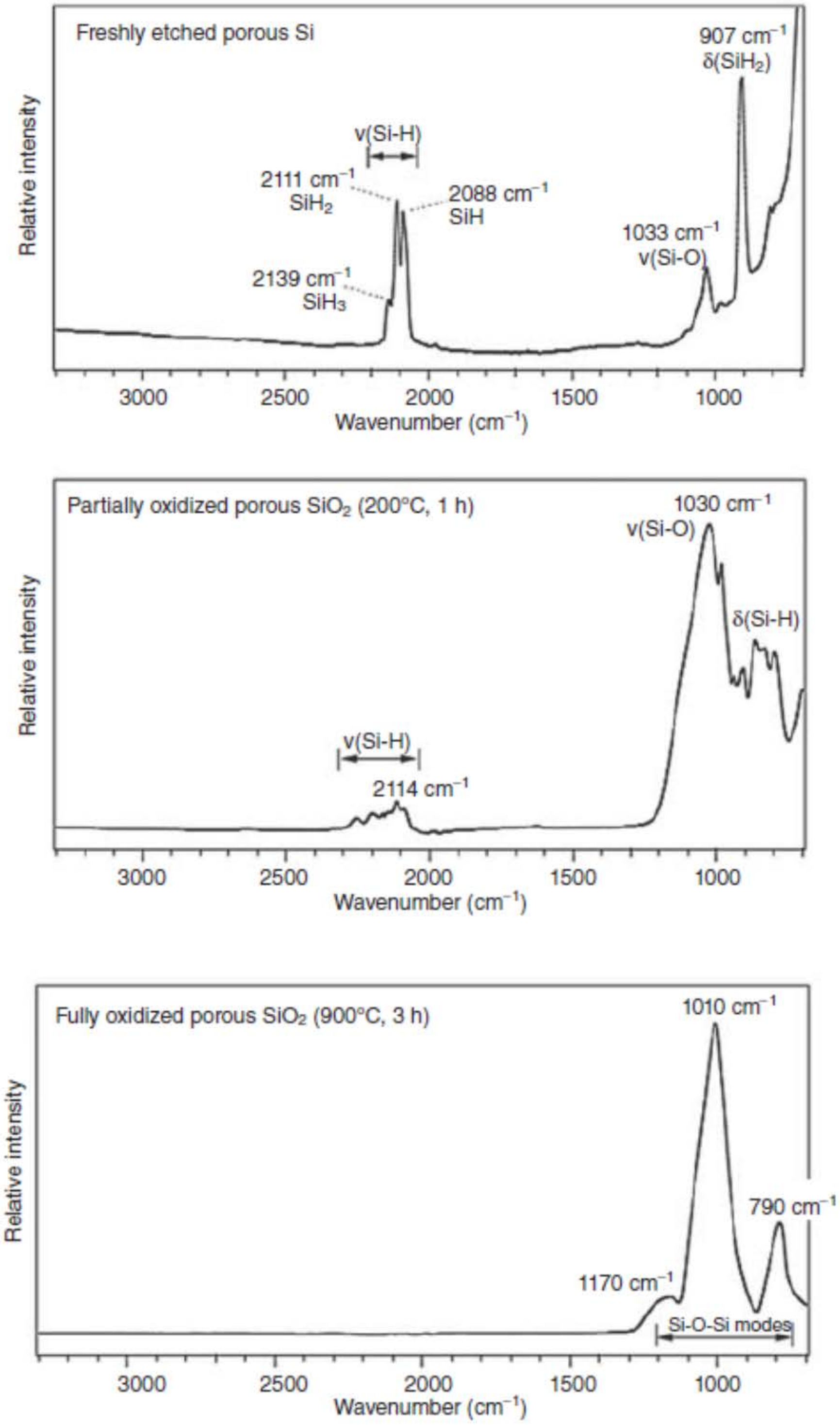


Fig.18.16 ATR - FTIR spectra of freshly etched and thermally oxidized PS samples

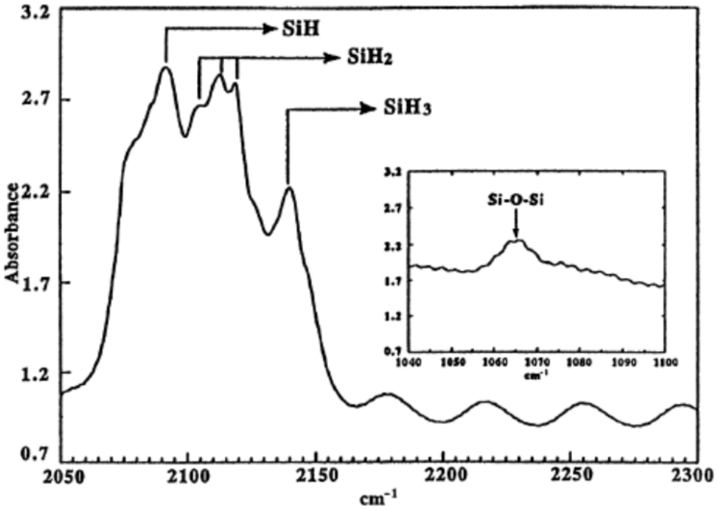


Fig. 18.17. IR absorption spectrum at 8 K of a free-standing p^+-type PS in the Si-H stretch mode. The thickness is 89 mm and the porosity 50%. The lines corresponding to OxSiH groups around 2200-2250 cm^{-1} and SiOSi groups around 1065 cm^{-1} are absent.

Table 18.4

Wave number positions and attributions of the absorption peaks observed in several p-Si samples

Peak position (cm^{-1})	Attribution	Peak position (cm^{-1})	Attribution
3610	OH stretching in SiOH	1463	CH_3 asymmetric deformation
3452	OH stretching in H_2O	1230	$SiCH_3$ bending
2958	CH stretching in CH_3	1056–1160	SiO stretching
2927	CH stretching in CH_2		in O–SiO and C–SiO
2856	CH stretching in CH	979	SiH bending in Si_2–H–SiH
2248	SiH stretching in O_3–SiH	948	SiH bending in Si_2–H–SiH
2197	SiH stretching in SiO_2–SiH	906	SiH_2 scissor
2136	SiH stretching in Si_2O–SiH	856	SiH_2 wagging
2116	SiH stretching in Si_2H–SiH	827	SiO bending in O–Si–O
2087	SiH stretching in Si_3–SiH	661	SiH wagging
1720	CO	624	SiH bending in Si_3–SiH

Other impurities are usually incorporated after the anodization process. The presence of carbon has been detected by several techniques. The carbon content can be as high as 10% in aged samples. The source of carbon is not ethanol, because carbon is present in similar amounts also in samples anodized in water/HF solutions. The source of carbon is the atmosphere. However, the adsorbed molecules are not CO or CO_2, because carbon and oxygen contents are not correlated. The carbon comes from hydrocarbon molecules present in the ambient air and often also in the residual gas in analysis chambers used for SIMS, IR, XPS, etc. Oxygen is the most important non-original impurity, and is normally adsorbed in a few minutes after drying in ambient air. The amount of oxygen can be as high as 1% after 15 min of air exposure, as confirmed by electron paramagnetic resonance (EPR) and increases to very high percentage values with aging. IR spectrum performed a few hours after anodization on a PS layer has also shown the presence of Si±O±Si group (mode at 1065 cmÿ1 in the inset in Fig. 29). The IR absorption due to the OxSH groups at 2200-2500 cm-1 is not present in this sample, which means that the oxidization is only partial. A few days after anodization further oxidization takes place, both through the formation of Si-O-Si groups and O-Si-H and O_3SiH groups. The existence of such groups does not modify the hydrogen passivation, and the possible presence of water simply accelerates the oxidization process. The relevance of the Si/SiOx interface in PS is also observed in EPR measurements. In Fig. 18.18 the EPR spectrum of a fresh n.-type doped PS layer which shows two absorption lines is reported. The broad line is due to the conduction electron spin resonance (CERS). EPR studies on n.-type doped and free standing PS layers have shown that the CERS signal is present only in the substrate, demonstrating that free carriers are absent in PS. Subtracting the signal corresponding to the free carriers, the signal due to the dangling bonds appears clearly. Dangling bonds are the most important paramagnetic defect at the Si/SiO2 interface (Pb centers), and have a •Si≡Si_3 structure. Another relevant paramagnetic center is the •Si≡SO_3 defect (E_0 center), which is formed especially in hydrogen depleted oxide layers.

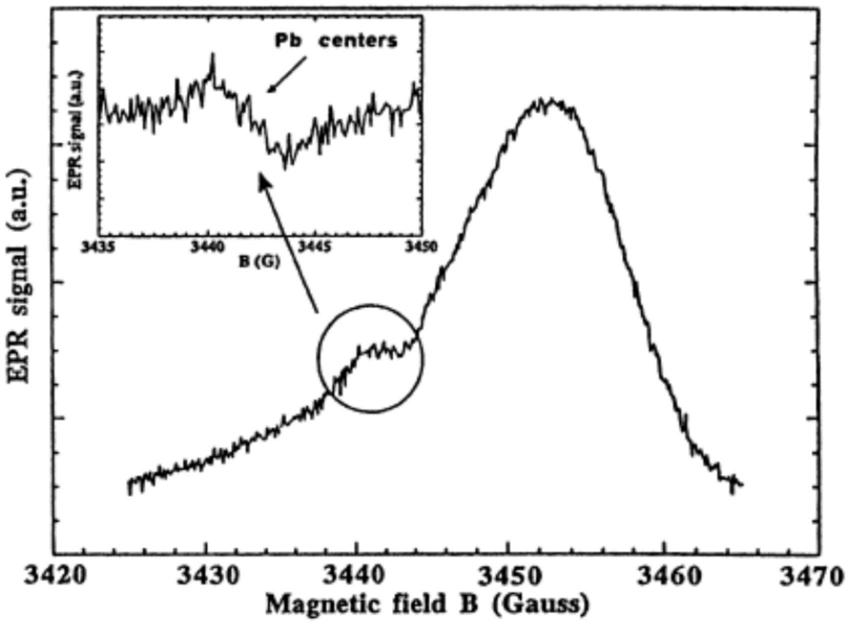


Fig. 18.18. Room temperature EPR spectrum of a (1 0 0) 11 mm thick n.-type PS sample. The magnetic field is in the (110) plane parallel to [001] orientation. In the inset the EPR spectrum of the Pb centers is shown, after subtraction of the large signal due to free carriers.

18.4.4. Intentional oxidization

A wide range of post-anodization treatments on PS have been developed for different purposes. Complete conversion into silica was the first broad application of PS for dielectric insulation in microelectronics and MEMS, but many oxidization treatments have been used recently as a means to improve photostability of luminescent films. Fig. 18.19 reports an idealized scheme of the thermal oxidization process from freshly etched PS to a wholly oxidized porous glass. Although in most cases the degree of oxidization has been revealed qualitatively, few data are available regarding the chemical composition of intentionally oxidized samples. Quantitative data exist only for anodic oxidization and thermal oxidization treatments.

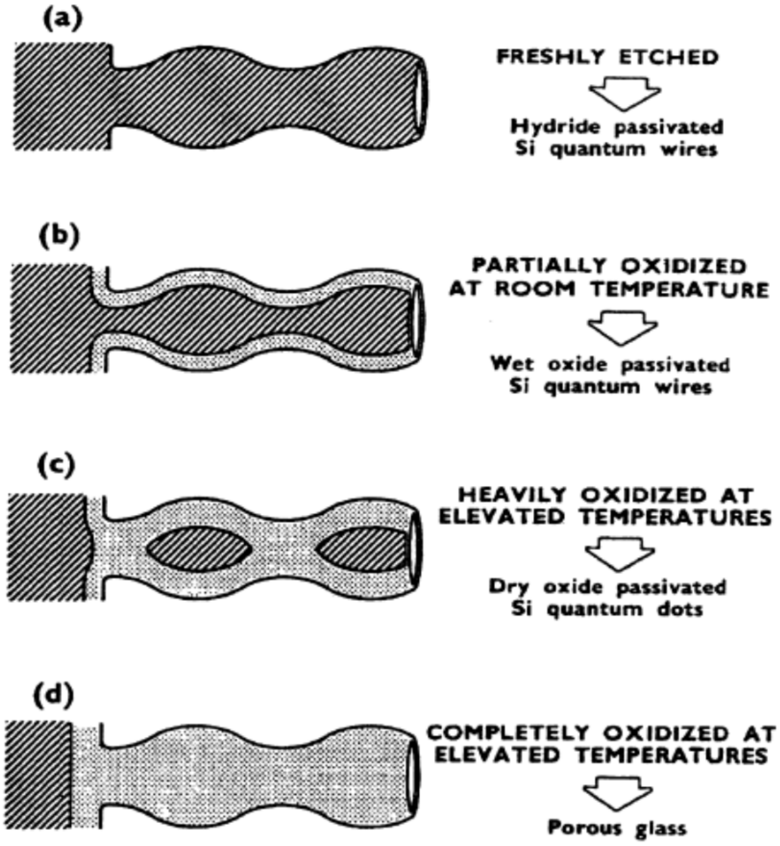


Fig. 18.19. Idealized schematic steps in the oxidization process of highly porous silicon

In the anodic oxidization process PS layers are oxidized uniformly, but complete oxidization is not achieved, because electrical isolation of parts of the Si skeleton occurs, preventing further current flow. In mesoporous samples of 65% porosity, formed on p$^-$-type doped substrates, about 40% of the Si remains unoxidized at the end of the process, corresponding to about a monolayer of oxide coating. In the case of thermal oxidization, a 300^{o}C stabilizer pretreatment prior to the high temperature oxidization is necessary to avoid pore coalescence. The content of oxygen at the end of the process varies with substrate, oxidization temperature and duration. In mesoporous material formed on n+-type doped substrates and oxidized at 1050^{o}C, more than 90% of the Si is oxidized within 30 s. However, heavily oxidized material is still susceptible to atmospheric ``aging" and contamination, simply because the oxide matrix is still porous. Its water content for instance, dependson ambient humidity. Chemical oxidization with hydrogen peroxide, nitric acid or boiling water has also been reported.

18.4.5. Electronic properties

In bulk crystalline Si the term band gap means the energy gap between the maximum of the valence band and the minimum of the conduction band, and its value can be measured by a simple transmission experiment. In PS a wide spread of band gaps exist, as demonstrated by the large and inhomogeneous lineshape of its visible PL band. Direct experimental measurements usually yield average values, which are still meaningful, because the intra-sample band gap variations are normally smaller than the intersample variations.

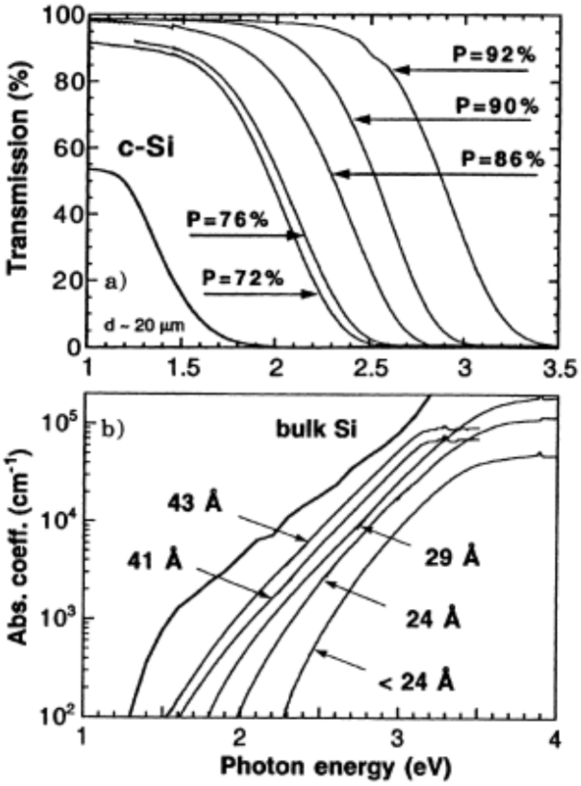


Fig. 18.20. (a) Transmission spectra of several 20 µm thick free-standing PS films (5-7 Om.cm p-type Si substrate) of various porosities and a 20 µm thick c-Si film. (b) Absorption spectra for Si nanocrystals with sizes from 4.3 to 2.4 nm deduced from the transmission spectra of (a). The absorption of c-Si is shown for comparison

Different techniques can be used to obtain experimental values of the PS band gap. Luminescence techniques measure the recombination energy of the carriers excited across the gap. Usually, PL maxima are used for sample-to-sample comparison, while the lineshape of the luminescence displays directly the intra-sample variation of the band gap. However, the direct attribution of the peak energy to the sample band gap is often questionable due to the neglection of the excitonic binding energy and to the fact that the luminescence could be due to other emissions than simple direct exciton recombination. Other techniques are correlated to the absorption across the gap, among them are the transmission, photothermal deflection spectroscopy, photoluminescence excitation spectroscopy (PLE) and photoconductivity. Fig. 18.20 shows typical transmission spectra which alow to recalculates the band gap of PS (fig. 18.21).

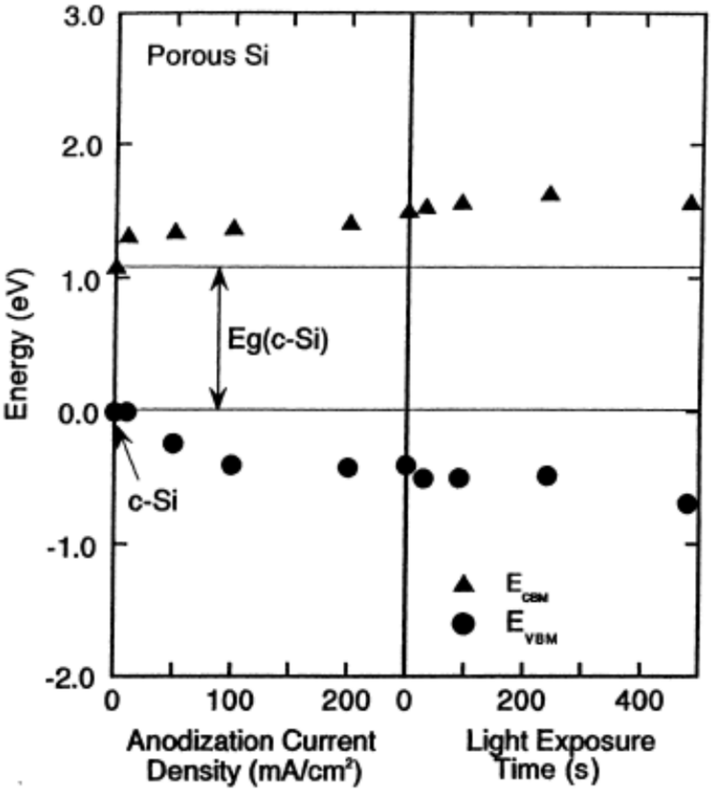


Fig. 18.21. Energy positions of conduction band minimum (CBM) and valence band maximum (VBM) obtained from PS with (right side of panel) and without (left side of panel) the light exposure treatment.

18.4.6. Refractive index and absorption coefficient

Avery simple method to evaluate the refractive index (n) is the measure of the interference fringes in a reflectance spectrum of a thin sample (see Fig. 18.22). The position of the interference maxima satisfies the following equation

$$2nd\left(\frac{1}{\lambda_r} - \frac{1}{\lambda_{r+1}}\right) = 0,$$

where d is the layer thickness and λr is the wavelength of the r-th reflectance maxima. If the layer thickness is known in an independent way, n is simply the ratio between the optical thickness nd and the layer thickness. This method can only be applied if the interference fringes are visible. If the layer is too achieved more easily with p-type substrates.

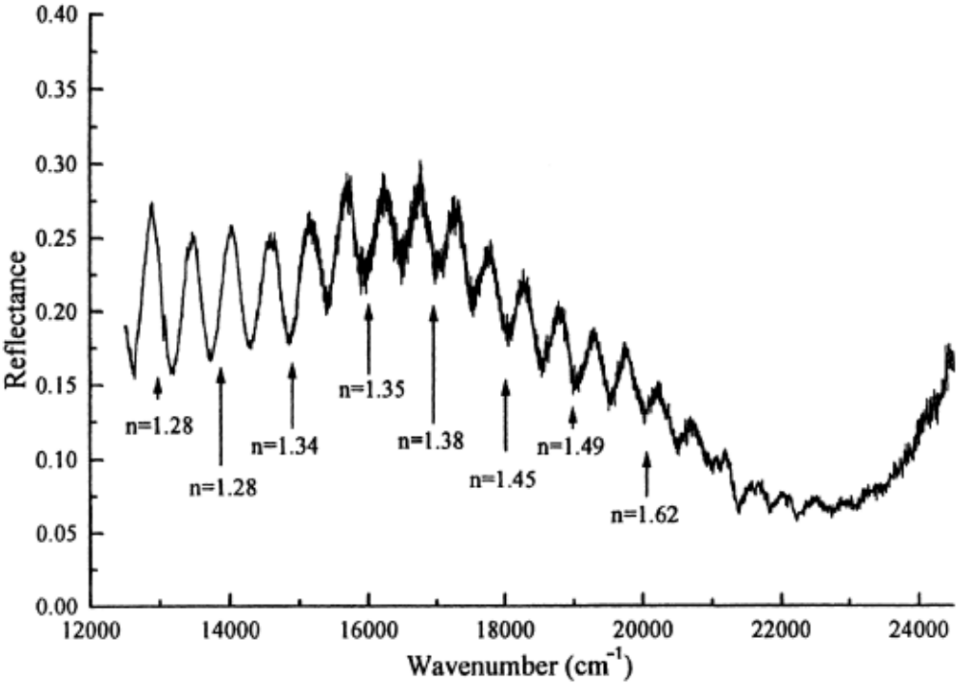


Fig. 18.22. Normal-incidence reflectance spectrum of a 75% porosity sample 6,8 µm thick

The refractive index of PS is expected to be lower than that of bulk Si, and decreases with increasing porosity, because PS is basically a mixture of air and Si (fig.18.23). However, the averaging of the dielectric functions of the individual components is not trivial, but depends on the microtopology of the material. The application of different effective medium theory leads to different formulas. For example, the widely used Bruggemann formula leads to the following symmetric expression

$$f\frac{\epsilon-\epsilon_{\mathrm{eff}}}{\epsilon+2\epsilon_{\mathrm{eff}}}+(1-f)\frac{\epsilon_{\mathrm{M}}-\epsilon_{\mathrm{eff}}}{\epsilon_{\mathrm{M}}+2\epsilon_{\mathrm{eff}}}=0,$$

where f describes the volumetric fraction, ε and ε_M are the dielectric functions of Si and of the embedding medium (air) and ε_{eff} is the effective dielectric function for PS. However, at least for microporous Si, the pore wall material cannot be considered as bulk Si, and different optical constants should be used.

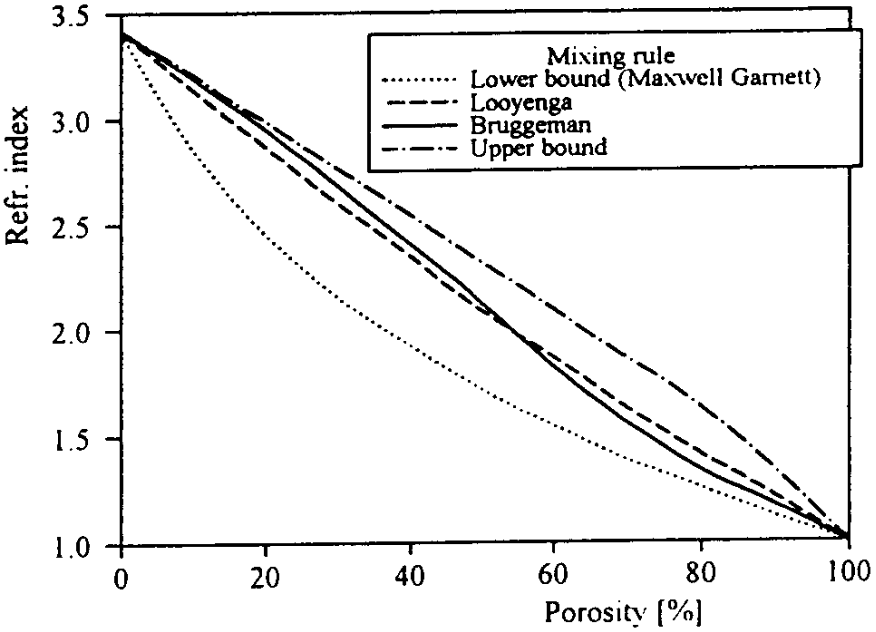


Fig. 18.23 Dependence of refractive index on porosity

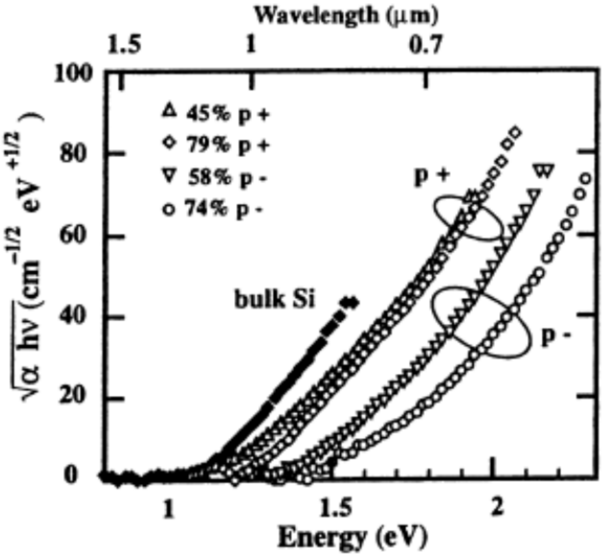


Fig. 18.24. Square root of the absorption coefficient times photon energy versus photon energy for various porous silicon sample.

A more precise lineshape analysis of the absorption coefficient shows that the energy dependence of a follows a trend like that of an indirect gap semiconductor very similar to that of Si, even though displaced to higher energy (see Fig. 18.24). It is observed that PS is well described by the Bruggemann result for energies higher than the Si direct energy gap. At lower energies, it departs from the Bruggemann result and shows an exponential decrease. This has been described by using the Urbach tail concept typical of Si or of amorphous semiconductors. A value of 180 meV is usually found for the Urbach energy in PS. The overall behavior of a in PS reflects the crystalline nature of the absorbing center. The Urbach tail is due to the random dispersion of sizes and shapes of these centers

18.5. Photoluminescence

18.5.1 General properties of PS photoluminescence

PS based structures have been reported to luminesce efficiently in the near infrared (0.8 eV), in the whole visible range and in the near UV (fig.18.25) at room temperature at UV excitation. Such a broad range of emission energies arises from a number of clearly distinct luminescent bands, which are listed in Table 18.5. In addition, PS has been used as an active host for rare earth impurity, e.g. Er, or dye solutions. Direct energy transfer between PS and the impurity or dye is demonstrated (fig.18.26). We will focus mainly on the properties of the so-called S-band, where S stands for slow. The S-band has been intensively studied up to now and has the most technological relevance since it can be electrically excited. Its main features are summarized in Table 18.6.

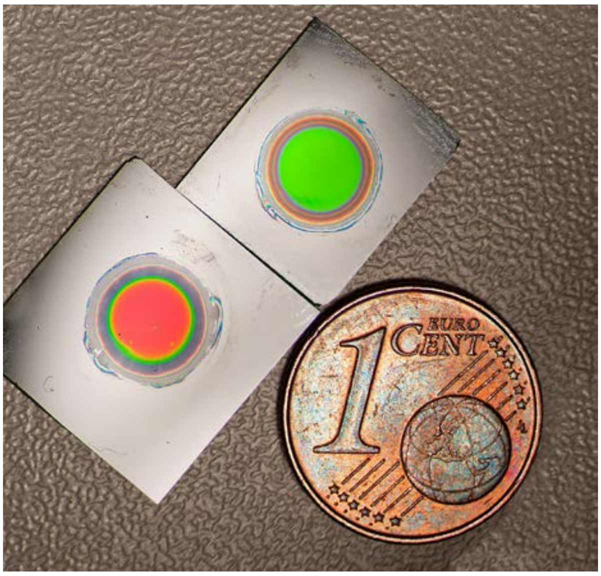


Fig.18.25 General view of visible PS photoluminescence at room temperature

Табл.18.5 Photoluminescence bands of PS

Spectral range	Peak wavelength	Label	PL	EL
UV	~350 nm	UV band	Yes	No
Blue–green	~470 nm	F band	Yes	No
Blue–red	400–800 nm	S band	Yes	Yes
Near IR	1100–1500 nm	IR band	Yes	No

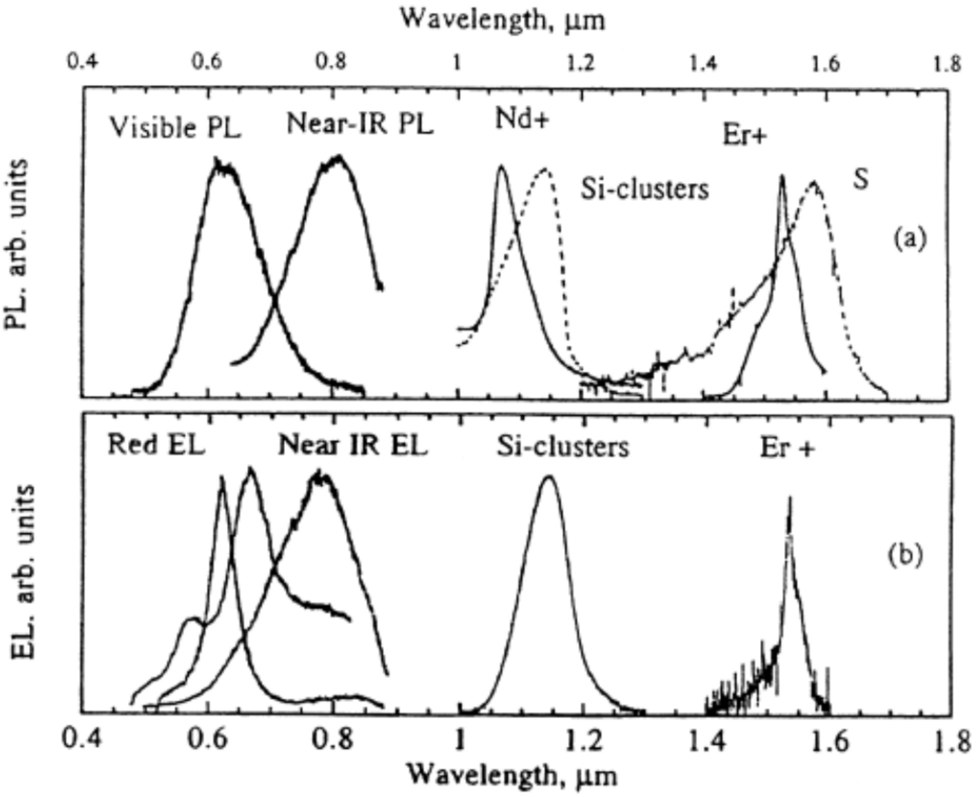
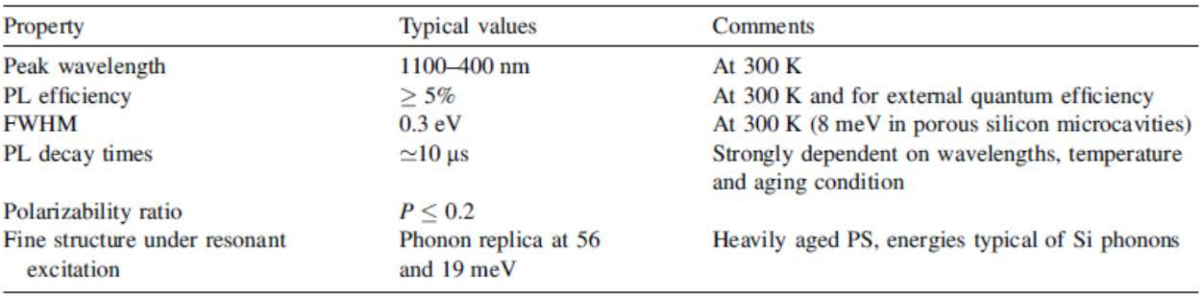


Fig.18.26. Room temperature photoluminescence and electroluminescence spectra for various PS structures which have been oxidized or implanted with some selected impurities

The S-band can be tuned from close to the bulk silicon band gap through the whole visible range. Fig. 18.27 displays the room temperature luminescence spectra of PS samples with different porosities. However, while the PL efficiency from red to yellow is high under blue or UV excitation, the blue emission one is rather low to date. The S-band large spectral width comes from inhomogeneous broadening, and its spectral position depends on porosity. It is important to note that not only the spectral position, but also the relative intensity of S-band changes with porosity (Fig. 18.28). Indeed, the S-band efficiency is not proportional to the inner surface area, but it seems that a ``threshold" porosity has to be exceeded to achieve an efficient luminescence. It has been found that postanodization chemical etching in HF, corresponding to a porosity increase, results in a strong rise in PL efficiency and a blue shift of the visible band. External quantum efficiencies higher than 10% are obtainable from high porosity PS layers.

Tabl.18.6 Properties of S band



Property	Typical values	Comments
Peak wavelength	1100–400 nm	At 300 K
PL efficiency	$\geq 5\%$	At 300 K and for external quantum efficiency
FWHM	0.3 eV	At 300 K (8 meV in porous silicon microcavities)
PL decay times	$\simeq 10$ μs	Strongly dependent on wavelengths, temperature and aging condition
Polarizability ratio	$P \leq 0.2$	
Fine structure under resonant excitation	Phonon replica at 56 and 19 meV	Heavily aged PS, energies typical of Si phonons

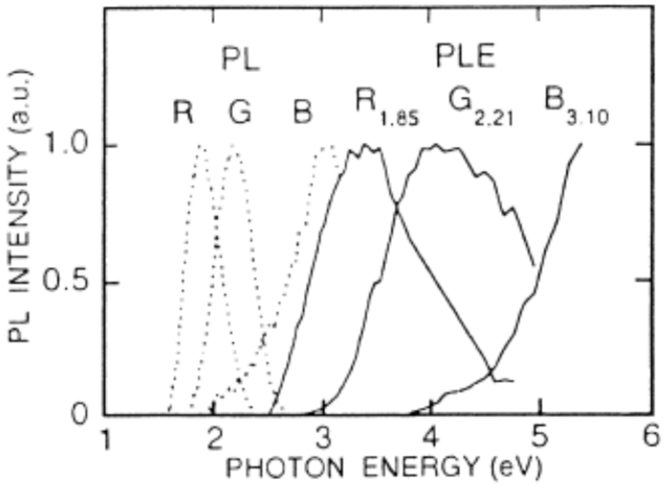


Fig. 18.27. The red (R), green (G) and blue (B) photoluminescence spectra (PL) tuned by post-anodization illumination (dashed curve) and the corresponding excitation spectra (PLE).

High porosity is essential for high visible PL efficiency, and it has been proven that the inefficient luminescence observed from inhomogeneous material of low porosity originates from microscopic areas of high porosity. In the macroporous, for instance, each macropore is coated with luminescent meso-and microporous material. This is further confirmed by studies on isolated PS nanoparticles produced by dispersing a colloidal suspension of PS fragments on a glass coverslip. The isolated nanocrystals show external quantum efficiencies ~88%, while the number of bright to dark nanocrystals in the suspension was only 2.8%. This means that the average quantum efficiency of a PS layer of 10% results from a statistical distribution of high and low quantum efficiency nanocrystals.

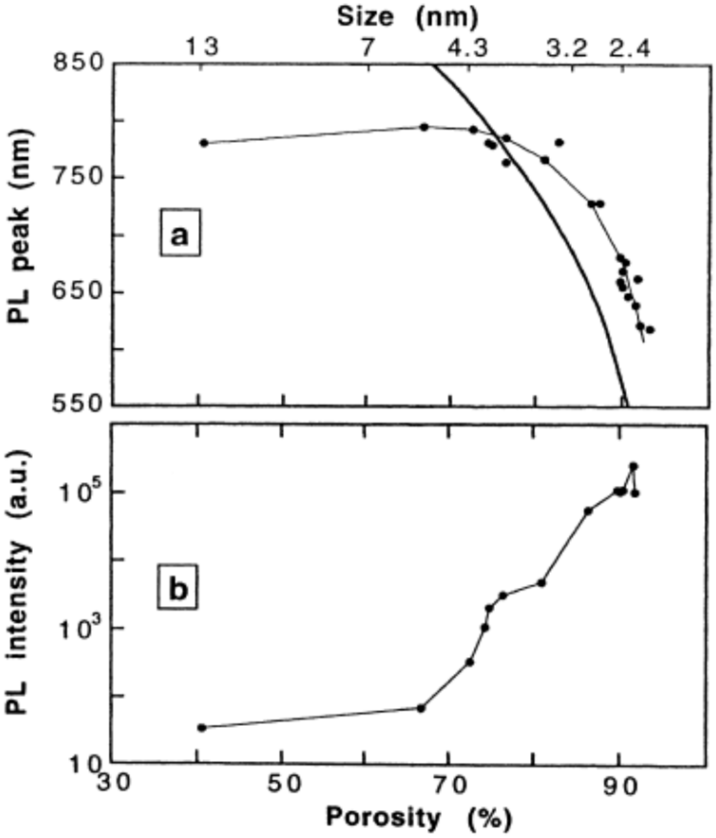


Fig. 18.28. Photoluminescence peak wavelength and intensity versus sample porosity. The thick line is obtained from first principle calculations.

18.5.2. Models for PS luminescence

Quantum confinement of carriers in Si wires was the first model proposed to explain PS luminescence. Afterwards, many other alternative explanations have been proposed. The various models can be grouped in six different categories, as illustrated in the scheme in Fig. 18.29. Except for the quantum confined model, all the others assume an extrinsic origin for the luminescence.

1)Hydrogenated amorphous silicon model (Fig.18.29b)

It has been proposed that PS luminescence is due to a hydrogenated amorphous phase (a-Si:H) which is formed during anodization. In fact, a-Si:H possesses a PL band in the visible range, and the tunability of the PL from PS can be in principle explained with alloying effects and with the variation of hydrogen and oxygen percentages. In addition the time resolved measurements indicate that the disorder plays a key role in the recombination dynamics. Against this model, recent TEM studies where sample damage has been minimized showed that there is little amorphous Si in PS. Moreover, the a-Si:H PL band is strongly quenched going from cryogenic to room temperature, while the PS visible band is enhanced as the temperature is raised. Finally strong spectroscopic evidences exist that the luminescence has both an electronic and a vibrational nature similar to that of crystalline Si.

2) Surface hydrides model (Fig. 18.29c)

Since the PS luminescence decreases dramatically if the hydrogen on the surface is thermally desorbed, and the PL intensity can be recovered with immersion in HF, which restores the hydrogen coverage, SiHx surface species were suggested to be responsible for luminescence in PS. Against this model there are many evidences. FT-IR studies have demonstrated that luminescence is totally quenched when the majority of hydrogen is still on the PS surface, and the luminescence loss is probably related to the formation of dangling bonds, efficient non-radiative decay channel. Another evidence comes from the simple fact that when the hydride coverage is replaced by a good quality oxide layer, the PL process is still efficient. Finally, the spectroscopic evidences support a crystalline nature of the emitting centers.

3)Defect models (Fig. 18.29d)

In defect models the luminescence originates from carriers localized at extrinsic centers, i.e. defects in the silicon or silicon oxide that covers the surface. However, luminescent nanocrystalline Si can be created in many different ways, and passivated either with hydrogen or with oxygen and it is then very unlikely that the same impurity or defect is always present. In any case the defects in the silicon oxide are ruled out, because SiO_2 is not present in fresh PS. Also the tunability of the PL band is difficult to justify, because the emission from defects expected to be almost insensitive to the size structure. Also the spectroscopic evidences are against this model.

4). Siloxene model (Fig. 8.29e)

Siloxene, an Si:H:O based polymer, supposedly created during PS anodization, was proposed as the origin of PS luminescence. This model is supported by the fact that the optical properties of siloxene are somehow similar to those of PS. Siloxene possesses a visible-red PL band and the IR spectrum closely relates to that of aged PS. Against this model there are many evidences. It is now well demonstrated that freshly etched PS has no detectable content in oxygen. PS can still be luminescent above 800^{o}C, while siloxene or other related molecules are totally decomposed at such a high temperature. In addition to the spectroscopic evidences which point to a crystalline nature of the emission, synchrotron radiation measurements show that SiO groups have no role in the emission process.

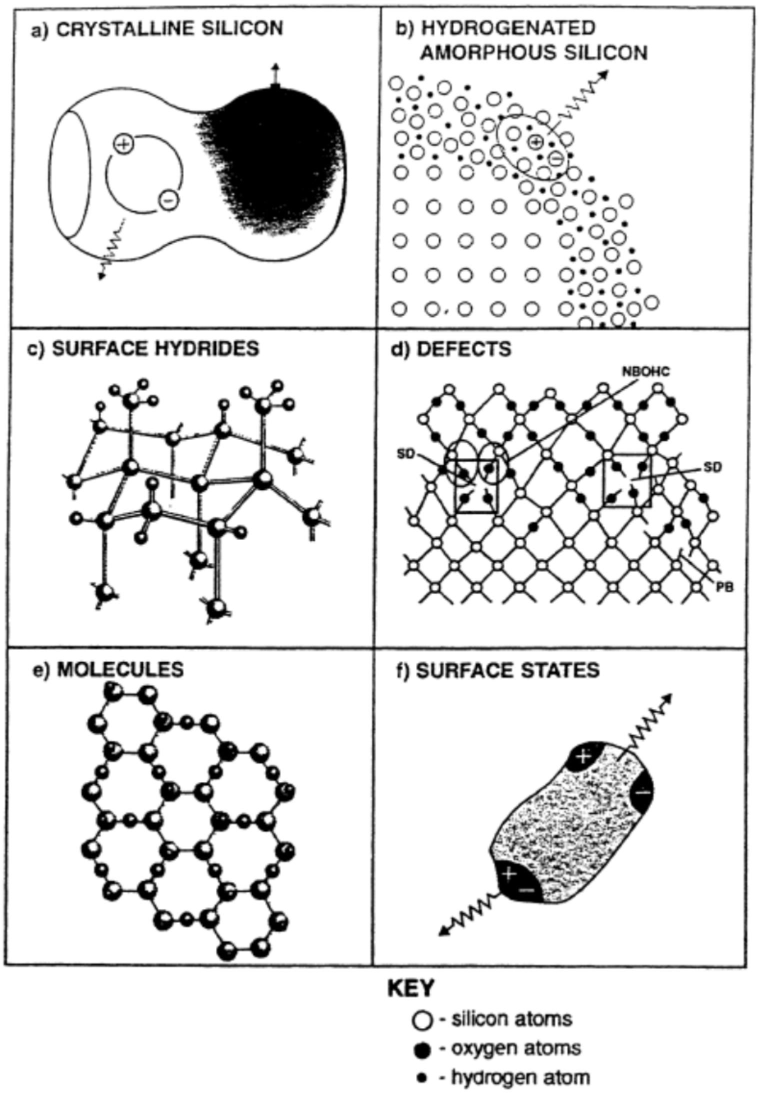


Fig. 18.29. The six groups of models proposed to explain PS luminescence. (a) Section of an undulating crystalline surface defect rends an undulation non-radiative, while an exciton localized in the neighboring undulation recombinesradiatively. (b) Crystalline Si covered by a layer of hydrogenated amorphous Si, where radiative recombination occurs. (c) Si surface passivated by SiHx terminations. Radiative recombinations occur at the Si-H bonds. (d) Partially oxidized Si containing defects proposed as radiative centers. (e) Siloxene molecule is proposed to exist on the large inner PS surface and acts as luminescence center. (f) Si dot with surface states that localize carriers and holes separately (upper part) or together (lowerpart, radiative recombination).

5) Surface states models (Fig. 18.29f)

The enormous inner surface of PS (~10% of the Si atoms in PS are surface atoms) has led to propose that it is involved in the luminescence process. The excitons, quantum confined in Si nanocrystals, should know the existence of the surface and of its reconstruction. For this reason the model has also been called smart quantum well model. In this model, absorption occurs in quantum confined structures, but radiative recombination involves localized surface states. Either the electron, or the hole, or both or none can be localized. Hence, a hierarchy of transitions is possible which explains the various emission bands of PS (see Fig. 18.30). The energy difference between absorption and emission peaks is well explained in this model, because photoexcited carriers relax intosurface states. The dependence of the luminescence from external factors or from the variation of thePS chemistry is naturally accounted for by surface states changes. Resonantly excited PL results are against the attribution of PL process to surface states, showing that PL arises from exciton coupling with momentum-conserving phonons. This means that the excitonwavefunction is extended over many Si atoms and not strongly localized, as it should be in the case ofdeep surface states. Furthermore, the values of the exchange splitting energy extracted from temperature dependent lifetime measurements also suggest that carriers are not localized on atomic scale, but in the whole volume of the Si nanocrystals, and that luminescence does not originate from localized states in the gap, but from extended states. Finally, the polarization measurements point to an extended nature of the luminescence states.

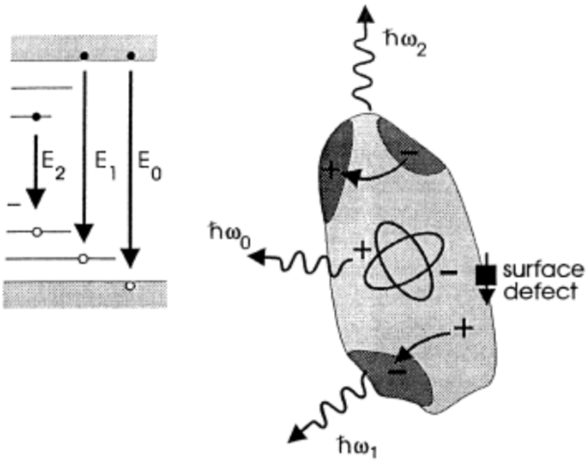


Fig. 18.30. Hierarchy of transitions in the surface-state model. The indices 0, 1 and 2 indicate the number of surface states involved

6) Quantum confinement model (Fig. 18.29a)

Quantum confinement in crystalline Si was the first model proposed to explain the efficient photoluminescence of PS. Quantum confinement effects result in an enlargement of the band gap, in a relaxation of the momentum-conserving rule, and in a size dependence of PL energy which naturally explains the efficient luminescence, the up-shift and the tunability of PL band in PS. Many other experimental evidences support the quantum confinement model. Structural characterization has proven that PS is crystalline in nature. Observations of nanocrystals of nanometric dimensions have been reported. The band gap up-shift is clearly visible in absorption spectra, and the luminescence blueshift after further chemical dissolution in HF is easily explained by further reduction of nanocrystals dimensions. Fresh structures can be theoretically modeled as QWW, while aged structures are usually more dot-like. Qualitative agreement with experiments has been obtained for calculated radiative lifetimes and for the observed splitting of the lowest lying exciton states.

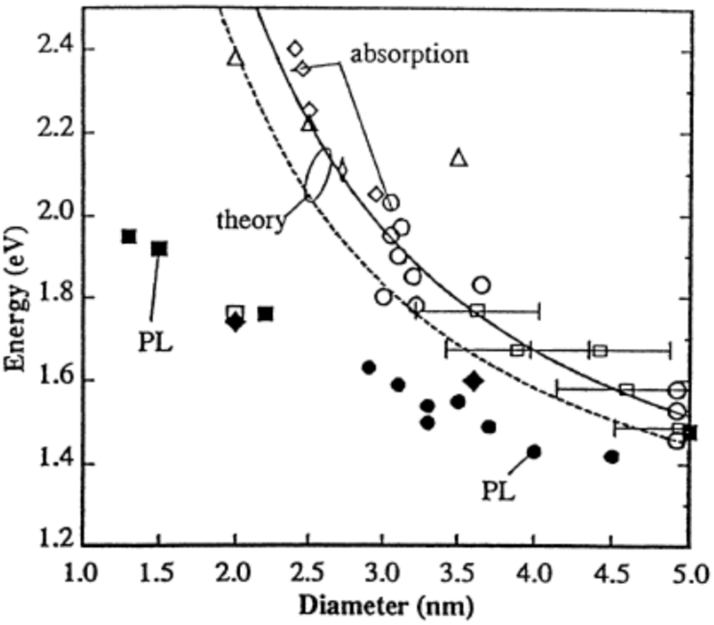


Fig. 18.31. Compilation of optical band gaps of silicon crystallites and PS samples obtained from optical absorption (unfilled symbols) and luminescence (filled symbols). The lines represent calculated values with (dashed line) or without (full line) excitonic correction

However, it is becoming increasingly clear that, even in the quantum confinement framework, the emission peak wavelength is not related only to size effects (see Fig. 18.31). Freshly etched and very high porosity samples which have not been exposed to the air have luminescence peak energies in the 3 eV range while as soon as they get into contact with air their luminescence peak moves to the usual 2 eV range. A change in the surface passivation, as well as dielectric effects, can produce wavelength shifts. In addition, all the spectroscopic studies reported in this section are not able todiscriminate between true extended states and shallow localized states, in which the carrier wavefunction extends over several lattice parameters. For these reasons, even though it is certain that quantum confinement plays a fundamental role in determining the peculiar properties of PS, somecomplements to the pure quantum confinement model are needed in order to take into account surface effects, especially when the nanocrystals have sizes smaller than 3 nm.

18.5.3 Aging and fatigue effects

A blue-shift in the S-band arises from simply storing PS in ambient air at room temperature, as shown in Fig. 18.32. However, the effect of aging on PS luminescence seems to be contradictory at fist sight. Even though blue-shifts are generally observed with aging, PL effiiencies are reported to drop in some cases and to rise in others. Such effects are, however, explained in terms of both surface passivation phenomena and carrier confinement.

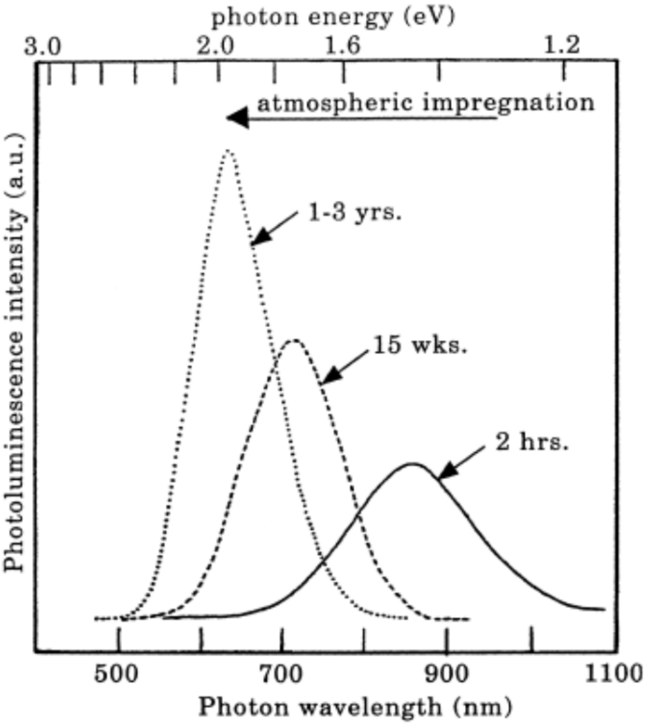


Fig. 18.32. PL spectra from a PS layer with 77% porosity and thickness 11,6 μm for various aging times.

Anodic oxidization can dramatically raise the PL efficiency in medium porosity PS giving an estimated external quantum efficiency (EQE) in the range 1-10%, while thermal oxidization at 400- 700^{o}C usually leads to a strong loss of PL efficiency. However, after oxidization at higher temperatures PS is still luminescent. The emission quenching at lower temperatures is due to the poor electronic quality of the Si/SiO_2 interface, as demonstrated by EPR spectroscopy. A clear anticorrelation between the dangling bond density and the luminescence intensity has been measured. This shows that oxidization can improve the luminescence due to an increased passivation.

18.6 Bragg reflectors and microcavities

A natural optical application of PS multilayers are dielectric Bragg reflectors (DBRs). These are stacks of alternate layers of high (n_H) and low (n_L) refractive indices. In order to achieve a very high reflectivity in a given range of wavelengths, the optical thickness nd of every single layer in the stack is 1/4 of the value of the central wavelength. The high reflectivity region, called stop-band, increases with the ratio n_H/n_L, and the number of layers. Fig. 18.33 (a) shows the reflectance spectra of PS Bragg reflectors. The best result and the highest value of the reflectance maximum is reached for the longest central wavelength. This effect is due to the absorbance of PS layers, which increases with decreasing wavelength . The number of periodic repetitions of the alternate layers in PS Bragg reflectors should be optimized taking into account the increase in reflectance for a large number of repetitions and the effects of depth inhomogeneity and additional chemical attach in the HF solution.

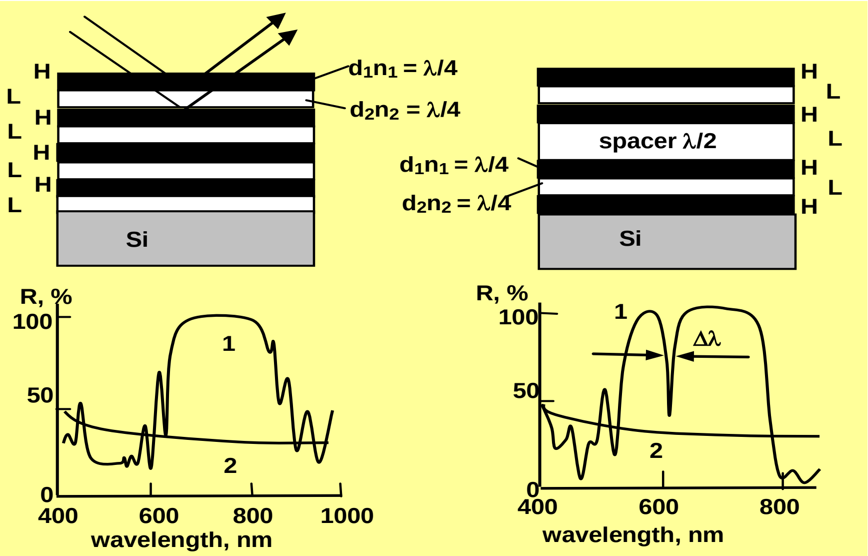
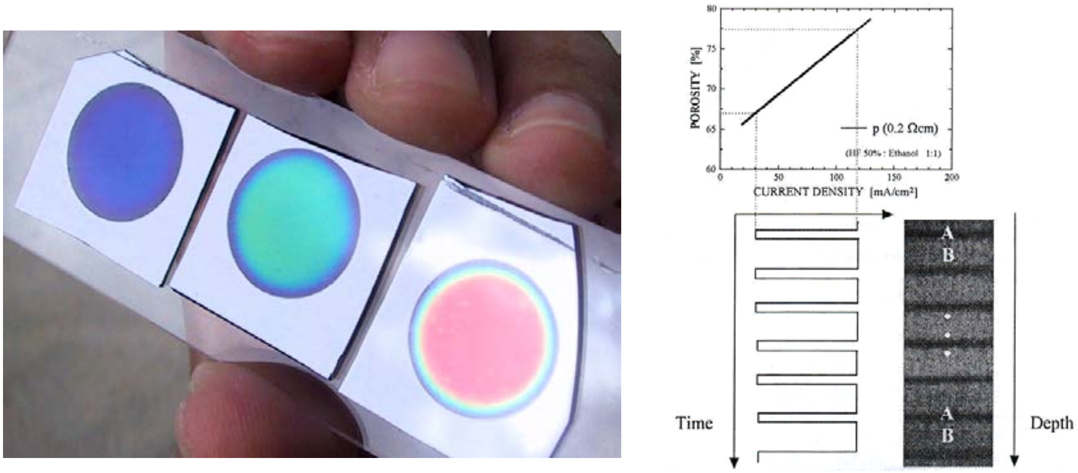


Fig.18.33 Bragg reflector (a) and microcavity (b)

PS Bragg reflector can be prepared by anodization process since refractive index depends on porosity and current density (fig.18.34).



Fig,18.34 Photo of Bragg reflectrors of different cental peaks (a) and method of it preparation and TEM photo.

In order to obtain a PS microcavity (PSM) the active PS layer must be placed in an optical cavity. This can be accomplished by using a Fabry-Perot (FP) where the spacer is the active PS layer, or, in other words, by embedding the central layer between two identical DBRs. The typical scheme of a PS microcavity is shown in Fig. 18.33b. The resonance wavelength is $\lambda_c = mn_c d$, where m is the order, n_c the active layer refractive index and d is its thickness. In PSM both the mirrors and the spacer are made by PS.

18.7 Porous Silicon Membranes

Solid silicon membranes are commercially available at micron scale thickness and are found in a number of MEMS applications like microvalves, micropumps, accelerometers, and microphones. Solid silicon membranes of less than ten to several hundred nanometer thickness (“nanomembranes”) are being researched for uses such as flexible electronics. Porosified silicon membranes first started to appear in the 1990s and have subsequently received research study for varied reasons. Figure 18.35 illustrates some examples of the types of membranes realized: free-standing structures, membranes within silicon chips, and membranes transferred to and supported on substrates of another material.

There are now three top-down techniques developed to realize porous silicon membranes from solid silicon: 1) electrochemical etching (anodization), 2) micromachining, and 3) thin film deposition/annealing. These techniques create different pore morphologies and are suited to different membrane thicknesses and porosity ranges. In this regard they are quite complementary.

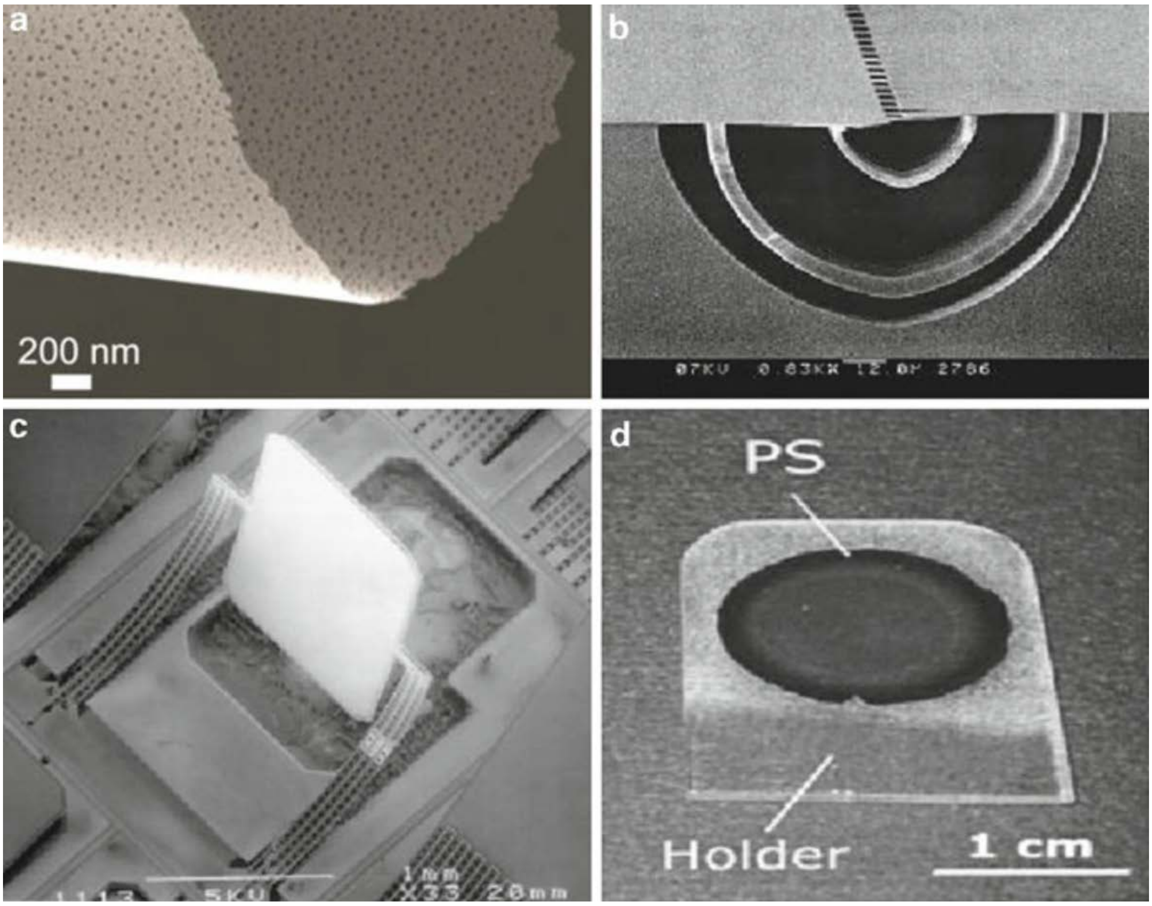


Fig. 18.35 Porous silicon membrane examples: (a) ultrathin membrane , (b)membrane in microfluidic channel , (c) suspended micromachined membrane, (d) membrane on silica support

1) Porous silicon membranes have been realized electrochemically by four routes:

- Anodization right through entire silicon wafers of standard thickness
- Anodization right through thin solid silicon membranes
- Anodization of part of the thickness of a wafer and then “lift off” by raising the current density sharply for a few seconds
- Pre-thinning of selected areas of wafers with subsequent anodization

2) Micromachining. This approach creates porosity with highly defined mesopore diameters at low concentrations. The slitlike pores result from removal of a sacrificial silicon oxide layer of highly controlled thickness (tunable in the mesopore to macropore range). Mesopore diameter is very accurately controlled (<5 % variation). The resulting membranes have thicknesses in the range 0.5–5 micron, but the overall mesoporosity is very low (~1 %). This provides excellent mechanical stability but at the cost of increased likelihood of pore blockage in complex biological media. Such structures are also expected to have similar biochemical properties to “bulk” crystalline silicon and thus not be medically biodegradable like highly porous silicon

3) Deposition/Annealing. This approach creates mesoporosity via thermal anneal treatments of ultrathin (10–50 nm) silicon films. A typical process flow is shown schematically in Fig. 18.36. For silicon films sandwiched between silica layers, porosities up to 15 % were achieved.

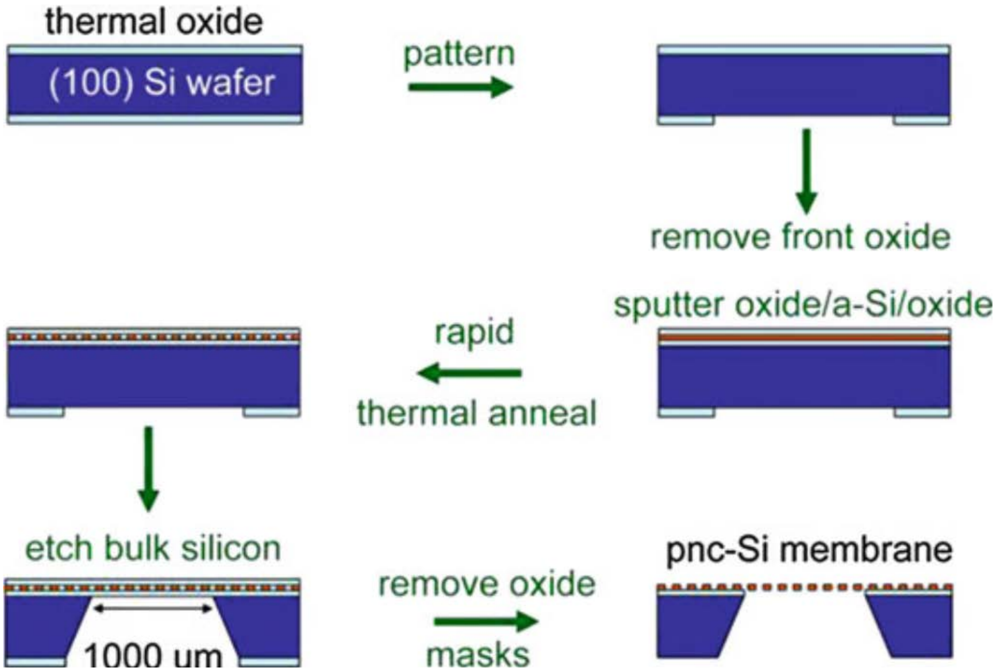


Fig. 18.36 Fabrication of mesoporous silicon membranes by thin layer deposition and thermal treatment

18.8 MACE Silicon Nanostructures.

The formation of porous silicon by metal-assisted chemical etch (MACE) was discovered in 1997 when patterned aluminum on silicon rapidly induced selective formation of porous silicon layers in stain-etch conditions. In a typical MACE process, a patterned metal deposited over a silicon substrate catalyzes etch of silicon in its vicinities, when placed in an oxidizing solution of hydrofluoric acid (HF). By anisotropically etching silicon in the vicinity of a metal mask, MACE generates silicon structures dictated by metal patterning and etching conditions: pores, nanowires, and porous silicon layers (Fig. 18.37). MACE combined with lithographic patterning generates high aspect ratio silicon nanostructures of tailored geometry. Under appropriate MACE conditions, porous silicon forms in the un-etched silicon around the metal . Combining anisotropic etching and porosification, MACE is a simple strategy to form high aspect ratio porous silicon nanostructures. MACE is a top-down nanofabrication technique that allows finer control than competing bottom-up strategies for growth of silicon nanowires.

MACE provides highly crystalline nanowires and grants control over their doping, surface roughness, and porosity. MACE is a low-cost strategy, which works at ambient temperature and pressure, to reproducibly generate wafer-scale porous silicon as well as ordered arrays of oriented nanowires, nanorods, and vias. A wide range of transition metals can catalyze MACE. Noble metals are especially favored for the formation of nanowires as well as for nanostructures with defined cross sections since they better preserve their structure during the etch, as they do not dissolve in HF. Non-noble transition metals have been mostly used to form sNP, pSi, and polished surfaces. Metals can be deposited on Si by a variety of methods, including electroless deposition.

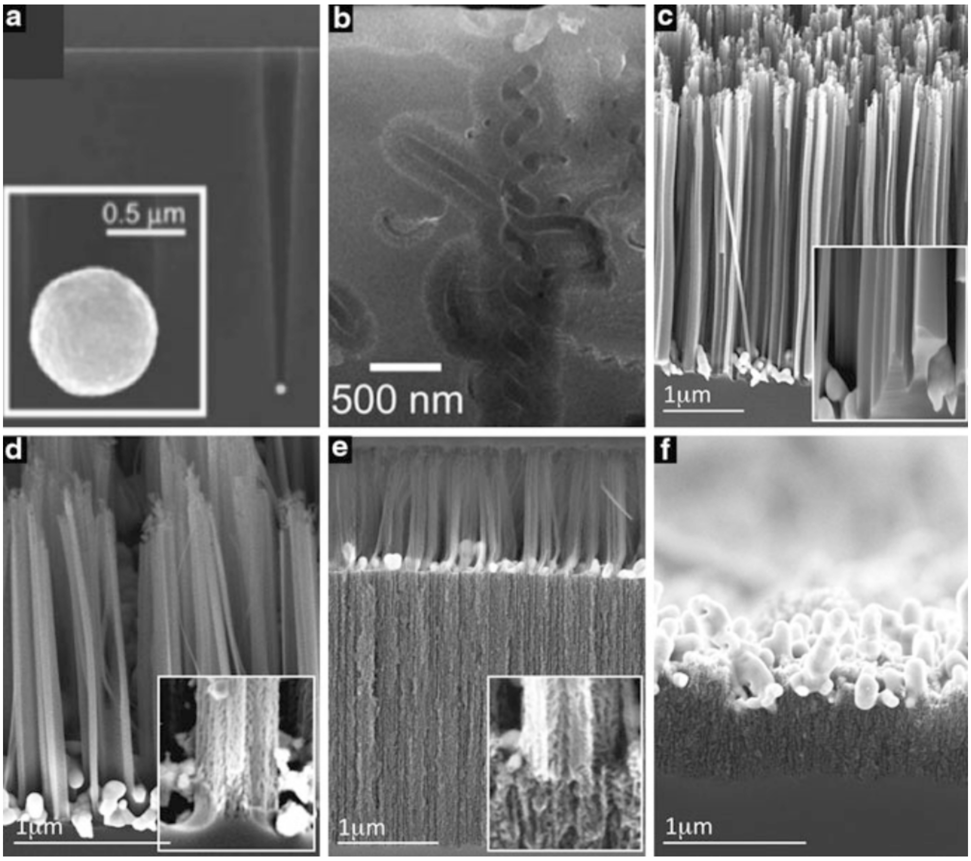


Fig. 18.37 Overview of nanostructures synthesized by metal-assisted chemical etch. (a) Nanopores with solid sidewalls (sNPs); (b) nanopores with mesoporous sidewalls (mNPs); (c) solid nanowires (sNWs); (d) mesoporous nanowires (mNWs); (e) porous nanowires over porous silicon; (f) porous silicon (pSi).

1-Step Metal-Assisted Etching.

The metal deposition and etching are carried out in the same chemical solution (Fig. 18.38). A typical solution is 0.02 M AgNO3 + 4.6 M HF. This method leads to the formation of porous Si or pillar-like or craterlike microstructures or arrays of silicon nanowires (SiNWs) standing vertically on the Si substrate . Metal ions are dissolved in HF. When the redox potential of the ions is more positive than the valence band of Si, a galvanic reaction occurs in which the ions are reduced to metal as particles, dendrites, and film, while the Si is oxidized and dissolved in HF (case of noble metals) following the reaction:

$$4\ M^{+}(aq) + Si^{0}(s) + 6F^{-} \rightarrow 4\ M(s) + SiF_6{}^{2-}(aq)$$

2-Step Metal-Assisted Etching.

The metal particles are deposited on the Si surface generally by electroless metal deposition (EMD) or chemical vapor deposition or sputtering, prior to the etching in the HF solution in the presence of an oxidizing agent. During the etching, dissolution of Si underneath the metal particles is strongly enhanced, and pores are formed while the particles sink into the Si pores (Fig. 18.38). The etching can be done in various HF solutions containing an oxidizing agent, typically H_2O_2

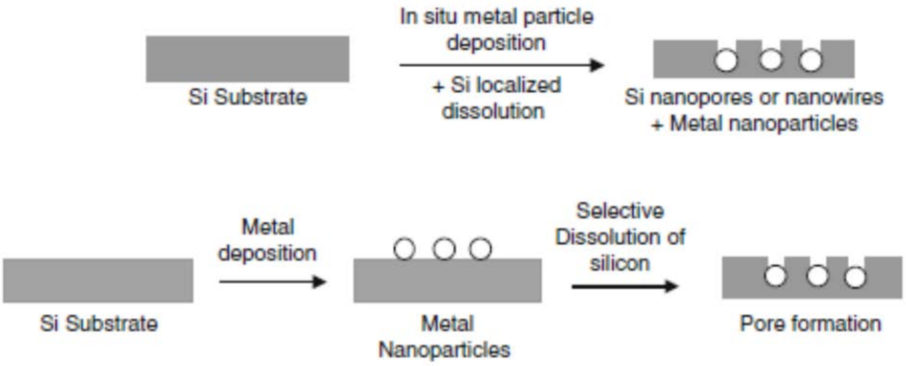


Fig.18.38 Schematic of 1-step and 2-sstep metal-assisted etching

18.9 Polymer - Porous Silicon Composites

The combination of polymers with nanostructured silicon scaffolds, in particular porous Si, into a single composite system opens vast opportunities for developing advanced functional materials. These composites display unique properties that are culminated by the characteristics of each building block, to allow the design of highly tunable nanomaterials. Over the past decade, various PS-polymer composites were introduced and their application as sensors, actuators, optical devices , drug delivery systems , and tissue-engineered scaffolds was demonstrated. Porous Si-polymer composites may be designed in diverse configurations. Figure 18.39 schematically illustrates the most common structures: PS infiltrated with a polymer, polymer-coated PS, polymer-capped PS, released PSi film supported by a polymer, PS particles encapsulated by a polymer, and composite microparticles.

Each of these structures possesses different properties, which can be further refined by a proper choice of the polymer constituent and the PS nanostructure. The simplest composite structure is that of a polymer-infiltrated PSi substrate (Fig. 18.39a), wherein the polymer fills the entire porous volume . The polymer is confined within the nanoscale pores, and its interaction with the pore wall can be enhanced by covalent attachment. To some extent, polymer-coated PS (Fig. 18.39b) has a similar design, where the polymer only forms a uniform layer onto the pore walls, resulting in an open porous structure . Polymer-capped PS is a more sophisticated architecture (Fig. 18.39c), in which the polymer only forms a blocking layer on top of the PS, leaving the greater fraction of the porous volume unoccupied. The fabrication of these composites is more challenging, as the degree of polymer penetration into the pores needs to be precisely controlled. Figure 18.39d presents a polymeric replica from PS. These replicas are usually prepared from polymer-infiltrated PS by selective removal of the Si scaffold . The previously described designs make use of intact PSi substrates. However, the following PS-polymer composite configurations require mechanical processing of the PS. A freestanding PS film supported by polymer is presented in Fig. 18.39e. In this case, separation of the porous layer from the bulk Si is typically achieved by electropolishing . Further fracturing of these porous films allows us to fabricate particulate composites . The resulting PSi micro-/nanoparticles can be embedded in a polymer matrix (Fig. 18.39f) or encapsulated individually by a polymer layer (Fig. 18.39g). The latter case is more synthetically challenging.

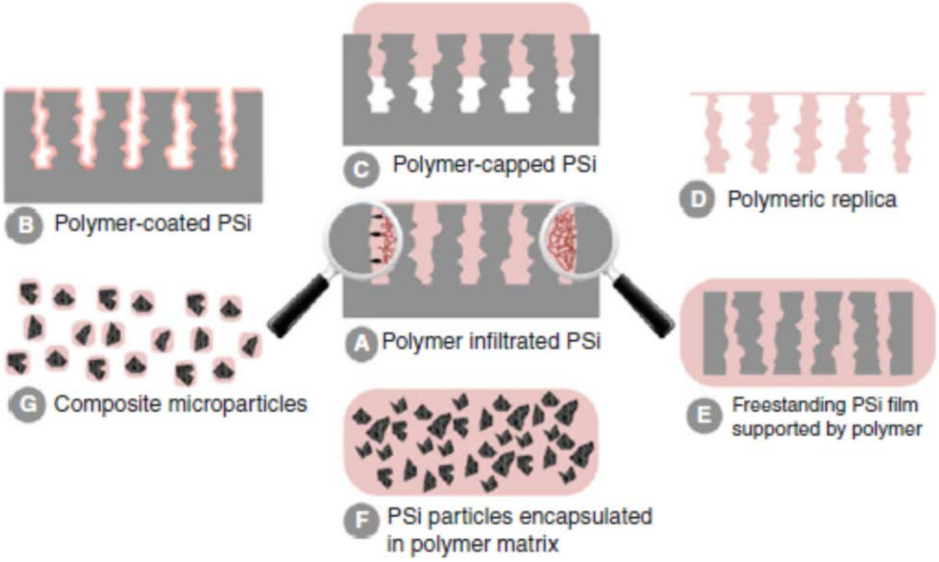


Fig. 18.39 Common structures of PS-polymer composites. Insets in A illustrate interfacial chemistry where the polymer is not attached to PS (right) and is covalently attached to the PSi surface through different linkers. Schematics are not drawn to scale

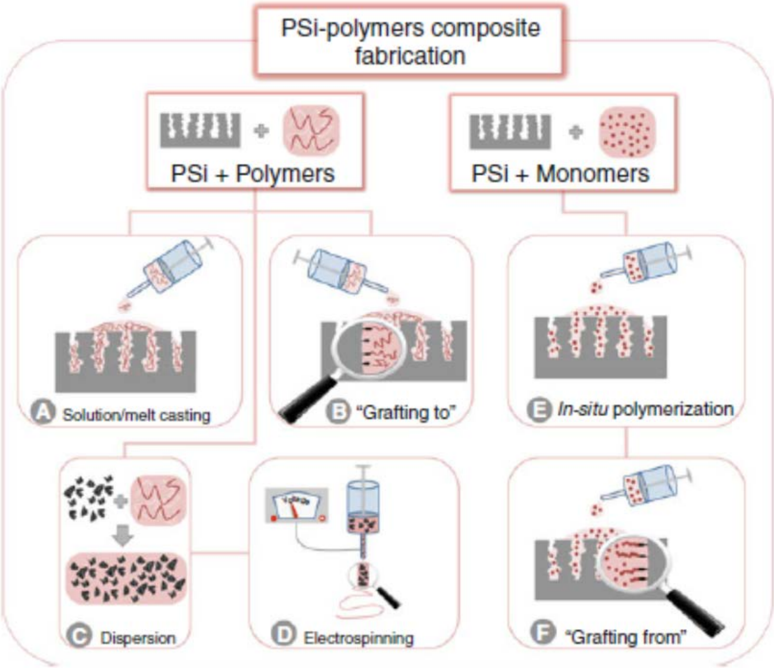


Fig. 18.40 Common practiced methods for the fabrication of PSi-polymer composites

There are many synthetic approaches for integrating polymers with PSi. Figure 18.40 summarizes the main methods that are practiced for the fabrication of PS-polymer composites. In general, these techniques can be divided into two main categories. The first is incorporating a preformed polymer with the Si scaffold. The second involves in situ polymerization of monomers within/on the PS.

18.10 Imprinting Porous Silicon

This chapter describes the capabilities of the direct imprinting of porous substrates (DIPS) technique for patterning and modifying the physical properties of porous silicon films. DIPS can achieve very high-resolution two-dimensional and three-dimensional patterning with feature sizes below 100 nm while eliminating the need for intermediate masking materials and etch recipes that complicate and increase the expense of other patterning techniques.

While several imprinting techniques have been established for polymers and other soft materials , imprinting has only recently been demonstrated as a high-fidelity, low-cost, and straightforward method for patterning porous silicon and other porous materials including nanoporous gold, porous titania, and porous alumina This process, referred to as "direct imprinting of porous substrates" (DIPS), overcomes several key challenges associated with applying conventional lithographic techniques to porous media. First, traditional lithography strategies (e.g., electron-beam lithography (EBL) or photolithography) are often expensive and limited by a trade-off between resolution and throughput. Second, resists and thermoplastics are likely to infiltrate the pores, potentially modifying the internal porous structure and often proving very difficult to remove, especially in deep porous networks. Third, wet etching is not always an option due to chemical incompatibilities, and dry-etching techniques can be expensive and require specific chemistries that must be appropriately characterized and tailored for each unique porous material. As an imprinting process, DIPS can achieve very high-resolution patterning while overcoming many of the aforementioned challenges through eliminating the need for intermediate masking materials and etch recipes.

Figure 18.41 illustrates the DIPS process. Similar to nanoimprint lithography , DIPS first relies on the fabrication of a reusable stamp. Silicon is a convenient stamp material owing to its large material hardness (~10 GPa) and to the large variety of imprintable patterns that can be readily fabricated through conventional lithographic techniques. Moreover, for patterning porous silicon films, using a silicon stamp minimizes contamination issues, which is especially important for nanoparticles designed for biomedical applications. After preparing the reusable stamp, the porous material must be prepared on a solid substrate support, for example, silicon, glass, or metal. For porous silicon, the parent silicon wafer is an excellent natural support, although it is possible to lift-off porous layers for anchoring to alterative substrates prior to or after performing DIPS. Imprinting is then performed by directly applying the stamp to the porous substrate at a pressure on the order of 100 N/mm2 (100 MPa). The straightforward process can be performed in a matter of seconds at room temperature without the need for any curing, developing, baking, or etching processes (which are often required in nanoimprint lithography or soft lithographic techniques).

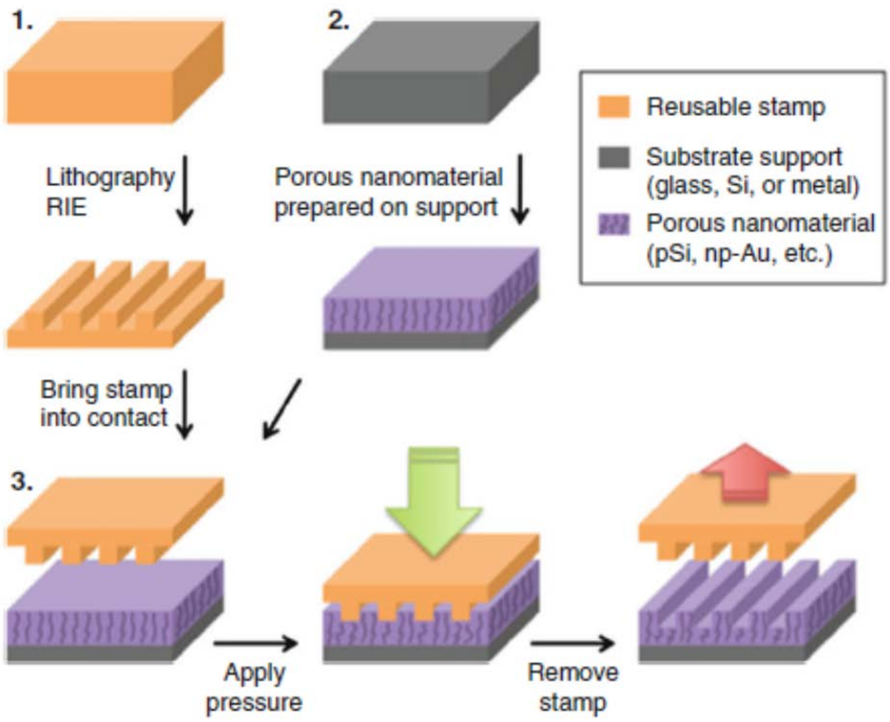


Fig. 18.41 Illustration of the direct imprinting of porous substrates (DIPS) process

18.11 Porous Silicon Micromachining Technology

In this chapter, silicon electrochemical micromachining (ECM) technology is eviewed with particular emphasis to the fabrication of complex microstructures and microsystems, as well as to their applications in optofluidics, biosensing, photonics, and medical fields. ECM, which is based on the controlled electrochemical dissolution of n-type silicon under backside illumination in acidic (HF-based) electrolytes, enables microstructuring of silicon wafers to be controlled up to the higher aspect ratios (over 100) with sub-micrometer accuracy, thus pushing silicon micromachining well beyond up-to-date both wet and dry microstructuring technologies. Both basic and advanced features of ECM technology are described and discussed by taking the fabrication of a silicon microgripper as case study.

ECM (fig.18.42) technology enables the low-cost fabrication of high-complexity silicon microstructures and microsystems with sub-micrometer accuracy at aspect-ratio values (about 100) that are well beyond up-to-date, both wet and dry, micromachining technologies. Among ECM main features there are (i) fabrication of microstructure with very high aspect ratio, which can be three times higher than that of state-of-the-art deep etching techniques employed for microsystem fabrication; (ii) tunability of the etching anisotropy (from zero to one) in real-time as the etching progresses, which enables 3D freestanding microstructure fabrication by one-step etching; (iii) fine control of the etching features at the sub-micrometer scale, for both small (features down to 1 μm in size) and large (length over 1 mm) areas; and (iv) high quality of etched structures in terms of both verticality (percentage error over a depth of 100 μm of about 0.04 %) and surface roughness (about 20 nm). In the near future, ECM technology is expected to be greatly exploited at the lab scale, although not limited to, for silicon microstructuring, with potential impact on both microdevices and applications in different research fields.

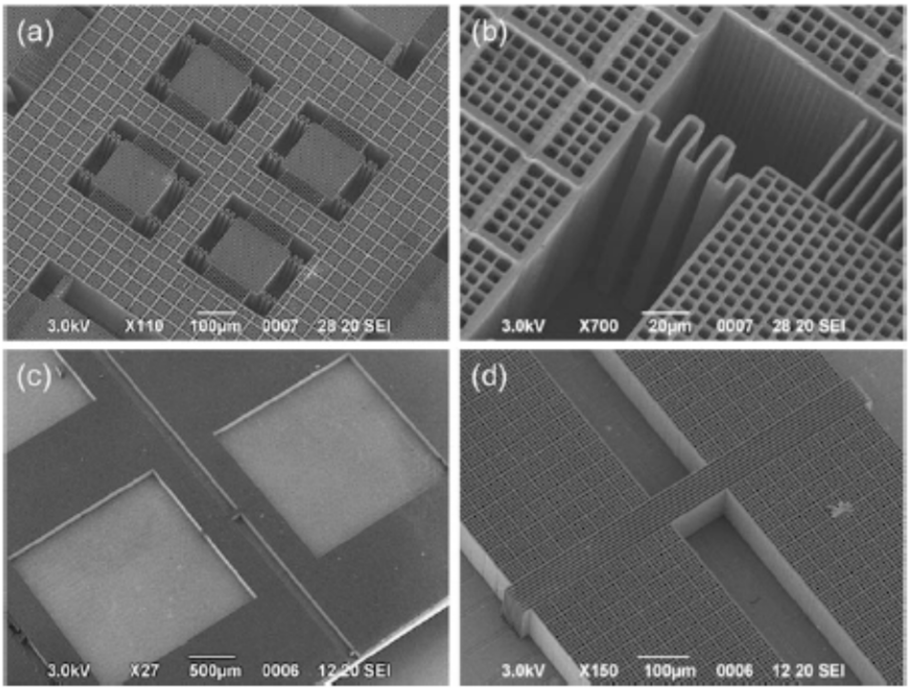
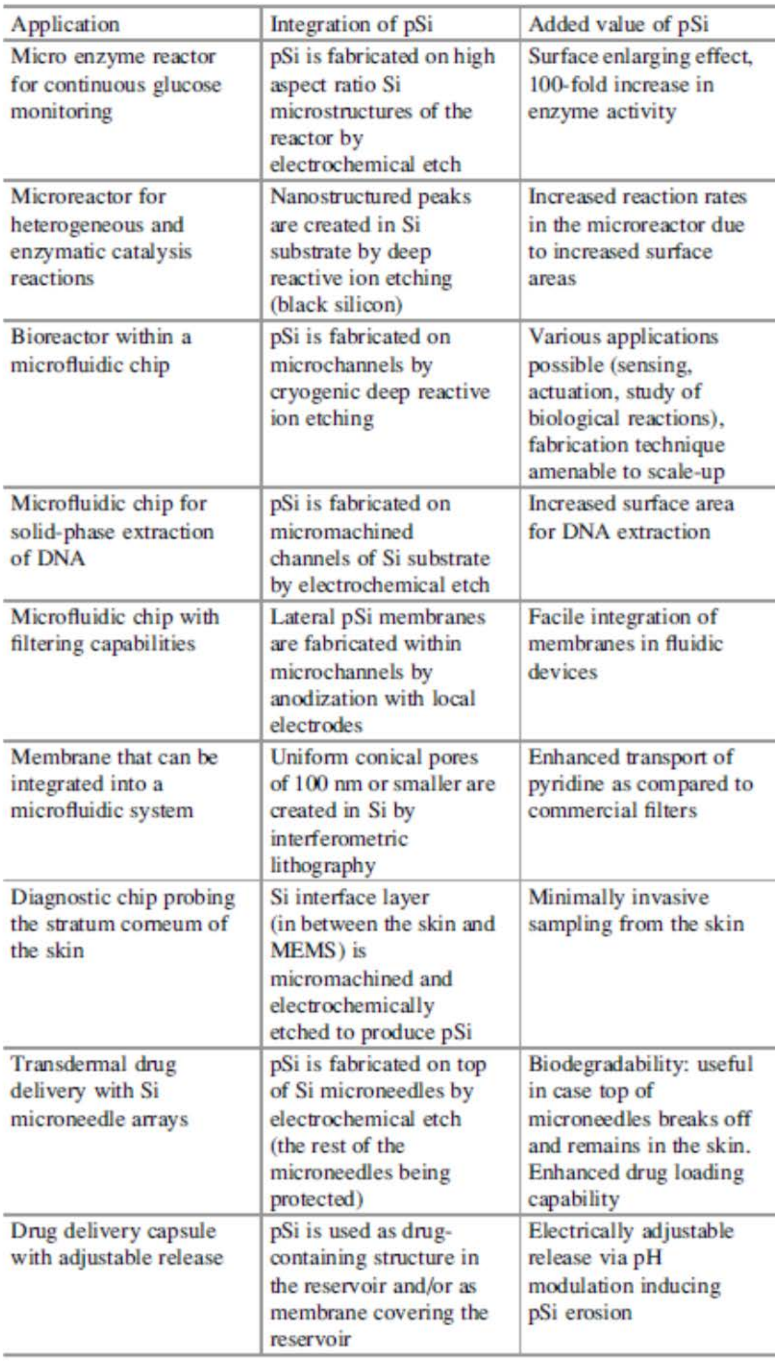


Fig. 18.42 Examples (SEM images) of two microsystems fabricated by ECM technology: (a) MEMS structures consisting of freestanding inertial masses suspended from the substrate by folded springs and featuring comb fingers for actuation; (b) magnification of a folded spring of the MEMS structures in (a); (c) optofluidic platform featuring a one-dimensional photonic crystal assensing transducer; (d) magnification of the photonic crystal transducer of the platform in (c)

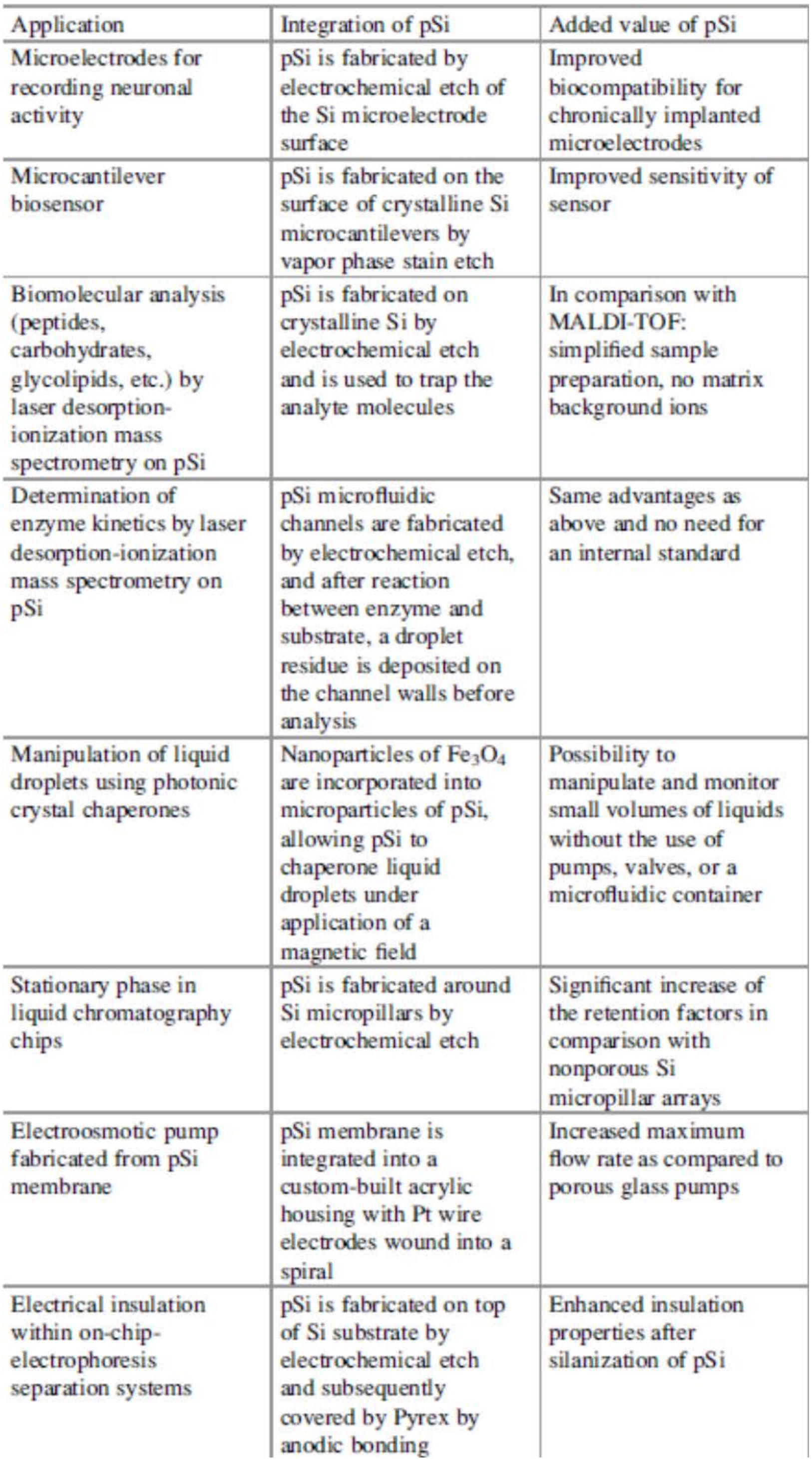
18.12 Porous Silicon Functionalities for BioMEMS

This chapter presents a literature survey of the applications of porous silicon in BioMEMS (biological/biomedical microelectromechanical systems). This material possesses properties particularly suitable for biomedical purposes: biocompatibility, biodegradability, photoluminescence, ability to precisely control the pore size and shape, and possibility to easily modify the surface chemistry. Many applications can, for instance, be found in the fields of sensing and delivery of therapeutics. It is expected that the number of BioMEMS using porous silicon will continue to increase in the future with the development of lab-on-a-chip/ microfluidic devices.

Table 18.7 BioMEMS applications using pSi as part of a chip-based system



Application	Integration of pSi	Added value of pSi
Micro enzyme reactor for continuous glucose monitoring	pSi is fabricated on high aspect ratio Si microstructures of the reactor by electrochemical etch	Surface enlarging effect, 100-fold increase in enzyme activity
Microreactor for heterogeneous and enzymatic catalysis reactions	Nanostructured peaks are created in Si substrate by deep reactive ion etching (black silicon)	Increased reaction rates in the microreactor due to increased surface areas
Bioreactor within a microfluidic chip	pSi is fabricated on microchannels by cryogenic deep reactive ion etching	Various applications possible (sensing, actuation, study of biological reactions), fabrication technique amenable to scale-up
Microfluidic chip for solid-phase extraction of DNA	pSi is fabricated on micromachined channels of Si substrate by electrochemical etch	Increased surface area for DNA extraction
Microfluidic chip with filtering capabilities	Lateral pSi membranes are fabricated within microchannels by anodization with local electrodes	Facile integration of membranes in fluidic devices
Membrane that can be integrated into a microfluidic system	Uniform conical pores of 100 nm or smaller are created in Si by interferometric lithography	Enhanced transport of pyridine as compared to commercial filters
Diagnostic chip probing the stratum corneum of the skin	Si interface layer (in between the skin and MEMS) is micromachined and electrochemically etched to produce pSi	Minimally invasive sampling from the skin
Transdermal drug delivery with Si microneedle arrays	pSi is fabricated on top of Si microneedles by electrochemical etch (the rest of the microneedles being protected)	Biodegradability: useful in case top of microneedles breaks off and remains in the skin. Enhanced drug loading capability
Drug delivery capsule with adjustable release	pSi is used as drug-containing structure in the reservoir and/or as membrane covering the reservoir	Electrically adjustable release via pH modulation inducing pSi erosion



Application	Integration of pSi	Added value of pSi
Microelectrodes for recording neuronal activity	pSi is fabricated by electrochemical etch of the Si microelectrode surface	Improved biocompatibility for chronically implanted microelectrodes
Microcantilever biosensor	pSi is fabricated on the surface of crystalline Si microcantilevers by vapor phase stain etch	Improved sensitivity of sensor
Biomolecular analysis (peptides, carbohydrates, glycolipids, etc.) by laser desorption-ionization mass spectrometry on pSi	pSi is fabricated on crystalline Si by electrochemical etch and is used to trap the analyte molecules	In comparison with MALDI-TOF: simplified sample preparation, no matrix background ions
Determination of enzyme kinetics by laser desorption-ionization mass spectrometry on pSi	pSi microfluidic channels are fabricated by electrochemical etch, and after reaction between enzyme and substrate, a droplet residue is deposited on the channel walls before analysis	Same advantages as above and no need for an internal standard
Manipulation of liquid droplets using photonic crystal chaperones	Nanoparticles of Fe_3O_4 are incorporated into microparticles of pSi, allowing pSi to chaperone liquid droplets under application of a magnetic field	Possibility to manipulate and monitor small volumes of liquids without the use of pumps, valves, or a microfluidic container
Stationary phase in liquid chromatography chips	pSi is fabricated around Si micropillars by electrochemical etch	Significant increase of the retention factors in comparison with nonporous Si micropillar arrays
Electroosmotic pump fabricated from pSi membrane	pSi membrane is integrated into a custom-built acrylic housing with Pt wire electrodes wound into a spiral	Increased maximum flow rate as compared to porous glass pumps
Electrical insulation within on-chip-electrophoresis separation systems	pSi is fabricated on top of Si substrate by electrochemical etch and subsequently covered by Pyrex by anodic bonding	Enhanced insulation properties after silanization of pSi

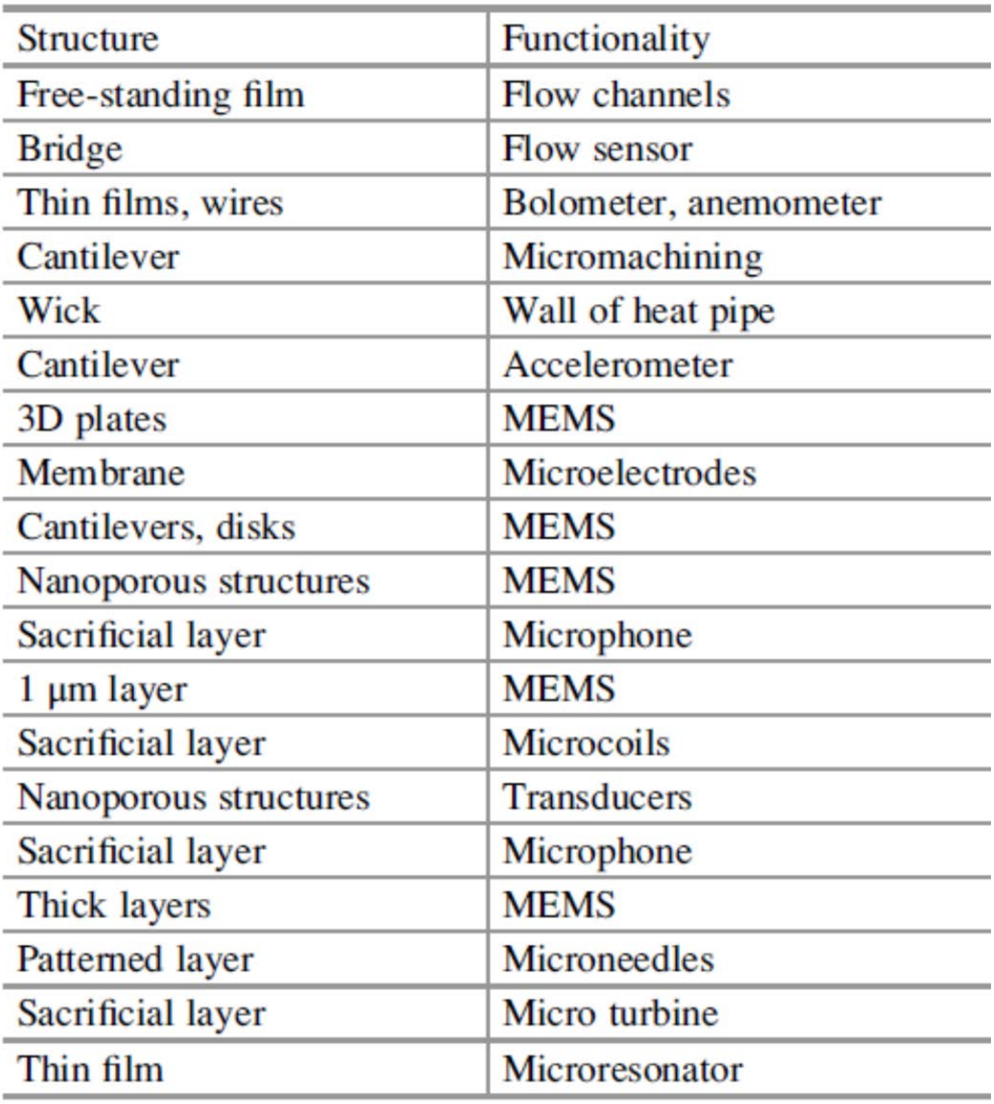
18.13 Porous Silicon for Microdevices and Microsystems

Macroporous silicon technology found its principal application in integrated microsystems as sacrificial layer: multilayered and suspended structures, such as bridges, membranes, and cantilevers, often require fabrication, almost always by isotropic etching and removing, by alkaline-based water solution, of a porous layer.

The thickness of this layer can be up to 100 µm, or more, which is very much greater with respect to those obtained by thin film deposition techniques (<10 µm): in this sense, porous silicon passive layer is an exclusive technology. In Table 18.8 are reported some references of porous silicon sacrificial layers together with the functionalities of the resulting microsystems.

Silicon on insulator technology and thermal insulation are other important fields where porous silicon morphology plays a key role, and thus, it is often used in complex microsystems: thermal properties of porous silicon layers can be strongly modulated by changing the porosity, i.e., the amount of air present in the silicon volume. On the other hand, pores can be completely filled by silicon dioxide, so that a nanocrystalline film can be transformed by thermal oxidation in an oxide layer preserving the desired geometry.

Table 18.8 Porous silicon as sacrificial layer



Structure	Functionality
Free-standing film	Flow channels
Bridge	Flow sensor
Thin films, wires	Bolometer, anemometer
Cantilever	Micromachining
Wick	Wall of heat pipe
Cantilever	Accelerometer
3D plates	MEMS
Membrane	Microelectrodes
Cantilevers, disks	MEMS
Nanoporous structures	MEMS
Sacrificial layer	Microphone
1 µm layer	MEMS
Sacrificial layer	Microcoils
Nanoporous structures	Transducers
Sacrificial layer	Microphone
Thick layers	MEMS
Patterned layer	Microneedles
Sacrificial layer	Micro turbine
Thin film	Microresonator

18.14 Porous Silicon Optical Waveguides

A waveguide structure consists of a high refractive index layer surrounded by lower refractive index media. In porous silicon, waveguides can be achieved by fabricating single-layer, double-layer, or triple-layer structures, as illustrated in Fig. 18.43; graded-index profile waveguides are also achievable. Conceptually, in traditional step index waveguides, light remains trapped in the high refractive index (low porosity) layer due to total internal reflection at the interfaces with the adjacent lower refractive index materials. Photonic crystal mirrors have also been utilized to provide light confinement in a porous silicon waveguide . Very recently, three-layer porous silicon antiresonant reflectingoptical waveguides (ARROW) have been demonstrated . A waveguide structure can support a discrete number of modes, each with a characteristic effective index (neff), based on the thickness and refractive index of each layer of the waveguide. While light is only guided in the low-porosity waveguide layer, an exponentially decaying field does exist in the regions directly above and below the low-porosity waveguide layer. This evanescent field is critical for allowing light to be efficiently coupled into and out of the waveguide, Scanning near-field optical microscopy has been used to directly probe the evanescent field and measure the propagation constant of a porous silicon waveguide. If a secondary material is incorporated into one or more of the porous silicon layers constituting the waveguide structure, then the field distribution, neff of each mode, and possibly the number of modes is changed. Note that while porous silicon double-layer structures, in general, have been intensively investigated over the past few years for size-selective and environmentally robust sensing applications , many of these structures are not waveguides because the appropriaterefractive index profile is not achieved and only the reflectance spectrum is interrogated.

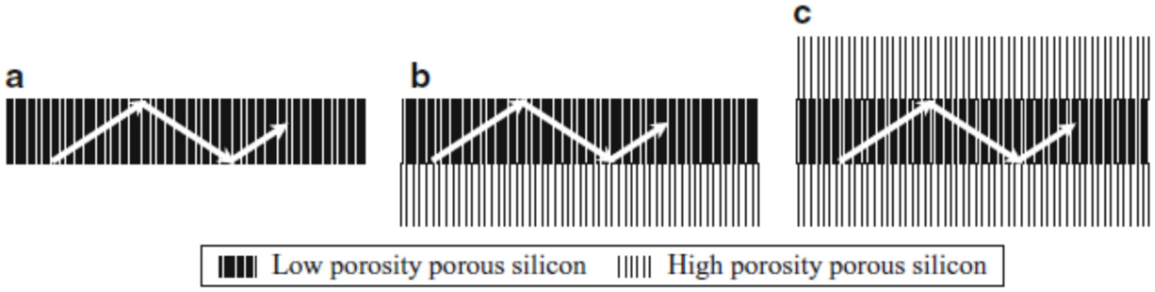


Fig. 18.43 Schematics of porous silicon waveguides consisting of (a) one, (b) two, or (c) three porous silicon layers. Light is confined in a high refractive index (low porosity) porous silicon layer when surrounded by lower refractive index media, which can include a higher porosity porous silicon layer, air, or another low index solid, liquid, or gaseous substance

Figure 18.44 illustrates some common porous silicon planar waveguide configurations. Rectangular waveguide geometries, such as rib or ridge waveguides, including Mach-Zehnder interferometer configurations, have also been studied and achieved in porous silicon using standard lithography techniques or direct laser writing

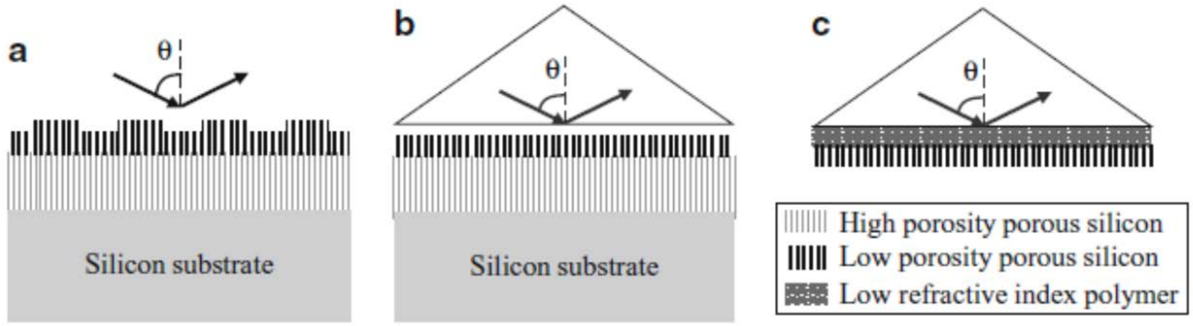


Fig. 18.44 Common porous silicon waveguide configurations: (a) grating-coupled two-layer waveguide, (b) prism-coupled two-layer waveguide in the Otto configuration with an air gap between the waveguide and the prism, and (c) prism-coupled single-layer waveguide in the Kretschmann configuration using a low index polymer layer or low index fluid to attach the porous silicon membrane to the prism. End-fire coupling using a microscope objective and butt coupling using an optical fiber (not shown) have also been employed

Chapter 19. Aerogels

19.1 Introduction

The term aerogel was first introduced by Kistler in 1932 to designate gels in which the liquid was replaced with a gas, without collapsing the gel solid network. While wet gels were previously dried by evaporation, Kistler applied a new supercritical drying technique, according to which the liquid that impregnated the gels was evacuated after being transformed to a supercritical fluid. In practice, supercritical drying consisted in heating a gel in an autoclave, until the pressure and temperature exceeded the critical temperature Tc and pressure Pc of the liquid entrapped in the gel pores. This procedure prevented the formation of liquid–vapor meniscuses at the exit of the gel pores, responsible for a mechanical tension in the liquid and a pressure on the pore walls, which induced gel shrinkage. Besides, a supercritical fluid can be evacuated as gas, which in the end lets the "dry solid skeleton" of the initial wet material. The dry samples that were obtained had a very open porous texture, similar to the one they had in their wet stage. Overall, aerogels designate dry gels with a very high relative or specific pore volume, although the value of these characteristics depends on the nature of the solid and no official convention really exists . Typically, the relative pore volume is of the order of 90% in the most frequently studied silica aerogels. This porous texture is well illustrated in Figure 19.1, which shows a Y_2O_3 aerogel, transmission electron micrographs (TEM) of this aerogel and of a SiO_2 aerogel, plus a technical art illustration of a local domain in an aerogel solid network, inspired from these two micrographs.

By considering recent developments in the synthesis of aerogels, it is indeed more realistic to define these materials with reference to the initial idea of Kistler, simply as gels in which the liquid has been replaced by air, with very moderate shrinkage of the solid network. This definition enlightens the main difference between aerogels and xerogels. The term xerogel is defined by IUPAC as an "Open network formed by the removal of all swelling agents from a gel". And indeed, the capillary stresses previously mentioned may contract a wet gel down to 30% or less of its initial volume and hence to a much lower value than in aerogels.

In his founding work, Kistler synthesized a series of aerogels of very different nature. Besides silica aerogels, which could be relatively easily handled without any fracture damage, he succeeded in making alumina aerogels, which proved to be mechanically very weak. He also synthesized more exotic aerogels of tungstic, ferric, or stannic oxide and nickel tartrate, a list to which organic aerogels of cellulose, nitrocellulose, gelatine, agar, or egg albumin must be added. Besides, the fluids in which supercritical drying was performed were also quite diverse. The silica gels were made from sodium silicate (water glass, Na_2SiO_3), and water was found to disperse the gel to a powder near the aqueous medium critical point. To maintain amonolithic character, the wet gelswere actually dialyzed to replace water by ethanol, so that Kistler initiated supercritical drying in this alcohol. But Kistler also applied the supercritical method to other fluids: ethyl ether, propane, and even liquid CO_2 in an unsuccessful attempt to make rubber aerogels

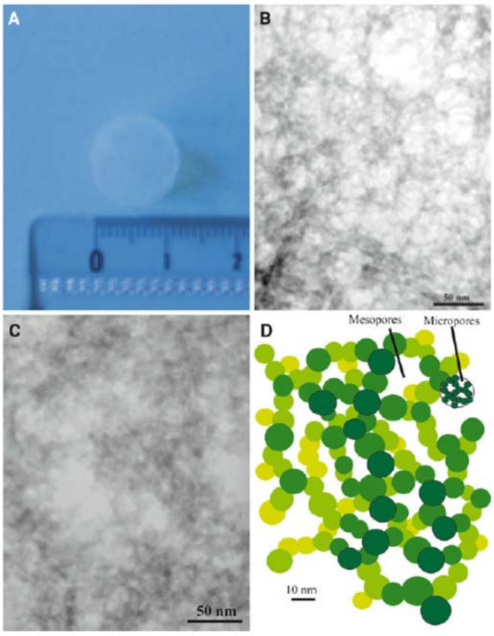


Fig. 19.1 A.Y_2O_3 aerogel dried by the CO_2 supercritical method; B. TEM micrograph of the aerogel in A. ; C. TEM micrograph of a SiO_2 aerogel made from 80% TMOS and 40% methyltrimethoxysilane, dried by the CO_2 supercritical method; D. Technical modeling of a local domain in an oxide aerogel, inspired from local zones in the TEM micrographs B. and C.

At last, Kistler rapidly perceived the potential industrial applications of aerogels, for instance, as catalysts, thickening agents, insulating materials, or water repellents when they were hydrophobic . He deposited several patents: one of them was assigned to the Monsanto Chemical Company, which started the industrial production of silica aerogels commercialized under the name of Santocel®. He also patented the first hydrophobic silica aerogels made by sylilation with trichloromethyl silane, for use as water repellents. Hence, Kistler initiated most of the directions in which aerogels were later developed.

After Kistler, the chemical composition of the materials that were made as aerogels has progressively diversified. A large range of simple oxides or binary oxides were investigated . Borate aerogels were prepared at the Sandia National Laboratories , at the Lawrence Livermore National Laboratory in Berkeley, developed new organic aerogels made by polycondensation of resorcinol–formaldehyde (RF) for a laser project . These aerogels had a low thermal conductivity of 0.012 W m^{-1} K^{-1} . Their pyrolysis in an inert atmosphere produced a carbon aerogel that conducted electricity.

Examples of an RF aerogel, the carbon aerogel derived from it, and the fractal structure of this carbon aerogel observed under a TEM are illustrated in Figure 19.2.

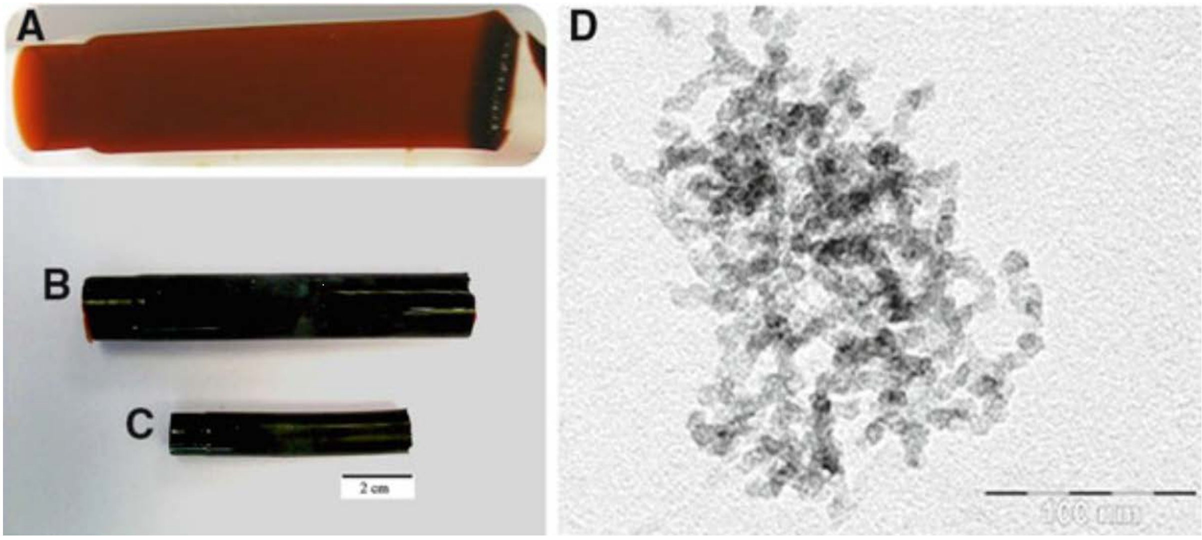


Fig.19.2 Resorcinol Formaldehyde (RF) and carbon aerogels. A. RF gel before drying; B. RF aerogel after direct CO2 supercritical drying; C. carbon aerogel derived from the RF aerogel in B. by pyrolysis at 1,050°C under nitrogen gas flow; D. transmission electron micrograph of the carbon aerogel in C.

19.2 SiO_2 Aerogels

This chapter focuses on one of the most studied aerogel materials, silica aerogels. Silica sol gels are generally formed via hydrolysis and polycondensation reactions of silica precursors in the presence of an acid and/or base catalyst (Figure 19.3). The resulting silica matrix is highly porous, and the pores of the sol gel are filled with the solvent byproducts of the hydrolysis and polymerization reactions. If the solvent mixture can be removed from the wet sol gel without substantial structural collapse, an aerogel is formed.

It aims at presenting a brief overview of the elaboration steps (sol–gel synthesis, aging, and drying), the textural and chemical characteristics (aggregation features, porosity, and surface chemistry), the main physical properties (from thermal, mechanical, acoustical, and optical, to biological, medical, etc.), and a rather broad panel of related potential applications of these fascinating nanostructured materials.

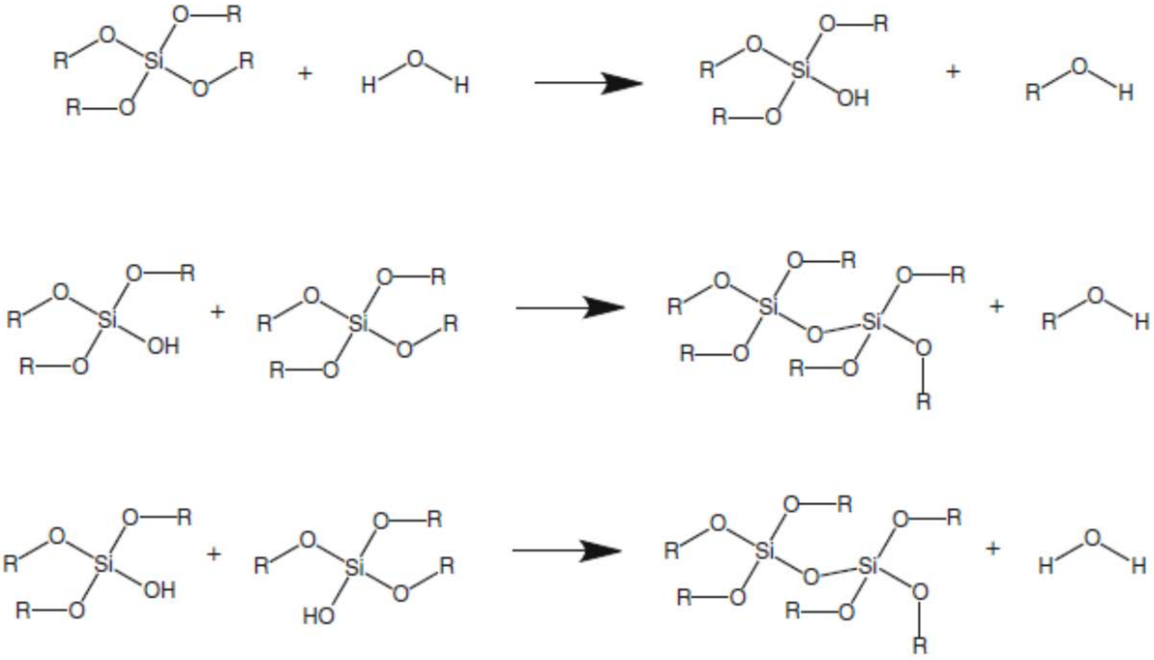


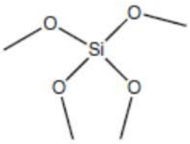
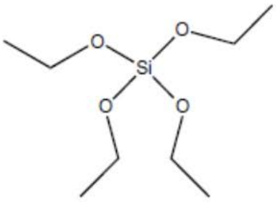
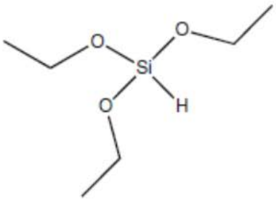
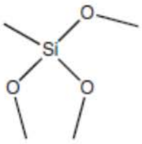
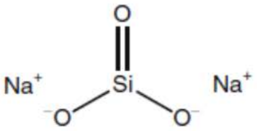
Figure 19.3. Schematic representation of reactions taking place during formation of silica sol gels.

19.2.1. Sol–Gel Synthesis

Schematically, a nanostructured solid network is formed in a liquid reaction medium as a result of a polymerization process, which creates siloxane bridges (≡Si–O–Si≡) between Si atoms delivered by precursor molecules. Such transformations are the equivalent of a polymerization process in organic chemistry, where direct bonds between the carbon atoms of organic precursors are established leading to linear chains or branched (crosslinked) structures, depending on the type of reactive monomers and crosslinkers used. Dispersed solid colloidal silica particles (i.e., nanoparticles with a size well below 1 mm) or “more or less” linear oligomers are formed in the early stage of the sol–gel process. In the second stage, these elementary objects can link with each other while still in the solvent, such as to make up a three-dimensional (3D) open network structure termed a gel, only limited by the container. The continuous transformation of a sol to a gel constitutes the gelation process. The brutal change from the liquid to the solid stage is termed the sol–gel transition. The gels that are obtained are termed either colloidal or polymeric depending on the nature of the building blocks of which the network is composed of and whether they are nanoparticulate or more linear (polymer-like). For both stages, the driving reactions are hydrolysis and water and/or alcohol condensation. For silica gels, a first important precursor is sodium metasilicate Na_2SiO_3, also termed waterglass , which was previously used by Kistler to produce the first silica aerogels reported in the literature. This precursor reacts with an acid such as HCl according to the reactions of the type shown in (19.1) below. A salt is produced, which must be eliminated bytedious dialysis or by exchange for H+ through an acidic ion exchange column

$$Na_2SiO_3 + 2HCl + (x-1)H_2O \rightarrow SiO_2 \cdot xH_2O + 2NaCl \qquad (19.1)$$

Table 19.1. Some precursor chemicals used in the production of silica aerogels

Precursor	Line structure	Abbreviation
Tetramethoxysilane $Si(OCH_3)_4$		*TMOS*
Tetraethoxysilane $Si(OC_2H_5)_4$		*TEOS*
Triethoxysilane $Si(OC_2H_5)_3H$		TriEOS
Methyltrimethoxysilane $Si(OCH_3)_3CH_3$		*MTMS*
Sodium Metasilicate Na_2SiO_3 and related salts		SS
Methyl Silicate 51 $CH_3OSi(OCH_3)_{24}OCH_3$		MS51
Polyethoxydisiloxane (E-40)		PEDS
Silbond H-5 (Ethyl Polysilicate) $C_2H_5O–Si(OC_2H_5)_2O_2–xC_2H_5$		SH5

However, the Si precursors most frequently used nowadays are alkoxides of the $Si(OR)_4$ type, in which R and OR designate alkyl and alkoxide groups, respectively. Often, R is a methyl group CH_3 (or Me). An often used precursor is hence termed TetraMethOxySilane (or TMOS). Another common material has four ethyl rests C_2H_5 (or Et) as R groups, in which case the precursor is termed TetraEthOxySilane (or TEOS). The first inorganic gels synthesized were indeed silica gels made from such precursors, except that they were not dried by a supercritical method to produce aerogels. A much larger list of alkoxide-derived precursors and mixtures of them are used today, comprising, for instance, polyethoxydisiloxane (PEDS), methyltrimethoxysilane (MTMS), methyltriethoxysilane (MTES), 3-(2-aminoethylamino) propyltrimethoxysilane (EDAS), Noctyltriethoxysilane, dimethyldiethoxysilane, and perfluoroalkysilane (PFAS).

Some precursors (such as EDAS) include built-in chemical functionality, which can then be used to modify the resulting gel materials with appropriate chemical synthetic strategies. All these precursors are characterized by the existence of Si–O polar covalent bonds. In real life, Si alkoxides are often available as complexes in solution in their parent alcohol and are typically polymerized to a smaller or larger extent. Even if some monolithic but quite dense aerogel-like materials can be synthesized through ultrasonic-assisted solventless sol–gel routes, their polymerization is mostly carried out in an organic solvent through simultaneous hydrolysis (19.2) and polycondensation of water (19.3) and alcohol (19.4) so that water becomes a reactant added in controlled proportion to drive the hydrolysis reaction (19.2).

$$\equiv \mathrm{Si} - \mathrm{OR} + \mathrm{H_2O} \rightarrow \equiv \mathrm{Si} - \mathrm{OH} + \mathrm{R} - \mathrm{OH}$$

$$\equiv \mathrm{Si} - \mathrm{OH} + \mathrm{HO} - \mathrm{Si} \equiv \rightarrow \equiv \mathrm{Si} - \mathrm{O} - \mathrm{Si} \equiv + \mathrm{H_2O}$$

$$\equiv \mathrm{Si} - \mathrm{OR} + \mathrm{HO} - \mathrm{Si} \equiv \rightarrow \equiv \mathrm{Si} - \mathrm{O} - \mathrm{Si} \equiv + \mathrm{R} - \mathrm{OH} \qquad (19.2\text{-}19.4)$$

Additional parameters that affect the properties of the final aerogel are the nature of the solvent, which most typically is an alcohol but can be acetone or ethyl acetoacetate. In addition, there is also the molar ratio “Si-precursor to water” as well as the concentration of Si precursor in the solvent and the catalyst nature and concentration, which are important parameters too. In short, these parameters rule the nanostructuration of the resulting gel (primary particles size, pores size distribution, fractality, tortuosity, density, etc.). To better control this nanostructuration, researchers have developed two-step processes so that they can successively favor one or another type of catalysis , a trick which is used, for example, to elaborate ultralight silica aerogels or to achieve a good compromise between low thermal conductivity and high optical transparency. The latter aerogels were synthesized first with H_2SO_4 catalysis of TEOS in ethanol under a substoichiometric molar ratio H_2O/TEOS = 1.8, followed by a second catalysis step with HF in ethylacetoacetate. Subsequent studies have shown that the intermediate precondensed species were rather polymeric, while the final gels were more colloidal (Figure 19.3)

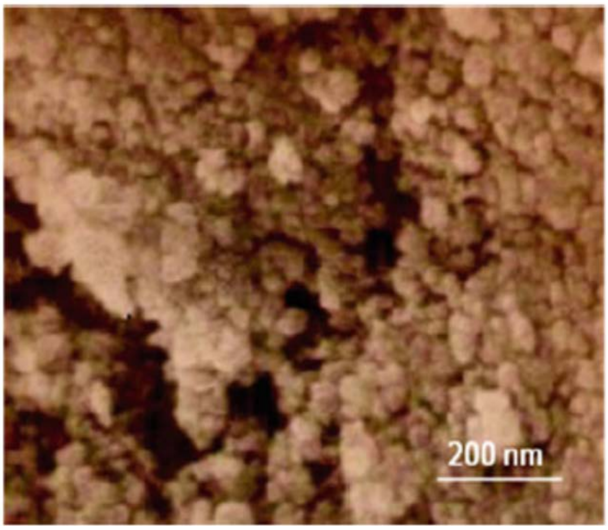


Figure 19.4. High-resolution scanning electron micrographs (HR-SEM) of silica aerogels synthesized by a two-step process (first catalysis with H_2SO_4 followed by HF catalysis)

19.2.2. Ageing

Before drying, silica gels are often aged via different processes. The aim of this step generally is to mechanically reinforce the tenuous solid skeleton generated during the sol–gel process. Aging schematically consists in taking advantage of syneresis and/or Ostwald ripening mechanisms by modifying the composition of the liquid phase contained in pores of the silica gel. Adding water and/or monomeric alkoxysilanes such as TEOS can significantly enhance surface reactions and primarily those involving the residual hydroxy/alcoxy groups. Consequently, supplementary condensation reactions and dissolution/reprecipitation of silica can occur. The associated kinetics depends on the pH and the nature of the solvent. Generally, the particles "neck" area, the average pore size, and the apparent density of the gel increase through aging treatments. If properly controlled, these morphological changes can significantly improve the mechanical properties (E, K, MOR. . .) and the liquid permeability (D). In one case, for example, an augmentation of the shear modulus by up to 23 times was reported . Recently, it has been shown that other types of successful aging treatments exist to simultaneously increase the permeability and mechanical properties. These include, for example, the addition of larger precursor molecules, e.g., polyethoxydisiloxanes or simply adding a dilute HF solution without additional silica precursor (Figure 19.5).

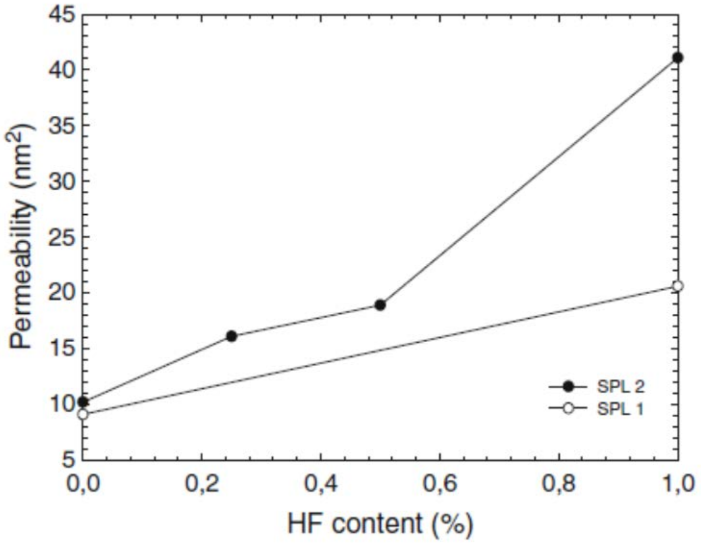


Figure 19.5. Permeability of silica wet gels (i.e., before drying) as a function of HF concentration in the aging bath composed of ethanol (SPL1, white marks) or ethylacetoacetate (SPL2, solid marks) containing 3 vol% water.

19.2.3. Drying

Capillary stresses inevitably occur whenever gas–liquid menisci appear at the pore boundaries, during evaporation of the pores' liquid. Even if the siloxane bonds have approximately 50% covalent character, which makes the silica gels much more capable of resisting to evaporative effects than other oxide gels, particular attention must be paid to this really tricky step. Three main routes are commonly used for drying:

1. Freeze-drying (which necessitates to bypass the triple point).
2. Evaporation (which implies crossing the liquid–gas equilibrium curve).
3. Supercritical drying (which necessitates to bypass the critical point).

In general, freeze-drying and evaporation applied to finely nanostructured silica gels have not yet proved to produce monolithic aerogels. Freeze-drying leads to cracked pieces or even powder-like products Evaporation without specific surface (e.g., sylilation and/or aging treatments results in "dense" (e.g., $\rho > 0.25$ g cm^{-3}) and even cracked materials, so-called xerogels, Figure 19.6). the densification during evaporation comes from the condensation of remaining reactive silica species. When submitted to capillary stresses, initially far distant hydroxyl and/or alkoxy groups can come close enough to one another to react and generate new siloxane bonds, thus leading to irreversible shrinkage (Figure 19.7), because of the inherent flexibility of the silica chains.

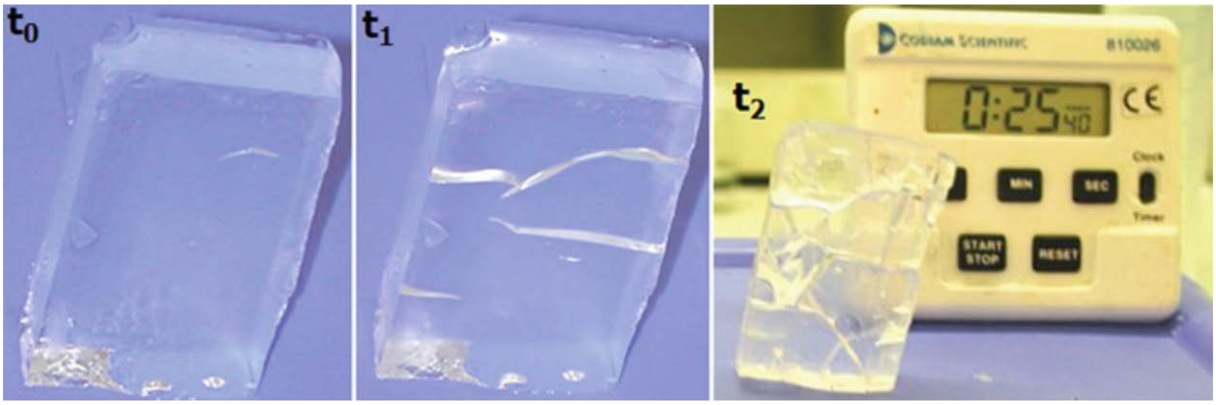


Figure 19.6. Appearance of cracks during evaporative drying under ambient conditions (here observed with mesoporous silica wet gels impregnated with ethylacetoacetate at times t_1 = 5 min and t_2= 20 min, respectively,after the beginning of evaporation.

.

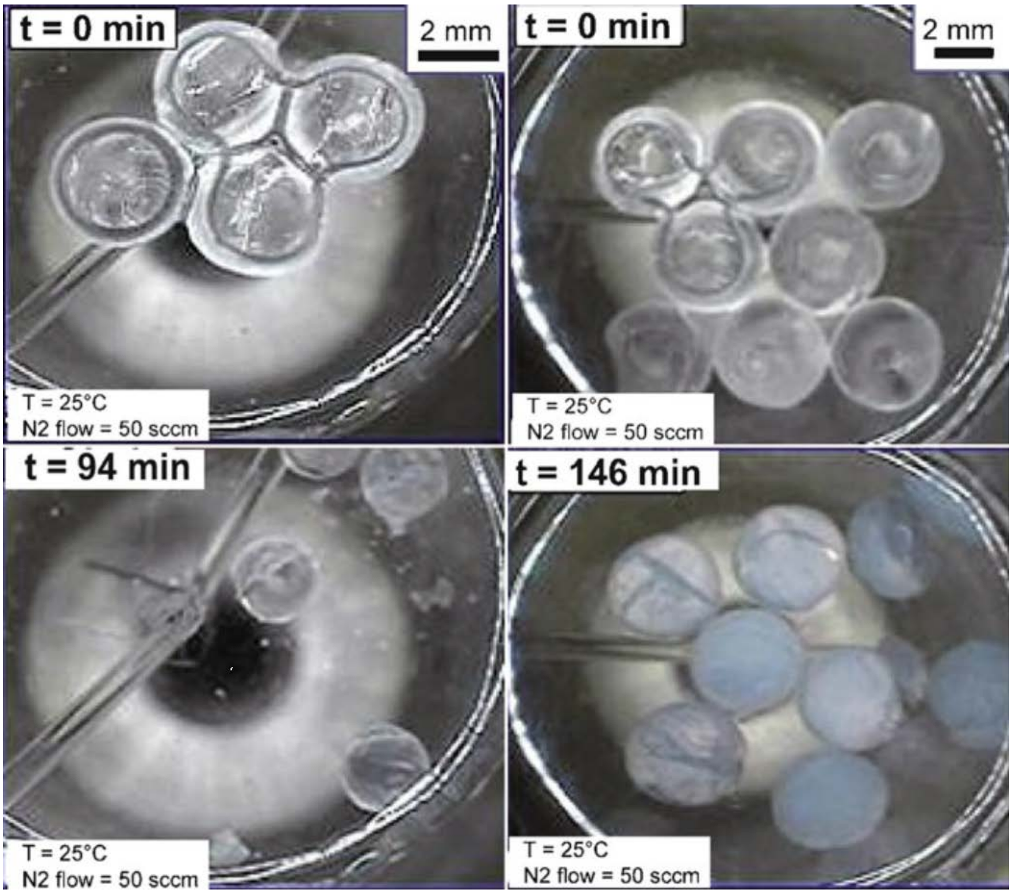


Figure 19.7. Comparison of the shrinkage behavior, from the wet (upper photographs) to the dry (lower photographs) states, occurring during evaporative drying of native, e.g., untreated (left-hand side) and sylilated (righthand side) silica gels.

Supercritical drying, on the other hand, permits to eliminate capillary stresses. Hence, this process produces monolithic silica aerogels of rather large dimensions (Figure 19.8), if required by a targeted application (Figure 19.9).

Figure 19.8. Monolithic silica aerogels obtained after supercritical CO_2 extraction

Figure19.9. Large monolithic silica aerogel monoliths integrated in demonstration glazing (left side) and window (right side) prototypes.

Supercritical drying can be performed (1) in organic solvents in their supercritical state (generally alcohol as the pore liquid and consequently above 260°C if ethanol is used) according to a so-called HOT process or (2) in supercritical CO_2 at a temperature slightly above the critical temperature of CO_2 ~31°C according to a so-called COLD process.

19.2.4. Synthesis Flexibility

Besides these synthesis and processingmethods, it must beemphasized that the flexibility of sol–gel processes permits to enlarge the selection of silica aerogel-based materials, which is currently accessible. Bulk architecture can be tailored by templating techniques. The gel chemistry can be modified by grafting, either during (Figure 19.10) or after gelation. Composites and nanocomposites can be elaborated by impregnation of foams (Figure 19.11) or fibrous networks, by dispersion of particles, powders, or polymers, or by synthesis of mixed silica-based oxides. Organic silica hybrids can also be made either by many techniques such as cogelation and crosslinking or by reaction with functionalized particles.

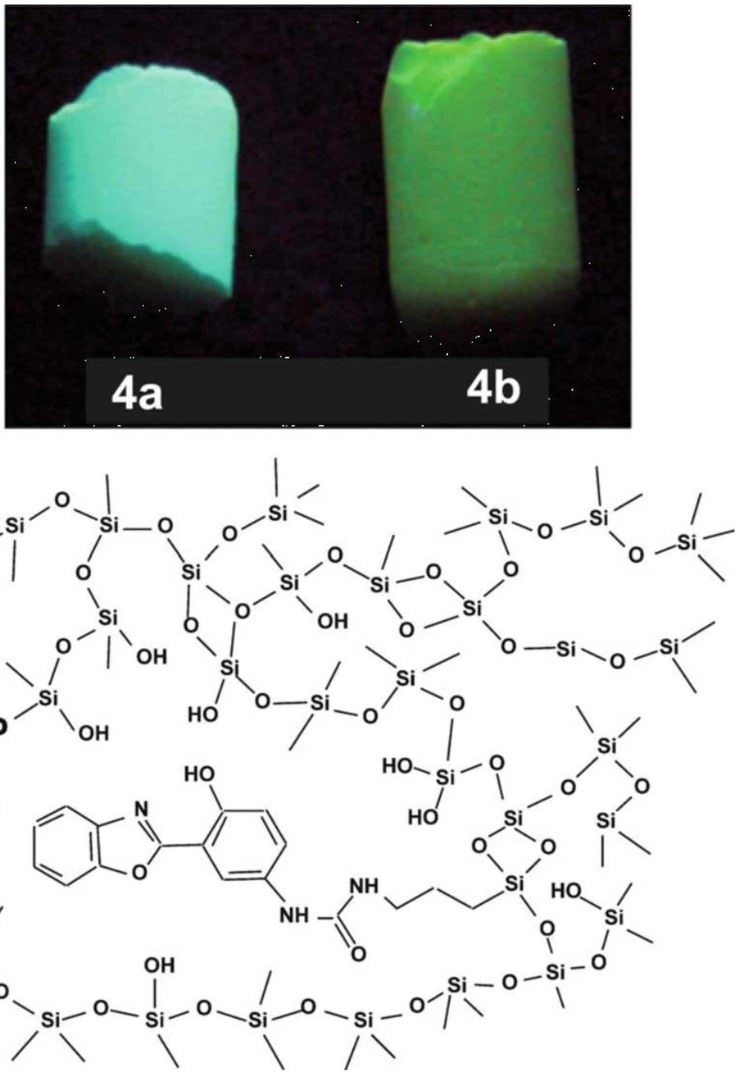


Figure 19.10. Monolithic fluorescent silica aerogels obtained by reaction of silyl-functionalized benzazoles dyes with polyethoxydisiloxane in isopropanol under HF catalysis.

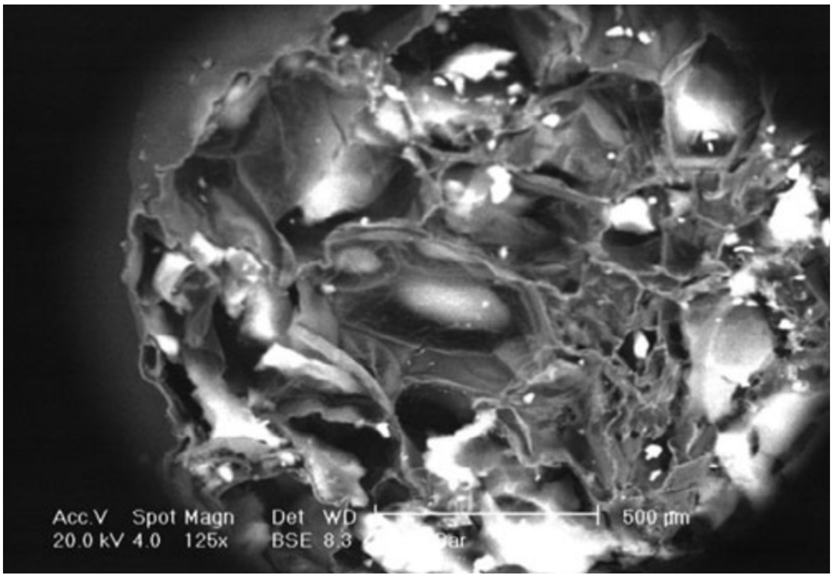


Figure 19.11. Scanning electron microscopy of an organic–inorganic composite obtained by impregnation of a cellular open-cell organic foam with a silica sol.

19.3. Other Aerogels

19.3.1 Hydrophobic Aerogels

Figure 19.12 shows the contact angle measure which allow to distiguish the hydrophilic and hydrophobic surfaces. For hydrophilic surface angle $\theta < 90^o$, for hydrophobic surfaces $\theta > 90^o$.

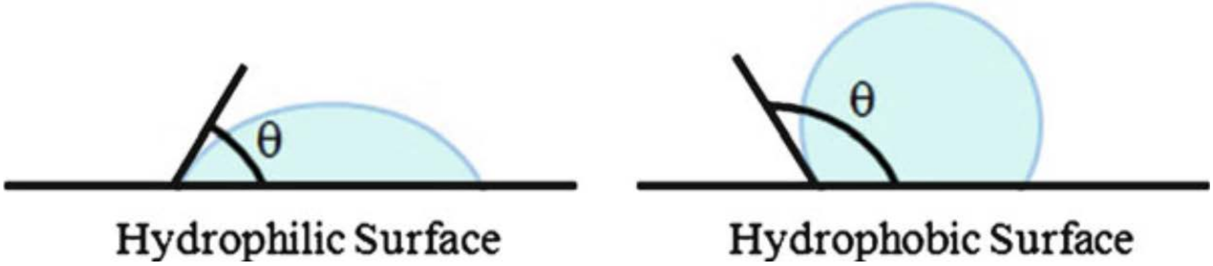


Figure 19.12. Representation of contact angles on hydrophilic and hydrophobic surfaces.

What Makes an Aerogel Hydrophobic?A hydrophobic surface is one that repels water, whereas a hydrophilic surface attracts water. But there is a wide range of surface behavior; most surfaces are neither fully hydrophilic nor absolutely hydrophobic. One of the main determinants of the overall hydrophobicity of a sample is the extent to which the functional groups on the surface of the material interact with water. These interactions include intermolecular forces (dipole–dipole interactions and hydrogen bonds), as well as acid/base chemistry and other types of surface reactions involving water. Water will readily wet a hydrophilic surface (see Figure 19.13A), but will bead up on a hydrophobic surface (Figure 19.13B).

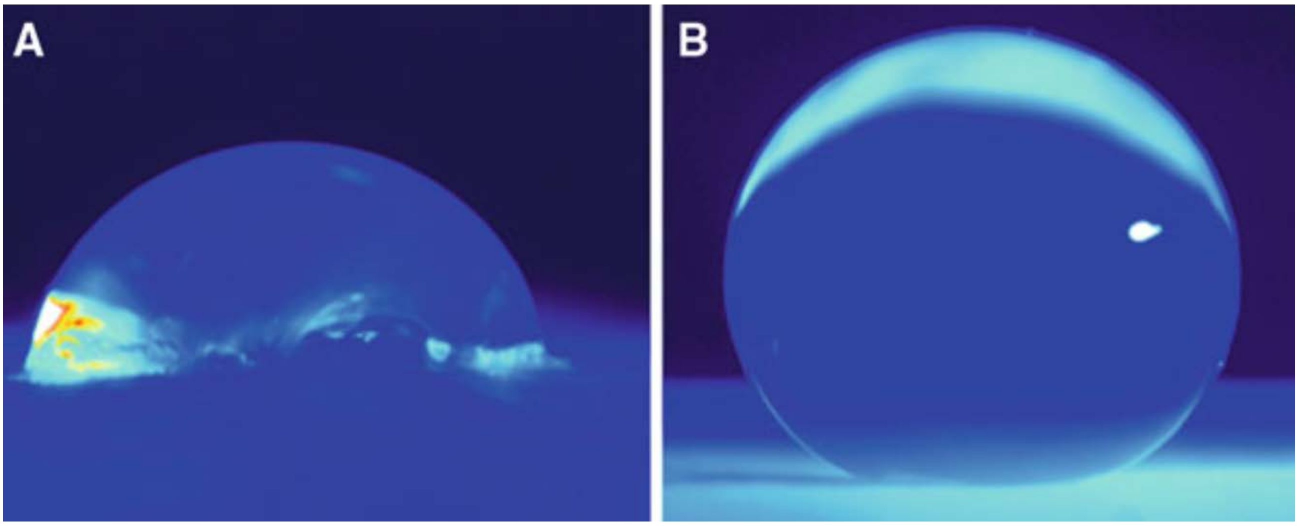


Figure 19.13. Water droplets on aerogel surfaces: A. a TMOS-based aerogel fabricated in an RSCE process and B. An aerogel made from a 25% by volume TMOS and 75% by volume MTMS mixture in an RSCE process. The water droplet on the TMOS aerogel spreads out and wets the surface, indicating a lack of hydrophobicity, whereas the droplet on the aerogel made with an MTMS co-precursor beads up, indicating that the surface is hydrophobic. We note that in some cases a water droplet on an unmodified TMOS aerogel is immediately taken up by the material. For the RSCE aerogel in A. the droplet remained on the surface of the aerogel.

Silica sol–gel matrices form via a series of condensation reactions of individual alkoxide molecules. The majority of the resulting material is silica (SiO2). However, some of the alkoxide side chains remain unreacted (Si–OR) and some are left partially reacted (Si–OH) at the end of the polymerization reaction. If the silica precursors contained groups that could not react via condensation (Si–R), those groups will also be present in the aerogel matrix. When these groups (Si–OR, Si–OH, Si–R) are found at the surface of the aerogel, they can come into contact with water from the environment. And, as aerogels have unusually high surface areas compared to many other materials, there are many possible points of contact with water from the environment.

Silica aerogels prepared from TMOS or TEOS have Si–OR and Si–OH groups. The Si–OH groups render the aerogels hydrophilic. The Si–OH groups can undergo strong hydrogen bonding with water. We have observed that when TMOS-based aerogels are placed in water, they undergo structural collapse, rendering them unsuitable for chemical sensor applications in aqueous samples.

Indeed, some silica aerogels are so hydrophilic as to be hygroscopic. Miner et al. [7] noted that significant uptake of water by silica aerogels in humid environments results in the aerogel materials becoming cloudier with time, and even fragmenting; this complicates the use of aerogel materials for thermal insulation. Moreover, the Si–OH group is a weak acid. Consequently, the average charge of these surface groups depends on the pH of the environment in which the gel is found.

This property of silica is exploited in chromatography and electrophoresis applications.
The presence of a significant number of non-polar side groups (Si–R) on the surface of a silica aerogel renders it hydrophobic. Water molecules undergo only weak intermolecular forces with hydrocarbons and fluorohydrocarbons, but very strong intermolecular forces with other water molecules. This is why water beads up on PTFE surfaces and on hydrophobic aerogels, and why researchers have observed hydrophobic aerogels floating on top of water for months at a time.

19.3.2. Superhydrophobic and Flexible Aerogels

Commonly, aerogels are brittle and fragile materials. In the case of aerogels which are made from precursors carrying internal hydrophobic groups such as MTMS and MTES, one observes a strong deviation from these classical properties of silica aerogels. The behavior of the resulting aerogels is closer to that of soft rubber or polymer-like materials. The synthesis of flexible superhydrophobic silica aerogels can occur in a single or two-step sol–gel process without the need for additional chemical modification of the gels. This new and fascinating class of aerogels offers similar network structure but vastly different mechanical properties when compared with standard silica aerogels. Also, they regain their original shape and size after loading and offer extremely low values of Young’s modulus Y.

Figure 19.14. Flexible MTMS aerogel sample which can bend about 90°

These aerogels also absorbs up to 15 times their own mass of organic solvents or oil and a moderate heating of the soaked material releases the solvent quantitatively. This in combination with their superhydrophobic and flexible nature makes them promising candidates for specialty spill clean-up applications. Little work has been done so far to produce these materials by ambient drying but ambient dried materials of this class would certainly raise their attractiveness for commercialization.

19.3.3 Sodium Silicate Based Aerogels via Ambient Pressure Drying

Waterglass-based silica aerogels are commonly fabricated by a typical sol–gel process exemplified in Figure 19.15.

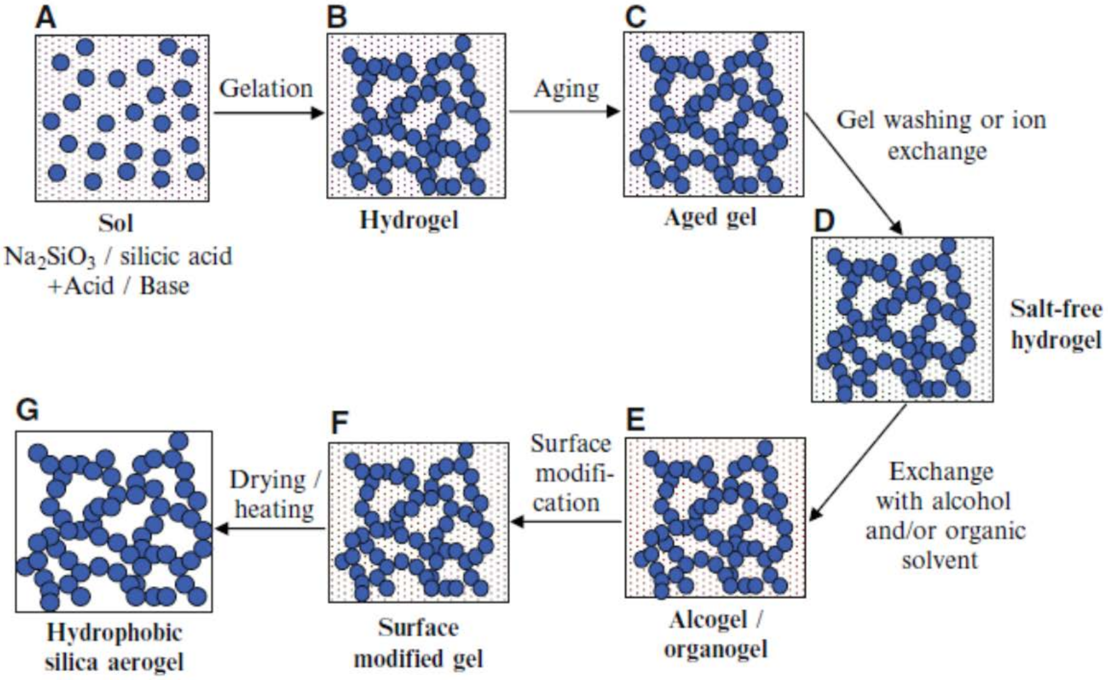


Figure 19.15. Schematic illustration of the synthetic steps used to manufacture silica aerogels from waterglass.

A number of different strategies are known today, which can produce ambient dried silica aerogels from waterglass. Generally, one can distinguish:

1. “Conventional” gelation followed by solvent exchange and hydrophobization and
2. Co-precursor methods.

In the latter, the surface modification agent is added together with the silica precursor (waterglass) during gelation. Looking back at the times of the discovery of aerogels by Kistler, there has been a substantial development in the process efficiency: all recently developed techniques are by far faster and more economic. However, most of the time required for preparation is still spent during washing, solvent exchange, surface modification, and drying steps, i.e., during the stages (c) through (g). The drying step at ambient pressure requires a more or less fixed amount of time and hence offers less potential for speeding up the synthesis. This leaves room for optimization for gel-washing, solvent exchange, and hydrophobization. It is in exactly those steps where most synthetic methods digress. Let us briefly peruse the most recent developments that were able to greatly reduce the overall processing time:

- The use of a combined solvent exchange and hydrophobization (CSH): Often, the hydrogel is immersed into a mixture of an alcohol (isopropyl alcohol (IPA)/alkane (hexane or heptane)/TMCS solution. In thisway, steps (e) and (f) are combined.
- In some instances, aging of the gels is done in a water alcohol or acetone mixture in order to facilitate and/or speed up the subsequent CSH treatment .
- waterglass/HMDS coprecursor method to synthesize superhydrophobic silica aerogel powders in a total preparation time of less than 5 h.

Silica gels are synthesized from molecular silicon-containing precursors. Two common methods are used to initiate the gelation of a waterglass solution:

1. Acidifcation or partial neutralization of a sodium silicate solution by addition of a Bronsted acid.
2. Replacement of sodium ions Na+ by protons H+ by means of an ion exchange resin in its acid form, thus forming a silicic acid solution and initiating gelation by addition of a Lewis base (F-) or Bronsted base (OH-).

Method (1) is a so-called single step process. Adjusting the pH to a value between 5 and 9 is equivalent with a partial neutralization of the sodium silicate. Commonly, the term acid catalysis is used to describe this process. Strictly speaking, this is only partially correct, because the addition of acid serves the primary purpose of partially neutralizing the alkaline sodium silicate solution and to lower the pH value. Method (2) is a classical two-step process. In the language of tetraalkoxysilane (TEOS, TMOS) chemistry, this is often referred to as an acid-base catalyzed synthesis. For the waterglass system, the same language is commonly encountered in the literature to describe the various stages of sol and gel formation. In the following, let us discuss the silica gel formation from waterglass. The two key steps in this process are neutralization and condensation. Figure 19.16 shows the neutralization of silicate to silicic acid H_2SiO_3. In a second step, the formation of dimer species with one equivalent of silicic acid (A) or sodium silicate (B) respectively is shown.

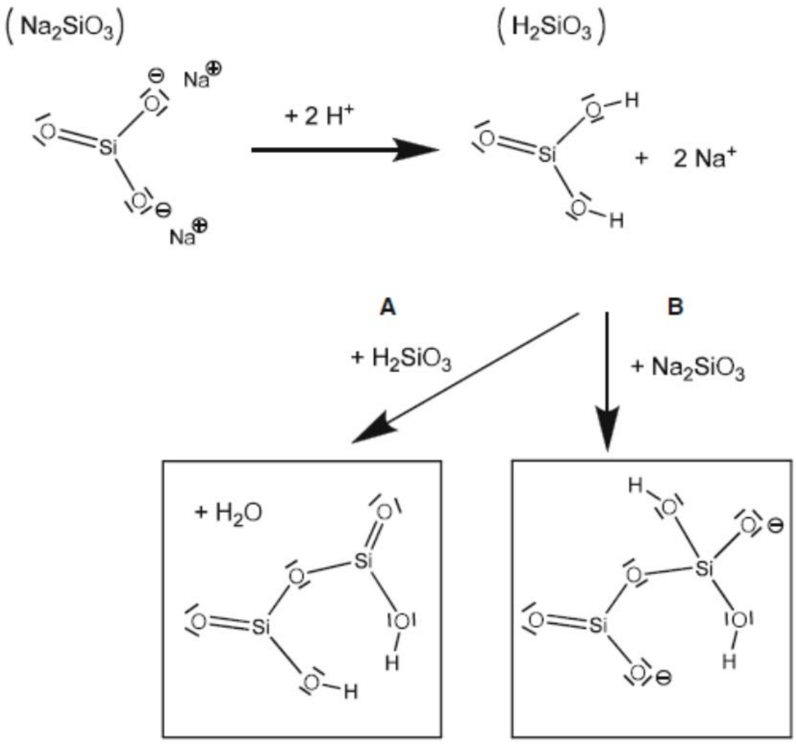
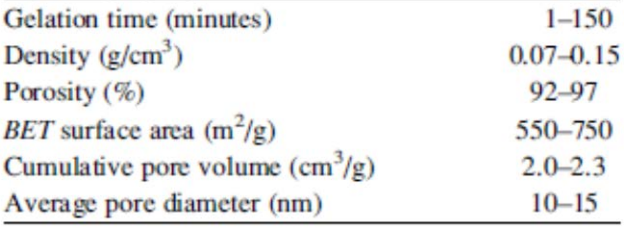


Figure 19.16. Acidification of a sodium silicate molecule to produce silicic acid and reaction with another molecule of A. silicic acid or B. sodium silicate.

Drying of the surface-modified gels from an organic solvent phase is typically done in two steps. The first is a low-temperature drying step at close to ambient temperatures (25ºC–50ºC) typically lasting one to two days. The second step is done at high temperatures between 150ºC and 200ºC. The last step is the most energy intensive of the APD process. A list of typical physical properties of ambient dried waterglass-based silica aerogels is given below in Table 19.2.

Today, the APD process yields aerogels that offer identical properties to those dried supercritically. However, the elaboration of every synthetic method starts with arduous trials and optimization studies until this point is reached. From that point of view, the supercritical drying method is certainly less labor-intensive because there are fewer steps to the final product and the drying is not as sensitive to process parameter changes as in the APD case.

Table 19.2 Collection of typical properties of silica aerogels derived from waterglass by APD



Physical property	
Gelation time (minutes)	1–150
Density (g/cm^3)	0.07–0.15
Porosity (%)	92–97
BET surface area (m^2/g)	550–750
Cumulative pore volume (cm^3/g)	2.0–2.3
Average pore diameter (nm)	10–15

In the following lines, the influence of a number of selected reaction parameters, demonstrating their effect on essential physical properties such as density, thermal conductivity, optical transmission, specific surface area, and hydrophobicity is described. Figure 19.17 below shows a typical transmission electron micrograph of an ambient dried silica aerogel made from waterglass. It clearly reveals the nanoscopic building blocks and the high porosity of the material.

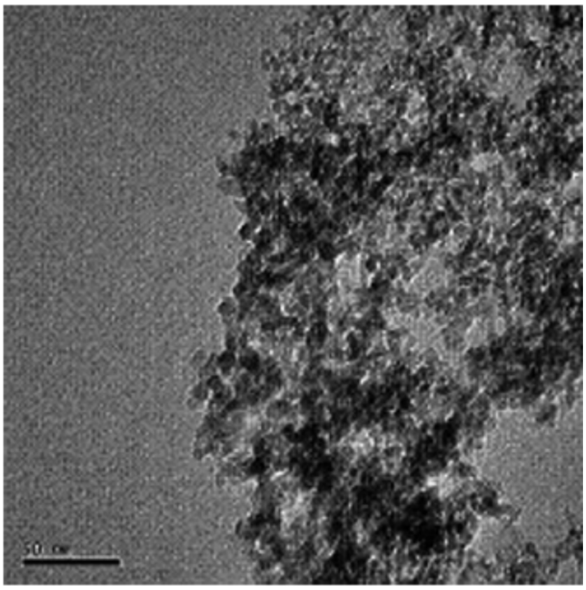
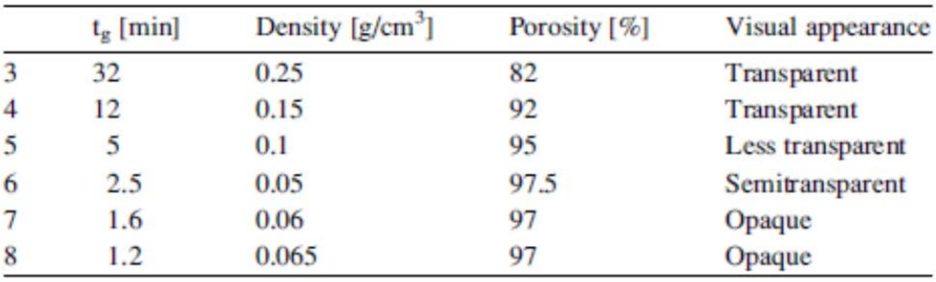


Figure 19.17. TEM Image of a waterglass-based silica aerogel revealing the mesoporous pore structure

Effects of various process parameters on the physicochemical properties of the aerogels presented in Figure 19.18 and table 19.3

Table 19.3 Effect of the gelation pH on the gelation time and aerogel propertie



	t_g [min]	Density [g/cm^3]	Porosity [%]	Visual appearance
3	32	0.25	82	Transparent
4	12	0.15	92	Transparent
5	5	0.1	95	Less transparent
6	2.5	0.05	97.5	Semitransparent
7	1.6	0.06	97	Opaque
8	1.2	0.065	97	Opaque

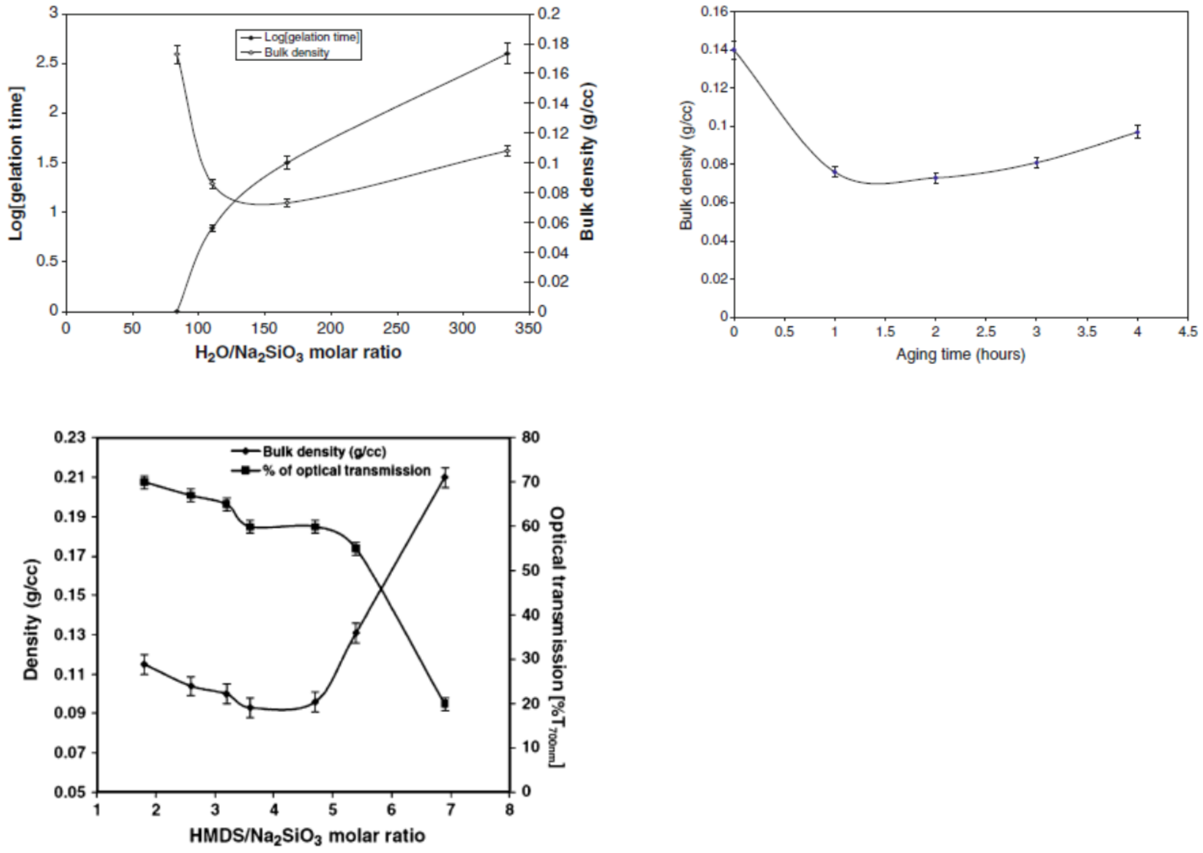


Figure 19.18 Effect of H_2O/Na_2SiO_3 molar ratio (A) on gelation and density of aerogels. (a) Effect of the aging period on the bulk density of the final aerogel (b), Effect of HMDS/Na_2SiO_3 molar ratio on bulk density and % of optical transmission of aerogels.

19.4. Aerogels Containing Metal, Alloy, and Oxide Nanoparticles Embedded into Dielectric Matrices.

Aerogels are regarded as ideal candidates for the design of functional nanocomposites based on supported metal or metal oxide nanoparticles. The large specificsurface area together with the open pore structure enables aerogels to effectively host finely dispersed nanoparticles up to the desired loading and to provide nanoparticle accessibility as required to supply their specific functionalities. The incorporation of nanoparticles as a way to increase the possibility of the use of aerogels as innovative functional materials The different strategies that have been adopted to synthesize nanocomposite aerogels follow two general approaches, depending on whether the nanophase (or its precursor) is added during or after the sol–gel process.

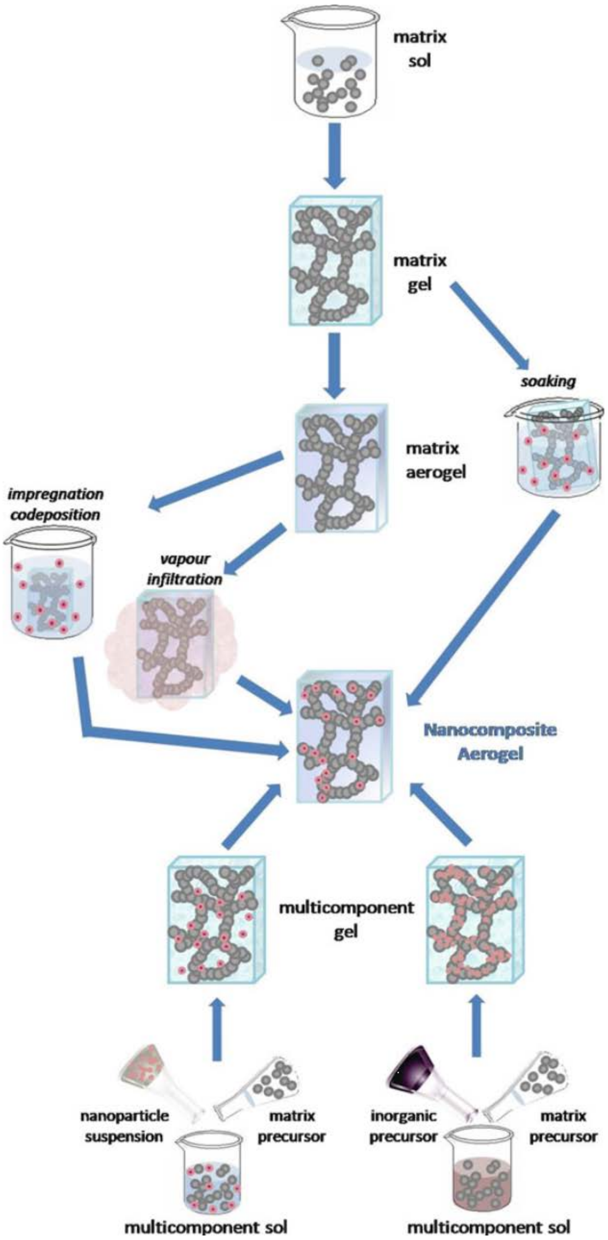


Figure 19.19. Schematic diagram of the different approaches for the chemical synthesis of nanocomposite aerogels.

The first approach includes cohydrolysis and cogelation of the nanoparticle and of the matrix precursors and cogelation of the matrix precursor together with preformed nanoparticles. This approach offers the advantage of producing materials with a controllable loading of nanoparticles throughout. On the other hand, several disadvantages should be taken into account: an accurate choice of the synthetic conditions has to be made in order to obtain a homogeneous multicomponent gel, and the nanoparticle precursors as well as the capping agents needed to stabilize preformed colloidal nanoparticles may affect the sol–gel synthesis of the matrix.

The second approach includes methods based on the addition of the nanophase after the sol–gel process and they should preserve the porous structure and morphology of the matrix. These methods include the deposition of the nanophase by impregnation, deposition–precipitation, and chemical vapor infiltration procedures. Deposition– precipitation synthesis, based on the deposition of metal hydroxides on the support by tuning the pH of the solution, as well as impregnation of a preformed aerogel by a solution of metal salts, are very straightforward and inexpensive routes.

These approaches, however, suffer from two major drawbacks: (1) the poor compositional homogeneity of the resulting nanocomposites and (2) the potential damage of the support in liquid media, aerogels being brittle and fragile. Tethering the metal to a gel matrix modified by coordinating groups and soaking the alcogel or aquogel into the metal solution prior to supercritical drying have been proposed as a way to overcome respectively drawbacks described in (1) and (2).

Deposition of nanoparticles from vapor phase, as opposed to wet impregnation ethods, does not alter the porous matrix and ensures that the guest phase will be distributed throughout the matrix, thanks to its open porous texture. The general applicability of this approach is limited by the availability of precursors with a sufficiently high vapor pressure to give rise to volatile by-products.

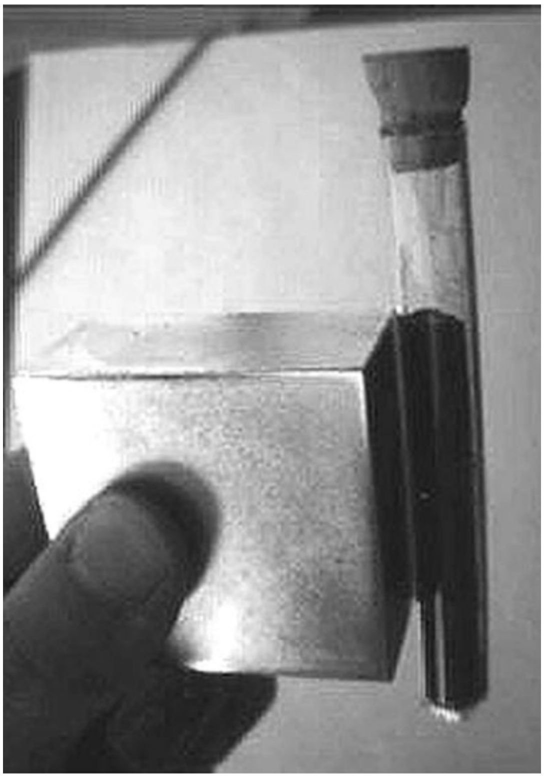
Figure 19.19 is a schematic summary of the different preparation approaches; it should be taken into account, however, that a postsynthesis treatment, most commonly a thermal treatment under controlled atmosphere, is required to induce phase separation, to promote the formation of the desired oxide or metallic nanophase, and to obtain the desired crystallinity.

Several examples are available in the literature on the synthesis of aerogels containing oxide nanoparticles embedded into dielectric matrices to be used for their magnetic, catalytic, and optical properties. While almost all of the examples refer to SiO_2-based nanocomposites, a large variety of oxide nanoparticles have been synthesized as a dispersed nanophase.

Different approaches have been tried for the synthesis of aerogels containing magnetic iron oxide nanoparticles dispersed into SiO_2 with particular emphasis on maghemite, g-Fe_2O_3, nanoparticles, which possess interesting magnetic properties. However, even if this metastable phase is frequently kinetically stable at the nanometer scale, it proved to be difficult to obtain. It was prepared several magnetic samples (see Figure 19.20) by varying different sol–gel parameters. In particular, they used different precursors for the silica (TEOS or TMOS) in alcoholic solution (ethanol or methanol) in the presence of different amounts of water and using acid or base catalyst. They also used two different iron oxide precursors: either $Fe(NO_3)_3.9H_2O$ or a (FeNa(EDTA). $2H_2O$) metallic complex in order to avoid strong bonding to the matrix and to increase pore diameters. The samples obtained with these procedures have surface areas ranging between 200 and 600 m^2 g^{-1}, while XRD does not permit a clear identification of peaks apart from one single case in which the presence of six-line ferrihydrite, a poorly crystalline iron oxide hydroxide, became apparent.

Mossbauer spectroscopy allowed identification of the iron oxide phase as ferrihydrite in some samples and magnetite in others and also indicated that the samples present superparamagnetic relaxation effects. TEM shows either spherical or acicular nanoparticles.

Fe_2O_3–SiO2 nanocomposite aerogels were prepared using TEOS and $Fe(NO_3)_3 . 9H_2O$ as precursors for the silica matrix and the iron oxide nanoparticles, respectively, while ethanol was used as solvent and nitric acid as catalyst. The obtained alcogels were submitted to sc-HT drying adopting two different conditions that lead to mesoporous and microporous aerogels, respectively. After supercritical drying, samples were calcined at elevated temperatures. XRD of the aerogels after supercritical drying showed faint peaks due to six-line ferrihydrite. After calcination, new peaks showed up and kept on growing. At 900°C, there was a mixture of different Fe(III) oxides present. Surface areas were quite high even after calcination at 900°C (500 m^2 g^{-1}) and TEM images showed a homogeneous distribution of the nanoparticles. TEM micrographs show nanoparticles with a narrow distribution of size (around 4 nm) located in large and irregular aggregates within the amorphous matrix, as shown in Figure 19.21.



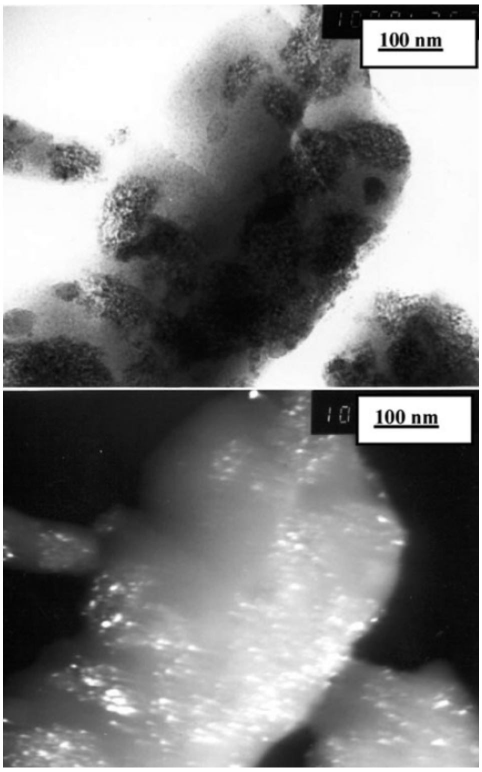


Figure 19.20 Magnetic aerogel attracted by a permanent magnet

Figure 19.21. BF (top) and DF (bottom) TEM micrographs of maghemite–SiO_2 aerogel

The preparation of nanocomposite aerogels made out of zero-valent metal or alloy nanoparticles in an insulating aerogel matrix has attracted a great deal of effort in view of the preparation of materials for catalysis, biomedicine, optoelectronics, and sensors. As it is the case for all sol–gel nanocomposites, silica is by far the most investigated matrix for supporting metal/alloy nanoparticles, although alumina, titania, chromia, and polymer supports have also been proposed.
On the other hand, a variety of metal nanoparticles have been considered either on the basis of the ease of reducibility, with the aim of developing optimized synthetic protocols, or in order to provide the desired functionality to the nanocomposite. Noble metal nanoparticles are the most investigated as they meet both requisites: they can be easily obtained and stored in the zero-valent state; they are catalytically active toward a wide range of chemical processes; and they can have biomedical applications such as antimicrobial activity. Moreover, metal nanoparticles exhibit characteristic surface plasmon bands and when hosted in dielectric matrices, due to the refractive index mismatch, can give rise to enhanced optical effects, namely, surface enhanced Raman scattering.

Transition metals and their alloys have also attracted much attention due to their magnetic properties and due to their potential in catalysis thanks to the occurrence of different oxidation states. The properties of sc-CO_2 in the development of novel processes for the synthesis of nanostructured materials were also exploited in the case of Pt–SiO_2 aerogels obtained by impregnation of preformed commercial silica aerogels. By making use of sc-CO_2 as the solvent for the deposition of the metal precursor (dimethyl cyclooctadiene platinum-II), Pt–SiO_2 nanocomposites with a narrower size distribution were obtained. However, metal nanoparticles seem to be mainly deposited on the outer surface of the silica aerogel (see Figure 19.22). By applying the same deposition procedure on different substrates (carbon black, alumina, silica, and silica aerogel), the authors were able to show that particle formation and growth was mainly governed by the strength of the interaction between the organometallic precursors and the support.

Extensive structural investigation by EXAFS, HREM, EDX, EELS has been used to provide compositional information at the nanometer scale for FeCo–SiO2 nanocomposites (Fig.19.23)

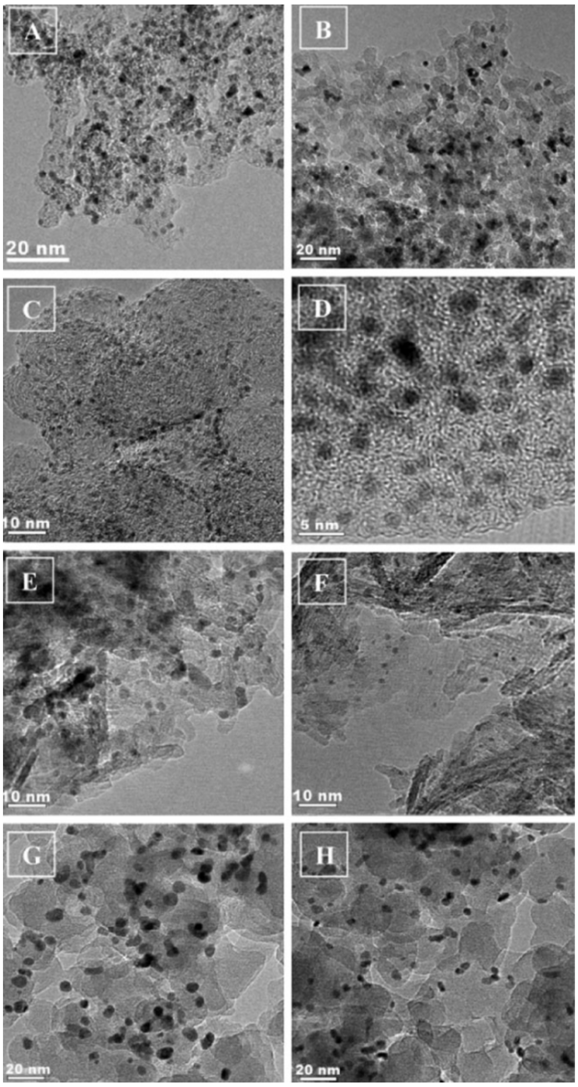


Figure 19.22. HRTEM micrographs of supported Ru and Pt nanoparticles on different substrates: A. Carbon aerogel, B. silica aerogel, C. carbon black, D. Nafion, E, F. γ-Al_2O_3, G, H. SiO_2

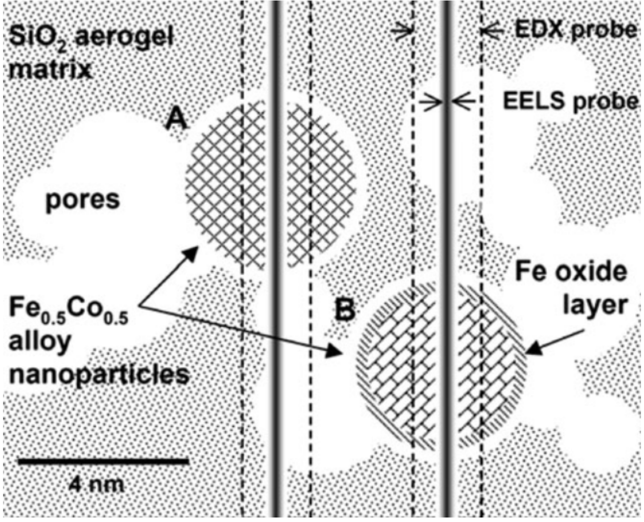


Figure 19.23. Possible scenarios for the microstructure of a 4 nm diameter FeCo alloy nanoparticle: (A) nanoparticle consists of a single homogenous alloy phase; (B) nanoparticle is covered with a 0.2 nm layer of iron oxide determining a slightly reduced concentration of Fe in the core of the alloy

19.5. Main Properties

The key structural parameters of a porous material are the total fraction of the pore and the solid phase, the typical extension of the backbone and the pore phase, the connectivity of the two phases, the characteristics of the interface between the phases, and the molecular structure of the backbone phase (see Figure 19.24). Related to these properties are different characterization techniques (Table 19.4).

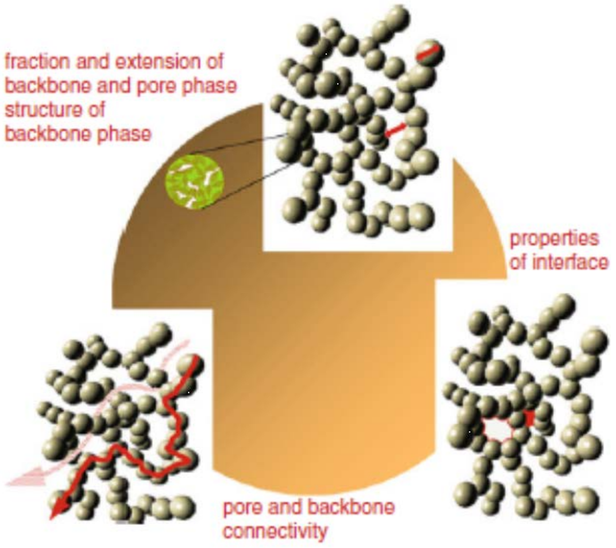


Figure 19.24. Overview over the key structural parameters of a porous solid

19.5.1. Texture

Silica aerogels are amorphous materials. They have a skeletal density, as measured by Helium pycnometry ~2 g cm^{-3}, close to that of amorphous silica (2.2 g cm^{-3}). They typically have a pore volume above 90% of their whole monolith volume. Some ultraporousand ultralight silica aerogels can be synthesized by two-step process, and a density as low as 0.003 g cm^{-3} has been reported : these are the lightest silica aerogels that can be found in the literature.

Silica aerogels are usually largely mesoporous, with interconnected pore sizes typically ranging from 5 to 100 nm and an average pore diameter between 20 and 40 nm. Micropores (i.e., pore sizes < 2 nm) become significant in aerogels synthesized under acid catalysis conditions or having undergone particular treatments . The associated specific surface area is rather high, typically from 250 to 800 $m^2\,g^{-1}$ and can exceed 1,000 $m^2\,g^{-1}$.

The architecture of silica gel networks is often described as that of a fractal geometry and it is possible to distinguish mass from surface fractalities. In the former case, the mass M of a gel inside a sphere of radius R, centered about a random point in the gel network, is a statistical function of R of the type (19.5).

Tabl.19.4 Overview of parameters and related experimental techniques

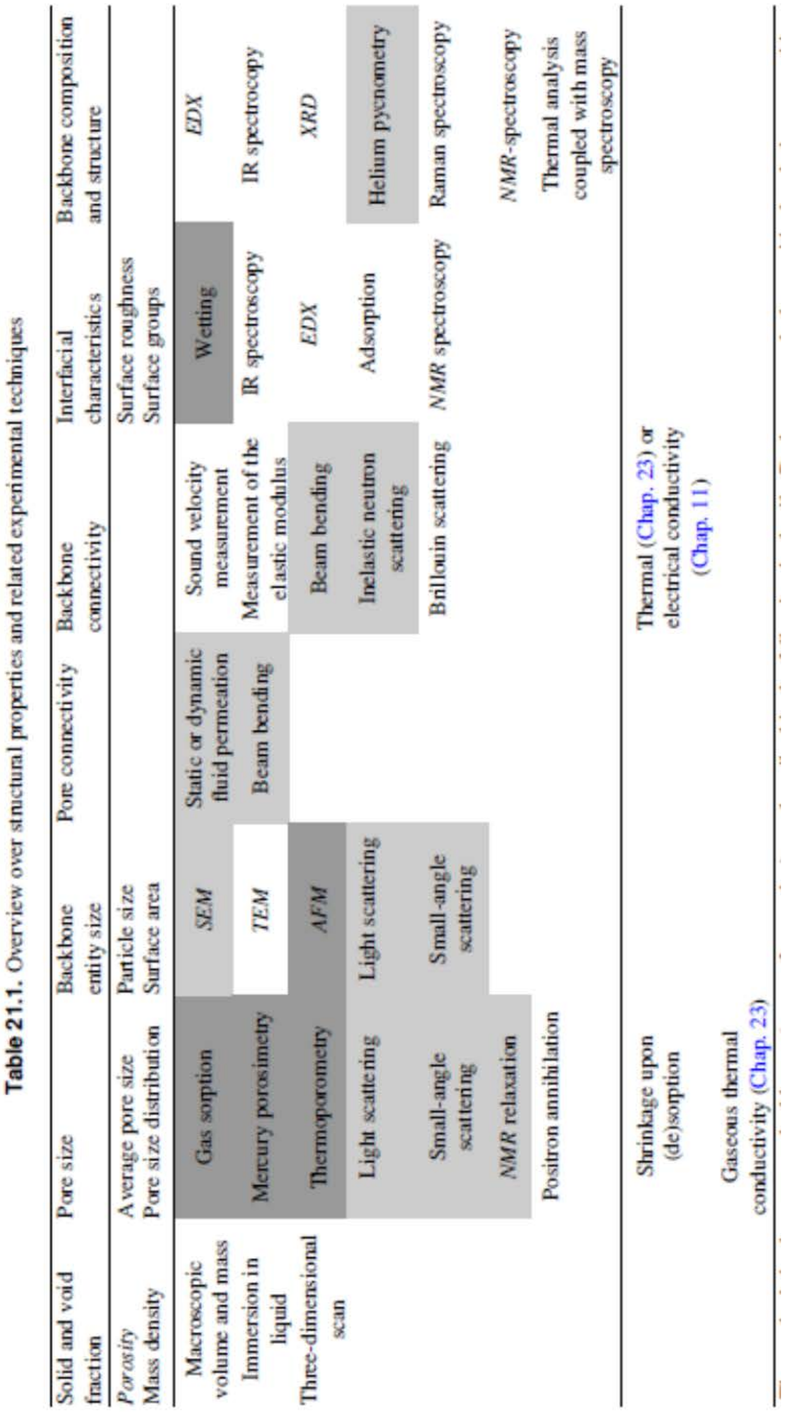


Table 21.1. Overview over structural properties and related experimental techniques

Solid and void fraction	Pore size	Backbone entity size	Pore connectivity	Backbone connectivity	Interfacial characteristics	Backbone composition and structure
Porosity Mass density	Average pore size Pore size distribution	Particle size Surface area			Surface roughness Surface groups	
Macroscopic volume and mass	Gas sorption	*SEM*	Static or dynamic fluid permeation	Sound velocity measurement	Wetting	*EDX*
Immersion in liquid	Mercury porosimetry	*TEM*	Beam bending	Measurement of the elastic modulus	IR spectroscopy	IR spectrocopy
Three-dimensional scan	Thermoporometry	*AFM*		Beam bending	*EDX*	*XRD*
	Light scattering	Light scattering		Inelastic neutron scattering	Adsorption	Helium pycnometry
	Small-angle scattering	Small-angle scattering		Brillouin scattering	*NMR* spectroscopy	Raman spectroscopy
	NMR relaxation					*NMR*-spectroscopy
	Positron annihilation					Thermal analysis coupled with mass spectroscopy
	Shrinkage upon (de)sorption			Thermal (Chap. 23) or electrical conductivity (Chap. 11)		
	Gaseous thermal conductivity (Chap. 23)					

$$M \approx R^f ,$$ (19.5)

where f is termed the fractal dimension because this is not an integer.

For a surface fractal object, the surface area A follows the law (19.6):

$$A \approx R^{fS} \qquad (19.6)$$

The fractal dimension can be experimentally determined by adsorption of molecules of different cross-sectional area, by small or ultra-small-angle X-ray scattering (SAXS/USAXS) or by small-angle neutron scattering (SANS) spectroscopy. SAXS/USAXS and SANS are also widely used to characterize elementary particles and/or distributions of clusters that constitute the silica skeleton as well as other structural features such as the specific surface area. These results generally give larger values than their BET analogues The porous network characteristics (specific pore volume, mean pore size, pore size distribution, etc.) reported in the literature are usually obtained either by adsorption of nitrogen or by non-intrusive mercury porosimetry. An example of pore size distribution obtained by non-intrusive mercury porosimetry is presented in Figure 19.25.

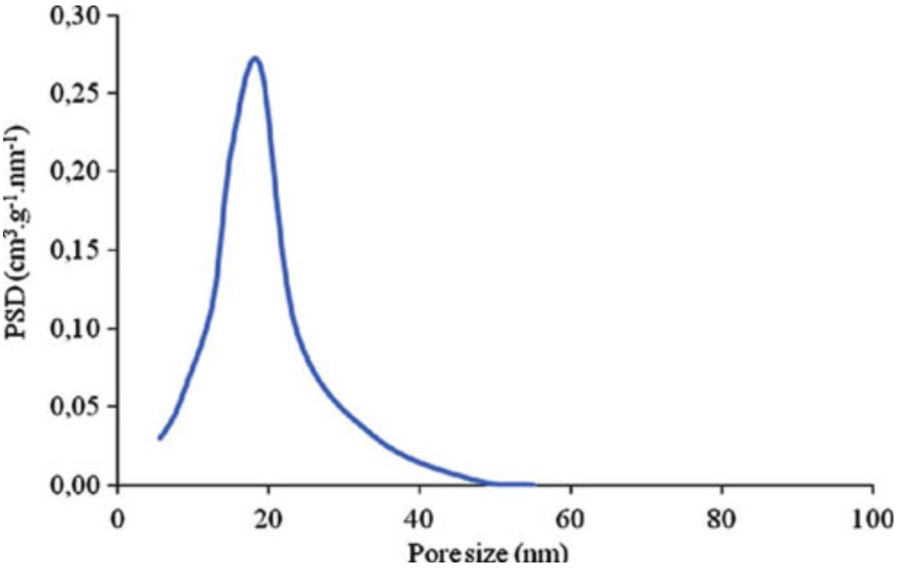


Figure 19.25. Pore size distribution of a 0.15 g cm^{-3} silica aerogel obtained by non-intrusive mercury porosimetry on samples synthesized in alcohol via a two-step process with TetraEthOxySilane (TEOS).

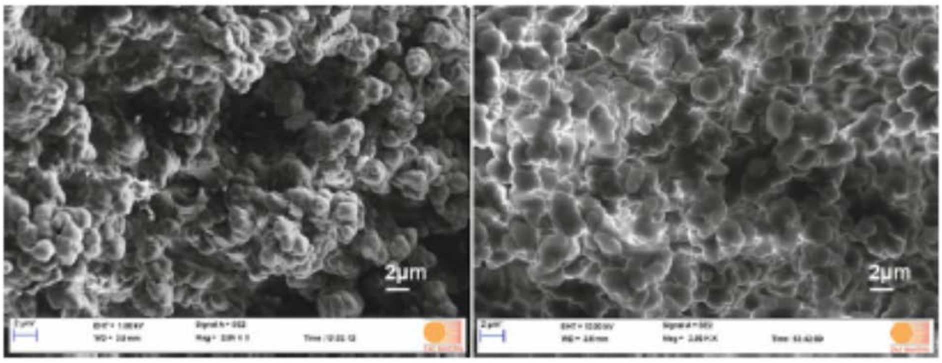


Figure 19.26. SEM image of the very same spot of an organic aerogel taken at low (1 kV) and high acceleration voltage (10 kV).

Figure 19.26 shows an example of the very same spot of an organic aerogel probed with different acceleration voltages at otherwise identical conditions; the images reveal the switching from a surface to a bulk sensitive imaging mode of SEM.

19.5.2. Chemical Characteristics of silica aerogels

After drying, untreated (e.g., not sylilated) contain residual reactive groups (e.g., alkoxy and hydroxyl groups). Fourier-transformed infrared spectra (FTIR) reveals a silanol surface coverage (-Si–OH) of typically 4–6 groups per nm^2 , which gives them a certain hydrophilic behavior. This parameter is influenced by the specific sol–gel conditions used as well as by the drying route. Concerning drying, silanol groups condense during evaporation while HOT supercritical processing induce reesterification of the silanol functionalities. This makes both xerogels and aerogels more hydrophobic after drying is complete. Indeed, for many applications, it is necessary that aerogels be hydrophobic to not absorb (even adsorb) water vapor. Briefly, to elaborate hydrophobic silica aerogels, the pore surface must be covered with nonpolar side functions such as $-Si–CH_3$, which can be introduced by several methods. For instance, a silicon precursor containing at least one nonpolar chemical group, of the type RSi(OR')3 where R and R' are alkyl groups, can be cogelled in various proportions with TMOS or TEOS. For example, when some hybrid gels are synthesized with TMOS and MTMS, the silanol concentration can be decreased. MTMS, for instance, is more difficult to condense than TMOS. Hence, the gel network is then mostly made by TMOS, while in the second stage $-Si–CH_3$ end groups cover the pore surface.

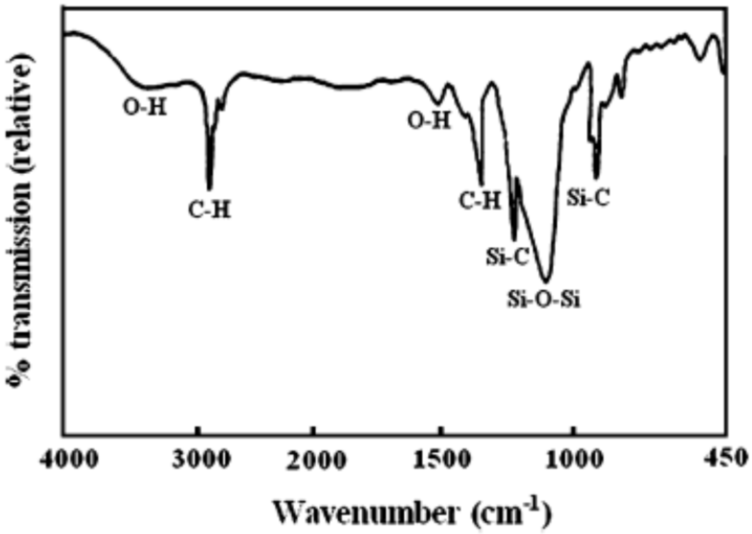


Figure 19.27. FTIR spectrum of an MTMS-based aerogel sample.

Infrared spectroscopy (FTIR) was used to identify the chemical bonding states within the material. The FTIR spectrum shown in Figure 19.27 indicates the presence of all chemical bonds which are expected: The Si–O–Si linkages of the silica network are represented by a strong and broad Si–O vibrational mode at 1100 cm^{-1}. The presence of Si–CH3 functionalities is confirmed by Si–C modes at 840 cm^{-1} and 1310 cm^{-1} as well as C–H stretching and bending vibrations at 2900 cm^{-1} and 1400 cm^{-1} corresponding to Si–C and C–H bonds, respectively. Very small O–H vibrations at 3500 cm^{-1} and 1650 cm^{-1} indicate the presence of residual free OH groups (or adsorbed water) within the aerogel. These aerogels offer unprecedented hydrophobic as well as mechanical character unlike native silica aerogels.

19.6. Physical Properties and Some Related Applications

19.6.1 Thermal Conductivity

One of the major characteristics of silica aerogels is their very low thermal conductivity, typically of the order of 0.015W m-1 K-1 at ambient temperature, pressure, and relative humidity. These values are significantly lower than the conductivity of air under the same conditions, e.g., 0.025W m^{-1} K^{-1} (see typical experiments whuch shows the low thermal conductivity of aerogels, Fig.19.28). Thus, silica aerogels are among the best-known thermal insulating materials . Besides, silica is nonflammable and silica aerogels are amorphous. Moreover, silica aerogels can be made optically transparent, although they are also very brittle (Fig.19.29). Consequently, they present an amazing application potential for opaque or transparent insulating components as well as daylighting devices, if they can be mechanically reinforced.

The physics governing their thermal properties, as well as their use in thermal insulation and the first related commercial products . As an example for a high-tech product, it shall be underlined that silica aerogels find ever increasing use as thermal insulator in the aeronautical and aerospace domains . They were, for instance, used in the recent PATHFINDER MARS mission to insulate the Sojourner Mars Rover. During the mission, the nocturnal temperature dropped down to -67^{o}C, while a stable inside temperature of 21^{o}C was maintained. This permitted to protect the Rover's very sensitive electronics from damage by the cold. For a similar program termed European Retrieval Carrier (EURECA) satellite, the use of aerogels has been investigated .

Figure 19.28. A crayons and flower is on a piece of aerogel which is suspended over a flame from a Bunsen nurner. Aerogel has excellent insulating properties, and the flower is protected from the flame

Aerogels are known for their outstanding thermal insulation properties combined with light weight, chemical resistivity, and nonflammability. The compressive strength, tensile strength, and elastic modulus of silica aerogels are very low and largely depend on the network connectivity and aerogel density. Indeed, silica aerogels can easily be elastically compressed when the porous texture is studied, for example, by the capillary stresses. Nevertheless it is ideal candidates for both ambient and high temperature insulation applications (Fig.19.29). In general, heat is transferred by radiation, conduction, and convection. Due to the small pore sizes of aerogels, the convective and gas conduction pathways are virtually eliminated. What remains are primarily radiative heat transfer and conduction through the silica network structure. Silica aerogels are the best-known aerogel insulation systems and so far also the only ones that are commercially exploited. Products by Aspen Aerogels and Cabot Corporation offer thermal conductivity values as low as 0.013 W/(mK).

Aerogel insulation products are successful in niche markets; however, with increasing popularity and production volumes, there will be room for growth of respective market shares, and new products will become affordable for applications that are not yet covered today. In this respect, APD waterglass-based aerogels have a tremendous potential for upscaling and cost reduction for the two main reasons widely discussed in this text: the process and cost advantage of ambient pressure drying and the low cost of waterglass as a precursor for the gel synthesis. Therefore, it seems reasonable to assume that there will be an ongoing interest in waterglass based aerogels in the near future, both academically and from a manufacturing point of view.

Of course, it must be underlined here that applications in space are not at all limited to thermal insulation. Indeed, silica aerogels can be applied to collect aerosol particles , to protect space mirrors, or to design tank baffles. The most recent project, Stardust, successfully returned to earth in January 2006 (Fig.19.30). This mission provided samples of a recently deflected comet named Wild-2, which are being examined in various laboratories all over the world.

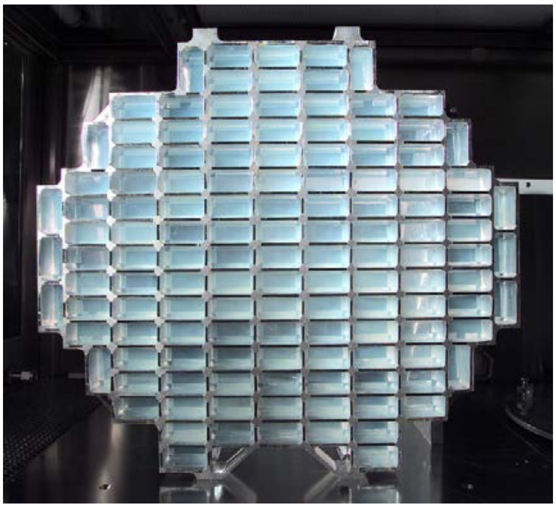


Fig.19.29. A 2.5 kg brick is supported by a piece of aerogel with a mass of only 2 g.

Fig.19.30 The “Stardust” dust collector with aerogel blocks. (NASA)

19.6.2 Optical and Dielectric Properties

The optical transmission and scattering properties of silica aerogels constitute another group of important characteristics, sometimes in conjunction with their thermal properties when a transparent thermal insulation is targeted, such as in windows (Figures 19.31, 19.32). The transparency and visible light transmittance of silica aerogels can be high, although they all tend to scatter the transmitted light to some extent, which reduces their optical quality. The Rayleigh scattering due to the solid gel network heterogeneities in the nanometer range is responsible for a yellowish coloration of silica aerogel observed in transmission and a bluish coloration when observed in reflection mode against a dark background. The scattering due to heterogeneities in the micrometer range is responsible for a blurred deformation of optical images .

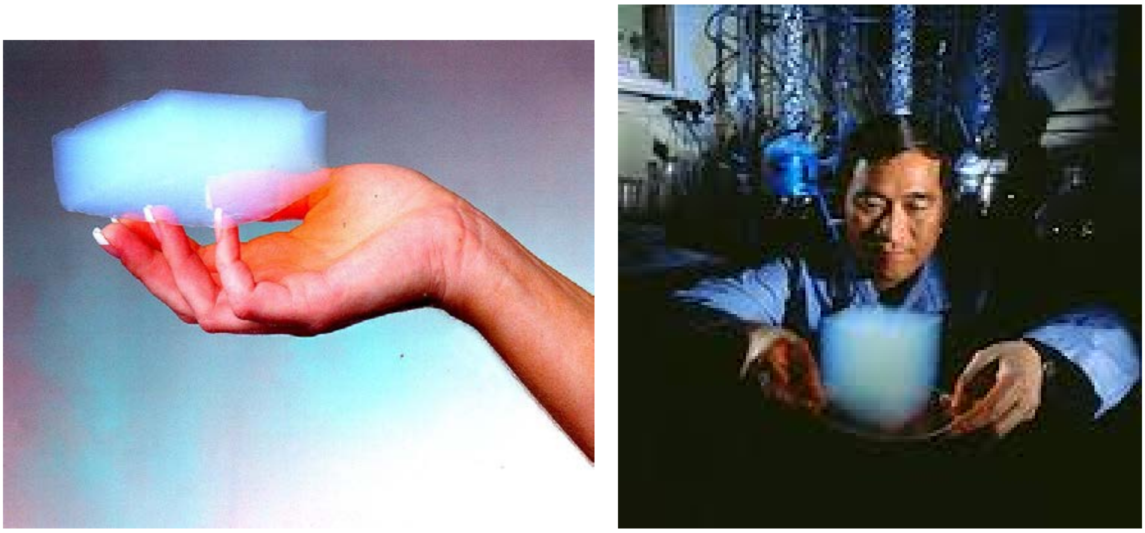


Figure 19.31 A block of aerogel in a person's hand (a), a sample of aerogel at Jet Propulsion Laboratory, California Institute of Technology

Figure 19.32 Examples of two aerogel insulation demonstration projects in Switzerland: Transparent insulation (right, high-school gymnastic hall in Z€urich) and facade blanket insulation (left, single family home in Meilen, ZH).

A twostep gelation catalysis procedure, previously described, was found to give satisfactory transparency results. Typically, aerogels made from TMOS in methanol can be obtained with an optical transmittance ratio up to 93% (for ~1 cm thick aerogels) at a wavelength of 900 nm . Silicon precursors prehydrolyzed with an acid catalyst and a substoichiometric water molar ratio w = n_{H2O}/n_{Si}

<2 are available on the market and make it possible to control a more uniform porous texture, as is needed for optical applications. Transparency ratio up to 90% together with specific extinction coefficient of the order of 15m^{-1} can be obtained with 1 cm thick aerogels synthesized with such prepolymerized precursors (Figure 19.33).

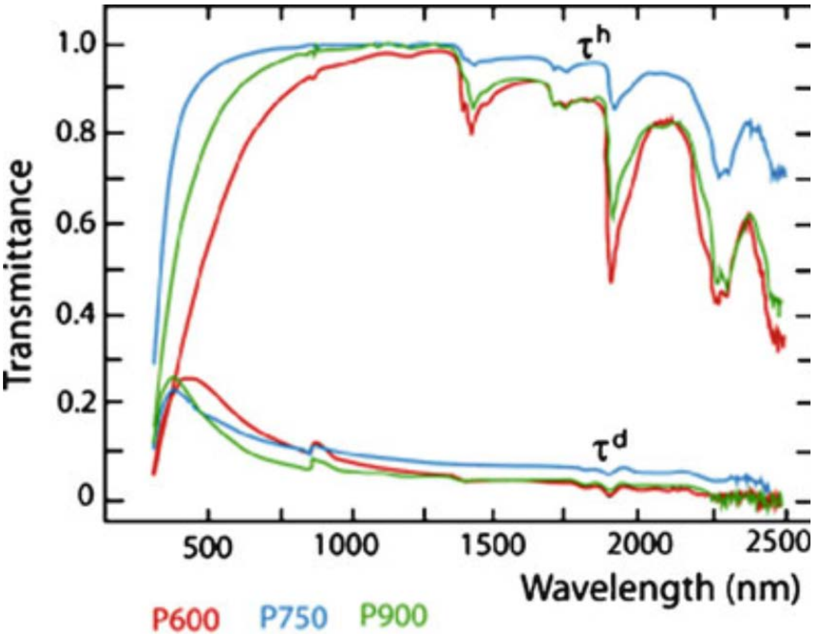


Figure 19.33. Hemispherial (th) and diffuse (td) transmittances of 1 cm thick plates of silica aerogels synthesized by a two-step process, from prepolymerized silica precursors

The refractive index of silica aerogels increases with their density r (kg m-3) according to (19.7)

$$n - 1 = 2.1 \times 10^{-4}\rho. \quad (2.7)$$

Hence, n is very close to 1, which makes aerogels excellent materials to apply in the radiator of Cerenkov counters . In a Cerenkov counter, the radiator is a low-density medium such as an aerogel slab, in which electrically charged elementary particles travel with a velocity “v” higher than the velocity of light “c” and from where they radiate (emit) photons. An analysis of these photons can therefore be used to derive the velocity of the particles and hence their nature. This was one of the well-known historical uses of silica aerogels.

The relative dielectric constant of silica aerogels can be as low as 1.1 . Hence, thin film silica aerogels could be and are being considered as super-low dielectric constant material for integrated circuits in computers. It is also possible to modify the surface of silica aerogel to obtain good electret materials (i.e., materials that produce a permanent external electric field).

19.6.3 Acoustic Properties

The acoustic properties of silica aerogels are closely related to their thermal insulation properties. The acoustic propagation in aerogels depends on the interstitial gas nature and pressure, the aerogel density, and more generally the texture. Silica aerogels are indeed excellent acoustic insulators. The propagation of an acoustic wave is attenuated both in amplitude and velocity because the wave energy is progressively transferred from the gas to the aerogel solid network, over the entire aerogel workpiece thickness. The longitudinal acoustic velocity is typically of the order of 100 m s^{-1} , which makes silica aerogels suitable for applications in acoustic devices.

19.6.4 Entrapment, Release, Sorption, and Storage. Environmental Clean-up and Protection

The combination of a high specific pore volume, with – in some specific cases – a relatively resistant solid SiO_2 network, can also be advantageously used to entrap a large variety of molecules or nanoparticles. The entrapped species can be luminescent molecules or particles . Hollow silica aerogel droplets were also proposed for the inertial confinement of fluids (ICF), namely, mixtures

of liquid deuterium and tritium, as the target in fusion experiments under a powerful laser. The droplets were generated by a double nozzle reactor; they were stable to radiation and could be wetted by deuterium and tritium. In aerogels, the gel network acts as an impurity for the fluids that strongly interact with the solid surface. Hence, large fluid NMR signals with long polarization lifetimes can be recorded, revealing a very weak specific wall relaxation . This effect can be used to study either the gel itself or the trapped fluid. The gel network was, for instance, studied by impregnation with liquid 131Xe near its critical temperature (Tc of approximately 289 K). Magnetic resonance images were obtained, which made it possible to differentiate between aerogels of different densities and hydration levels . As far as studies on fluid flow are concerned, the interaction between superfluid

3He and the silica aerogel network has received significant attention . The solid network introduced disorders in these interactions, thus allowing us to suppress the superfluid transition temperature.

Similarly, when an aerogel is impregnated with a nematic liquid crystal, the gel network randomness destroys the crystal long-range order and transforms the liquid crystal into a glassy state. Biomaterials can also be successfully immobilized inside aerogels. Inversely, aerogels can be used to adsorb or extract some chemical compounds, for instance, to treat waste water, to confine radioactive waste , or to filter gases.

Investigators have studied the use of aerogels in environmental clean-up or protection applications. These applications take advantage of the porous, high surface area nature of the aerogel material that can be used to absorb contaminants. Figure 19.34 illustrates the oleophilic and hydrophobic nature of TMOS/MTMS aerogels.

Figure 19.34. Demonstration of the oleophilic and hydrophobic nature of TMOS/MTMS aerogels. The aerogel monolith was placed in a beaker containing oil and water. Note that the aerogel has adsorbed the yellow oil, and is floating at the top of the water. The top part of the silica aerogel retains its optical properties, indicating that this aerogel had capacity to absorb additional oil. No oil layer is visible on the water surface.

Superhydrophobic MTMS- or MTES-derived aerogels are certainly a class of remarkable materials, but also, they are not ready for large-scale fabrication. As far as potential applications are concerned they can only compete in certain well-defined niche markets. Among the most likely applications are the removals of oil spills from water surfaces, where the oil exists in the form of a thin film, or the clean-up of organic compounds which need to be retrieved no matter the cost. Some examples could involve spills of extremely precious compounds in the chemical industry, such as active drug ingredients and fine chemicals, for example, pure enantiomer forms of selected organic compounds. Perhaps, even the removal of extremely toxic or hazardous metalloorganic compounds (e.g., alkyl derivates of heavy metals) can be absorbed by superhydrophobic aerogels and in some cases must be retrieved quantitatively. Hence, an application of this type of aerogel compound only makes sense, where its performance exceeds state of the art by a considerable amount and if the added cost of the material is compensated at least in part by its superior performance. Some of the main advantages of these materials are their tremendous uptake capacity on a per weight basis and their low density (weight). However, conventional silica-based superhydrophobic aerogels offer similar density and uptake capacities. Here, the flexible nature of alkyltrialkoxysilane based aerogels offers an additional advantage but it also comes with a higher price tag, at least when dried supercritically. Because of their low density and Young’s modulus, these materials also are promising dampers for ultrasensitive measurement equipment such as scanning microscopes (AFM, STM) or for microoptics and MEMS applications.

19.6.5. Anisotropic Aerogels by Photolithography

A general method is presented that allows fabrication of sol–gel materials with anisotropic physical properties. The gelation solvent is exchanged with a suitable solution of precursors and suitable chemical reactions are triggered in irradiated regions of the monoliths. The physical properties of the exposed regions can be varied almost at leisure by changing the precursors. For example, metal and sulfide nanoparticles can be formed inside the pores of the matrices and these nanoparticles change optical absorption, emission and index of refraction of the exposed regions. Polymers can be attached to the walls of the matrix pores, and this allows modulation of mechanical strength, hydrophobicity and optical properties. The character of the patterns can be adapted to the specific applications by varying the precursor solution. Single-photon reactions are used to generate patterns that start on the surface of the monoliths and extend within the bulk of the monoliths. Precursors that react when exposed to ionizing radiation are employed to create high aspect ratio patterns, and precursors that are dissociated by multiphoton processes are used to produce three-dimensional architectures.

19.7. Aerogels as Platforms for Chemical Sensors

The unusual properties of aerogels – very high surface area, high porosity, low density – render them particularly appealing for sensing applications.

One of the first reported uses of silica aerogels is in optical sensing. Photoluminescent silica aerogels were generated by exposing the “standard” aerogels to energized ammonia, a reducing gas. Oxygen quenches the luminescence of the aerogel. As the collision frequency between gas-phase oxygen molecules and the luminescent carriers generated in the aerogel increases, the emission intensity decreases due to energy transfer to the oxygen molecules.

If an emission signal is due to a single type of luminescent species experiencing one microenvironment, the emission signal can be related to the concentration of quencher ([Q]) by the Stern–Volmer equation :

$$I_0/I = 1 + K_{SV}[Q] \qquad (19.8)$$

where I_0 is the luminescence intensity in the absence of quencher; I is the luminescence intensity at a given concentration of quencher; and K_{SV} is the Stern–Volmer quenching constant. If the luminescent species is present in two different environments, each of which has a different accessibility to the quencher, a somewhat more complicated equation is required. Figure 19.35 shows Stern–Volmer plots for simulated data.

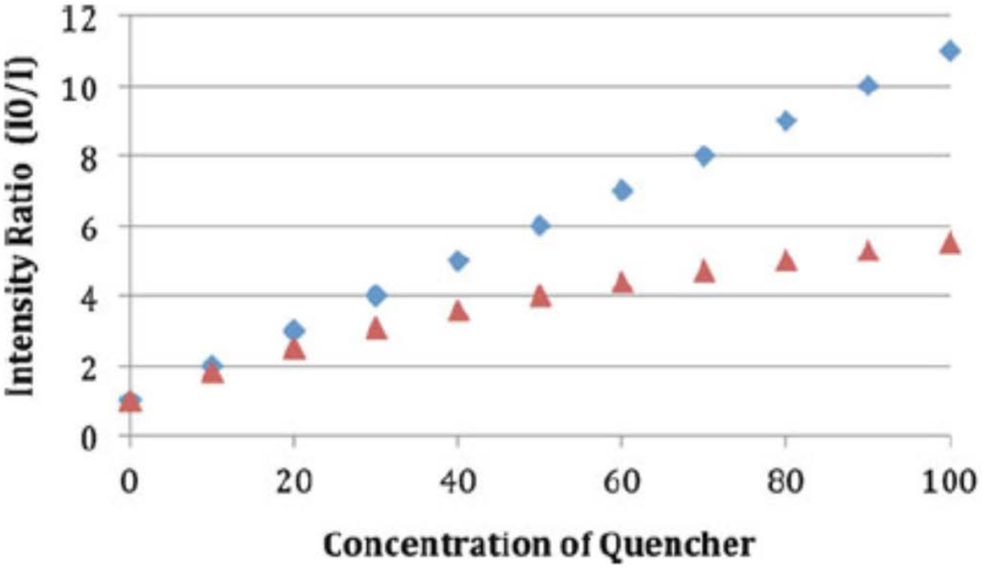


Figure 19.35. Stern–Volmer plots of simulated data. Data shown as blue diamonds are for a luminescent species in one microenvironment with K_{SV} = 0.10. Red triangles are for the same luminescent species in two microenvironments with 90% of the species having KSV =0.10 and the other 10% being nearly inaccessible to the quencher, with K_{SV}= 0.0001.

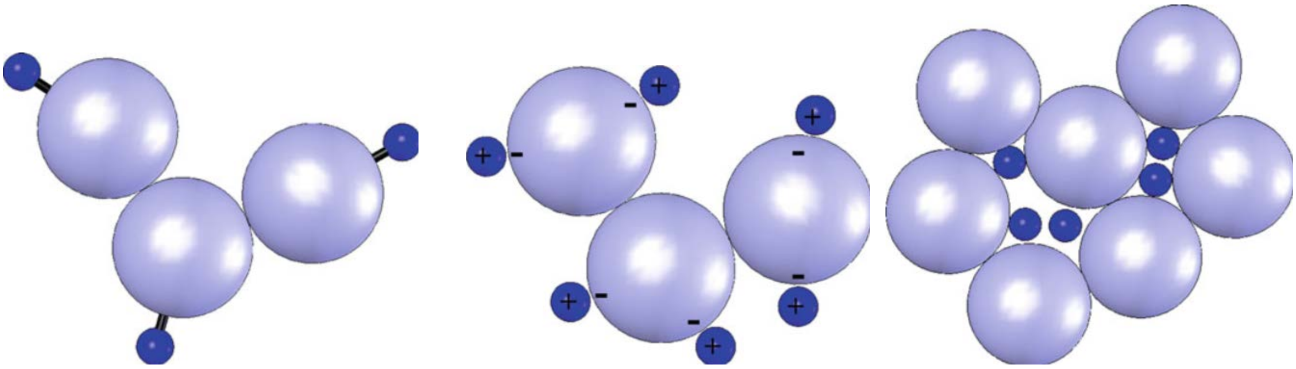


Figure 19.36 a). Schematic representation of covalent attachment of a sensing species (dark blue spheres) to the silica sol (light blue spheres). B) Schematic representation of electrostatic attachment of a cationic sensing species (dark blue spheres) to the anionic silica sol (light blue spheres). C) Schematic representation of entrapment of a sensing species (dark blue spheres) in the silica sol (light blue spheres). Not to scale.

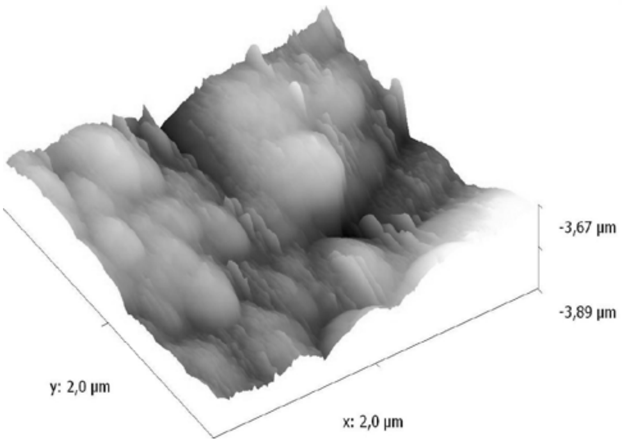
A way to produce an aerogel-based sensor is to incorporate the sensing moiety directly into the backbone of the aerogel through ***covalent attachment***, as shown schematically in Figure 19.36a. If probe moieties are covalently attached to the silica matrix of the aerogel they cannot leach from the aerogel. This is a major advantage.

Another approach is to exploit the negative charges on the silica aerogel surface to attract and hold probe species via ***electrostatic attraction***. Cations are attracted to the surface via electrostatic force, as is shown schematically in Figure 19.36b.

Physically entrapping probe species within the matrix is another way to incorporate probes into aerogels. Because this approach does not require synthetic modification of either the probe or the precursor used, it is potentially a more universal method of incorporating probes. In this case, the luminescent probe is simply dissolved in one of the reagents used to prepare the sol-gel. A probe-doped gel forms. This is an extremely common means of preparing doped xerogels. When dried under ambient conditions, the solvent ithin the pores evaporate, but the nonvolatile probes remain entrapped within the xerogel, as shown schematically in Figure 19.36c. The solvent extraction methods employed when preparing aerogels can make entrapment a more challenging approach for doped aerogel preparation.

Nanocomposite materials based on silicon nanocrystals embedded into the dielectric SiO_x matrix have been intensively investigated for applications in different technical areas. It is considered that similar materials can be used as a comprehensive matrix for nanoparticles incorporation and molecular adsorption in order to create the luminescent nanocomposites, photodetectors, catalysts and gas sensors

The photoluminescence method has been used to create the chemical sensor based on pressed composite systems based on silica aerogel and nanosilicon powders. The effect of PL quenching of considered composite structures in the presence of deionized water and ethyl alcohol solutions in water has been studied for sensor is observed. The powdered silica aerogel (ar-SiO_x) preparation method is based on the gradual water removing from the silicon acid gel. The silica gel, prepared from the sodium silicate water solution, was converted into an aerogel by a supercritical drying process. The drying process was carried out in an autoclave at supercritical parameters ($T=240°C$ and $p=8\cdot10^6$ Pa) in the presence of ethanol. These conditions are maintained for a short time with subsequent slow pressure reduction down to athmospheric pressure. As prepared material was then annealed at 450°C at oxygen athmosphere. Nanocrystalline silicon (nc-Si) powder was produced by electrochemical etching of p-Si plates in $HF:C_2H_5OH:H_2O$ electrolyte at constant current mode of 50 $mAcm^{-2}$ during 2,5 hours. Porous powder was then rinsed in ethanol, dried in air and thoroughly grated. The final product were characterized to be the silicon nanocrystallites surrounded by the porous silicon shell with the mean diameters of (2-3) nm [5]. Their porosity was about 60-70%. The samples used for PL measurements were prepared by pressing the mixture on nc-Si and ar-SiO_x powders with the different nc-Si - ar-SiO_x concentration ratio into pellets with diameter of about 12 mm and typical thickness of 0,2÷0,3 mm at a pressure of 10 MPa. Fig.19.37a shows the image of nanocomposite surface obtained using AFM.



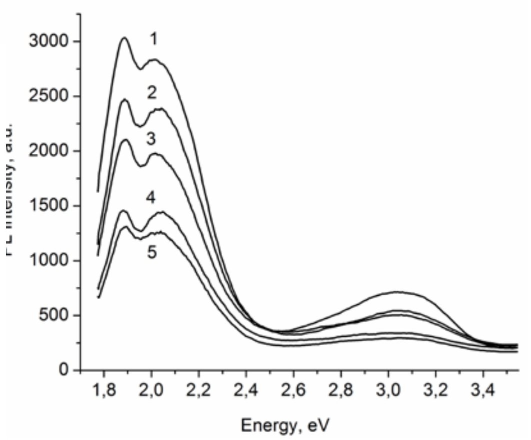


Fig. 19.37a AFM image of nc-Si/ar-SiO_x nanocomposite: the surface morphology

Fig.19.37b Evolution of non-correlated PL spectra of nc-Si/SiO_x sample treated by ethyl alcohol solutions with different C_2H_5OH concentrations: 2 – 96%; 3 – 76%; 4 – 57 %; 5 – 9%; 1 – initial sample.

Fig. 19.37b presents the typical PL spectra evolution for nc- Si/ar-SiO_x sample treated by ethyl alcohol solutions with different concentration. It shows that ethyl alcohol effectively quenched the photoemission, the most stipulated effect is observed at low-energy region. It's clearly seen that the PL quenching effect strongly depends on ethyl alcohol concentration. In particular, the pure ethyl alcohol (96%) absorption causes the 27% PL quenching, whereas the pure water (0%) absorption - 60% PL quenching

19.8. Aerogel Cherenkov Counter

When a charged particle passes through a transparent medium with a velocity faster than that of light in the material, Cherenkov light is emitted. The Cherenkov light is radiated at the Huygens wavefront as shown in Figure 19.38a. Here, the particles passes through a medium with speed v and the ratio between the speed of the particle and the speed of light is defined as n=v/c, where c is the speed of light. n is the refractive index of the medium; therefore, the emitted light waves has a speed of c/n. The angle between the direction of the emitted light and that of the particle is defined as y (Cherenkov angle) and expressed as follows:

$$\cos\theta = \frac{(c/n)\Delta t}{v\ \Delta t} = \frac{1}{n\beta}. \tag{19.9}$$

In high-energy physics research, observations and analysis of the Cherenkov light produced in transparent silica aerogel blocks can identify charged particles. Since each particle has a characteristic mass, any particle can be identified by a simultaneo measurement of its velocity and momentum. Hence, information of the velocity obtained by the Cherenkov light is directly translated into mass information, and these identifications have led to the success of high-energy physics experiments. There are two methods that make use of Cherenkov light for particle identification: one is known as the “threshold” type and the other is the “Ring Imaging” type (RICH). The former method identifies particle species according to whether Cherenkov light is emitted or not in the medium, while the latter measures the Cherenkov angle by observing the position of light on a detector.

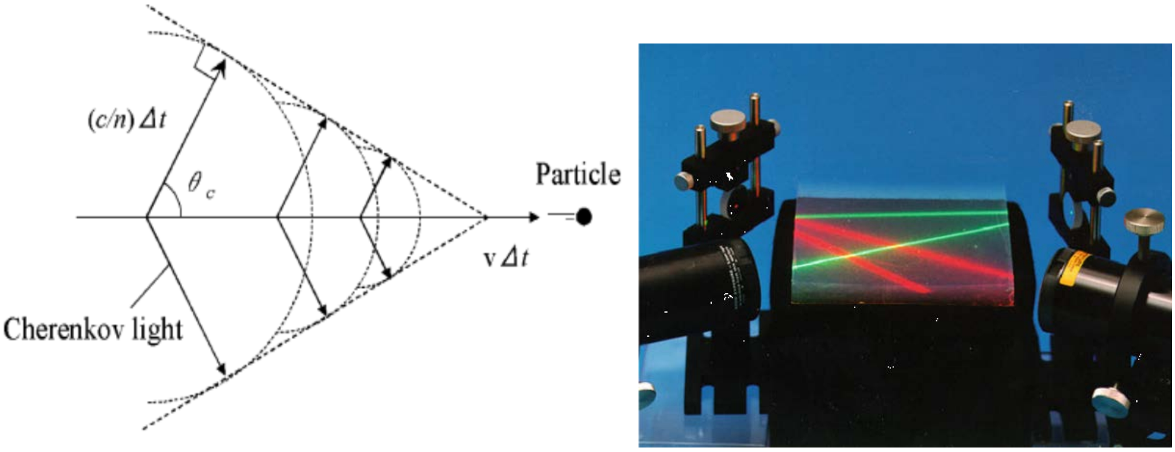


Figure 19.38. a) Cherenkov light radiation in a transparent medium. B)Photograph of 110 x 110 x10 mm^3 hydrophbic silica aerogel block.

Figure 19.39 shows a typical model of a threshold-type Cherenkov counter. Silica aerogel is filled in a box to which photomultiplier tubes (PMTs) are attached. The refractive index of aerogel is set so that the Cherenkov light is emitted for pions but not for kaons. In this way, it is known that kaon/pion/proton separation in the momentum region of 1.0–3.5 GeV/c is possible using aerogel with low refractive index such as 1.01–1.03. This type of TMSA Cherenkov counter module was tested using a 3.5 GeV/c negative pion beam at KEK. TMSA with refractive index is 1.015 and two 2.500 PMTs were arranged.

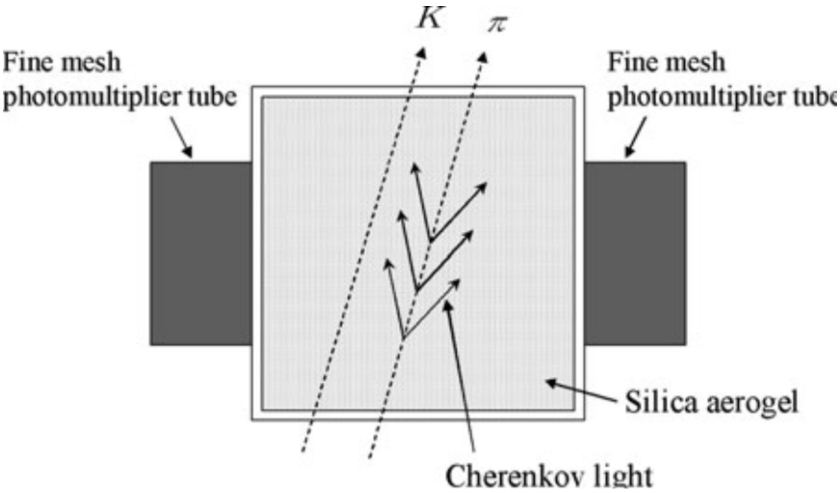


Figure 19.39. Schematic view of a typical threshold-type silica aerogel Cherenkov counter.

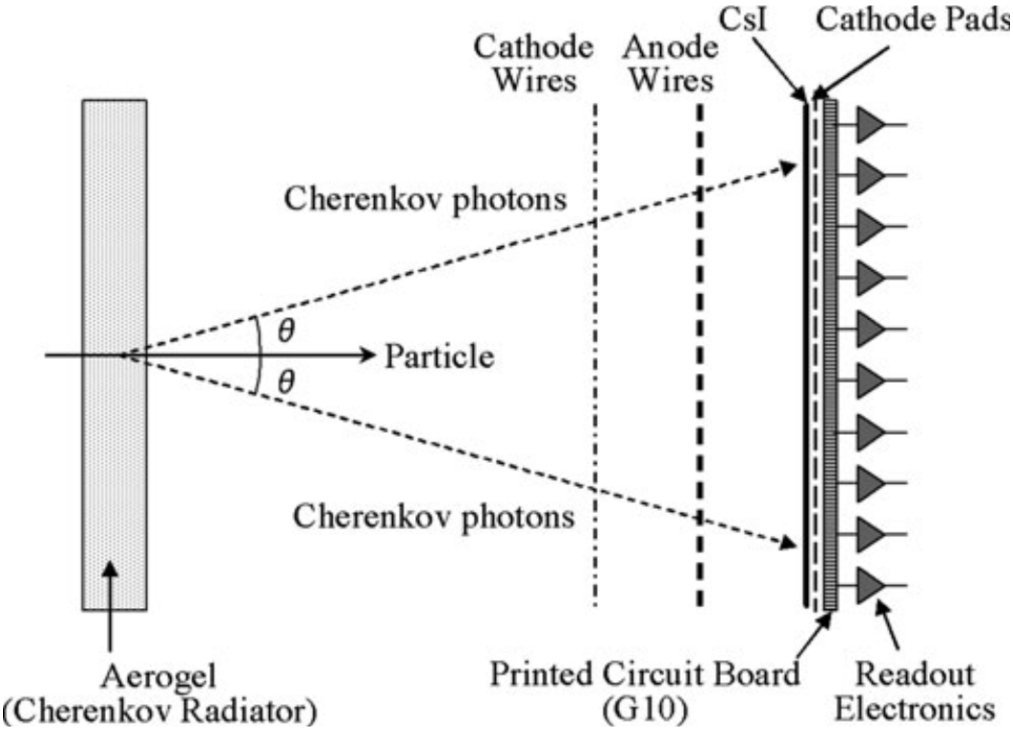


Figure 19.40 Schematic view of a ring image Cherenkov counter using silica aerogel blocks.

Figure 19.40 shows a schematic view of a typical RICH. Its principle is as follows: Cherenkov photons emitted from the radiator with Cherenkov angle (y) hit the photocathode. Then photoelectrons are released from the CsI photocathode surface. The photoelectrons drift to the anode wire and are multiplied around the wire. Then electrical charges are induced on segmented cathode pads. From the radius of a reconstructed ring image, particle information (Cherenkov angle (θ), i.e., particle speed (β)) can be deduced.

Laboratory Work #1

Calculations of Doping Impurity Concentration and Distribution in Silicon

The aim of this laboratory work is to obtain practical experience in calculation and modeling of doping impurity distribution in silicon ingots and wafers.

The work includes some concrete tasks and examples of their solutions.

Task 1.

Plot the doping distribution of arsenic at distances of 10, 20, 30, 40, and 45 cm from the seed in a silicon ingot 50 cm long that has been pulled from a melt with an initial doping concentration of 10^{17} cm^{-3}.

Solution 1.

$C_0=10^{17}$ cm^{-3} is initial doping concentration;

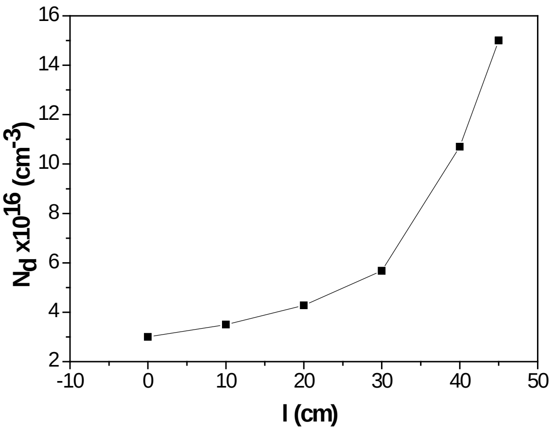
k_0 (As in Si) = 0.3 is the segregation coefficient of As in Si.

$C_s = k_0C_0(1-M/M_0)^{k0-1}$

where M_0 is the initial weight of the a crystal being grown from a melt; M is the mass crystal at a given point of growth.

$C_s = 0.3\times10^{17}(1-x)^{-0.7} = 3\times10^{16}/(1-l/50)^{0.7}$

x	0	0.2	0.4	0.6	0.8	0.9
l (cm)	0	10	20	30	40	45
C_s (cm^{-3})	3×10^{16}	3.5×10^{16}	4.28×10^{16}	5.68×10^{16}	1.07×10^{17}	1.5×10^{17}



Task 2.

In silicon, the lattice constant is 5.43 °A**.** Assume a hard-sphere model. (a) Calculate the radius of a silicon atom. (b) Determine the density of silicon atoms in atoms/cm^3. (c) Use the Avogadro constant to find the density of silicon.

Solution 2.

(a) The radius of a silicon atom can be expressed as

$r=\sqrt{3}/8\ a$

where a is the lattice constant

so $r=\sqrt{3}/8\times5.43 = 1.175$ °A

(b) The numbers of Si atom in its diamond structure are 8.

So the density of silicon atoms is

$n=8/a^3=8/(5.43)^3 = 5.0\times10^{22}$ atoms/cm^3.

(c) The density of Si is

$\rho = (M/6.02\times10^{23})/(1/n) = (28.09\ \times5\times10^{22})/(6.02\times10^{23} = 2.33$ g/cm^3.

where M is atomic mass of Si; $R=6.02\times10^{23}$ is Avogadro constant.

Task 3.

Assuming that a 10 kg pure silicon charge is used, what is the amount of boron that must be added to get the boron-doped silicon having a resistivity of 0.01 Ω-cm when one half of the ingot is grown?

Solution 3.

k_0 (B in Si) = 0.8 is the segregation coefficient of B in Si.

$M/M_0 = 0.5$

The density of Si is 2.33 g/cm^3.

The acceptor concentration for $\rho = 0.01$ Ω-cm is 9×10^{18} c rn^{-3}.

The doping concentration Cs is given by

$C_s = k_0C_0(1-M/M_0)^{k0-1}$

Therefore

$C_0 = C_s /[k_0(1-M/M_0)^{k0-1}] \quad = 9\times10^{18}/[0.8(1-0.5)^{-0.2}] = 9.8\times10^{18}$ cm^{-3}

The amount of boron required for a 10 kg charge is

$10\ 000 / 2.338 \times 9.8\text{x}10^{18} = 4.2\times10^{22}$ boron atoms

So that

10.8 g/mole × $(4.2\times10^{22}$ atoms $/6.02\times10^{23}$ atoms/mole$) = 0.75$ g boron.

Task 4.

A silicon wafer 1-mm thick having a diameter of 200 mm contains 5.41 mg of boron uniformly distributed in substitutional sites. Find (a) the boron concentration in atoms/cm^3 and, (b) the average distance between boron atoms.

Solution 4.

(a) The molecular weight of boron is 10.81

The boron concentration can be given as

n_b = number of boron atoms / volume of silicon wafer

$= [(5.41\times10^{-3}\text{ g} / 10.81\text{ g}) \times 6.02\times10^{23}] / (10.02 \times 3.14 \text{ x } 0.1) = 9.78 \times10^{18}$ atoms/cm^3.

(b) The average occupied volume of everyone boron atoms in the wafer is

$V = 1/n_b = 1 / 9.78\times10^{18}\text{ cm}^3$

We assume the volume is a sphere, so the radius of the sphere (r) is the average distance between two boron atoms. Then

$r = \sqrt{(3V/4\pi)} = 2.9\times10^{-7}$ cm.

Task 5.

The seed crystal used in Czochralski process is usually necked down to a small diameter (5.5 mm) as a means to initiate dislocation-free growth. If the critical yield strength of silicon is $2 \textbf{ x } 10^6$ g/cm^2, calculate the maximum length of a silicon ingot 200 mm diameter that can be supported by such a seed.

Solution 5.

The cross-section area of the seed is

$\pi(0.55 / 2)^2 = 0.24\text{ cm}^2$.

The maximum weight that can be supported by the seed equals the product of the critical yield strength and the seed's cross-sectional area:

$(2\times10^6) \times 0.24 = 4.8\times10^5$ g = 480 kg

The corresponding weight of a 200-mm-diameteinr got with length *l* is

$(2.33\ \mathrm{g/cm^3})\pi(20.0/2)^2\ l = 480\ 000\ \mathrm{g}$

$l = 656\ \mathrm{cm} = 6.56\ \mathrm{m}$.

Task 6.

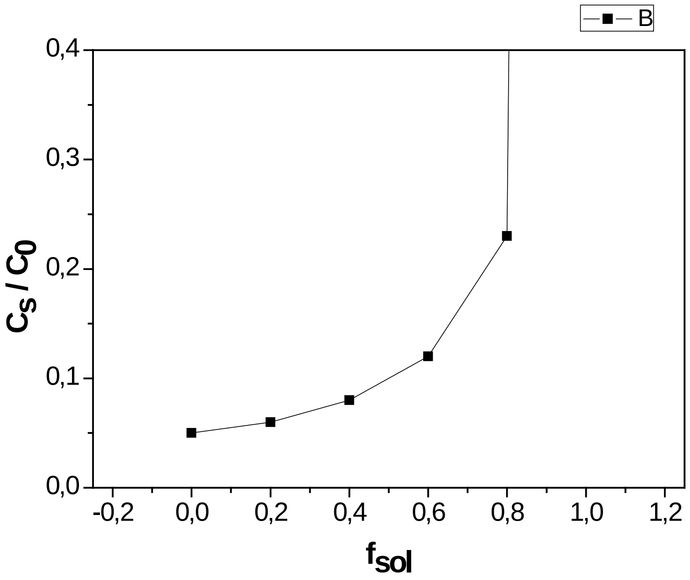
Plot the curve of C_s/C_0 value for $k_0 = 0.05$ in Czochralski technique.

Solution 6.

We have

$C_s / C_0 = k_0\, (1-M/M_0)^{k0-1}$

Fractional solidified	0	0.2	0.4	0.6	0.8	1.0
C_s / C_0	0.05	0.06	0.08	0.12	0.23	∞



Task 7.

A Czochralski grown crystal is doped with boron. Why is the boron concentration larger at the tail end of the crystal than at the seed end?

Solution 7.

The segregation coefficient of boron in silicon is 0.72. It is smaller than unity, so the solubility of B in Si under solid phase is smaller than that of the melt. Therefore, the excess B atoms will be thrown-off into the melt, then the concentration of B in the melt will be increased. The tail end of the crystal is the last to solidify. Therefore, the concentration of B in the tail-end of grown crystal will be higher than that of seed-end.

Task 8.

Why is the impurity concentration larger in the center of the wafer than at its perimeter?

Solution 8.

The reason is that the solubility in the melt is proportional to the temperature, and the temperature is higher in the center part than at the perimeter. Therefore, the solubility is higher in the center part, causinga higher impurity concentration there.

Control questions:

1. Describe the silicon growth by Czochralski method.
2. What is equilibrium segregation coefficient?
3. Describe the chemical reaction for Si obtaining.
4. Describe the silicon float-zone process.
5. A silicon ingot, which should contain 1016 boron atoms/cm3, is to be grown by the Czochralski technique. What concentration of boron atoms should be in the melt to give the required concentration in the ingot?

Laboratory Work #2

Modelling of Silicon Oxide Growth

The aim of this laboratory work is to obtain practical experience in modelling of silicon oxide growth kinetics in dependence on technological conditions.

The work includes some concrete tasks for modelling silicon oxide growth and examples of their solutions.

Task 1.

A p-type <l00>-oriented, silicon wafer with a resistivity of 10 Ω-cm is placed in a wet oxidation system to grow a field oxide of 0.45 μm at 1050°C. Determine the time required to grow the oxide.

Solution 1.

Equation for growth kinetics is (with τ=0) is

$x^2 + Ax = Bt$

where x is the oxide thickness; t is the time; A and B are the coefficients.

For the given temperature T=1050°C.

B/A=1.5 μm/hr, B=0.47 μm^2/hr, (taken from Figs showing dependence of these coefficients on temperature for case of T=1050 °C). Therefore A = 0.31 μm.

The time required to grow 0.45 μm oxide is

$t = 1/B(x^2 + Ax) = 1/0.47 \times (0.45^2 + 0.31\times0.45) = 0.72$ hr = 44 min.

Task 2.

After the first oxidation as given in Task 1, a window is opened in the oxide to grow a gate oxide at 1000°C for 20 minutes in dry oxidation. Find the thicknesses of the gate oxide and the total field oxide.

Solution 2.

Equation for growth kinetics in general case is

$x^2 + Ax = B(t+\tau)$

where τ represents a time coordinate shift to account for the initial oxide layer d_0.

After a window is opened in the oxide for a second oxidation, the rate constants are B: 0.01 μm^2/hr, A= 0.116 μm (B/A = 6×10^{-2} μm/hr). (Taken from Figs showing dependence of these coefficients on temperature for case of T=1000 °C).

If the initial oxide thickness is 20 nm = 0.02 μm for dry oxidation, the value of τ can be obtained as followed:

$(0.02)^2 = 0.116\ (0.02) = =0.01\ (0+\tau)$

or

τ= 0.372 hr

For an oxidation time of 20 min (=1/3 hr), the oxide thickness in the window area is

$x^2 + 0.166x = 0.01(0.333+0.372) = 0.007$

or

$x = 0.0350 = 35$ nm (gate oxide).

For the field oxide with an original thickness 0 .45 μm, the effective τ is given by

$\tau = 1/B \times (x^2 + Ax) = 1/0.01\ (0.45^2 + 0.166 \times 0.45) = 27.71$ hr

$x^2 + 0/166x = 0.01(0.333 + 27.72) = 0.28053$

or

$x = 0.4530$ μm ((an increase of 0.003 μm only for the field oxide).

Task 3.

Show that equation $x^2 + Ax = B(t+\tau)$ reduces to $x^2 = Bt$ for long times and to $x = B/A \times (t + \tau)$ for short times.

Solution 3.

$x^2 + Ax = B(t+\tau)$

$(x + A/2)^2 - A^2/4 = B(t + \tau)$

$(x + A/2)^2 = B[A^2/4B + (t + \tau)]$

when $t >> \tau$, $t >> A^2/4B$

then, $x^2 = Bt.$

when $t \geq \tau$, $t < A^2/4B$

then, $x = B/A \times (t+\tau).$

Task 4.

Determine the diffusion coefficient D for dry oxidation of <l00>-oriented silicon samples at 980°C and 1 atm.

Solution 4.

At 980 °C (=1253 K) and 1 atm, B=8.5×10^{-3} μm^2/hr , B/A = 4×10^{-2} μm (taken from Figs showing dependence of these coefficients on temperature for case of T=980 °C). Since $A=2D/k$, $B/A = kC_0/C_1$, C_0=5.2×10^{16} molecules/cm^3 and C_1 = 2.2×10^{22} cm^{-3}.

Here D is the diffusion coefficient; C_0 is the molecules concentration in dry oxygen C_0= 5.2 × 10^{16} molecules/cm^3, C_1 is the number of molecules of the oxidizing species in a unit volume of the oxide, C_1=2.2×10^{22} cm^{-3}; k is the surface reaction rate constant for oxidation.

The diffusion coefficient is given by

$D=Ak/2 = A/2\ (B/A \times C_1/C_0) = B/2(C_1/C_0) = 8.5\times10^{-3}/2 \times 2.2\times10^{22}/5.2\times10^{16}$ μm^2/hr =1.79×10^3 μm^2/hr = 4.79×10^{-9} cm^2/s.

<u>Control questions:</u>

1. What is the general equation for oxide growth?
2. What view has equation of oxide growth for short times of oxidation?
3. What view has equation of oxide growth for long times of oxidation?
4. What does parameter τ in the general oxidation equation mean?
5. What parameters determine the diffusion coefficient?

Numerical Lab #3

Quantum-size effects in semiconductor heterostructures

Quantum tunneling. Resonant tunneling

Task: Demonstration the effect of resonant tunneling through two-barrier structures.

Staps:

1.Open file "quantum tunneling"

2. Write the parameters set with the same quantum barriers: V_1=V_2=0.83 eV, barriers positions left side 0 and 2; right side 2 and 3.

3. Choose the regime "plane wave"

4. Calculate the value of local energy level inside quantum well (for the barriers structure it is about V= 0.65 eV)

5. Sure that similar quantum double-barrier structure has transmission coefficient $T \neq 0$ (T=0.23 or more)

6. Find the total electron energy at which the transmission coefficient will be maximum.

quantum-tunneling_en2b.jnlp
quantum-tunneling_en.jnlp

References

Ch.1

1. Veikko Lindroos, Markku Tilli , Ari Lehto and Teruaki Motooka, Handbook of Silicon Based MEMS Materials and Technologies, 2010 , Oxford , Elsevier
2.Toward Functional Nanomaterials, Lecture Notes in Nanoscale Science and Technology, Zhiming M. Wang (Ed.) Springer Dordrecht. 2009
3. Smart Materials-Based Actuators at the Micro/Nano-Scale Characterization, Control, and Application, Micky Rakotondrabe Ed. Springer New York, 2013
4. Advanced Materials andTechnologies for Micro/Nano-Devices, Sensors and Actuators
Ed Evgeni Gusev, Eric Garfunkel and Arthur Dideikin, Springer Science + Business Media B.V. 2010
5. Nano- and Micromaterials. Jun Takeda, Yoshiyuki Kawazoe, Kaoru Ohno, Masatoshi Tanaka (Eds.), Springer Berlin Heidelberg 2008
6. Nano and microstructural design of Advanced materials, M.A. Meyers, R.O. Ritchie, M. Sarikaya, Elsevier, Amsterdam, 2003
7. Handbook of nanophysics. Functional nanomaterials / editor, Klaus D. Sattler., CRC Press Taylor & Francis Group, 2010

Chapter 4.

1. L.A. Corathers. Minerals Yearbook, Vol. I, Metals & Minerals:. [Internet]. 2013 :144. Available from:http://minerals.usgs.gov/minerals/pubs/commodity/silicon/mcs-2013-simet.pdf
2. Anon. Silicon and Ferrosilicon: Global Industry Markets and Outlook, 13th edition. London: Roskill; 2011.
3. S. Pizzini. Towards solar grade silicon: Challenges and benefits for low cost photovoltaics. Solar Energy Materials and Solar Cells. **94(9)**, 1528 – 1533 (2010).
4. S. Pizzini, C. Calligarich. On the Effect of Impurities on the Photovoltaic Behavior of Solar-Grade Silicon. Journal of The Electrochemical Society. **131(9)**, 2128 (1984).
5. C. Smith, A. Barron. Synthesis and Purification of Bulk Semiconductors. [Internet]. 2012 . Available from:http://cnx.org/content/m23936/1.7/
6. P.A. Basore. Defining terms for crystalline silicon solar cells. Progress in Photovoltaics: Research and Applications, **2**, 177-179 (1994).
7. M. A. Green, Recent Developments in Photovoltaics, Solar Energy, **76** (1–3), 3–8 (2004).
8. S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd ed.), New York: Oxford University Press (2001).
9. B.G. Streetman, and S. Banerjee, *Solid State Electronic Devices* (5th ed.), New Jersey: Prentice Hall, (2000).
10. A.V. Shah, et al., Material and solar cell research in microcrystalline silicon, Solar Energy Materials and Solar Cells**, 78** (1–4), 469–491 (2003).
11. N.B. Hannay, *Treatise on Solid State Chemistry,* Vol 1, *The Chemical Structure of Solids*, Plenum Press, New York (1973).
12. J. Czochralski, Ein neues Verfahren zur Messung der Kristallisationsgeschwindigheit der Metalle, Zeitschrift für physikalische Chemie, **92**, 219–221 (1918).
13. S.W. Glunz, S. Rein, W. Warta, J. Knobloch, W. Wettling. Degradation of carrier lifetime in Cz silicon solar cells. Solar Energy Materials and Solar Cells, **65**, 219 – 229 (2001).
14. K. Bothe, R. Sinton, J. Schmidt, Fundamental boron-oxygen-related carrier lifetime limit in mono- and multicrystalline silicon. Progress in Photovoltaics: Research and Applications, **13**, 287 – 296 (2005).
15. J. Zhao, A. Wang, M.A. Green, 24.5% efficiency PERT silicon solar cells on SEH MCZ substrates and cell performance on other SEH CZ and FZ substrates. Solar Energy Materials and Solar Cells, **66**, 27 – 36 (2001).

16. W.G. Pfann, Zone-refining, Trans. AIME, 194 (1952).
17. M.A. Green, The path to 25% silicon solar cell efficiency: History of silicon cell evolution. Progress in Photovoltaics: Research and Applications, **17**, 183-189 (2009).
18. W.M. Bullis, *Silicon materials properties* in *Handbook of Semiconductor Silicon Technology*, p. 347-450.
19. W. Shockley, *Electrons and Holes in Semiconductors*, Van Nostrand, Princeton, N.J. (1950).
20. W.L. Bond, and W. Kaiser, Interstitial versus Substitutional Oxygen in Silicon, J. Phys. Chem. Solids **16**, 44-45 (1960)
21. R.D. Deslattes, The Avogadro Constant, Ann. Rev. Phys. Chem. **31**, 435-461 (1980).
22. R.G. Rhodes, *Imperfections and Active Centers in Semiconductors*, Pergamon Press, Oxford, U.K. (1964).
23. Antoniadis, D.A., and Moskowitz, I., "Diffusion of Substitutional Impurities in Silicon at Short Oxidation Times-An Insight into Point Defect Kinetics," J. Appl. Phys. 53, 6788-6796 (1982).
24. Gosele, U., and Tan, T.Y., "The Nature of Point Defects and Their Influence on Diffusion Processes in Silicon at High Temperatures," in Defects in Semiconductors - II, S. Mahajan and J.W. Corbett, eds, North-Holland, New York (1983), pp. 45-59.
25. Trumbore, F.A., "Solid Solubilities of Impurity Elements in Germanium and Silicon," Bell Syst. Tech. J. 39, 205-234 (1 960).
26. Boltaks, B.I., and Hsueh, S.-Y., "Diffusion, Solubility and the Effect of Silver Impurities on Electrical Properties of Silicon," Sov. Phys. - Solid State 2, 2383-2388 (1960-61)
27. Yoshida, M., and Furusho, K., "Behavior of Nickel as an Impurity in Silicon," Japan J. Appl. Phys. 3521-529 (1964).
28. Svob, L., "Solubility and Diffusion Coefficient of Sodium and Potassium in Silicon," Solid-State Electron. 10, 991-996 (1967).
29. Weber, E.R., "Transition Metals in Silicon," Appl. Phys. A 30, 1-22 (1983).
30. Neuberger, M., and Welles, S.J., "Silicon," DS- 162, October 1969.1 364 Handbook of Semiconductor Silicon Technology.
31. Nozaki, T., Yatsurugi, Y., Akiyama, N., Endo, Y., and Makida, Y., "Behavior of Light Impurity Elements in the Production of Semiconductor Silicon," J. Radioanal. Chem. 19, 109-128 (1974).
32. Scott, W., Ynfrared Spectra of New Acceptor Levels in Indium- or Aluminum-Doped Silicon," Appl. Phys. Lett. 32, 540-542 (1978).
33. Poindexter, E.H., Gerardi, G.J., Rueckel, M.-E., Captan, P.J., Johnson, N.M., and Biegelsen, D.K., "Electronic Traps and Pb Centers at the Si/SiO2 Interface: Band-Gap Energy Distribution," J. Appl. Phys. 56, 2844-2849 (1984). Johnson, N.M., and Biegelsen, D.K., "Identification of Deep-Gap States in a-Si:H by Photodepopulation-Induced Electron- Spin Resonance," Phys. Rev. B 31, 4066-4069 (1985).
34. Johnson, N.M., Herring, C., and Chadi, D.J., "Interstitial Hydrogen and Neutralization of Shallow-Donor Impurities in Single-Crystal Silicon," Phys. Rev. Lett. 56, 769-772 (1986).
35. Pankove, J.I., Carlson, D.E., Berkeyheiser, J.E., and Wance, R.O., "Neutralization of Shallow Acceptor Levels in Silicon by Atomic Hydrogen," Phys. Rev. Lett., 51, 2224- 2225 (1983); Johnson, N.M., "Mechanism for Hydrogen Compensation of Shallow-Acceptor Impurities in Single-Crystal Silicon," Phys. Rev., B 31, 5525-5528 (1985); DeLeo, G.G., and Fowler, W.G., "Hydrogen- Acceptor Pairs in Silicon: Pairing Effect on the Hydrogen Vibrational Frequency," Phys. Rev. B 31, 6861-6864 (1985).
36. Johnson, N.M., and Hahn, S.K., "Hydrogen Passivation of the Oxygen-Related Thermal-Donor Defect in Silicon," Appl. Phys. Lett. 48, 709-711 (1986).
37. Pensl, G., Roos, G., Holm, C., Sirtl, E., and Johnson, N.M., "Hydrogen Neutralization of Chalcogen Double Donors in Silicon," Appl. Phys. Lett. 51, 451-453 (1987).
38. Schnegg, A., Grunder, M., and Jacob, H., "On the Polishing of Silicon: Acceptor Compensation by Atomic Hydrogen,"in Semiconductor Silicon 1986, H.R. Huff, T. Abe, and B.

Kolbesen, eds, The Electrochemical Society, Pennington, NJ (1986), pp. 198-205
39. Fonash, S.J., and Rohaatgi, A., "RIE Damage and its Control in Silicon Processing," in Emerging Semiconductor Technology, ASTM STP 960, D.C. Gupta and P.H. Langer, Silicon Material Properties 439 eds, American Society for Testing and Materials, Philadelphia, (1986).
40. Horn, M.W., Heddleson, J.M., and Fonash, S.J., "Permeation of Hydrogen into Silicon during Low-Energy Hydrogen Ion Beam Bombardment," Appl. Phys. Lett. 51, 490-492 (1987).
41. Stein, H.J., "Infrared Absorption Band for Substitutional Nitrogen in Silicon," Appl. Phys. Lett. 47, 1339-1341 (1985).
42. Abe, T., Masui, T., Harada, and H., Chikawa, J., "The Characteristics of Nitrogen in Silicon Crystals," in VLSI Science and Technology/l985, W.M. Bullis and S. Broydo, eds, The Electrochemical Society, Pennington, NJ (1985).
43. Watanabe, M., Usami,T. Muraoka, H., Matsuo, S., Imanishi. Y., and Nagashima, H., "Oxygen-Free Silicon Crystal Grown from Silicon Nitride Crucible," in Semiconductor Silicon 1981, H.R. Huff, R.J. Kriegler, Y. Takeishi, eds, The Electrochemical Society, Pennington, NJ (1981).
44. Conzelmann, H., Graff, K., and Weber, E.R., Thromium and Chromium-Boron Pairs in Silicon," Appl. Phys. A 30, 169-175 (1983).
45. Ourmazd, A., and Schrdter, W., "Gettering of Metallic Impurities in Silicon," in Impurity Diffusion and Getteriirg in Silicon, R.B. Fair, C.W. Pearce, and J. Washburn, eds, Materials Research Society, Pittsburgh (1985).
46. Ward, P.J., "A Survey of Iron Contamination in Silicon Substrates and Its Impact on Circuit Yield," J. Electrochem. Soc. 129, 2573-2576 (1982).
47. Stewart, D.W., and Newton, D.C., "Determination of Iron in Semiconductor-Grade Silicon by Furnace Atomic Absorption Spectrometry," Analyst 108, 1450- 1458 (1983).
48. McGillivray, I.G., Robertson, J.M., and Walton, A.J., "Effects of Process Chemical Purity on MOS Capacitor Electrical Parameters," in Semiconductor Silicon 1986, H.R. Huff, T. Abe, and B. Kolbesen, eds, The Electrochemical Society, Pennington, NJ (1986), pp. 999-1010
49. Takizawa, R., Nakanishi, T., and Ohsawa, A., "Degradation of Metal-Oxide-Semiconductor Devices Caused by Iron Impurities on the Silicon Wafer Surface," J. Appl. Phys. 62, 4933-4935 (1987).
50. Huber, A., Rath, H.J., and Eichinger, P., "Sub-ppm Monitoring of Transition Metal Contamination on Silicon Wafer Surfaces by VPO-TXRF," in Diagnostic Techniques for Semiconductor Materials and Devices, T.J. Schaffner and D.K. Schroder, eds, The Electrochemical Society, Pennington, NJ (1988); Hockett, R.S., Baumann, S., and Schemmel, E., "An Evaluation of Ultra-Surface (<3 nm), Trace (10^{11}/cm^2) Impurity Analysis of Silicon Using a New XRay Technique," ibid., pp. 113-130,
51. Grimmeiss, H.G., Janzen, E., Ennen, H., Schirmer, O., Schneider, J., Warmer, R., Holm, C., Sirtl, E., and Wagner, P., "Tellurium Donors in Silicon," Phys. Rev. B 24, 4571-4580 (1981).
52. Salih, A.S.M., Kim, H.J., Davis, R.F., and Rozgonyi, G.A., "Extrinsic Gettering via the Controlled Introduction of Misfit Dislocations," in Semiconductor Processing, ASTM STP 850, D.C. Gupta, ed, American Society for Testing and Materials, Philadelphia (1984).
53. Bean, K.E., Lindberg, K., and Rozgonyi, G.A., "Defect Engineering," in Proc. First Int. Symp. on Advanced Materials for ULSI. M.P. Scott, Y. Ahasaka, and R. Reif, eds, The Electrochemical Society, Pennington, NJ (1988).
54. Rhodes, R.G., Imperfections and Active Centers in Semiconductors, Pergamon Press, Oxford, U.K. (1964).
55. Frank, F.C., "Crystal Dislocations - Elementary Concepts and Definitions," Phil. Mag. 42, 809-819 (1951).
56. Dash, W.C., "Copper Precipitation on Dislocations in Silicon,"J. Appl. Phys. 27, 1193-1195 (1956).

57. ASTM Standard Test Method F 47 for Crystallographic Perfection of Silicon by Preferential Etch Techniques, Annual Book of ASTM Standards, Vol. 10.05, American Society for Testing and Materials, Philadelphia, PA.

58. Dash, W.C., "Silicon Crystals Free of Dislocations," J. Appl. Phys. 29:736-737 (1958); "Growth of Silicon Crystals Free from Dislocations," ibid. 30, 459-474 (1959).

59. Dash, W.C., "Dislocations in Silicon and Germanium Crystals," in Properties of Elemental and Compound Semiconductors, Metallurgical Society of the AIME, Warrendale, PA (1960).

60. Inoue, N., Osaka, J., and Wada, K., "Oxide Micro-Precipitates in As-Grown Cz Silicon," J. Electrochem. SOC. 129, 2780-2788 (1982).

61. ASTM Standard Test Method F 416 for Detection of Oxidation Induced Defects in Polished Silicon Wafers, Annual Book of ASTM Standards, Vol. 10.05, American Society for Testing and Materials, Philadelphia, PA.

62. Claassen, A.P., Bloem, J., Valkenburg, W.G.J.N. and van den Brekel, C.H.J., The deposition of silicon from silane in a low-pressure hot-wall system, J. Crystal Growth 57, 259-266 (1982).

63. Coltrin, M.E., Kee, R.J. and Miller, J.A., A mathematical model of the coupled fluid mechanics and chemical kinetics in a chemical vapor deposition reactor, J. Electrochem. SOC. 131, 425-434 (1984).

64. Meyerson, B.S. and Jasinski, J.M., Silane pyrolysis rates for the modeling of chemical vapor deposition, J. Appl. Phys. 61, 785-787 (1987).

65. Claassen, W.A.P. and Bloem, J., The nucleation of CVD silicon on SiO, and Si,N, substrates. I. The SiH_4-HCl- H_2 system at high temperatures, J. Electrochem. Soc. 127, 194-202 (1980).

66. Kamins, T.I., Chemically vapor deposited polycrystalline-silicon films, IEEE Trans. Parts, Hybrids, and Packaging, PHP-IO, 221-229 (1974).

67. Faggin, F. and Klein, T., Silicon gate technology, Solid-State Electron. 13, 1125-1144 (1970).

68. Peters, D., Implanted-silicided polysilicon gates for VLSI transistors, IEEE Trans. Electron Devices, ED-33, 1391-1393 (1986).

69. Lifshitz, N., Dependence of the work-function difference between the polysilicon gate and silicon substrate on the doping level in polysilicon, IEEE Trans. Electron Devices, ED-32, 617-621 (1985).

70. Lifshitz, N., Luryi, S. and Sheng, T.T., Influence of the grain structure on the Fermi level in polycrystalline silicon: A quantum size effect? Appl. Phys. Lett. 51, 1824- 1826 (1987).

71. Sunami, H., Cell structure for future DRAMS, Tech. Digest International Electron Devices Meeting (Washington DC), paper 29.1, pp. 694-697 (1985).

72. McKenny, V.G., A 5 V-only 4-K static RAM, Digest 1977 IEEE International Solid-state Circuits Conference, paper WAM-1.3, pp. 16-17 (1977).

73. Fa, C.H. and Jew, T.T., The poly-silicon insulated-gate field-effect transistor, IEEE Trans. Electron Devices, ED- 13, 290-29 1 (1966).

74. Kamins, T.I., Field-effects in polycrystalline-silicon films, Solid-state Electron. 15, 789-799 (1972).

75. Kamins, T.I., and Marcoux, P.J., Hydrogenation of transistors fabricated in polycrystalline-silicon films, IEEE Electron Device Lett. EDL-1, 159-161 (1980).

76. Malhi, S.D.S., Shah, R.R., Shichijo, H., Pinizzotto, R.F., Chen, C.E., Chatterjee, P.K. and Lam, H.W., Effects of grain boundary passivation on the characteristics of p-channel MOSFETs in LPCVD polysilicon, Electronics Lett. 19, 993-994 (1983).

77. Lee, K.F., Gibbons, J.F., Saraswat, K.C. and Kamins, T. I., Thin film MOSFET's fabricated in laser-annealed polycrystalline silicon, Appl. Phys. Lett. 35, 173- 175 (1979).

78. Johnson, N.M., Biegelsen, D.K. and Moyer, M.D., Grain boundaries in p-n junction diodes fabricated in laser-recrystallized silicon thin films, Appl. Phys. Lett. 38, 900-902 (1981).

79. Tihanyi, J. and Schlotterer, H., Properties of ESFI MOS transistors due to the floating substrate and the finite volume, IEEE Trans. Electron Devices, ED-22, 1017- 1023 (1975).

80. Sasaki, N., Charge pumping in SOS-MOS transistors, IEEE Trans. Electron Devices, ED-28, 48-52 (1981).
81. Lim, H.K., and Fossum, J.G., Transient drain current and propagation delay in SO1 CMOS, IEEE Trans. Electron Devices, ED-31, 1251-1258 (1984).
82. Ashburn, P. and Soerowirdjo, B., Comparison of experimental and theoretical results on polysilicon emitter bipolar transistors, IEEE Trans. Electron Devices, ED-31, 853-860 (1984).
83. Ning, T.H. and Isaac, R.D., Effect of emitter contact on current gain of silicon bipolar devices, IEEE Trans. Electron Devices, ED-27, 2051 -2055 (1980).
84. Matsushita, T., Oh-uchi, N., Hayashi, H. and Yamoto, H., A silicon heterojunction transistor, Appl. Phys. Lett. 35, 549-550 (1979).
85. Manoliu, J. and Kamins, T.I., p-n junctions in polycrystalline- silicon films, Solid-state Electron. 15, 1103- 1106 (1972).
86. Dutoit, M. and Sollberger, F., Lateral polysilicon p-n diodes, J. Electrochem. Soc. 125, 1648-1651 (1978).
87. Davidsohn, U.S. and Lee, F., Dielectric isolated integrated circuit substrate processes, Proc. IEEE, 57, 1532- 1537 (1969).
88. van Dijk, H.J.A. and de Jonge, J., Preparation of thin silicon crystals by electrochemical thinning of epitaxially grown structures, J. Electrochem. Soc. 117, 553-554 (1970).
89. Meek, R.L., Electrochemically thinned N/N+ epitaxial silicon-Method and application, J. Electrochem. Soc. 118, 1240-1246 (1971).
90. Kamins, T.I., A new dielectric isolation technique for bipolar integrated circuits using thin single-crystal silicon films, Proc. IEEE, 60, 915-916 (1972).
91. Rung, R.D., Momose, H. and Nagakubo, Y., Deep trench isolated CMOS device, Tech. Digest 1982 International Electron Devices Meeting, paper 9.6, pp. 237-240 (1982).
92. Sunami, H., Kure, T., Hashimoto, N., Itoh, K., Toyabe, T. and Asai, S., A corrugated capacitor cell (CCC) IEEE Trans. Electron Devices, ED-31, 746-753 (1984).
93. Richardson, W.F., Bordelon, D.M., Pollack, G.P., Shah, A.H., Malhi, S.D.S., Shichijo, H., Banerjee, S.K., Elahy, M., Womack, R.H., Wang, C.P., Gallia, J., Davis, H.E. and Chatterjee, P.K., A trench transistor cross-point dRAM cell, Tech. Digest 1985 IEEE International Electron Devices Meeting, paper 29.6, pp. 714-717 (1985).
94. Silvestri, V.J., Si selective epitaxial trench refill, Fall 1986 Electrochemical Society Meeting, abstract 269, pp.402-403 (1986).
95. T.D. Kamins, “Preparation and properties of Polycrystalline Silicon Films”, in “Handbook of Semiconductor Silicon Technology” eds. W.C. O’Mara, R.B. Herring, L.P. Hunt, Noyes Publ., Park Ridge, New Jersey, USA, 1990, pp. 640-730.
96. R.W. Collins, et al., "Evolution of microstructure and phase in amorphous, protocrystalline, and microcrystalline silicon studied by real time spectroscopic ellipsometry", Solar Energy Materials and Solar Cells, 78(1-4), 143-180 (2003).
97. S. Y. Myong, S. W. Kwon, J. H. Kwak, K. S. Lim, J. M. Pearce, and M. Konagai, "Good stability of protocrystalline silicon multilayer solar cells against light irradiation originating from vertically regular distribution of isolated nano-sized silicon grains", 4th World Conference on Photovoltaic Energy Conversion Proceedings, p. 492, 2006.
98. C. R. Wronski, et al., "Intrinsic and Light Induced Gap States in a-Si:H Materials and Solar Cells - Effects of Microstructure", Thin Solid Films, 451-452, 470-475 (2004).
99. "Towards very low-cost mass production of thin-film silicon photovoltaic (PV) solar modules on glass". Elsevier B.V. 2005-09-02.
100. M. A. Kreiger, et al., Life Cycle Analysis of Silane Recycling in Amorphous Silicon-Based Solar Photovoltaic Manufacturing, Resources, Conservation & Recycling, 70, 44-49 (2013).
101. R. A. Street, Hydrogenated Amorphous Silicon, Cambridge Solid State Science Series (1986).
102. A. Madan, M.P. Shaw, The Physics an Applications of Amorphous Semiconductors, Academic Press, Inc. San Diego, California (1998).

103. J.H. Moller, Semiconductors for Solar Cells, Artech House Inc., Boston, MA (1993).
104. W. Luft and Y. Tsuo, Hydrogenated Amorphous Silicon Alloy Deposition Processes, Marcel Dekker, Inc., New York, (1993).
105. E. A. Davis and N. F. Mott, Conduction in non-crystalline systems. v. conduction, optical absorption and photoconductivity in amorphous semiconductors, Phil. Mag., 22(1970) 903-922.
106. Kim, Ji Cheol and Schwartz, Richard J., Parameter Estimation And Modeling of Hydrogenated Amorphous Silicon, ECE Technical Reports, Paper 89 (1996).
107.. M. H. Cohen, H. Fritzsch, and S. R. Ovshinsky, Simple band model for amorphous semiconducting alloys, Phys. Rev. Lett., 22 (1969) 1065-1068.
108. T. Tiedje, J. M. Cebuka, D. L. Morel, and B. Abels, Evidence for exponential band tail in amorphous silicon hydride, Phys. Rev. Lett., 46 (1981) 1425- 1428.
109. H. Dersch, J. Stuke, and J. Beichler, Electron spin resonance of doped glow discharge amorphous silicon, Phys. Status Solidi (B), 105(1981) 265-274.
110. Y. Kazama, K. Seki, W. Kim, M. K. S. Yamanaka, and K. Takahashi, High efficiency amorphous silicon solar cells with delta-doped p-layer, Jap. J. Appl. Phys., 28(July 1989) 1160-1164.
111. D. Adler, Semiconductors and Semimetals, 21 Academic Press, New York, (1984).
112. N. F. Mott and E. A. Davis, Electronic processes in non-crystalline materials, Clarendon Press, (1979).
113. E. A. Davis and N. F. Mott, Conduction in non-crystalline systems. v. conduction, optical absorption and photoconductivity in amorphous semiconductors, Phil. Mag., 22(1970) 903-922.
114. J. M. Marshall and A. E. Owen, Drift mobility studies in vitreous arsenic triselenide, Phil. Mag., 24(1971) 1281-1301.
115. R. A. Street, Luminescence and recombination in hydrogenated amorphous silicon, Adv. Phys., 30(5) (1981) 593-676.
116. B. A. Wilson, A. M. Sergent, and J. P. Harbison, Radiative recombination at dangling bonds in a-Si:H, Phys. Rev. B, 30(1984) 2282-2285.
117. R. A. Street, D.K. Biegelsen, and R. L. Weisfield, Recombination in A-Si:H: Transitions through defect states , Phys. Rev. B, 30(1)(1984) 5861-5870.
118. K. Morigaki, Y. Sano, and I. Hirabayashi, Radiative and non Radiative recombination processes in hydrogenated amorphous silicon as elucidated by optically detected magnetic resonance, Solid St. Commun. , 39 (1981) 947-951.
119. J. C. Knights, G. Lucovsky, and R. J. Nemanich, Defects in plasma-deposited a-Si:H, J. of Non-Crys. Sol., 32 (1979) 393-403.
120. K. Maschke, E. Merk, and W. Czaja, Thermalization and Radiative recombination of photo-excited carriers in a-Si:H, Phil. Mag. Lett., 56(4) (1987) 457-470.
121. W. M. Pontuschka, W. W. Carlos, P. C. Taylor, and R. W. Griffith, Radiation induced paramagnetism in a-Si:H, Phys. Rev. B, 25 (1982) 4362-4376.
122. A. Morimoto, T. Miura, M. Kumeda, and T. Shimizu, Defects in hydrogenated amorphous silicon-carbon alloy films prepared by glow discharge decomposition and sputtering, J . Appl. Phys., 53 (1982) 7299-7305.
123. M. Stutzmann, Occupancy of dangling bond defects in doped hydrogenated amorphous silicon, Solid St. Commun. 62 (3)(1987) 153-157.
124. S. R. Elliot, Defect states in amorphous silicon, Phil. Mag. B, 38(4) (1978) 325-334.
125. S. R. Elliot, Photo-induced changes in glow-discharge-deposited amorphous silicon: the Staebler-Wronski effect, Phil. Mag. B, 39(4) (1979) 349-356.
126. D. Adler, Handbook on semiconductors, North-Holland Publ. Co., Amsterdam, 1(1982) 805-841.
127. Y. Bar-Yam and J. D. Joannopoulos, Dangling bond in a-Si:H, Phys. Rev. Lett., 56 (May 1986) 2203-2206.
128. D.L. Staebler and C.R. Wronski, Appl. Phys. Lett., 31 (1977) 292.

130. M. Stutzmann, Weak bond-dangling bond conversion in amorphous silicon, Phil. Mag. B, 569(1) (1987) 63-70.
129. V.L. Dalal et al., Mat. Res. Soc. Symp. Proc. 149 (1989) 601.
131. H. M. Branz, Phys. Rev. B, 59 (8), (1999) 5498.
132. J.B. Dresner, B. Goldstein and D. Szostak, Appl. Phys. Lett., 38 (1981) 998.
133. R. Biswas, B.C. Pan, and Y. Ye, Metastability of amorphous silicon from silicon network rebonding, Phys. Rev. Lett., 88 (2002) 205502.
134. R. Biswas, B.C. Pan, Defect kinetics in new model of metastability in a-Si:H, Journal of Non-Crystalline Solids, 299-302(2002) 507.
135. H. Branz, H collision model, Phys. Rev. B 59(1999) 5498.
136. R. Biswas and B.C. Pan, Microscopic nature of Staebler-Wronski defect formation in amophous silicon, Appl. Phys. Lett., 72(1998) 371.
137. R. Biswas, Q. Li, B.C. Pan, Y. Yoon, Mechanism for hydrogen diffusion in amorphous silicon, Phys. Rev. B., 57(1998) 2253.
138. Q. Li and R. Biswas, Hydrogen Rebonding and defect formation in a-Si:H, Phys. Rev. B., 52(1995) 10705.
139. R. Biswas, and Y.P. Li, H-flip model of light induced changes of amorphous silicon, Phys. Rev. Lett. ,82(1999) 2512.
140. V.L. Dalal, E. Ping et al., Appl. Phys. Lett., 64 (1994) 1862.
141. V.L. Dalal et al., J. J. of Non-Crystalline Solids, 227-230 (1998) 1257.
142. A. H. Mahan, J. Carapella et al., Appl. Phys. Lett., 64 (1991) 1862.
143. E. Maruyama, S. Tsuda, S. Nakano, Hydrogenated amorphous silicon, Neber-Aeschbacher H. (Ed.), Science Publication, Zurich, Part 2, 44-46(1995) 863.
144. V.L. Dalal and Greg Baldwin, Mat. Res. Soc. Symp. Proc., 297 (1993) 833.
145. R.E.I. Schropp, Hydrogenated amorphous silicon, Neber-Aeschbacher H. (Ed.), Science Publication, Zurich, Part 2, 44-46 (1995) 853-861.
146. S. Guha, in: R. Street (Ed.), Technology and applications of amorphous silicon, Spring, Berlin, 37 (1999) 252.
147. X. Xu, J. Yang, S. Guha, Thin solid films 198-200 (1996) 60.
148. N. Souffi, A. Matsuda et al., Solid State Commun. 122 (2002) 259.
149. G. Gaungly, A. Matsuda, Phys. Rev. B 49 (1994) 10986.
150. T. Nishimoto, T. Takagi et al., Solar Energy Materials & Solar Cells, 66 (2001) 179.
151. K. Ohkawa, S. Shimizu et al., Solar Energy Materials & Solar Cells, 66 (2001) 297.

Ch5

1. Wide Bandgap Semiconductors. Fundamental Properties and Modern Photonic and Electronic Devices, eds.: K. Takahashi, A. Yoshikawa, and A. Sandhu, Springer-Verlag Berlin Heidelberg 2007.
2. W. Walukiewicz, S.X. Li, J. Wu, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, and W.J. Schaff: J. Cryst. Growth 269 (2004) 119.
3. W. Walukiewicz: Physica B Condens. Matter 302 (2001) 123.
4. W. Faschinger: in Wide Bandgap Semiconductors – Growth, Processing and Applications (S.J. Pearton, ed.), William Andrew Publishing/Noyes Publications, Norwich, NY, 2000, pp. 1–41.
5. W.A. Harrison and J. Tersoff: J. Vac. Sci. Technol. B 4 (1986) 1068.
6. H. Matsunami (ed.): Technology of Semiconductor SiC and Its Application, The Nikkan Kogyo Shimbun, Tokyo, 2003 (in Japanese).
7. K. Arai and S. Yoshida: Fundamentals and Applications of SiC Devices, Ohmsha, Tokyo, 2003 (in Japanese).
8. W.J. Choyke, H. Matsunami, and G. Pensl (eds.): Silicon Carbide – A Review of Fundamental Questions and Applications to Current Device Technology, Wiley, New York, 1997.

9. W.J. Choyke, H. Matsunami, and G. Pensl (eds.): Silicon Carbide – Recent Major Advances, Springer, Berlin Heidelberg New York, 2003.
10. S. Yoshida: OYO BUTURI 68 (1999) 787.
11. S. Yoshida: IEICE Trans. Electron. (Japanese Edition) J86-C (2003) 412.
12. Lei, T., Fanciulli, M., Molnar, R.J., Moustakas, T.D., Graham, R.J. and Scanlon, J. (1991) Applied Physics Letters, 59, 944.
13. Paisley, M.J., Sitar, Z., Posthill, J.B. and Davis, R.F. (1989) Journal of Vacuum Science & Technology, 7, 701.
14. Powell, R.C., Lee, N.E., Kim, Y.W. and Greene, J.E. (1993) Journal of Applied Physics, 73, 189.
15. Mizita, M., Fujieda, S.,Matsumoto, Y. And Kawamura, T. (1986) Japanese Journal of Applied Physics, 25, L945.
16. Xia, Q., Xia, H. and Ruoff, A.L. (1993) Journal of Applied Physics, 73, 8198.
17. Perlin, P., Jauberthie-Carillon, C., Itie, J.P., San Miguel, A., Grzegory, I. And Polian, A. (1992) Physical Review B: Condensed Matter, 45, 83.
18. Ueno, M., Yoshida, M., Onodera, A., Shimommura, O. and Takemura, K. (1994) Physical Review B: Condensed Matter, 49, 14.
19. Pirouz, P. and Yang, J.W. (1993) Ultramicroscopy, 51, 189.
20. Ruterana, P., S_anchez, A.M. and Nouet, G. (2003) Extended defects in wurtzite GaN layers: atomic structure, formation and interaction mechanisms, in Nitride Semiconductors – Handbook on Materials and Devices (eds P. Ruterana, M. Albrecht and J. Neugebauer), Wiley-VCH Verlag GmbH, Weinheim, Germany.
21. Harris, W.A. (1980) Electronic Structure and Properties of Solids, Dover, NY, pp. 174–179.
22. Yeh, C.-Y., Lu, Z.W., Froyen, S. And Zunger, A. (1992) Physical Review B: Condensed Matter, 46, 10086.
23. Wu, J., Walukiewicz, W., Yu, K.M., Ager, J.W., III, Haller, E.E., Lu, H., Schaff,W.J., Saito, Y. and Nanishi, Y. (2003) Applied Physics Letters, 80, 3967.
24. Tanaka, M., Nakahata, S., Sogabe, K., Nakata, H. and Tabioka, M. (1997) Japanese Journal of Applied Physics, 36, L1062.
25. Angerer, H., Brunner, D., Freudenberg, F., Ambacher, O., Stutzmann, M.,H€opler, R., Metzger, T., Born, E., Dollinger, G., Bergmaier, A., Karsch, S. and K€orner, H.-J. (1997) Applied Physics Letters, 71, 1504.
26. Domagala, J., Leszczynski, M., Prystawko, P., Suski, T., Langer, R., Barski, A. and Bremser, M. (1999) Journal of Alloys and Compounds, 286, 284.
27. Kim, K., Lambrecht,W.R.L. and Segall, B. (1996) Physical Review B: Condensed Matter, 53, 16310.
28. Wright, A.F. and Nelson, J.S. (1995) Physical Review B: Condensed Matter, 51, 7866.
29. Leszczynski, M., Suski, T., Perlin, P., Teisseyre, H., Grzegory, I., Bockowski, M., Jun, J., Porowski, S., Pakula, K., Baranowski, J.M., Foxon, C.T. and Cheng, T.S. (1996) Applied Physics Letters, 69, 73.
30. Detchprohm, T., Hiramatsu, K., Itoh, K. and Akasaki, I. (1992) Japanese Journal of Applied Physics, 31, L1454.
31. Paszkowicz,W. (1999) Powder Diffraction, 14, 258.
32. Leszczynski, M., Teisseyre, H., Suski, T., Grzegory, I., Bockowski, M., Jun, J., Porowski, S. andMajor, J. (1995) Journal of Physics D: Applied Physics, 69, A149.
33. Deguchi, T., Ichiryu, D., Toshikawa, K., Sekiguchi, K., Sota, T., Matsuo, R., Azuhata, T., Yamaguchi, M., Yagi, T., Chichibu, S. and Nakamura, S. (1999) Journal of Applied Physics, 86, 1860.
34 Phillips, J.C. (1973) Bonds and Bands in Semiconductors, Academic Press, New York, p. 31.
35. Xu, Y.-N. and Ching, W.Y. (1993) Electronic, optical, and structural properties of some wurtzite crystals. Physical Review B: Condensed Matter, 48, 4335–4350.

36. Morko_c, H., Strite, S., Gao, G.B., Lin, M.E., Sverdlov, B. and Burns, M. (1994) Large-band-gap SiC, III–V nitride, and II–VI ZnSe-based semiconductor device technologies. Journal of Applied Physics, 76 (3), 1363–1398.
37. Akasaki, I. and Amano, H. (1994) Properties of Group III Nitrides (ed. J.H. Edgar), EMIS Data Review Series, No. 11, INSPEC, The Institution of Electrical Engineers, Stevenage, UK, pp. 30–34.
38. Christensen, N.E. and Gorczyca, I. (1994) Optical and structural properties of III–V nitrides under pressure. Physical Review B: Condensed Matter, 50, 4397–4415.

Chapter 6

1. B.E. Deal and A.S. Grove, General relationship for the thermal oxidation of silicon, J. Appl. Phys. **36**, 3770 (1965).
2. S. K. Ghandhi: *VLSI Fabrication Principles – Silicon and Gallium Arsenide* (Wiley, New York 1983)
3. W. A. Pilskin: Comparison of properties of dielectric films deposited by various methods, J. Vac. Sci. Technol. **21,** 1064–1081 (1977).
4. J. S. Danel, F. Michel, G. Delapierre: Micromachining of quartz and its application to an acceleration sensor, Sens. Actuators A **21-23,** 971–977 (1990).
5. A. Yasseen, J. D. Cawley, M. Mehregany: Thick glass film technology for polysilicon surface micromachining, J. Microelectromech. Syst. **8,** 172–179 (1999).
6. R. Liu, M. J. Vasile, D. J. Beebe: The fabrication of nonplanar spin-on glass microstructures, J. Microelectromech. Syst. **8,** 146–151 (1999).
7. J.D. Meindl, et. al. Silicon epitaxy and oxidation, in *Process and device modeling for integrated circuits design*, Eds. F. Van de Wiele, W.L. Engl, and P.O. Jespers, Noorhoff, Leyden, 1977.
8. A.S. Grove, *Physics and technology of semiconductor devices*, Wiley, New York, 1967.
9. G.S. May, S.M. Sze, *Fundamentals of semiconductor fabrication*, John Wiley & Sons, Inc., New York, 2004.
10. S.M. Sze, *Semiconductor devices: Physics and technology*, John Wiley & Sons, Inc., New York, 2002.
11. G.S. May, C.J. Spanos, *Fundamentals of semiconductor manufacturing and process control*, John Wiley & Sons, Inc., New York, 2006.
12. A.G. Revesz, J. Non-cryatalline solids **4**, 347 (1970).
13. A.G. Revesz, Phys. Stat. Sol (a) **57**, 235 (1980).
14 J.C. Bravman, and R. Sinclair, Transmission electron microscopy studies of the polycrystalline silicon-SiO_2 interface, Thin Solid Films, **104**, 153-161 (1983).
15. T.I. Kamins, Oxidation of phosphorus-doped low pressure and atmospheric pressure CVD polycrystalline- silicon films, J. Electrochem. Soc. **126**, 838-844 (1979).
16. T.I. Kamins, Preparation and properties of polycrystalline silicon films, in *Handbook of semiconductor silicon technology*, eds. W.C. O'Mara, R.B. Herring, L.P. Hunt, P.640, Noyes Publ., New Jersey, USA (2009).
17. Sunami, H., Thermal oxidation of phosphorus-doped polycrystalline silicon in wet oxygen, J. Electrochem. Soc. **125,** 892-897 (1978).
18. E.A. Irene, E. Tierney, and D.W. Dong, Silicon oxidation studies: Morphological aspects of the oxidation of polycrystalline silicon, J. Electrochem. Soc. **127**, 705-7 13 (1980).
19. A.C. Adams, "Dielectric and polisilicon films deposition," in S.M. Sze, Ed., VLSI technology, McGraw-Hill, New York, 1983.
20. K. Eujio, et. al. "Doped silicon oxide deposition by atmospheric pressure and low temperature chemical vapor deposition using tetraethoxysilane and ozone," J. Electrochem. Soc. **138**, 3019 (1991).
21. K. P. Jayadevan, T. Y. Tseng, Oxide Nanoparticles in *Encyclopedia of Nanoscience and Nanotechnology* Edited by H. S. Nalwa, Vol. 8, pp. 333–376, (2004).

22. G. B. Basim, J. J. Adler, U. Mahajan, R. K. Singh, and B. M. Moudgil, J. Electrochem. Soc. 147, 3523 (2000).
23. H. K. Bowen, Mater. Sci. Eng. **44,** 1 (1980).
24. S. K. Friedlander, "Smoke, dust and haze: Fundamentals of aerosol behavior," Wiley Interscience, New York 1977.
25. R. W. Siegel, in "Physics of new materials," edited by F. E. Fujita, Springer-Verlag, Berlin, Germany, 1994.
26. C. G. Grandqvist and R. A. Buhrman, J. Appl. Phys. **47**, 2200 (1976).
27. A. Gurav, T. Kodas, T. Pluym, and Y. Xiong, Aerosol Sci. Technol. **19**, 411 (1993).
28. S. R. Nagel, J. B. MacChesney, and K. L. Walder, in "Optical fiber communications," edited by L. Tingye, Academic Press, NY, Vol. 1, 1985.
29. S. E. Pratsinis and S. V. R. Mastrangelo, Chem. Eng. Prog. **85**, 62 (1989).
30. S. E. Pratsinis and T. T. Kodas, in "Aerosol measurement," edited by K. Willeke and P. A. Baron, Van Nostrand Reinhold, New York, 1993.
31. G. D. Ulrich, Chem. Eng. News **62**, 22 (1984).
32. M. K. Wu, R. S. Windeler, C. K. R. Steiner, T. Bors, and S. K. Friedlander, Aerosol Sci. Technol. **19**, 527 (1993).
33. T. T. Kodas and M. J. H. Smith, "Aerosol processing of materials," Wiley-VCH, New York, 1999.
34. M. T. Harris, T. C. Scott, and C. H. Byers, Mater. Sci. Eng. A **168**, 125 (1993).
35. M. T. Harris, W. G. Sisson, T. C. Scott, O. A. Basaran, C. H. Byers, W. Renand, and T. Meek, Mater. Res. Soc. Symp. Proc. **271**, 945 (1992).
36. M. T. Harris, T. C. Scott, O. A. Basaran, and C. H. Byers, AIChE Symp. Series, Superconducting Engineering **88**, 44 (1992).
37. M. T. Harris, T. C. Scott, O. A. Basaran, and C. H. Byers, Mater. Res. Soc. Symp. Proc. **180**, 153 (1990).
38. J. Livage, M. Henry, and S. Sanchez, Prog. Solid State Chem. **18**, 259 (1988).
39. C. J. Brinker and G. W. Scherer, "Sol-gel science," Academic Press, San Diego, 1990.
40. D. C. Bradely, R. C. Mehrotra, and D. P. Gaur, "Metal alkoxides," Academic Press, London, 1978.
41. I. A. Aksay, in "Forming of ceramics," edited by J. A. Mangels and G. L. Messing, American Ceramic Society, Columbus, OH, 1984.
42. C. J. Brinker and G.W. Scherer, J. Non-Cryst. Solids **70**, 301 (1985).
43. J. A. Lewis, J. Am. Ceram. Soc. 83, 2341 (2000).
44. M. D. Sacks and T. Y. Tseng, J. Am. Ceram. Soc. **67**, 526 (1984).
45. M. D. Sacks and T. Y. Tseng, J. Am. Ceram. Soc. **67**, 532 (1984).
46. T. Y. Tseng and J. J. Yu, J. Mater. Sci. **21**, 3615 (1986).
47. T. Y. Tseng and J. C. Lee, Am. Ceram. Soc. Bull. **52**, 443 (1987).
48. M. Nayak and T. Y. Tseng, Thin Solid Films **408**, 194 (2002).

Chapter 7

1. J. Homola, S. S. Yee, and G. Gauglitz, "Surface plasmon resonance sensors: review," Sensor. Actuat. B 54, 3–15 (1999).
2. J. Homola, "Surface plasmon resonance (SPR) biosensors and their applications to food safety and security," in Frontiers in Planar Lightwave Circuit Technology,, S. Janz, J. Ctyroky, and S. Tanev, eds. (Springer Netherlands, 2006), pp. 101-118.
3. M. Quinten, A. Leitner, J. R. Krenn, and F. R. Aussenegg, "Electromagnetic energy transport via linear chains of silver nanoparticles," Opt. Lett. 23, 1331–1333 (1998).

4. S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. G. Requicha, B. E. Koel, and H. A. Atwater, “Plasmonics –a route to nanoscale optical devices,” Adv. Mater. 13, 1501–1505 (2001).
5. S. J. Oldenburg, C. C. Genick, K. A. Clark, and D. A. Schultz, “Base pair mismatch recognition using plasmon resonant particle labels,” Anal. Biochem. 309, 109–116 (2002).
6. J. M. K¨ohler, A. Cs´aki, J. Reichert, R. M¨oller, W. Straube, and W. Fritzsche, “Selective labeling of oligonucleotide monolayers by metallic nanobeads for fast optical readout of DNAchips,” Sensor. Actuat. B 76, 166–172 (2001).
7. S. Schultz, D. R. Smith, J. J. Mock, and D. A. Schultz, “Single target molecule detection with nonbleaching multicolor optical immunolabels,” Proc. Natl. Acad. Sci. USA 97, 996–1001 (2000).
8. G. Bohren and D. Huffman, Absorbtion and Scattering of Light by Small Particles (John Wiley & Sons, Inc., New York, 1998).
9. E. Hecht, Optics (Addison–Wesley, San Fransisco, 2002).
10. H. A. Kramers, “La deffusion de la lumiere par les atomes,” Atti del Congresso Internazionale dei Fisici 2, 545–557 (1927).
11. R. de L. Kronig, “On the theory of dispersion of X-rays,” J. Opt. Soc. Am. 12, 547–557 (1926).
12. L. Landau and E. Lifshitz, Electrodynamics of ContinuousMedia (Pergamon Press, Oxford, 1975).
13. M. Dressela and G. Gr ¨uner, Electrodynamic of Solids: Optical Properties of Electrons in Matter (Cambridge University Press, Cambridge, 2003).
14. M. Fox, Optical Properties of Solids (Oxford University Press, Oxford, 2001).
15. F. Wooten, Optical Properties of Solids (Academic Press, New York, 1972).
16. P. Drude, The Theory of Optics (Dover Publications, New York, 1959).
17. D. R. Lide, ed., Handbook of Chemistry and Physics (CRC Press LLC, Boca Raton, 2005).
18. A. Hohenau, A. Leitner, and F. R. Aussenegg, “Near-field and far-field properties of nanoparticle arrays,” in Surface Plasmon Nanophotonics, M. L. Brongersma and P. G. Kik, eds. (Springer, 2007).
19. N. Ashcroft and N. Mermin, Solid State Physics (Holt- Saunders, New York, 1976).
20. I. Lindau and P. O. Nilsson, “Experimental verification of optically excited longitudinal plasmons,” Phys. Scripta 3, 87–92 (1971).
21. S. A. Maier, Plasmonic Fundamentals and Applications (Springer science + Business Media LLC, New York, 2007).
22. R. B. M. Schasfoort and A. J. Tudos, Handbook of Surface Plasmon Resonance (The Royal Society of Chemistry, Cambridge, 2008).
23. E. Kretschmann and H. Reather, “Radiative decay of nonradiative surface plasmons excited by light,” Z. Naturforsch., Tail A 23, 2135–2136 (1968).
24. G. Mie, “Beitr¨age zur optik tr ¨uber medien, speziell kolloidaler metall ¨osungen,” Ann. Phys. 25, 377–455 (1908).
25. K. L. Kelly, E. Coronado, L. L. Zhao, and G. C. Schatz, “The optical properties of metal nanoparticles: the influence of size, shape, and idelectric environment,” J. Phys. Chem. B 107, 668–677 (2003).
26. L. Novotny and B. Hecht, Principles of Nano-Optics (Cambridge University Press, Cambridge, 2006).
27. C. Evans, Precision Engineering: An Evolutionary View (Cranfield University Press, Cranfield, 1989).
28. S. Choi, M. Yan, L. Wanga, and I. Adesida, “Ultra-dense hydrogen silsesquioxane (HSQ) structures on thin silicon nitride membranes,” Microelectron. Eng. 86, 521–523 (2009).
29. A. E. Grigorescu, M. C. van der Krogt, C. W. Hagen, and P. Kruit, “10 nm lines and spaces written in HSQ, using electron beam lithography,” Microelectron. Eng. 84, 822–824 (2007).

30. J. Zheng, P. E. Constantinou, C. Micheel, A. P. Alivisatos, R. A. Kiehl, and N. C. Seeman, “Two-Dimensional Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs,” Nano Lett. 6, 1502–1504 (2006).
31. T. Aaltonen, Atomic Layer Deposition of Noble Metal Thin Films, PhD thesis (University of Helsinki, Helsinki, 2005).
32. D. M. Mattox, Handbook of Physical Vapor Deposition (PVD) Processing (Noyes Publications, Westwood, 1998).
33. R. P. Netterfeld, “Review of thin film deposition techniques,” Optical Interference Coatings, Proc. OSA, MA2–3 (2001).
34. N. Kaiser, “Review of fundamentals of thin film growth,” Optical Interference Coatings, Proc. OSA, MA1–3 (2001).
35. A. Lehmuskero, M. Kuittinen, and P. Vahimaa ”Refractive index and extinction coefficient dependence of thin Al and Ir films on deposition technique and thickness” Optics Express, 15, pp. 10744–10752, 2007.
36. M.J. Buiting, A.F. Otterllo, and A.H. Montree, “Kinetical aspects of the LPCVD of titanium nitride from titanium tetrachloride and ammonia”, J. Electrochem. Soc., 138, 500 (1991).
37. R. Tobe, et. al. “Plasma-enhanced CVD of TiN and Ti using low-pressure and high-density helicon plasma”, Thin Solid Films, 281-282, 155 (1996).
38. J. Hu et. al., “Electrical properties of Ti/TiN films prepared by chemical vapor deposition and their applications in submicron structures as contact and barrier materials”, Thin Solid Films, 308, 589 (1997).
39. M. Leskel¨a and M. Ritala, “Atomic layer deposition (ALD): from precursors to thin film structures,” Thin Solid Films 409, 138–146 (2002).
40. T. Alasaarela, T. Saastamoinen, J. Hiltunen, A. Saynatjoki, A. Tervonen, P. Stenberg, M. Kuittinen, and S. Honkanen, “Atomic layer deposited titanium dioxide and its application in resonant waveguide grating,” Appl. Opt. 49, 4321–4325 (2010).
41. M. Ritala and M. Leskel¨a, “Atomic layer deposition,” in Handbook of Thin Film Materials, H. S. Nalwa, ed. (Academic Press, 2002), pp. 103–159.
42. M. Ohring, Materials Science of Thin Films: Deposition & Structure (Academic Press, San Diego, 2002).
43. R. H. Doremus, “Optical properties of thin metallic films in island form,” J. Appl. Phys. 37, 2775–2781 (1966).
44. C. V. Thompson, “Structure evolution during processing of polycrystalline films,” Annu Rev. Mater. Sci. 30, 150–190 (2000).
45. C. R. M. Grovenor, H. T. G. Hentzell, and D. A. Smith, “The development of grain structure during growth of metallic films,” Acta Metall. 32, 773–781 (1984).
46. L. B. Freund and S. Suresh, Thin Film Materials: Stress, Defect Formation and Surface Evolution (Cambridge University Press, Cambridge, 2003).
47. M. Hansen and A. Anderko, Constitution of Binary Alloys, McGraw-Hill, New York, 1958.
48. D. Pramanik and A. N. Saxena, "VLSI Metallization Using Aluminum and Its Alloys," Solid State Tech., 26, No. 1, 127 (1983), 26. No. 3, 131 (1983).
49. C. L. Hu, and J. M. E. Harper, "Copper Interconnections and Reliability," Matter. Chern. Phys., 52, 5 (1998).
50. P. C. Andricacos, et al., "Damascene Copper Electroplating for Chip Interconnects," 193rd Meet. Electrochem. Soc., 1998, p. 3.
51. J.M. Steigerwald et. Al. “Chemical mechanical planarization of microelectronic materials”, Wiley, New York, 1997.
52. Guide to Using WVASE32 (J.A Woollam Co., Lincoln, 2003).
53. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, Numerical Recipes in C, The Art of Scientific Computing (Cambridge University Press, Cambridge, 1988).
54. E. Collet, Polarized Light: Fundamentals and Applications (Marker Dekker, Cambridge, 1993).

55. E. Palik, ed., Handbook of Optical Constants of Solids I (Academic Press, San Diego, 1985).
56. G. Hass and J. E.Waylonis, “Optical constants and reflectance and transmittance of evaporated aluminium in the visible and ultraviolet,” J. Opt. Soc. Am. 51, 719–722 (1961).
57. L. W. Bos and D. W. Lynch, “Optical properties of antiferromagnetic chromium and dilute Cr-Mn and Cr-Re alloys,” Phys. Rev. B 2, 4567–4577 (1970).
58. R. F. Fane and W. E. J. Neal, “Optical constants of aluminum films related to the vacuum environment,” J. Opt. Soc. Am. 60, 790–793 (1970).
59. H.-J. Hagemann, W. Gudat, and C. Kunz, “Optical constants from the far infrared to the X-ray region: Mg, Al, Cu, Ag, Au, Bi, C, and Al2O3,” J. Opt. Soc. Am. 65, 742–744 (1975).
60. H. Ehrenreich and H. R. Philipp, “Optical properties of Ag and Cu,” Phys. Rev. 128, 1622–1629 (1962).
61. L. G. Schulz and F. R. Tangherlini, “Optical constants of silver, gold, copper and aluminum,” J. Opt. Soc. Am. 44, 362–368 (1954).
62. E. Shiles, T. Sasaki, M. Inokuti, and D. Y. Smith, “Selfconsistency and sum-rule tests in the Kramers-Kronig analysis of optical data: applications to aluminum,” Phys. Rev. B 22, 1612–1628 (1980).
63. J. H. Weaver, C. G. Olson, and D. W. Lynch, “Optical investigation of the electronic structure of bulk Rh and Ir,” Phys. Rev. B 15, 4115–4118 (1977).
64. D. E. Aspnes, “Optical properties of thin films,” Thin Solid Films 89, 249–262 (1982).
65. O. Hunderi, “Influence of grain boundaries and lattice defects on the optical properties of some metals,” Phys. Rev. B 7, 3419–3429 (1973).
66. L.Ward, “The effective optical constants of thin metal films in island form,” Brith. J. Appl. Phys. (J. Phys. D) 2, 123–125 (1969).
67. W. Zhang, S. H. Brongersma, O. Richard, B. Brijs, R. Palmans, L. Froyen, and K. Maex, “Influence of the electron mean free path on the resistivity of thin metal films,” Microelectron. Eng. 76, 146–152 (2004).
68. D. Hyman, J. Lam, B. Warneke, A. Schmitz, T. Y. Hsu, J. Brown, J. Schaffner, A. Walson, R. Y. Loo, M. Mehregany, J. Lee: Surface micromachined RF MEMS switches on GaAs substrates, Int. J. Radio Frequency Microwave Commun. Eng. 9 (1999) 348–361.
69. C. Chang, P. Chang: Innovative micromachined microwave switch with very low insertion loss, Sens. Actuators 79 (2000) 71–75.
70. C. L. Shih, B. K. Lai, H. Kahn, S. M. Phillips, A. H. Heuer: A robust co-sputtering fabrication procedure for TiNi shape memory alloys for MEMS, J. Microelectromech. Syst. 10 (2001) 69–79.
71. G. Hahm, H. Kahn, S. M. Phillips, A. H. Heuer: Fully microfabricated silicon spring biased shape memory actuated microvalve, Technical Digest – Solid State Sens. Actuator Workshop, Hilton Head Island 2000, ed. by L. Bousse (Transducers Research Foundation, Cleveland Heights Ohio 2000) 230–233.
72. S. D. Leith, D. T. Schwartz: High-rate through-mold electrodeposition of thick (> 200 micron) NiFe MEMS components with uniform composition, J. Microelectromech. Syst. 8 (1999) 384–392.
73. N. Rajan, M. Mehregany, C. A. Zorman, S. Stefanescu, T. Kicher: Fabrication and testing of micromachined silicon carbide and nickel fuel atomizers for gas turbine engines, J. Microelectromech. Syst. 8 (1999) 251–257.
74. T. Pornsin-sirirak, Y. C. Tai, H. Nassef, C. M. Ho: Titanium-alloy MEMS wing technology for a microaerial vehicle application, Sens. Actuators A 89 (2001) 95–103.
75. U.S. Department of Energy Report on “Nanoscale Science, Engineering and Technology Research Directions,” available at http:// www.sc.doe.gov/bes/reports/files/ NSET_rpt.pdf.
76. T.-C. Chiang, Surf. Sci. Rep. 39, 181 (2000).
77. T.-C. Chiang, Science 306, 1900 (2004).
78. F. J. Himpsel, J. E. Ortega, G. J. Mankey, and R. F. Willis, Adv. Phys. 47, 511 (1998).

79. S-Å. Lindgren and L. Walldén, Handbook of Surface Science, Vol. 2, Electronic Structure, ed. S. Holloway, N. V. Richardson, K. Horn, and M. Scheffler (Elsevier, Amsterdam, 2000).
80. M. Milun, P. Pervan, D. P. Woodruff, Rep. Prog. Phys. 65, 99 (2002).
81. L. Aballe, C. Rogero, P. Kratzer, S. Gokhale, and K. Horn, Phys. Rev. Lett. 87, 156801 (2001).
82. I. Matsuda, T. Ohta, and H. W. Yeom, Phys. Rev. B 65, 085327 (2002).
83. A. Mans, J. H. Dil, A. R. H. F. Ettema, and H. H. Weitering, Phys. Rev. B 66, 195410 (2002).
84. Y. Z. Wu, C. Y. Won, E. Rotenberg, H. W. Zhao, F. Toyoma, N. V. Smith, and Z. Q. Qiu, Phys. Rev. B 66, 245418 (2002).
85. J. J. Paggel, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 81, 5632 (1998).
86. J. J. Paggel, T. Miller, and T.-C. Chiang, Science 283, 1709 (1999).
87. J. J. Paggel, T. Miller, and T.-C. Chiang, Phys. Rev. B 61, 1804 (2000).
88. M. H. Upton, T. Miller, and T.-C. Chiang, Appl. Phys. Lett. 85, 1235 (2004).
89. M. Upton, C. M. Wei, M. Y. Chou, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 93, 026802 (2004).
90. H. Hong, R. D. Aburano, D.-S. Lin, H. Chen, T.-C. Chiang, P. Zschack, and E. D. Specht, Phys. Rev. Lett. 68, 507 (1992).
91. J. E. Ortega and F. J. Himpsel, G. J. Mankey, and R. F. Willis, Phys. Rev. B 47, 1540 (1993).
92. A. Danese and R. A. Bartynski, Phys. Rev. B 65, 174419 (2002).
93. Z. Q. Qiu and N. V. Smith, J. Phys. Condensed Matter 14, R169 (2002).
94. J. J. Paggel, C. M. Wei, M. Y. Chou, D.-A. Luh, T. Miller, and T.-C. Chiang, Phys. Rev. B 66, 233403 (2002).
95. C. M. Wei and M. Y. Chou, Phys. Rev. B 66, 233408 (2002).
96. F. K. Schulte, Surf. Sci. 55, 427 (1976).
97. X. Ma et al., Proc. Nat. Acad. Sci. 104, 9204 (2007).
98. D.-A. Luh, T. Miller, J. J. Paggel, M. Y. Chou, and T.-C. Chiang, Science 292, 1131 (2001).
99. C. M. Wei and M. Y. Chou, Phys. Rev. B 68, 125406 (2003).
100. P. Czoschke, L. Basile, H. Hong, and T.-C. Chiang, Phys. Rev. Lett. 93, 036103 (2004).
101. P. Czoschke, H. Hong, L. Basile, and T.-C. Chiang, Phys. Rev. B 72, 075402 (2005).
102. M. Hupalo, S. Kremmer, V. Yeh, L. Berbil-Bautista, E. Abram, and M. C. Tringides, Surf. Sci. 493, 526 (2001).
103. H. Hong, C.-M. Wei, M. Y. Chou, Z. Wu, L. Basile, H. Chen, M. Holt, and T.-C. Chiang, Phys. Rev. Lett. 90, 076104 (2003).
104. P. Czoschke, H. Hong, L. Basile, and T.-C. Chiang, Phys. Rev. Lett. 91, 226801 (2003).
105. P. Czoschke, H. Hong, L. Basile, 035305 (2005).
106. T.C. Chiang, “Quantum physics of thin metal films”, AAPPS Bulletin, 18, 2-10 (2008).
107. G. Grimvall, The Electron-Phonon Interaction in Metals (North Holland, New York, 1981).
108. T. Balasubramanian, E. Jensen, X. L. Wu, and S. L. Hulbert, Phys. Rev. B 57, R6866 (1998).
109. M. Hengsberger, D. Purdie, P. Segovia, M. Garnier, and Y. Baer, Phys. Rev. Lett. 83, 592 (1999).
110. D.-A. Luh, T. Miller, J. J. Paggel, and T.-C. Chiang, Phys. Rev. Lett. 88, 256802 (2002).
111. J. J. Paggel, D.-A. Luh, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 92, 186803 (2004).
112. Y. Guo et al., Science 306, 1915 (2004).
113. M. M. Özer, Y. Jia, Z. Zhang, J. R. Thompson, and H. H. Weitering, Science 316, 1594 (2007).
114. D. Eom, S. Qin, M.-Y. Chou, and C. K. Shih, Phys. Rev. Lett. 96, 027005 (2006).
115. S.-J. Tang, L. Basile, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 93, 216804 (2004).
116. N. J. Speer, S.-J. Tang, T. Miller, and T.-C. Chiang, Science 314, 804 (2006).
117. S.-J. Tang, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 96, 036802 (2006).

118. S.-J. Tang, Y.-R. Lee, S.-L. Chang, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. 96, 216803 (2006).

Ch 8

1. Robert Dorey, Ceramic Thick Films for, MEMS and Microdevices, AMSTERDAM, Elesevier, 2012
2.NANOCRYSTALLINE METALS AND OXIDES, Selected Properties and Applications EDs Philippe Knauth, Joop Schoonman, KLUWER ACADEMIC PUBLISHERS, NEW YORK,2002
3. Adaptive oxide electronics: A review. Sieu D. Ha and Shriram Ramanathan J. Appl. Phys. 110, 071101 (2011); doi: 10.1063/1.3640806
4. Metal Oxides Chemistry and Applications, J. L. G. Fierro, Taylor & Francis, NY, 2006
5. Ceramics Science and Technology: Volume3: Synthesis and Processing, First Edition. Edited by Ralf Riedel and I-Wei Chen, 2012 Wiley-VCH Verlag GmbH
6. Handbook of Nanophysics, Functional Nanomaterials, Ed. Klaus D. Sattler, Taylor & Francis, NY, 2011
7. Shriram Ramanathan, Ed. Thin Film Metal-Oxides Fundamentals and Applications in Electronics and Energy, Springer, 2009
8. Functional Metal Oxides, New Science and Novel Applications, Eds. Satishchandra B. Ogale, Thirumalai V. Venkatesan,and Mark G. Blamire,, Wiley, NY, 2013

Ch 9

1.Strobl, G., The Physics of Polymers: Concepts for Understanding Their Structures and Behavior,Springer-Verlag Berlin Heidelberg, 2007.
2. Cowie, J.M.G. and Arighi, V., Chemistry and Physics of Modern Materials, 3rd Edition, 2008
3. Drobny, Jiri George. Polymers for electricity and electronics : materials, properties, and applications, John Wiley & Sons, Inc., Hoboken, New Jersey, 2012
4. Bansi D. Malhotra. Handbook of Polymers in Electronics, Rapra Technology Limited, Shawbury, Shrewsbury, Shropshire, SY4 4NR, UK
5. Mittal, Vikas. Advances in polymer nanocomposite technology, 2010, Nova Science Publishers, Inc.
6. Organic Electronics. Eds: Gregor Meller, · Tibor Grasser, Springer-Verlag Berlin Heidelberg 2010.
7. Polymer Electronics, Hsin-Fei Meng ed., CRC Press, Taylor & Francis Group, NW, 2012
8. Applied methodologies in polymer research and technology, eds. Abbas Hamrang, Devrim Balköse, Gennady E. Zaikov, and A. K. Haghi, CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway NW, 2014

Ch10

1. L. B. Kong et al., Waste Energy Harvesting, Lecture Notes in Energy 24, Springer-Verlag, Berlin Heidelberg 2014.
2. S. Roundy, R.K. Wright, J. Rabaey, A study of low level vibrations as a power source for wireless sensor nodes. Comput. Commun. 26, 1131–1144 (2003).
3. S.P. Beeby, M.J. Tudor, N.M. White, Energy harvesting vibration sources for microsystems applications. Measur. Sci. Technol. 17, R175–R195 (2006).
4. K.A. Cook-Chennault, N.Thambi, A.M. Sastry, Powering MEMS portable devices - a review of non-regenerative and regenerative power supply systems with emphasis on piezoelectric energy harvesting systems. Smart Mater. Struct. 17, 043001 (2008).
5. S.R. Anton, H.A. Sodano, A review of power harvesting using piezoelectric materials (2003–2006). Smart Mater. Struct. 16, R1–R21 (2007).

6. S. Priya, Advances in energy harvesting using low profile piezoelectric transducers. J.Electroceram. 19, 167–184 (2007).
7. C.B. Williams, R.B. Yates, Analysis of a micro-electric generator for microsystems. Sens. Actuators A 52, 8–11 (1996).
8. P. Glynne-Jones, M.J. Tudor, S.P. Beeby, N.M. White, An electromagnetic, vibrationpowered generator for intelligent sensor systems. Sens. Actuators A 110, 344–349 (2004).
9. D. Arnold, Review of microscale magnetic power generation. IEEE Trans. Magn. 43, 3940–3951 (2007).
10. P. Mitcheson, P. Miao, B. Start, E. Yeatman, A. Holmes, T. Green, MEMS electrostatic micropower generator for low frequency operation. Sens. Actuators A 115, 523–529 (2004).
11. Y.B. Jeon, R. Sood, J.H. Jeongand, S. Kim, MEMS power generator with transverse mode thin film PZT. Sens. Actuators A 122, 16–22 (2005).
12. L. Wang, F.G. Yuan, Vibration energy harvesting by magnetostrictive material. Smart Mater. Struct. 17, 045009 (2008).
13. W.J. Choi, Y. Jeon, J.H. Jeong, R. Sood, S.G. Kim, Energy harvesting MEMS device based on thin film piezoelectric cantilevers. J. Electroceram. 17, 543–548 (2006).
14. G.H. Haertling, Ferroelectric ceramics: history and technology. J. Am. Ceram. Soc. 82(4), 797–818 (1999).
15. D. Damjanovic, Ferroelectric, dielectric and piezoelectric properties of ferroelectric thin films and ceramics. Rep. Prog. Phys. 61, 1267–1324 (1998).
16. L.B. Kong, J. Ma, H.T. Huang, W. Zhu, O.K. Tan, Lead zirconate titanate ceramics derived from oxide mixture treated by a high-energy ball milling process. Mater. Lett. 50, 129–133 (2001).
17. L.B. Kong, T.S. Zhang, J. Ma, Y.C.F. Boey, Progress in synthesis of ferroelectric ceramic materials via high-energy mechanochemical techniques. Prog. Mater. Sci. 53(2), 207–322 (2008).
18. L.E. Cross, R.E. Newnham, in History of ferroelectrics, Ceramics and Civilization, vol 3, ed. by W.D. Kingery. High-Technology Ceramics—Past, Present and Future (American Ceramic Society, Westerville, 1987), pp. 289–305.
19. Y. Saito, H. Takao, T. Tani, T. Nonoyama, K. Takatori, T. Homma, T. Nagaya, M. Nakamura, Lead-free piezoceramics. Nature 432, 84–87 (2004).
20. J. Rodel, W. Jo, K.T.P. Seifert, E.-M. Anton, T. Granzow, Perspective on the development of lead-free piezoceramics. J. Am. Ceram. Soc. 92(6), 1153–1177 (2009).
21. S. Roundy, P.K. Wright, A piezoelectric vibration based generator for wireless electronics. Smart Mater. Struct. 13, 1131–1144 (2004).
22. M. Birkholz, Crystal-field induced dipoles in heteropolar crystals – II. physical significance, J. Phys. B **96**(3), 333–340 (1995).
23. S. Trolier-McKinstry, Chapter3: Crystal Chemistry of Piezoelectric Materials. In A. Safari, E.K. Akdo˘gan, *Piezoelectric and Acoustic Materials for Transducer Applications*. New York: Springer, (2008).
24. Sensor Sense: Piezoelectric Force Sensors. Machinedesign.com (2008-02-07). Retrieved on 2012-05-04.
25. D. Damjanovic, Ferroelectric, dielectric and piezoelectric properties of ferroelectric thin films and ceramics, *Reports on Progress in Physics* **61**(9): 1267–1324 (1998).
26. V Kochervinskii, Piezoelectricity in Crystallizing Ferroelectric Polymers, Crystallography Reports **48**(4): 649–675 (2003).
27. S.O.R. Moheimani and A.J. Fleming, Piezoelectric Transducers for Vibration Control and Damping", Springer, Germany, 2006.
28. D. Radusinović, and C. Markov, "Macedonite - lead titanate: a new mineral", *American Mineralogist* **56**, 387-394 (1971).
29. E.A.J. Burke, and C. Kieft, "Second occurrence of makedonite, $PbTiO_3$, Långban, Sweden", *Lithos* **4**, 101-104 (1971).

30. M. Akizuki, M.S. Hampar, and J. Zussman, "An explanation of anomalous optical properties of topaz". *Mineralogical Magazine* **43** (326): 237–241 (1979).
31. M. Schwartz, Smart Materials, CRC Press., United States of America, 2009.
32. M. Migliorato, et al. A Review of Non Linear Piezoelectricity in Semiconductors. *AIP Conf Proc* (AIP Publishing) **1590**(N/A): 32–41.
33. G. Arlt, The influence of microstructure on the properties of ferroelectric ceramics.Ferroelectrics 104, 217–227 (1990).
34. A.D. Polli, F.F. Lange, C.G. Levi, C G. Metastability of the fluorite, pyrochlore, and perovskite structures in the PbO-ZrO_2-TiO_2 system. J. Am. Ceram. Soc. 83(4), 873–881 (2000).
35. C.A. Randall, N. Kim, J.P. Kucera, W. Cao, T.R. Shrout, Intrinsic and extrinsic size effects in fine-grained morphotropic-phase-boundary lead zirconate titanate ceramics. J. Am. Ceram. Soc. 81(3), 677–688 (1998).
36. B. Jaffe, W.R. Cook, H. Jaffe, Piezoelectric Ceramics (Academic Press Limited, London, 1971).
37. I.N. Andryushinan, L.A. Reznichenko, V.A. Alyoshin, L.A. Shilkina, S.V. Titov, V.V. Titov, K.P. Andryushin, S.I. Dudkina, The PZT system (PbZr1-xTixO3, 0.0 B x B 1.0): specific features of recrystallization sintering and microstructures of solid solutions (part 1).Ceram. Int. 39, 753–761 (2013).
38. B.W. Lee, Synthesis and characterization of compositionally modified PZT by wet chemical preparation from aqueous solution. J. Eur. Ceram. Soc. 24, 925–929 (2004).
39. S.R. Shannigrahi, F.E.H. Tay, K. Yao, R.N.P. Choudhary, Effect of rare earth (La, Nd, Sm, Eu, Gd, Dy, Er and Yb) ion substitutions on the microstructural and electrical properties ofsol-gel grown PZT ceramics. J. Eur. Ceram. Soc. 24, 163–170 (2004.)
40. L.B. Kong, J. Ma, T.S. Zhang, W. Zhu, O.K. Tan, Pb(ZrxTi1-x)O3 ceramics via reactive sintering of partially reacted mixture produced by a high-energy ball milling process. J. Mater. Res. 16(6), 1636–1643 (2001).
41. N. Izyumskaya, Y.I. Alivov, S.J. Cho, H. Morkoc, H. Lee, Y.S. Kang, Processing, structure, properties, and applications of PZT thin films. Crit. Rev. Solid State Mater. Sci. 32, 111–202 (2007).
42. L.B. Kong, A new solution method to deposit thick PZT films, unpublished works.
43. B.M. Xu, Y.H. Ye, L.E. Cross, Dielectric properties and field-induced phase switching of lead zirconate titanate stannate antiferroelectric thick films on silicon substrates. J. Appl. Phys. 87, 2507–2515 (2000).
44. B.M. Xu, L.E. Cross, D. Ravichandran, Synthesis of lead zirconate titanate stannate antiferroelectric thick films by sol-gel processing. J. Am. Ceram. Soc. 82(2), 306–332 (1999).
45. B.M. Xu, Y.H. Ye, Q.M. Wang, N.G. Pai, L.E. Cross, Effect of compositional variations on electrical properties in phase switching (Pb, La)(Zr, Ti, Sn)O3 thin and thick films. J. Mater. Sci. 35, 6027–6033 (2000).
46. Z.H. Wang, C.L. Zhao, W.G. Zhu, O.K. Tan, W. Liu, X. Yao, Processing and characterization of Pb(Zr, Ti)O3 thick films on platinum-coated silicon substrate derived from sol-gel deposition. Mater. Chem. Phys. 75, 71–75 (2002).
47. W.G. Zhu, Z.H. Wang, C.L. Zhao, O.K. Tan, H.H. Hng, Low temperature processing of nanocrystalline lead zirconate titanate (PZT) thick films and ceramics by a modified solgel route. Jpn. J. Appl. Phys. 41, 6969–6975 (2002).
48. Z.H. Wang, W.G. Zhu, C.L. Zhao, O.K. Tan, Dense PZT thick films derived from sol-gel based nanocomposite process. Mater. Sci. Eng., B 99, 56–62 (2003).
49. C.L. Zhao, Z. Wang, W.G. Zhu, O.K. Tan, H.H. Hng, Microstructure and properties of PZT53/47 thick films derived from sols with submicron-sized PZT particle. Ceram. Int. 30, 1925–1927 (2004).
50. Z.H. Wang, W.G. Zhu, C.L. Chao, X.F. Chen, Characterization of composite piezoelectric thick film for MEMS application. Surf. Coat. Technol. 198, 384–388 (2005).
51. R.G. Kepler, R.A. Anderson, Ferroelectric polymers. Adv. Phys. 41(1), 1–57 (1992).

52. V.V. Kochervinskii, Piezoelectricity in crystallizing ferroelectric polymers: poly(vinylidene fluoride) and its copolymers (a review). Crystallogr. Rep. 48(4), 649–675 (2003).
53. T. Furukawa, Piezoelectricity and pyroelectricity in polymers. IEEE Trans. Electr. Insul. 24, 375–394 (1989).
54. J.F. Tressler, S. Alkoy, A. Dogan, R.E. Newnham, Functional composites for sensors, actuators and transducers. Compos. Part A 30, 477–482 (1999).
55. B. Hilczer, J. Kulek, E. Markiewicz, M. Kosec, B. Malic, Dielectric relaxation in ferroelectric PZT-PVDF nanocomposites. J. Non-Cryst. Solids 305, 167–173 (2002).
56. "Market Report: World Piezoelectric Device Market". Acmite Market Intelligence.
57. Moubarak, et al., A Self-Calibrating Mathematical Model for the Direct Piezoelectric Effect of a New MEMS Tilt Sensor, IEEE Sensors Journal, 12 (5), 1033 – 1042 (2012).
58. S.P. Beeby, M.J. Tudor, N.M. White, Energy harvesting vibration sources for Microsystems applications. Meas. Sci. Technol. 17, R175-R195 (2006).
59. S. Xu, B.J. Hansen, Z.L. Wang, Piezoelectricnanowire-enabled power source for driving wireless microelectronics. Nat. Commun. 93, 1-5 (2011).
60. S. Roundy, E.S. Leland, J. Baker, E. Carleton, E. Reilly, E. Lai, B. Otis, J.M. Rabaey, P.K. Wright, V. Sundarajan, Improving power output for vibration-based energy scavengers. Pervasive computing, IEEE, 4(1), 28-36, (2005).
61. S. Chalasani, J.M. Conrad, A survey of energy harvesting sources for embedded systems. Souteastcon, IEEE, 442-447, (2008).
62. Z. Fan, J.G. Lu, Nanostructured ZnO: building blocks for nanoscale devices. Int. J. Hi. Spe. Ele. Syst. 16(4), 883-896 (2006).
63. C. Xu, X. Wang, Z.L. Wang, Nanowire structured hybrid cell for concurrently scavenging solar and mechanical energies, J. Am. Chem. Soc. 131, 5866-5872 (2009).
64 L.K. Mende, J.L.M. Driscoll, ZnO nanostructures, defects and devices, Mater. Today 10(5), 40-48 (2007).
65. D. Guyomar, A. Badel, E. Lefeuvre, and C. Richard, Toward energy harvesting using active materials and conversion improvement by nonlinear processing. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 52, 584-594 (2005).
66. J. Qiu, H. Jiang, H. Ji, and K. Zhu, Comparison between four piezoelectric energy harvesting circuits. Frontiers of Mechanical Engineering in China, 4, 153-159 (2009).
67. J. Qiu and H. Ji, The application of piezoelectric materials in smart structures in China, Int’l J. of Aeronautical & Space Sci. 11(4), 266–284 (2010).
68. H. L. Ji, Y. Ma, J. H. Qiu, H. Jiang, H. Shen, and K. J. Zhu, Optimal design of high efficiency piezoelectric energy harvester. Guangxue Jingmi Gongcheng/Optics and Precision Engineering, 16, 2346-2351 (2008).
69. E. Lefeuvre, A. Badel, C. Richard, L. Petit, and D. Guyomar, A comparison between several vibration-powered piezoelectric generators for standalone systems. Sensors and Actuators, A: Physical, 126, 405-416 (2006).
70. J.R. Phillips, "Piezoelectric Technology: A Primer". *eeProductCenter*. TechInsights (2000-08-10).
TreeHugger. Discovery Communications, LLC, (2006-08-04).
72. S.H. Wright, "MIT duo sees people-powered "Crowd Farm".*MIT news*. Massachusetts Institute of Technology, (2007-07-25).
73. A. Kannampilly, "How to Save the World One Dance at a Time". *ABC*. ABC, (2008-07-11).
74. F. J. Holler, D.A. Skoog, S.R. Crouch, "Chapter 1". *Principles of Instrumental Analysis* (6th ed.). Cengage Learning. p. 9 (2007).
75. I.A. Ivan, M. Rakotondrabe, J. Agnus, R. Bourquin, N. Chaillet, P. Lutz, J.-C. Poncot, R. Duffait and O. Bauer, Comparative material study between PZT ceramic and newer crystalline PMN-PT and PZN-PT materials for composite bimorph actuators, Rev. Adv. Mater. Sci. 24, 1-9 (2010).

76. M. Sitti, D. Campolo, J. Yan, R.S. Fearing, T. Su, D. Taylor and T. Sands, In: *Development of PZT and PZN-PT Based Unimorph Actuators for Micromechanical Flapping Mechanisms* (IEEE Int. Conf. Robotics and Automation, Seoul Korea, 2001), p. 3839.
77. B. Ko and B. H. Tongue, Smart Mater. Struct. **15,** 1912 (2006).
78. R. Le Letty, F. Barillot, N. Lhermet, F. Claeyssen, M. Yorck, J. Gavira Izquierdo, H. Arends, Barillot, Lhermet, Claeyssen, Yorck, Gavira Izquierdo, Arends, "The scanning mechanism for ROSETTA/MIDAS from an engineering model to the flight model". *Proceedings of the 9th European Space Mechanisms and Tribology Symposium, 19–21 September 2001, Liège, Belgium. Compiled by R. A. Harris. ESA SP-480, Noordwijk, Netherlands: ESA Publications Division* **480,** 75–81 (2001).
79. J.M. Jou, A Study on the Composite Type Piezoelectric Motor, Open Journal of Acoustics Vol.3, No.3 (2013).
80. A. Agarwal, Foundations of analog and digital electronic circuits. Department of electrical engineering and computer science, Massachusetts Institute of Technology, 2005, p. 43.
81. http://www.ndt-ed.org/EducationResources/CommunityCollege/Ultrasonics/EquipmentTrans/emats.htm, as referenced on 11th August 2009.
82. P.J. Shull, "Non destructive evaluation: Theory, techniques and applications," *CRC Press,* 2002.
83. http://www.ndt-ed.org/EducationResources/CommunityCollege/Ultrasonics/EquipmentTrans/piezotransducers.htm, as referenced on 11th August 2009.
84. R.L. Goldberg, S.W. Smith, J. Mottley, and K.W. Ferrara, "The Biomedical Engineering Handbook," *CRC Press*, 2000.
85. N.F. Foster, "Handbook of Thin Film Technology," *McGraw-Hill*, 1970.
86. F. Akasheh, J.D. Fraser, S. Bose, and A. Bandyopadhyay, "Piezoelectric micromachined ultrasonic transducers: modeling the influence of structural parameters on device performance," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 52, pp. 455-468, 2005.
87. P.C. Eccardt and K. Niederer, "Micromachined ultrasonic transducers with improved coupling factors from a CMOS compatible process," *Ultrasonics*, 38, pp. 744-780, 2000.
88. P. Muralt and J. Baborowski, "Micromachined ultrasonic transducers and acoustic sensors based on piezoelectric thin films," *Journal of Electroceramics*, 12, pp. 101-108, 2004.
89. M. Torndahl, M. Almgvist, L. Wallman, H. W. Persson, and K. Lindstorm, "Characterisation and comparison of a CMUT versus a piezoelectric transducer for air applications," *Proceedings of IEEE Ultrasonics Symposium*, 2, pp. 1023-1026, 2002.
90. B.T. Khuri-Yakub, C.H. Cheng, F.L. Degertekin, S. Ergun, S. Hansen, X.C. Jin, and O. Oralkan, "Silicon micromachined ultrasonic transducers," *Japan Journal of Appied Physics*, 39, pp. 2883-2887, 2000.
91. A.S. Ergun, H. Yongli, C.H. Cheng, O. Oralken, J. Johnson, H. Jagannathan, U. Demirci, G.G. Yaraliogiu, M. Karaman, and B.T. Khuri-Yakub, "Broadband capacitive micromachined ultrasonic transducers ranging from 10 KHz to 60 MHz for imaging and more," *Proceedings of IEEE Ultrasonics Symposium*, Munich, pp. 1039-1043, 2002.
71. M.G. Richard, "Japan: Producing Electricity from Train Station Ticket Gates".
92. O. Oralkan, A.S. Ergun, C.H. Cheng, J.A. Johnson, M. Karaman, T.H. Lee, and B.T. Khuri-Yakub, "Volumetric ultrasound imaging using 2-D CMUT arrays," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 50, pp. 1581-1594, 2003.
93. P. Reynolds and J.D. Fraser, "Finite element determination of effective electromechanical coupling coefficient with applications to piezoelectric and electrostatic transducers," *Ultrasonics International Conference*, 2001.
94. P. Osbond, C.M. Beck, C.J. Brierley, M.R. Cox, S.P. Marsh, and N.M. Shorrocks, "The influence of ZnO and electrode thickness on the performance of thin film bulk acoustic wave resonators," *Proceedings of IEEE Ultrasonics Symposium*, pp. 911-914, 1999.

95. G. Percin and B.T. Khuri-Yakub, “Piezoelectrically actuated flextensional micromachined ultrasonic transducers II: Fabrication and experiments,” *IEEE Transaction on Ultrasonics, Ferroelectrics and Frequency Control*, 49, pp. 585-595, 2002.
96. J.L. Vernet, W. Steichen, R. Lardat, O. Garcia, and J.F. Gelly, “PMUTs design optimization for medical probes applications,” *Proceedings of IEEE Ultrasonic Symposium*, pp. 899-902, 2001.
97. A. Manbachi, and R.S.C. Cobbold, "Development and Application of Piezoelectric Materials for Ultrasound Generation and Detection". *Ultrasound* **19**(4), 187–196 (2011).
98. D.J. Hoigne, S. Stubinger, O. Von Kaenel, S. Shamdasani, P. Hasenboehler, "Piezoelectic osteotomy in hand surgery: first experiences with a new technique". *BMC Musculoskelet Disord* **7,** 36 (2006).
99. M. Labanca, F. Azzola, Vinci R, L.F. Rodella, "Piezoelectric surgery: twenty years of use". *Br J Oral Maxillofac Surg* **46**(4): 265–9 (2008).
100. "Good vibrations lead to efficient excitations in hybrid solar cells". Gizmag.com. Retrieved 2013-11-11.
101. H.S. Tzou, Piezoelectric shells: distributed sensing & control, Kluwer Academic Publishers, London, 1993.
102. J. Nuffer, T. Bein, Application of piezoelectric materials in transportation industry**,** Global Symposium on Innovative Solutions for the Advancement of the Transport Industry, 4.-6. October 2006, San Sebastian, Spain
103. E. Pipitone, L D'Acquisto, “Development of a low-cost piezo film-based knock sensor”. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 217(8), 691-699 (2003).
104. G. Merker, S. Meyer, „Piezo–Aktuatorik in einem Forschungs–Common- Rail-Einspritzsystem“.Symposium ‚Keramik im Fahrzeugbau’, Eds. J. Heinrich, H. Gasthuber, Deutsche Keramische Gesellschaft, Köln 15-20, (2003).
105. F. Böcking, B. Sugg, “Piezo Actuators: A Technology prevails with injection valves for combustion engines”, Actuator 2006, 10th International Conference on New Actuators, (2006) 171-176 .
106. M. Bacher – Höchst, A. Scheffold, „Systemverhalten und Betriebsfestigkeit von Komponenten in modernen Einspritzsystemen und Bremsanlagen“. Konstruktion 7/8 (2005). 76 - 81
107. E. Bischur, N. Schwesinger, "Organic Piezoelectric Energy Harvesters in Floor". *Advanced Materials Research*. 433-440, 5848–5853 (2012).
108. "Introducing our new Pavegen technology". *http://www.pavegen.com/*. Retrieved 23 July 2014.
109. F. Duarte, F. Casimiro, D. Correia, R. Mendes, A. Ferreira, "A new pavement energy harvest system". *Renewable and Sustainable Energy Conference (IRSEC), 2013 International*: 408–413.
110. S. Cafiso, M. Cuomo, A. Di Graziano, C. Vecchio, "Experimental Analysis for Piezoelectric Transducers Applications into Roads Pavements". *Advanced Materials Research* **684,** 253–257.
111. A. Arjun; A. Sampath; S. Thiyagarajan; V. Arvind, "A Novel Approach to Recycle Energy Using Piezoelectric Crystals". *International Journal of Environmental Science and Development* **2,** 88–492 (2011).
112. Li, Xiaofeng; Strezov, Vladimir. "Modelling piezoelectric energy harvesting potential in an educational building". *Energy Conversion and Managemen* **85,** 435–442.
113. X. Li, V. Strezov, "Modelling piezoelectric energy harvesting potential in an educational building". *Energy Conversion and Management* **85**: 435–442.

Chapter 11.

1. A. Gruverman, Ferroelectric Nanodomains in *Encyclopedia of Nanoscience and Nanotechnology* Edited by H. S. Nalwa Volume 3: Pages (359–375), 2004.
2. M. E. Lines and A. M. Glass, "Principles and Applications of Ferroelectric and Related Materials." Clarendon, Oxford, 1977.
3. P. Ayyub, Ferroelectric Nanomaterials, in "Encyclopedia of Nanoscience and Nanotechnology" Edited by H. S. Nalwa, Volume 3: Pages (377–400), 2004.
4. M. E. Lines and A. M. Glass, "Principles and Applications of Ferroelectrics and Related Materials." Clarendon Press, Oxford, 1977.
5. T. Mitsui, I. Tatsuzaki, and E. Nakamura, "An Introduction to the Physics of Ferroelectrics." Gordon and Breach, London, 1976.
6. R. Blinc and B. Zeks, "Soft Modes in Ferroelectrics and Antiferroelectrics." Elsevier, New York, 1974.
7. Y. Xu, "Ferroelectric Materials and Their Applications." North-Holland, Amsterdam, 1991.
8. C. Jaccard, W. Kanzig, and M. Peter, *Helv. Phys. Acta* 26, 521 (1953).
9. M. Anliker, H. R. Brugger, and W. Kanzig, *Helv. Phys. Acta* 27, 99 (1954).
10. W. Kanzig, *Phys. Rev.* 98, 549 (1955).
11. K. Okazaki and K. Nagata, *J. Am. Ceram. Soc.* 56, 82 (1973).
12. H. T. Martrena and J. C. Burfoot, *J. Phys. C* 7, 3182 (1974).
13. H. Diamond, *J. Appl. Phys.* 32, 909 (1961).
14. K. Kinoshita and A. Yamaji, *J. Appl. Phys.* 47, 371 (1976).
15. G. Arlt, D. Hennings, and G. de With, *J. Appl. Phys.* 58, 1619 (1985).
16. K. Kanata, T. Yoshikawa, and K. Kubota, *Solid State Commun.* 62, 765 (1987).
17. K. Uchino, E. Sadanaga, and T. Hirose, *J. Am Ceram. Soc.* 72, 1555 (1989).
18. S. Schlag, H. F. Eicke, and W. B. Stern, *Ferroelectrics* 173, 351 (1995).
19. A. I. Frenkel, M. H. Frey, and D. A. Payne, *J. Synchrotron Radiation* 6, 515 (1999).
20. M. H. Frey and D. A. Payne, *Phys. Rev. B* 54, 3158 (1996).
21. R. Bottcher, C. Klimm, D. Michel, H. C. Semmelhack, G. Volkel, H. J. Glasel, and E. Hartmann, *Phys. Rev. B* 62, 2085 (2000).
22. R. Bachmann and K. Вдrner, *Solid State Commun.* 68, 865 (1988).
23. C. A. Randall, D. E. McCauley, and D. P. Cann, *Ferroelectrics* 206, 325 (1998).
24. K. Ishikawa, K. Yoshikawa, and N. Okada, *Phys. Rev. B* 37, 5852 (1988).
25. G. Burns and B. A. Scott, *Phys. Rev. Lett.* 25, 167 (1970); *Phys. Rev. B* 7, 3088 (1973).
26. W. L. Zhong, B. Jiang, P. L. Zhang, J. M. Ma, H. M. Cheng, Z. H. Yang, and L. X. Li, *J. Phys.: Cond. Mat.* 5, 2619 (1993).
27. B. Jiang, J. L. Peng, L. A. Bursill, and W. L. Zhong, *J. Appl. Phys.* 87, 3462 (2000).
28. S. Chattopadhyay, P. Ayyub, V. R. Palkar, and M. S. Multani, *Phys. Rev. B* 52, 13177 (1995).
29. S. Chattopadhyay, P. Ayyub, V. R. Palkar, A. V. Gurjar, R. M. Wankar, and M. S. Multani, *J. Phys.: Cond. Mat.* 9, 8135 (1997).
30. K. Binder, *Ferroelectrics* 73, 43 (1987).
31. S. K. Mishra and D. Pandey, *J. Phys.: Cond. Mat.* 7, 9287 (1995).
32. C. Kittel, *Phys. Rev.* 70, 965 (1946).
33. W. Cao and C. A. Randall, *J. Phys. Chem. Solids* 57, 1499 (1996).
34. C. A. Randall, N. Kim, J. P. Kucera, W. Cao, and T. R. Shrout, *J. Am. Ceram. Soc.* 81, 677 (1998).
35. M. R. Srinivasan, P. Ayyub, M. S. Multani, V. R. Palkar, and R. Vijayaraghavan, *Phys. Lett. A* 101, 435 (1984).
36. K. Okazaki, H. Igarashi, K. Nagata, and A. Hasegaka, *Ferroelectrics* 7, 153 (1974).
37. L. Zhang, W.-L. Zhong, C.-L. Wang, P.-L. Zhang, and Y.-G. Wang, *J. Phys. D* 32, 546 (1999).

38. J. Meng, G. Zou, J. Li, Q. Cui, X. Wang, Z. Wang, and M. Zhao, *Solid State Commun.* 90, 643 (1994).
39. Y. Park, K. M. Knowles, and K. Cho, *J. Appl. Phys.* 83, 5702 (1998).
40. A. J. Burggraaf and K. Keizer, *Matter. Res. Bull.* 10, 521 (1975).
41. P. C. Osbond, R. W. Whatmore, V. A. Griffiths, and F. W. Ainger, *Ferroelectrics* 54, 203 (1984).
42. K. Keizer and A. J. Burggraaf, *Phys. Status Solidi A* 26, 561 (1974).
43. Q. F. Zhou, H. L. W. Chan, Q. Q. Zhang, and C. L. Choy, *J. Appl. Phys.* 89, 8121 (2001).
44. X. H. Wang, T. L. Ren, Z. L. Gui, and L. T. Li, *Ferroelectrics* 262, 1219 (2001).
45. Z. X. Yue, X. L. Wang, L. Y. Zhang, and X. Yao, *Acta Phys. Sinica* 6, 913 (1995).
46. J. F. Meng, R. S. Katiyar, and G. T. Zou, *J. Phys. Chem. Solids* 59, 1161 (1998).
47. L. Zhang, W.-L. Zhong, C.-L. Wang, Y.-P. Peng, and Y.-G. Wang, *Europhys. J. B* 11, 565 (1999).
48. S. G. Lu, C. L. Mak, and K. H. Wong, *J. Am. Ceram. Soc.* 84, 79 (2001).
49. V. V. Shvartsman, A. Yu. Emelyanov, A. L. Kholkin, and A. Safari, *Appl. Phys. Lett.* 81, 117 (2002).
50. P. Marquardt and H. Gleiter, *Phys. Rev. Lett.* 48, 1423 (1982).
51. S. V. Pan’kova, V. V. Poborchii, and V. G. Solov’ev, *J. Phys.: Cond. Mat.* 8, L203 (1996).
52. A. V. Fokin, Yu. A. Kumzerov, N. M. Okuneva, A. A. Naberezhnov, S. B. Vakhrushev, I. V. Golosovsky, and A. I. Kurbakov, *Phys. Rev. Lett.* 89, 175503 (2002).
53. M. J. Westphal, *J. Appl. Phys.* 74, 6107 (1993).
54. P. Ayyub, V. R. Palkar, S. Chattopadhyay, and M. S. Multani, *Phys. Rev. B* 51, 6135 (1995).
55. P. Ayyub, M. S. Multani, M. Barma, V. R. Palkar, and R. Vijayaraghavan, *J. Phys. C* 21, 2229 (1988).
56. M. Ya. Gamarnik, *Phys. Status Solidi* 178, 59 (1993).
57. P. Ayyub, in “Frontiers in Materials Modelling and Design” (V. Kumar, S. Sengupta, and B. Raj, Eds.), p. 228. Springer-Verlag, Heidelberg, 1997.
58. V. R. Palkar, P. Ayyub, S. Chattopadhyay, and M. S. Multani, *Phys. Rev. B* 53, 2167 (1996).
59. W. L. Zhong, Y. G. Wang, P. L. Zhang, and B. D. Qu, *Phys. Rev. B* 50, 698 (1994).
60. I. Rychetskэ and O. Hudбk, *J. Phys.: Cond. Mat.* 9, 4955 (1997).
61. W. Y. Shih, W.-H. Shih, and I. A. Aksay, *Phys. Rev. B* 50, 15575 (1994).
62. B. Jiang and L. A. Bursill, *Phys. Rev. B* 60, 9978 (1999).
63. M. G. Cottam, D. R. Tilley, and B. Zeks, *J. Phys. C* 17, 1793 (1984).
64. C. L. Wang, Y. Xin, X. S. Wang, and W. L. Zhong, *Phys. Rev. B* 62, 11432 (2000).
65. K. Sheshadri, R. Lahiri, P. Ayyub, and S. Bhattacharya, *J. Phys.: Cond. Mat.* 11, 2459 (1999).
66. P. Ayyub, S. Chattopadhyay, K. Sheshadri, and R. Lahiri, *Nanostruct. Mater.* 12, 713 (1999).
67. G. Smolenskii, *J. Phys. Soc. Jpn. Suppl.* 28, 26 (1970).
68. K. Uchino and S. Nomura, *Ferroelectrics Lett.* 44, 55 (1982).
69. L. E. Cross, *Ferroelectrics* 76, 241 (1987).
70. V. S. Tiwari and D. Pandey, *J. Am. Ceram. Soc.* 77, 1819 (1994).
71. S. Li, J. A. Eastman, R. E. Newnham, and L. E. Cross, *Phys. Rev. B* 55, 12067 (1997).
72. J. F Scott, *Ferroelectrics Rev.* 1, 1 (1998).
73. T. M. Shaw, S. Trolier-McKinstry, and P. C. McIntyre, *Annu. Rev. Mater. Sci.* 30, 265 (2000).
74. M. H. Francombe, “Physics of Thin Films: Mechanic and Dielectric Properties.” Academic, San Diego, 1993.
75. S. Trolier-McKinstry, P. Chindaudom, K. Vedam, and R. E. Newnham, in “Optical Characterization of Real Surfaces and Films” (M. H. Francombe and J. L. Vossen, Eds.). Academic, New York, 1994.
76. S. Ducharme, V. M. Fridkin, A. V. Bune, S. P. Palto, L. M. Blinov, N. N. Petukhova, and S. G. Yudin, *Phys. Rev. Lett.* 84, 175 (2000).

77. J. F. Scott, *Phase Transitions* 30, 107 (1991).
78. B. Qu, W. Zhong, K. Wang, and Z. Wang, *Integ. Ferroelectrics* 3, 7 (1993).
79. S. B. Ren, C. J. Lu, H. M. Shen, and Y. N. Wang, *Phys. Rev. B* 55, 3485 (1997).
80. V. P. Dudkevich, V. A. Bukreev, Vl. M. Mukhortov, Yu. I. Golovko, Yu. G. Sindeev, V. M. Mukhortov, and E. G. Fesenko, *Phys. Status Solidi A* 65, 463 (1981).
81. I. P. Batra, P. Wurfel, and B. D. Silverman, *Phys. Rev. Lett.* 30, 384 (1973).
82. P. Ayyub, S. Chattopadhyay, R. Pinto, and M. S. Multani, *Phys. Rev. B* 57, R5559 (1998).
83. J. Mendiola and B. Jimenez, *Ferroelectrics* 53, 159 (1984).
84. R. E. Newnham, D. P. Skinner, and L. E. Cross, *Mater. Res. Bull.* 13, 525 (1978).
85. E. Venkatragavaraj, B. Satish, P. R. Vinod, and M. S. Vijaya, *J. Phys. D* 34, 487 (2001).
86. M. C. Cheung, H. L. W. Chan, Q. F. Zhou, and C. L. Choy, *Nanostruct. Mater.* 11, 837 (1999).
87. Q. F. Zhou, H. L. W. Chan, and C. L. Choy, *Thin Solid Films* 375, 95 (2000).
88. C. K. Wong, Y. M. Poon, and F. G. Shin, *J. Appl. Phys.* 90, 4690 (2001).
89. K. Rittenmyer, T. Shrout, W. A. Schulze, and R. E. Newnham, *Ferroelectrics* 41, 189 (1982).
90. A. Safari, R. E. Newnham, L. E. Cross, and W. A. Schulze, *Ferroelectrics* 41, 197 (1982).
91. J. E. Smay, J. Cesarano III, B. A. Tuttle, and J. A. Lewis, *J. Appl. Phys.* 92, 6119 (2002).
92. C. K. Chiang and R. Popielarz, *Ferroelectrics* 275, 1 (2002).
93. Y. Bai, Z. Y. Cheng, V. Bharti, H. S. Xu, and Q. M. Zhang, *Appl. Phys. Lett.* 76, 3804 (2000).
94. Z.-M. Dang, Y. Shen, and C.-W. Nan, *Appl. Phys. Lett.* 81, 4814 (2002).
95. C. Wang, Q. F. Fang, and Z. G. Zhu, *Appl. Phys. Lett.* 80, 3578 (2002).
96. N. Duan, J. E. ten Elshof, H. Verweij, G. Greuel, and O. Dannapple, *Appl. Phys. Lett.* 77, 3263 (2000).
97. T. K. Kundu, M. Mukherjee, D. Chakravorty, and L. E. Cross, *J. Appl. Phys.* 83, 4380 (1998).
98. D. McCauley, R. E. Newnham, and C. A. Randall, *J. Am. Ceram. Soc.* 81, 979 (1998).
99. A. S. Bhalla, R. E. Newnham, L. E. Cross, W. A. Schulze, J. P. Dougherty, and W. A. Smith, *Ferroelectrics* 33, 139 (1981).
100. Y. Wang, W. Zhong, and P. Zhang, *J. Appl. Phys.* 74, 521 (1993).
101. B. Ploss, F. G. Shin, H. L. W. Chan, and C. L. Choy, *IEEE Trans. Dielectric Electrical Insulation* 7, 517 (2000).
102. G. S. Murugan, K. B. R. Varma, Y. Takahashi, and T. Komatsu, *Appl. Phys. Lett.* 78, 4019 (2001).
103. A. A. Berezhnoi, M. I. Vasilev, A. O. Volchek, A. V. Dotsenko, and V. A. Tsekhomskii, *Opt. Spectrosc.* 92, 230 (2002).
104. C. W. Nan, M. Li, and J. H. Huang, *Phys. Rev. B* 63, 144415 (2001).
105. C. W. Nan, L. Liu, N. Cai, J. Zhai, Y. Ye, Y. H. Lin, L. J. Dong, and C. X. Xiong, *Appl. Phys. Lett.* 81, 3831 (2002).
106. V. R. Palkar and S. C. Purandare, *Ferroelectrics Rev.* 2, 169 (2000).
107. L. L. Hench and J. K. West, *Chem. Rev.* 90, 33 (1990).
108. R. C. Mehrotra, “Structure and Bonding,” Vol. 77. Springer-Verlag, Berlin, 1992.
109. Y. Ozaki, *Ferroelectrics* 49, 285 (1983).
110. K. H. Jo and K. H. Yoon, *Matter. Res. Bull.* 24, 1 (1989).
111. T. Fukui, C. Sakurai, and M. Okuyama, *J. Mater. Res.* 7, 791 (1992).
112. H. S. Potdar, P. Singh, S. B. Deshpande, P. D. Godbole, and S. K. Date, *Mater. Lett.* 7, 791 (1992).
113. V. Pillai, P. Kumar, M. J. Hou, P. Ayyub, and D. O. Shah, *Adv. Colloid Interface Sci.* 55, 241 (1995).
114. J. Fendler and I. Dekany, “Nanoparticles in Solids and Solutions.” Kluwer, Dordrecht, 1996.

115. H. Herrig and R. Hemplemann, *Mater. Lett.* 27, 287 (1996).
116. C.-H. Lu and Y.-P. Wu, *Mater. Lett.* 27, 13 (1996).
117. A. T. Chien, J. S. Speck, F. F. Lange, A. C. Daykin, and C. G. Levi, *J. Mater. Res.* 10, 1784 (1995).
118. J. Y. Choi, C. H. Kim, and D. K. Kim, *J. Am. Ceram. Soc.* 81, 1353 (1998).
119. S. Komarneni, Q. Li, K. M. Stefansson, and R. Roy, *J. Mater. Res.* 8, 3176 (1993).
120. O. Auciello, C. M. Foster, and R. Ramesh, *Annu. Rev. Mater. Sci.* 28, 501 (1998).
121. P. Ayyub, R. Chandra, P. Taneja, A. K. Sharma, and R. Pinto, *Appl. Phys. A* 73, 67 (2001).
122. H. Sugiyama, K. Kodama, T. Nakaiso, M. Noda, and M. Okuyama, *Integr. Ferroelectrics* 34, 1521 (2001).
123. T. Nakaiso, H. Sugiyama, M. Noda, and M. Okuyama, *Jpn. J. Appl. Phys. Pt.1* 39, 5517 (2000).
124. R. Klarmann, J. Hemberger, W. Biegel, and B. Stritzker, *Ferroelectrics* 241, 1777 (2000).
125. F. Craciun, P. Verardi, M. Dinescu, and G. Guidarelli, *Thin Solid Films* 344, 90 (1999).
126. J. F. M. Cillessen, M. W. J. Prins, and R. M. Wolf, *J. Appl. Phys.* 81, 2777 (1997).
127. Z. H. Chen, D. F. Cui, J. Zhang, H. B. Lu, Y. L. Zhou, L. Li, and G. Z. Yang, *Ferroelectrics* 196, 343 (1997).
128. Q. X. Jia, A. T. Findikoglu, R. Zhou, S. R. Foltyn, and X. D. Wu, *Integr. Ferroelectrics* 14, 167 (1997).
129. V. R. Palkar, S. C. Purandare, S. P. Pai, S. Chattopadhyay, P. R. Apte, R. Pinto, and M. S. Multani, *Appl. Phys. Lett.* 68, 1582 (1996).
130. C. M. Foster, Z. Li, M. Buckett, D. Miller, P. M. Baldo, L. E. Rehn, G. R. Bai, D. Guo, H. You, and K. L. Merkle, *J. Appl. Phys.* 78, 2607 (1995).
131. A. C. Jones and P. R. Chalker, *J. Phys. D* 36, R80 (2003).
132. T. Tohma, H. Masumoto, and T. Goto, *Mater. Trans.* 43, 2880 (2002).
133. T. J. Leedham, A. C. Jones, H. O. Davies, N. L. Tobin, P. A. Williams, J. F. Bickley, and A. Steiner, *Integr. Ferroelectrics* 36, 111 (2001).
134. K. Sreenivas, M. Sayer, and P. Garrett, *Thin Solid Films* 172, 251 (1989).
135. D. Remiens, E. Cattan, C. Soyer, and T. Haccart, *Mater. Sci. Semicond. Proc.* 5, 123 (2002).
136. S. Kikuchi, Y. Miyaguchi, T. Jimbo, I. Kimura, M. Tanimura, K. Suu, and M. Ishikawa, *Integr. Ferroelectrics* 46, 105 (2002).
137. Q. Zhao, Z. X. Fan, Z. S. Tang, X. J. Meng, J. L. Song, G. S. Wang, and J. H. Chu, *Surface Coatings Tech.* 160, 173 (2002).
138. Y. Nishioka and H. Ishiwara, *Integr. Ferroelectrics* 46, 265 (2002).
139. S. F. Karmanenko, A. I. Dedyk, N. N. Isakov, Y. J. Oh, V. I. Sakharov, and I. T. Serenkov, *Integr. Ferroelectrics* 46, 265 (2002).
140. S. B. Krupanidhi, H. Hu, and V. Kumar, *J. Appl. Phys.* 71, 376 (1992).
141. S. P. Beeby, A. Blackburn, N. M. White: Processing of PZT piezoelectric thick films on silicon for microelectromechanical systems, J. Micromech. Microeng. **9** (1999) 218–229.
142. “Nanoelectronics and Information Technology, Advanced Electronic Materials and Novel Devices” (R. Waser Ed.). Wiley-VCH Verlag GmbH, Weinheim, 2003.
143. D. Sarid, “Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces,” Oxford Series in Optical and Imaging Sciences. University Press, Oxford, 1991.
144. R. Wiesendanger, “Scanning Probe Microscopy and Spectroscopy: Methods and Applications.” Cambridge University Press, Cambridge, UK, 1994.
145. G. Binnig, C. F. Quate, and Ch. Gerber, *Phys. Rev. Lett.* 56, 930 (1986).
146. A. Gruverman, O. Auciello, and H. Tokumoto, *Annu. Rev. Mater. Sci.* 28, 101 (1998).
147. Y. Martin, C. C. Williams, and H. K. Wickramasinghe, *J. Appl.Phys.* 61, 4723 (1987).
148. J. E. Stern, B. D. Terris, H. J. Mamin, and D. Rugar, *Appl. Phys. Lett.* 53, 2717 (1988).
149. Y. Martin, D.W. Abraham, and H. K.Wickramasinghe, *Appl. Phys. Lett.* 52, 1103 (1988).

150. M. Nonnenmacher, M. P. O’Boyle, and H. K. Wickramasinghe, *Appl. Phys. Lett.* 58, 2921 (1991).
151. R. C. Barrett and C. F. Quate, *J. Appl. Phys.* 70, 2725 (1991).
152. P. Lehnen, J. Dec, and W. Kleemann, *J. Phys. D: Appl. Phys.* 33, 1932 (2000).
153. S. Buhlmann, B. Dwir, J. Baborowski, and P. Muralt, *Integr. Ferroelectrics* 50, 261 (2002).
154. M. J. Higgins, A. Krishnan, M. M. J. Treacy, and S. Bhattacharya, *Appl. Phys. Lett.* 80, 3373 (2002).
155. A. Gruverman, K. Hironaka, Y. Ikeda, K. M. Satyalakshmi, A. Pignolet, M. Alexe, N. D. Zakharov, and D. Hesse, *Integr. Ferroelectrics* 27, 1203 (1999).
156. W. S. Yun, J. J. Urban, Q. Gu, and H. Park, *Nano Lett.* 2, 589 (2002).
157. O. Auciello, J. F. Scott, and R. Ramesh, *Phys. Today* 51, 22 (1998).
158. L. M. Eng, M. Friedrich, J. Fousek, and P. Gunter, *J. Vac. Sci. Technol., B* 14, 1191 (1996).
159. E. Z. Luo, Z. Xie, J. B. Xu, and I. H. Wilson, *Phys. Rev. B* 61, 203 (2000).
160. S. V. Kalinin and D. A. Bonnell, *Phys. Rev. B* 63, 125411 (2001).
161. S. V. Kalinin and D. A. Bonnell, *Appl. Phys. Lett.* 78, 1116 (2001).
162. A. Correia, J. Massanell, N. Garcia, A. P. Levanyuk, A. Zlatkin, and J. Przeslawski, *Appl. Phys. Lett.* 68, 2796 (1996).
163. M. Takashige, S. Hamazaki, Y. Takahashi, et al., *Ferroelectrics* 240, 1359 (2000).
164. V. Likodimos, M. Labardi, and M. Allegrini, *Phys. Rev. B* 66, 024104 (2002).
165. K. Franke, J. Besold, W. Haessle, and C. Seegebarth, *Surf. Sci. Lett.* 302, L283 (1994).
166. T. Hidaka, T. Maruyama, I. Sakai, M. Saitoh, L. A. Wills, R. Hiskes, S. A. Dicarolis, and J. Amano, *Integrated Ferroelectrics* 17, 319 (1997).
167. A. Gruverman, H. Tokumoto, S. A. Prakash, S. Aggarwal, B. Yang, M. Wuttig, R. Ramesh, O. Auciello, and V. Venkatesan, *Appl. Phys. Lett.* 71, 3492 (1997).
168. L. M. Eng, H.-J. Guntherodt, G. A. Schneider, U. Kopke, and J. M. Saldana, *Appl. Phys. Lett.* 74, 233 (1999).
169. A. Roelofs, U. Bцttger, R. Waser, F. Schlaphof, S. Trogisch, and L. M. Eng, *Appl. Phys. Lett.* 77, 3444 (2000).
170. A. Gruverman, O. Auciello, and H. Tokumoto, *Annu. Rev. Mater. Sci.* 28, 101 (1998).
171. A. Gruverman, O. Auciello, and H. Tokumoto, *Integrated Ferroelectrics* 19, 49 (1998).
172. S. V. Kalinin and D. A. Bonnell, *Phys. Rev. B* 65, 125408 (2002).
173. C. S. Ganpule, A. L. Roytburd, V. Nagarajan, B. K. Hill, S. B. Ogale, E. D. Williams, R. Ramesh, and J. F. Scott, *Phys. Rev. B* 65, 014101 (2002).
174. O. Auciello, A. Gruverman, H. Tokumoto, S. A. Prakash, S. Aggarwal, and R. Ramesh, *MRS Bull.* 23, 33 (1998).
175. C. Harnagea, A. Pignolet, M. Alexe, and D. Hesse, *Integrated Ferroelectrics* 38, 667 (2001).
176. M. Abplanalp, L. M. Eng, and P. Gunter, *Appl. Phys. A* 66, Part 1, Suppl., S231 (1998).
177. P. Paruch, T. Tybell, and J. M. Triscone, Appl. Phys. Lett. 79, 530 (2001).
178. T. Tybell, C. H. Ahn, and J. M. Triscone, *Appl. Phys. Lett.* 75, 856 (1999).
179. J. Woo, S. Hong, D.-K. Min, H. Shin, and K. No, *Appl. Phys. Lett.* 80, 4000 (2002).
180. S. Hong, J. Woo, H. Shin, J. U. Jeon, Y. E. Pak, E. L. Colla, N. Setter, E. Kim, and K. No, *J. Appl. Phys.* 89, 1377 (2001).
181. J. Hong, H. W. Song, S. Hong, H. Shin, and K. No, *J. Appl. Phys.* 92, 7434 (2002).
182. H. Y. Guo, J. B. Xu, I. H. Wilson, Z. Xie, E. Z. Luo, S. Hong, and H. Yan, *Appl. Phys. Lett.* 81, 715 (2002).
183. J. A. Christman, S.-H. Kim, H. Maiwa, J.-P. Maria, B. J. Rodriguez, A. I. Kingon, and R. J. Nemanich, *J. Appl. Phys.* 87, 8031 (2000).
184. M. Abplanalp, J. Fousek, and P. Gunter, *Phys. Rev. Lett.* 86, 5799 (2001).
185. C. Harnagea, A. Pignolet, M. Alexe, D. Hesse, and U. Gцsele, *Appl. Phys. A* 70, 261 (2000).

186. A. Pignolet, C. Schafer, K. M. Satyalakshmi, C. Harnagea, D. Hesse, and U. Gцsele, *Appl. Phys. A* 70, 283 (2000).
187. C. Harnagea, A. Pignolet, M. Alexe, K. M. Satyalakshmi, D. Hesse, and U. Gцsele, *Jpn. J. Appl. Phys.* 38, Part 2, L1255 (1999).
188. S. Hong and N. Setter, *Appl. Phys. Lett.* 81, 3437 (2002).
189. S. Hong, E. L. Colla, E. Kim, D. V. Taylor, A. K. Tagantsev, P. Muralt, K. No, and N. Setter, *J. Appl. Phys.* 86, 607 (1999).
190. M. Alexe, C. Harnagea, D. Hesse, and U. Goesele, *Appl. Phys. Lett.* 75, 1793 (1999).
191. M. Alexe, A. Gruverman, C. Harnagea, N. D. Zakharov, A. Pignolet, D. Hesse, and J. F. Scott, *Appl. Phys. Lett.* 75, 1158 (1999).
192. C. S. Ganpule, A. Stanishevsky, S. Aggarwal, J. Melngailis, E. Williams, R. Ramesh, V. Joshi, and C. A. Paz de Araujo, *Appl. Phys. Lett.* 75, 3874 (1999).
193. O. Auciello, A. Gruverman, and H. Tokumoto, *Integrated Ferroelectrics* 15, 107 (1997).
194. J. A. Christman, R. Woolcott, Jr., A. Kingon, and R. J. Nemanich, *Appl. Phys. Lett.* 73, 3851.
195. S. Hong, H. Shin, J. Woo, and K. No, *Appl. Phys. Lett.* 80, 1453 (2002).
196. S. Tiedke, T. Schmitz, K. Prume, A. Roelofs, T. Schneller, U. Kall, R. Waser, C. S. Ganpule, V. Nagarajan, A. Stanishevsky, and R. Ramesh, *Appl. Phys. Lett.* 79, 3678 (2001).
197. Y. Cho, K. Fujimoto, Y. Hiranaga, Y. Wagatsuma, A. Onoe, K. Terabe, and K. Kitamura, *Appl. Phys. Lett.* 81, 4401 (2002).
198. A. Gruverman, C. Isobe, and M. Tanaka, *Mater. Res. Soc. Symp. Proc.* 655, CC8.5 (2001).
199. E. Fatuzzo and W. J. Merz, "Ferroelectricity." North-Holland, Amsterdam, 1967.
200. O. Lohse, S. Tiedke, M. Grossmann, and R. Waser, *Integrated Ferroelectrics* 22, 123 (1998).
201. T. Tybell, P. Paruch, T. Giamarchi, and J.-M. Triscone, *Phys. Rev. Lett.* 89, 097601 (2002).
202. E. J. Mele, *Am. J. Phys.* 69, 557 (2001).
203. S. V. Kalinin, D. A. Bonnell, T. Alvarez, X. Lei, Z. Hu, J. H. Ferris, Q. Zhang, and S. Dunn, *NanoLetters* 2, 589 (2002).
204. V. M. Fridkin, "Photoferroelectrics." Springer-Verlag, Berlin, 1979.
205. A. Gruverman, O. Auciello, and H. Tokumoto, *J. Vac. Sci. Technol., B* 14, 602 (1996).
206. G. Rosenman, P. Urenski, A. Agronin, Y. Rosenwaks, and M. Molotski, *Appl. Phys. Lett.* 82, 103 (2003).
207. K. Terabe, S. Takekawa, M. Nakamura, K. Kitamura, S. Higuchi, Y. Gotoh, and A. Gruverman, *Appl. Phys. Lett.* 81, 2044 (2002); K. Terabe, M. Nakamura, S. Takekawa, K. Kitamura, S. Higuchi, Y. Gotoh, and Y. Cho, *Appl. Phys. Lett.* 82, 433 (2003).
208. A. Gruverman, J. Hatano, and H. Tokumoto, Jpn. J. Appl. Phys. 36, Part 1, 2207 (1997).
209. A. Gruverman, *Appl. Phys. Lett.* 75, 1452 (1999).
210. S. K. Streiffer, J. A. Eastman, D. D. Fong, C. Thompson, A. Munkholm, M. V. Ramana Murty, O. Auciello, G. R. Bai, and G. B. Stephenson, *Phys. Rev. Lett.* 89, 67601 (2002).
211. C. F. Quate, *Surf. Sci.* 386, 259 (1997).
212. P. Vettiger, J. Brugger, M. Despont, U. Drechsler, U. Durig, W. Haberle, M. Lutwyche, H. Rothuizen, R. Stutz, R. Widmer, and G. Binnig, *Microelectron. Eng.* 46, 11 (1999).
213. J. F. Scott, in "Ferroelectric Memories." Springer-Verlag, New York, 2000.
214. W. S. Trimmer, "Micromechanics and MEMS: Classic and Seminal Papers to 1990." IEEE Press, Piscataway, NJ, 1997.
215. U. Chon, H. M. Jang, M. G. Kim, and C. H. Chang, *Phys. Rev. Lett.* 89, 087601 (2002).
216. M. Alexe, C. Harnagea, A. Visinoiu, A. Pignolet, D. Hesse, and U. Gцsele, *Scripta Mater.* 44, 1175 (2001).
217. H. Uchida, N. Soyama, K. Kageyama, K. Ogi, and C. A. Pax de Araujo, *Integ. Ferroelectrics* 16, 41 (1997).
218. K. Ammanuma and T. Kunio, *Jpn. J. Appl. Phys. Part 1* 35, 1400 (1996).

219. C. S. Ganpule, A. Stanishevsky, Q. Su, S. Aggarwal, J. Melngailis, E. Williams, and R. Ramesh, *Appl. Phys. Lett.* 75, 409 (1999).
220. M. Alexe, C. Harnagea, D. Hesse, and U. Gцsele, *Appl. Phys. Lett.* 75, 1793 (1999).
221. J. F. Scott, *Annu. Rev. Mater. Sci.* 28, 79 (1998).
222. V. R. Palkar, S. C. Purandare, and R. Pinto, *J. Appl. Phys.* 32, R1 (1999).
223. H. Ishiwara, *J. Semicond. Technol. Sci.* 1, 1 (2001).
224. T. P. Ma and J. P. Han, *IEEE Electron Device Lett.* 23, 386 (2002).

Ch12

1. Kumar V ed. (2007) Nanosilicon, Amsterdam, Elsevier Ltd.
2. Geszke-Moritz M, Moritz M (2013) Quantum dots as versatile probes in medical sciences: Synthesis, modification and properties, Materials Science and Engineering C 33: 1008–1021
3. Fan J and Chu PK (2010) Group IV Nanoparticles: Synthesis, Properties, and Biological Applications, Small. 6: 2080–2098
4. FRONTIERS OF NANOSCIENCE, Series Editor: Richard E. Palmer, Vol. 3 Metal Nanoparticles and Nanoalloys ed. Roy L. Johnston and J.P. Wilcoxon, Elsevier , 2012
5. A L. Rogach (Ed.) Semiconductor Nanocrystal Quantum Dots Synthesis, Assembly, Spectroscopy and Applications, Springer-Verlag/Wien, 2008
6. Sustainable Preparation of Metal Nanoparticles Methods and Applications, Ed. Rafael Luque Rajender S Varma, The Royal Society of Chemistry, 2013, Cambridge CB4 0WF, UK
7. Semiconductor and Metal Nanocrystals, Synthesis and Electronic and Optical properties, ed. V.I.Klimov Copyright 2004 by Marcel Dekker, Inc. NY, USA
8. Markus Niederberger, Nicola Pinna, Metal Oxide Nanoparticles in Organic Solvents Synthesis, Formation, Assembly and Application, Springer-Verlag, London Limited, 2009
9. Metal nanoparticles. Synthesis, Characterisation and Applications, Daniel L. Feldheim Colby A. Foss, Marcel Dekker, Inc. Madison Avenue, New York, 2002
10. Hossam Haick, Introduction to Nanotechnology, Technion, Israel Institute of Technology, 2009.

Ch13

1. Kumar V ed. (2007) Nanosilicon, Amsterdam, Elsevier Ltd.
2. Geszke-Moritz M, Moritz M (2013) Quantum dots as versatile probes in medical sciences: Synthesis, modification and properties, Materials Science and Engineering C 33: 1008–1021
3. Michalet X, Pinaud FF, Bentolila LA, Tsay JM, Doose S, Li JJ, Sundaresan G, Wu AM, Gambhir SS, Weiss S (2005) Quantum Dots for Live Cells, in Vivo Imaging, and Diagnostics, Science. 307: 538-544
4. Fan J and Chu PK (2010) Group IV Nanoparticles: Synthesis, Properties, and Biological Applications, Small. 6: 2080–2098
5. Alivisatos AP (2004) The use of nanocrystals in biological detection. Nat.Biotechnol. 22: 47–52
6. Parak WJ, Pellegrino T, Plank C (2005) Labelling of cells with quantum dots, Nanotechnology. 16: R9–R25
7. Bonacina L (2013) Nonlinear Nanomedecine: Harmonic Nanoparticles toward Targeted Diagnosis and Therapy, Mol. Pharmaceutics. 10: 783–792
8. Bluet J-M, Botsoa J, Zakharko Y, Géloën A, Alekseev S, Marty O, Mognetti B, Patskovsky S, Rioux D, Lysenko V (2012) SiC as a Biocompatible Marker for Cell Labeling, Ch.11, In: Silicon Carbide Biotechnology, Ed. Stephen Saddow, Elsevier Inc.: 327-429
9. Kang Z, Liu Y and Lee S-T (2011) Small-sized silicon nanoparticles: new nanolights and nanocatalysts, Nanoscale. 3: 777-791

10. Dabbousi BO, Rodriguez-Viejo J, Mikulec FV, Heine JR, Mattoussi H, Ober R, Jensen KF, Bawendi MG (1997) CdSe-ZnS core-shell quantum dots: synthesis and characterization of a size series of highly luminescent nanocrystallites, J Phys Chem B. 101: 9463–9475

11. Nozik AJ. Spectroscopy and hot electron relaxation dynamics in semiconductor quantum wells and quantum dots. Annu Rev Phys Chem 2001;52:193–231.

12. Giljohann DA, Mirkin CA. Drivers of biodiagnostic development. Nature 2009;462:461–4.

13 Hines MA, Guyot-Sionnest P (1996) Synthesis and characterization of strongly luminescing ZnS-capped CdSe nanocrystals, J Phys Chem. 100: 468–471

14. Bruchez Jr M, Moronne M, Gin P, Weiss S, Alivisatos AP (1998) Semiconductor nanocrystals as fluorescent biological labels, Science. 281: 2013–2016

15. Derfus AM, Chan WCW (2004) Bhatia SN. Probing the cytotoxicity of semiconductor quantum dots, Nano Lett. 4: 11–18

16 Kirchner C, Liedl T, Kudera S, Pellegrino T, Munoz Javier A, Gaub HE, Stölzle S, Fertig N, Parak WJ (2005) Cytotoxicity of colloidal CdSe and CdSe/ZnS nanoparticles, Nano Lett. 5: 331–338

17. Fojtik A and Henglein A (1994) Luminescent colloidal silicon particles, Chem. Phys. Lett. 221 : 363-367

18. Erogbogbo F, Yong K-T, Roy I, Hu R, Law W-C, Zhao W, Ding H, Wu F, Kumar R, Swihart MT, and Prasad PN (2011) In Vivo Targeted Cancer Imaging, Sentinel Lymph Node Mapping and Multi-Channel Imaging with Biocompatible Silicon Nanocrystals, ACS NANO. 5: 413–423

19. Knipping J, Wiggers H, Rellinghaus B, Roth P, Konjhodzic D, and Meier C (2004) Synthesis of High Purity Silicon Nanoparticles in a Low Pressure Microwave Reactor, J. Nanosci. Nanotechn. 4: 1039-1044

20. Sankaran RM, Holunga D, Flagan RC, and Giapis KP (2005) Synthesis of Blue Luminescent Si Nanoparticles Using Atmospheric-Pressure Microdischarges, NANO LETTERS. 5: 537-541.

21. Rowsell BD and Veinot JGC (2005) Reductive thermolysis of a heterocyclic precursor: a convenient method for preparing luminescent, surfactant-stabilized silicon nanoparticles, Nanotechnology. 16: 732-736

22. Baldwin RK, Pettigrew KA, Ratai E, Augustine MP and Kauzlarich SM (2002) Solution reduction synthesis of surface stabilized silicon nanoparticles, Chem. Commun. 1822-1823.

23. Zou J, Baldwin RK, Pettigrew KA and Kauzlarich SM (2004) Solution Synthesis of Ultrastable Luminescent Siloxane-Coated Silicon Nanoparticles, Nano Lett. 4: 1181-1186

24. Rao S, Sutin J, Clegg R, Gratton E, Nayfeh MH, Habbal S, Tsolakidis A, and Martin RM (2004) Excited states of tetrahedral single-core Si_{29} nanoparticles, Phys. Rev. B . 69: art.205319

25. Tanaka A, Saito R, Kamikake T, Imamura M, and Yasuda H (2006) Electronic structures and optical properties of butyl-passivated Si nanoparticles, Solid State Commun. 140: 400-403

26. Lysenko V, Onyskevych V, Marty O, Skryshevsky VA, Chevolot Y, and Bru-Chevallier C (2008) Extraction of ultraviolet emitting silicon species from strongly hydrogenated nanoporous silicon, App Phys Let. 92: art.251910

27. Lysenko V, Bidault F, Alekseev S, Turpin C, Geobaldo F, Rivolo P, Garrone E, Zaitsev V, and Barbier D (2005) Study of Porous Silicon Nanostructures as Hydrogen Reservoirs, J. Phys. Chem. B . 109: 19711-19718

28. Nychyporuk T, Lysenko V, and Barbier D (2005) Fractal nature of porous silicon nanocrystallites, Phys. Rev. B. 71: art. 115402

29. Serdiuk T, Lysenko V, Alekseev S, Skryshevsky VA (2011) Size tuning of luminescent silicon nanoparticles with meso-porous silicon membranes, J Colloid and Interface Science. 364: 65–70

30. Serdiuk T, Skryshevsky VA, Ivanov II, Lysenko V (2011) Storage of luminescent nanoparticles in porous silicon: Toward a solid state “golden fleece”, Materials Letters. 65: 2514–2517

31. Belomoin G, Therrien J, Smith A, Rao S, Twesten R, Chaieb S, Nayfeh MH, Wagner L, Mitas L (2002) Observation of a magic discrete family of ultrabright Si nanoparticles, Appl. Phys. Lett. 80: art. 841
32. Patrone L, Nelson D, Safarov VI, Sentis M, Marine W (2000) Photoluminescence of silicon nanoclusters with reduced size dispersion produced by laser ablation, J. Appl. Phys. 87:art. 3829
33. Ledoux G, Guillois O, Porterat D, Reynaud C (2000) Photoluminescence properties of silicon nanocrystals as a function of their size, Phys.Rev.B. 62: 15942–15951
34. Feng ZC, Tsu R (Eds.) (1994) Porous Silicon, World Scientific, Singapore
35. Delerue C, Allan G, Lannoo M(1993) Theoretical aspects of the luminescence of porous silicon, Phys. Rev. B. 48: 11024-11036
36. Kang ZH, Tsang CHA, Zhang ZD, Zhang ML, Wong NB, Zapien JA, Shan YY and Lee ST (2007) A Polyoxometalate-Assisted Electrochemical Method for Silicon Nanostructures Preparation: From Quantum Dots to Nanowires, J.Am. Chem. Soc. 129: 5326-5327
37. Kang ZH, Tsang CHA, Wong NB, Zhang ZD and Lee ST (2007) Silicon Quantum Dots: A General Photocatalyst for Reduction, Decomposition, and Selective Oxidation Reactions, J. Am. Chem. Soc. 129: 12090-12091
38. Heinrich JL, Curtis CL, Credo GM, Kavanagh KL and Saylor MJ (1992) Luminescent Colloidal Silicon Suspensions from Porous Silicon, Science. 255: 66-68
39. Buriak JM, Allen MJ (1998) Lewis Acid Mediated Functionalization of Porous Silicon with Substituted Alkenes and Alkynes, J.Am.Chem.Soc. 120: 1339–1340
40. Manilov AI, Skryshevsky VA (2013) Hydrogen in porous silicon — A review, Materials Science and Engineering B. 178 :942–955
41. Anglin EJ, Cheng L, Freeman WR, Sailor MJ (2008) Porous silicon in drug delivery devices and materials, Advanced Drug Delivery Reviews. 60: 1266–1277
42. Boukherroub R, Petit A, Loupy A, Chazalviel JN, Ozanam F (2003) Microwave-Assisted Chemical Functionalization of Hydrogen-Terminated Porous Silicon Surfaces, J. Phys. Chem. B. 107: 13459–13462
43. Wu EC, Park J-H, Park J, Segal E, Cunin F, Sailor MJ (2008) Oxidation-Triggered Release of Fluorescent Molecules or Drugs from Mesoporous Si Microparticles, ACS Nano. 2: 2401–2409
44. Stewart MP, Buriak JM (1998) Photopatterned Hydrosilylation on Porous Silicon**,** Angewandte Chemie International Ed. 37: 3257–3260
45. Sam S, Chazalviel JN, Gouget-Laemmel AC, Ozanam F, Etcheberry A, Gabouze N (2011) Peptide immobilisation on porous silicon surface for metal ions detection, Nanoscale Research Letters. 6: art.412.
46. Hua F, Swihart MT, and Ruckenstein E (2005) Efficient Surface Grafting of Luminescent Silicon Quantum Dots by Photoinitiated Hydrosilylation, Langmuir. 21: 6054-6062
47. Ruizendaal L, Pujari SP, Gevaerts V, Paulusse JMJ, and Zuilhof H (2011) Biofunctional Silicon Nanoparticles by Means of Thiol-Ene Click Chemistry, Chem. Asian J. 6: 2776 – 2786
48. Sweryda-Krawiec B, Cassagneau T and Fendler J H (1999) Surface Modification of Silicon Nanocrystallites by Alcohols, J. Phys. Chem. B. 103: 9524-9529
49. Hallmann S, Fink MJ (2011) Mechanochemical synthesis of functionalized silicon nanoparticles with terminal chlorine groups, J. Mater. Res. 26: 1052-1060
50. Rosso-Vasic M, Spruijt E, Popovic Z, Overgaag K, Lagen B, Grandidier B, Vanmaekelbergh D, Dominguez-Gutierrez D, Cola L De and Zuilhof H (2009) Amine-terminated silicon nanoparticles: synthesis, optical properties and their use in bioimaging, J. Mater. Chem. 19: 5926–5933
51. Kang ZH, Liu Y, Tsang CHA, Ma DDD, Fan X, Wong NB, Lee ST (2009) Water-Soluble Silicon Quantum Dots with Wavelength-Tunable Photoluminescence, Adv. Mater. 21: 661-664

52. He Y, Su YY, Yang XB, Kang ZH, Xu TT, Zhang RQ, Fan C, Lee ST (2009) Photo and pH Stable, Highly-Luminescent Silicon Nanospheres and Their Bioconjugates for Immunofluorescent Cell Imaging, J. Am. Chem. Soc. 131: 4434-4438
53. He Y, Fan C, Lee S-T, (2010) Silicon nanostructures for bioapplications, Nano Today. 5: 282—295
54. Wang J, Liu Y, Peng F, Chen C, He Y, Ma H, Cao L, and Sun S (2012) A General Route to Efficient Functionalization of Silicon Quantum Dots for High-Performance Fluorescent Probes, Small. 8: 2430-2435
55. Gongalsky MB, Kharin AYu, Osminkina LA, Timoshenko VYu, Jeong J, Lee H and Chung BH (2012) Enhanced photoluminescence of porous silicon nanoparticles coated by bioresorbable polymers, Nanoscale Research Letters 7: art.446
56. Li Z F and Ruckenstein E (2004) Water-Soluble Poly(acrylic acid) Grafted Luminescent Silicon Nanoparticles and Their Use as Fluorescent Biological Staining Labels, NANO Letters. 4: art. 14631467
57. Zhang X, Neiner D, Wang S, Louie A Y and Kauzlarich S M (2007) A new solution route to hydrogen-terminated silicon nanoparticles: synthesis, functionalization and water stability, Nanotechnology. 18: art. 095601
58. Sudeep P K, Page Z and Emrick T (2008) PEGylated silicon nanoparticles: synthesis and characterization, Chem. Commun. 6126–6127
59. Wang K, Wang Q, Chao Y (2012) Simple fabrication of nanostructured silicon and photoluminescence, Appl Phys A. 109: 437–440
60. Ge J, Liu W, Zhao W, Zhang H, Zhuang X, Lan M, Wang P, Li H, Ran G, and Lee S-T (2011) Preparation of Highly Stable and Water-Dispersible Silicon Quantum Dots by Using an Organic Peroxide, Chem. Eur. J. 17: 12872 – 12876
61. Erogbogbo F, Chang C-W, May JL, Liu L, Kumar R, Law W-C, Ding H, Yong K T, Roy I, Sheshadri M, Swihart MT, Prasad PN (2012) Bioconjugation of Luminescent Silicon Quantum Dots to Gadolinium Ions for Bioimaging Applications, Nanoscale. 4: 5483-5489
62. May J L, Erogbogbo F, Yong K-T, Ding H, Law W-C, Swihart M T, Prasad P N (2012) Enhancing silicon quantum dot uptake by pancreatic cancer cells via pluronic® encapsulation and antibody targeting, Journal Solid Tumors. 2: 24-37
63. Fan J Y, Wu X L, Chu P K (2006) Low-dimensional SiC nanostructures: Fabrication, luminescence, and electrical properties, Prog. Mat. Sci., 51: 983 -1031
64. Matsumoto T, Takahashi J, Tamaki T, Futagi T, Mimura H, and Kanemitsu Y (1994) Blue-green luminescence from porous silicon carbide, Appl. Phys. Lett. 64: art. 226
65. Petrova-Koch V, Sreseli O, Polisski G, Kovalev D, Muschik T, and Koch F (1995) Luminescence enhancement by electrochemical etching of SiC(6H),Thin Solid Films. 255: 107-110
66. Rossi AM, Murphy ThE, and Reipa V (2008) Ultraviolet photoluminescence from 6H silicon carbide nanoparticles, Appl. Phys Lett. 92: art.253112
67. Botsoa J, Bluet J-M, Lysenko V, Marty O, Barbier D, and Guillot G (2007) Photoluminescence of 6H–SiC nanostructures fabricated by electrochemical etching, J. Appl. Phys. 102: art. 083526
68. Wu X L, Fan J Y, Qiu T, Yang X, Siu G G, and Chu P K (2005) Experimental Evidence for the Quantum Confinement Effect in 3*C*-SiC Nanocrystallites, Phys. Rev Lett. 94: art. 026102
69. Fan J Y, Wu X L, Li H X, Liu H W, Siu G G, Chu P K (2006) Luminescence from colloidal 3C-SiC nanocrystals in different solvents, Appl.Phys.Lett., 88: art. 041909
70. Zakharko Yu, Botsoa J, Alekseev S, Lysenko V, Bluet J-M, Marty O, Skryshevsky VA, and G Guillot (2010) Influence of the interfacial chemical environment on the luminescence of 3C-SiC nanoparticles, J.Appl.Phys. 107: art. 013503
71. Serdiuk T, Alekseev SA, Lysenko V, Skryshevsky VA and Géloën A (2012) Charge-driven selective localization of fluorescent nanoparticles in live cells, Nanotechnology. 23: art.315101

72. Alekseev S, Botsoa J, Zaitsev VN and Barbier D (2007) Fourier transform infrared spectroscopy and temperature programmed desorption mass spectrometry study of surface chemistry of porous 6H-SiC, Chem. Mater. 19: 2189-2194
73. Shin W, Seo W, Takai O, and Koumoto K (1998) Surface chemistry of porous silicon carbide, J. Electron. Mater. 27: 304-307
74. Beke D, Szekre'nyes Z, Balogh I, Veres M, Fazakas E, Varga LK, Kamara's K, Cziga'ny Z, and Gali A (2011) Characterization of luminescent silicon carbide nanocrystals prepared by reactive bonding and subsequent wet chemical etching, Appl. Phys. Lett. 99: art. 213108
75. Li Y, Chen C, Li J-T, Yang Y and Lin Z-M (2011) Surface charges and optical characteristic of colloidal cubic SiC nanocrystals, Nanoscale Res.Lett. 6: art. 454
76. Zhang N, Dai D, Zhang W, and Fan J (2012) Photoluminescence and light reabsorption in SiC quantum dots embedded in binary-polyelectrolyte solid matrix, J.Appl.Phys. 112: art. 094315
77. Che J, Wang X, Xiao Y, Wu X, Zhou L, Yuan W (2007) Effect of inorganic–organic composite coating on the dispersion of silicon carbide nanoparticles in non-aqueous medium, Nanotechnology, 18: art.135706
78. Iijima M, Kamiya H (2008) Surface Modification of Silicon Carbide Nanoparticles by Azo Radical Initiators, J. Phys. Chem. C. 112: 11786-11790
79. Pud AA, Noskov YV, Kassiba A, Fatyeyeva KY, Ogurtsov NA, Makowska-Janusik M, Bednarski W, Tabellout M, Shapoval G S (2007) New Aspects of the Low-Concentrated Aniline Polymerization in the Solution and in SiC Nanocrystals Dispersion, J. Phys. Chem. B. 111: 2174-2180
80. Serdiuk T, Lysenko V, Skryshevsky VA, Géloën A (2012) Vapor phase mediated cellular uptake of sub 5 nm nanoparticles, Nanoscale Research Letters, 7: art. 212
81. Serdiuk T, Lysenko V, Mognetti B, Skryshevsky V, and Géloën A (2013) Impact of cell division on intracellular uptake and nuclear targeting with fluorescent SiC-based nanoparticles, J. Biophotonics. 6: 291–297
82. Tagliazucchi M, Peleg O, Kroger M, Rabin Y, and Szleifer I (2013) Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex, PNAS. 110: 3363–3368
83. Weisburg JH, Weissman DB, Sedaghat T, Babich H (2004) In vitro cytotoxicity of epigallocat echin gallate and tea extracts to cancerous and normal cells from the human oral cavity, Basic Clin. Pharmacol. Toxicol. 95: 191–200
84. Zakharko Yu, Serdiuk T, Nychyporuk T, Geloen A, Lemiti M, and Lysenko V (2012) Plasmon-enhanced photoluminescence of SiC quantum dots for cell imaging applications, Plasmonics 7: 725-732.

Ch 14

1. H. W. Kroto, J. R. Heath, S. C. Obrien, R. F. Curl, and R. E. Smalley, Nature **318**, 162 (1985).
2. W. Kratschmer, L.D. Lamb, K. Foristopoulos, D.R. Huffman, Solid C_{60}: A new form of carbon. Nature 1990, **347**, 354.
3. S. Iijima, Helical microtubules of graphitic carbon, Nature, **354**, 56 (1991).
4. S. Iijima, T. Ichihashi, Single-shell carbon nanotubes of 1-nm diameter. Nature 1993, **363**, 603.
5. D.S. Bethune, C.H. Kiang, M.S. DeVries, G. Gorman, R. Savoy, R. Beyers, Cobalt-catalysed growth of carbon nanotubes with single-atomiclayer walls. Nature 1993, **363**, 605.
6. M.S. Dresselhaus, G. Dresselhaus, P.C. Eklund, Science of Fullerenes and Carbon Nanotubes; Academic Press: San Diego, 1996.
7. Carbon Nanotubes: Synthesis, Structure, Properties and Applications. M.S. Dresselhaus, G. Dresselhaus, Ph. Avouris, Eds.; Topics in Applied Physics, Springer: Berlin, 2001; vol. 80.
8. N. Hamada, S. Sawada, A. Oshiyama, New onedimensional conductors: Graphite microtubules. Phys. Rev. Lett. 1992, **68**, 1579.

9. A. Oberlin, M. Endo, and T. Koyama, Journal of Crystal Growth **32**, 335 (1976).
10. S. Iijima and T. Ichihashi, Nature **363**, 603 (1993).
11. D. S. Bethune, C. H. Kiang, M. S. Devries, G. Gorman, R. Savoy, J. Vazquez, and R. Beyers, Nature **363**, 605 (1993).
12. K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).
13. K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proceedings of the National Academy of Sciences of the United States of America **102**, 10451 (2005).
14. A. K. Geim and K. S. Novoselov, Nature Materials **6**, 183 (2007).
15. M. Laszlo, Fullerenes and Carbon Nanotubes in *Dekker Encyclopedia of Nanoscience and Nanotechnology*, 1203-1211, 2004.
16. C. Piskoti, J. Yager, A. Zettl, C36: A new carbon solid. Nature 1998, 393, 771.
17. A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y.H. Lee, S.G. Kim, D.T. Colbert, G. Scuseria, D. Toma´nek, J.E. Fischer, R.E. Smalley, Crystalline ropes of metallic carbon nanotubes. Science 1996, 273, 483.
18. J. Kong, A.M. Cassell, H.J. Dai, Chemical vapor deposition of methane for single-walled carbon nanotubes. Chem. Phys. Lett. 1998, **292**, 567.
19. T.A. Murphy, Th. Pawlik, A. Weidinger, M. Hohne, R. Alcala, J.M. Spaeth, Observation of atomlike nitrogen in nitrogen-implanted solid C_{60}. Phys. Rev. Lett. 1996, **77**, 1075.
20. C.S. Yannoni, M. Hoinkins, M.S. de Vries, D.S. Bethune, J.R. Salem, M.S. Crowder, R.D. Johnson, Scandium clusters in fullerene cages. Science 1992, **256**, 1191.
21. Dietel, E.; Hirsch, A.; Pietzak, B.; Waiblinger, M.; Lips, K.; Weidinger, A.; Gruss, A.; Dinse, K.P. Atomic nitrogen encapsulated in fullerenes: Effects of cage variations. J. Am. Chem. Soc. 1999, 121, 2432.
22. Hummelen, J.C.; Knight, B.; Pavlovich, J.; Gonzalez, R.; Wudl, F. Isolation of the heterofullerene $C_{59}N$ and its dimmer $(C_{59}N)_2$. Science 1995, 269, 1554.
23. Faigel, G.; Bortel, G.; Oszla´nyi, G.; Pekker, S.; Stephens, P.W.; Liu, D. Low-temperature phase transition in C60-n-pentane. Phys. Rev., B 1994, 49, 9186.
24. Forro, L.; Mihaly, L. Electronic properties of doped fullerenes. Rep. Prog. Phys. 2001, 64, 649.
25. Stephens, P.W.; Mihaly, L.; Lee, P.A.; Whetten, R.L.; Huang, S.M.; Kaner, R.; Diederich, F.; Holczer, K. Structure of single phase, superconducting K3C60. Nature 1991, 351, 632.
26. Tycko, R.; Dabbagh, G.; Fleming, R.M.; Haddon, R.C.; Makhija, A.V.; Zahurak, S.M. Molecular dynamics and the phase transition in solid C60. Phys. Rev. Lett. 1991, 67, 1886.
27. Heiney, P.A.; Fischer, J.E.; McGhie, A.R.; Romanow, W.J.; Denenstein, A.M.; McCauley, J.P., Jr.; Smith, A.B.; Cox, D.E. Orientational ordering transition in solid C60. Phys. Rev. Lett. 1991, 66, 2911. reply to comment 1991, 67, 1468.
28. David, W.I.F.; Ibberson, R.M.; Dennis, T.J.S.; Hare, J.P.; Prassides, K. Structural phase-transitions in the fullerene C60. Europhys. Lett. 1991, 18, 735.
29. Harris, A.B.; Sachidanandam, R. Orientational ordering of icosahedra in solid C60. Phys. Rev., B 1992, 46, 4944.
30. Stephens, P.W.; Bortel, G.; Faigel, G.; Tegze, M.; Ja´nossy, A.; Pekker, S.; Oszla´nyi, G.; Forro´, L. Polymeric fullerene chains in RbC_{60} and KC_{60}. Nature 1994, 370, 636.
31. Yildirim, T.; Fischer, J.E.; Harris, A.B.; Stephens, P.W.; Liu, D.; Brard, L.; Strongin, R.M.; Smith, A.B., III Orientational phase transition in Na_xC_{60} ($1< x <3$). Phys. Rev. Lett. 1993, 71, 1383.
32. Stephens, P.W.; Mihaly, L.; Lee, P.L.; Whetten, R.L.; Huang, S.-M.; Kaner, R.; Diederich, F.; Holczer, K. Structure of single-phase superconducting K_3C_{60}. Nature 1991, 351, 632.
33. Rosseinsky, M.J.; Murphy, D.W.; Fleming, R.M.; Tycko, R.; Ramirez, A.P.; Siegrist, T.; Dabbagh, G.; Barrett, S.E. Structural and electronic properties of sodium-intercalated C-60. Nature 1992, 356, 416.

34. Forro, L.; Mihaly, L. Electronic properties of doped fullerenes. Rep. Prog. Phys. 2001, 64, 649.
35. http://buckminster.physics.sunysb.edu.
36. Haddon, R.C.; Hebard, A.F.; Rosseinsky, M.J.; Murphy, D.W.; Duclos, S.J.; Lyons, K.B.; Miller, B.; Rosamilia, J.M.; Fleming, R.M.; Kortan, R.M.; Glarum, S.H.; Makhija, A.V.; Muller, A.J.; Eick, R.H.; Zahurak, S.M.; Tycko, R.; Dabbagh, G.; Thiel, F.A. Conducting films of C_{60} and C_{70} by alkali metal doping. Nature 1991, 350, 320.
37. Hebard, A.F.; Rosseinsky, M.J.; Haddon, R.C.; Murphy, D.W.; Glarum, S.H.; Palstra, T.T.M.; Ramirez, A.P.; Kortan, A.R. Superconductivity at 18K in potassium doped C_{60}. Nature 1991, 350, 600.
38. Holczer, K.; Klein, O.; Huang, S.-M.; Kaner, R.B.; Fu, K.-J.; Whetten, R.L.; Diederich, F. Alkalifulleride superconductors: Synthesis, composition and diamagnetic shielding. Science 1991, 252, 1154.
39. Allemand, P.-M.; Khemani, K.C.; Koch, A.; Wudl, F.; Holczer, K.; Donovan, S.; Gruner, G.; Thompson, J.D. Organic molecular soft ferromagnetism in C_{60}. Science 1991, 253, 301.
40. Stephens, P.W.; Cox, D.; Lauher, J.W.; Mihaly, L.; Wiley, J.B.; Allemand, P.M.; Hirsch, A.; Holczer, K.; Li, Q.; Thompson, J.D.; Wudl, F. Lattice structure of the fullerene ferromagnet TDAE-C_{60}. Nature 1991, 355, 331.
41. A.M. Rao, P. Zhou, K. Wang, G.T. Hager, J.M. Holden, Y. Wang, W.-T. Lee, X. Bi, C. Eklund, D.S. Cornett, M.A. Duncan, I.J. Amster, Photoinduced polymerization of solid C_{60} films. Science 1993, 259, 955.
42. M. Nunez-Regueiro, L. Marques, J.-L. Hodeau, O. Bethoux, M. Perroux, Polymerized fullerite structures. Phys. Rev. Lett. 1995, 74, 278.
43. S. Pekker, A. Janossy, L. Mihaly, O. Chauvet, M. Carrard, L. Forro, Single crystalline $(KC_{60})_n$—A conducting linear alkali fulleride polymer. Science 1994, 265, 1077.
44. J.W.G. Wildoer, L.C. Venema, A.G. Rinzler, R.E. Smalley, C. Dekker, Electronic structure of atomically resolved carbon nanotubes. Nature 1998, 391, 59.
45. H.W. Zhu, C.L. Xu, D.H. Wu, B.Q. Wei, R. Vajtai, P.M. Ajayan, Direct synthesis of long single-walled carbon nanotube strands. Science 2002, 296, 884.
46. P. M. Ajayan, and T. W. Ebbesen, Rep.Prog.Phys., **60**, (1025-1065, 2003.
47. Yasuda, Ayumu, Kawase, Noboru, and Mizutani, Wataru, Journal of Physical Chemistry B, 106, (51), 2002.
48. S. B. Sinnot, R. Andrews, D. Qian, A. M. Rao, Z. Mao, E. C. Dickey, and F. Derbyshire, Chem.Phys.Lett., **315**, (25-30, 1999.
49. S. H. Jung, M. R. Kim, S. H. Jeong, S. U. Kim, O. J. Lee, K. H. Lee, J. H. Suh, and C. K. Park, Applied Physics A-Materials Science & Processing, **76**, (2)**,** 285-286, 2003.
50. T. W. Ebbesen, and P. M. Ajayan, Nature, **358**, (220-222, 1992.
51. C. Journet, and P. Bernier, Applied Physics A-Materials Science & Processing, **67**, (1)**,** 1-9, 1998.
52. S. Farhat, M. L. La Chapelle, A. Loiseau, C. D. Scott, S. Lefrant, C. Journet, and P. Bernier, Journal of Chemical Physics, **115**, (14)**,** 6752-6759, 2001.
53. H. J. Huang, J. Marie, H. Kajiura, and M. Ata, Nano Letters, **2**, (10)**,** 1117-1119, 2002.
54. Takikawa, Hirofumi, Tao, Yoshitaka, Hibi, Yoshihiko, Miyano, Ryuichi, Sakakibara, Tateki, Ando, Yoshinori, Ito, Shigeo, Hirahara, Kaori, and Iijima, Sumio, AIP Conference Proceedings, 590, (Nanonetwork Materials), 2001.
55. H. Takikawa, M. Ikeda, K. Hirahara, Y. Hibi, Y. Tao, P. A. Ruiz, T. Sakakibara, S. Itoh, and S. Iijima, Physica B: Condensed Matter (Amsterdam, Netherlands), 323, (1-4), 2002
56. Anazawa, Kazunori, Shimotani, Kei, Manabe, Chikara, Watanabe, Hiroyuki, and Shimizu, Masaaki, Applied Physics Letters, 81, (4), 2002.
57. Lee, Seung Jong, Baik, Hong Koo, Yoo, Jae eun, and Han, Jong hoon, Diamond and Related Materials, 11, (3-6), 2002.

58. T. Guo, P. Nikolaev, A. Thess, D. T. Colbert, and R. E. Smalley, Chemical Physics Letters, 243, (1,2), 1995.
59. M. Yudasaka, R. Yamada, N. Sensui, T. Wilkins, T. Ichihashi, and S. Iijima, Journal of Physical Chemistry B, 103, (30), 1999.
60. P. C. Eklund, B. K. Pradhan, U. J. Kim, Q. Xiong, J. E. Fischer, A. D. Friedman, B. C. Holloway, K. Jordan, and M. W. Smith, Nano Letters, 2, (6), 2002.
61. W. K. Maser, E. Munoz, A. M. Benito, M. T. Martinez, G. F. de la Fuente, Y. Maniette, E. Anglaret, and J. L. Sauvajol, Chemical Physics Letters, 292, (4,5,6), 1998.
62. A. P. Bolshakov, S. A. Uglov, A. V. Saveliev, V. I. Konov, A. A. Gorbunov, W. Pompe, and A. Graff, Diamond and Related Materials, 11, (3-6), 2002.
63. C. D. Scott, S. Arepalli, P. Nikolaev, and R. E. Smalley, Applied Physics A: Materials Science &Processing, 72, (5), 2001.
64. Z. F. Ren, Z. P. Huang, J. W. Xu, J. H. Wang, P. Bush, M. P. Siegel, and P. N. Provencio, Science (Washington, D.C.), 282, (5391), 1998.
65. Z. F. Ren, Z. P. Huang, D. Z. Wang, J. G. Wen, J. W. Xu, J. H. Wang, L. E. Calvet, J. Chen, J. F. Klemic, and M. A. Reed, Applied Physics Letters, 75, (8), 1999.
66. Yudasaka, Masako, Kikuchi, Rie, Matsui, Takeo, Ohki, Yoshimasa, Yoshimura, Susumu, and Ota, Etsuro, Applied Physics Letters, 67, (17), 1995.
67. Yudasaka, Masako, Kikuchi, Rie, Ohki, Yoshimasa, Ota, Etsuro, and Yoshimura, Susumu, Applied Physics Letters, 70, (14), 1997.
68. M. Chen, C. M. Chen, and C. F. Chen, Journal of Materials Science, **37**, (17)**,** 3561-3567, 2002.
69. Z. P. Huang, D. Z. Wang, J. G. Wen, M. Sennett, H. Gibson, and Z. F. Ren, Applied Physics AMaterials Science & Processing, **74**, (3)**,** 387-391, 2002.
70. J. B. Park, G. S. Choi, Y. S. Cho, S. Y. Hong, D. Kim, S. Y. Choi, J. H. Lee, and K. I. Cho, Journal of Crystal Growth, **244**, (2)**,** 211-217, 2002.
71. R. L. V. Wal, L. Randall, L. J. Hall, J. Lee, and G. M. Berger, Journal of Physical Chemistry B, 106, (51), 2002.
72. R. L. V. Wal, L. Randall and T.M. Ticich, Journal of Physical Chemistry B, 105, (42), 2001.
73. R. L. V. Wal, G. M. Berger, and L. J. Hall, Journal of Physical Chemistry B, **106**, (14)**,** 3564-3567, 2002.
74. P. Avouris, Chemical Physics, **281**, (2-3)**,** 429-445, 2002.
75. S.J. Tans, M.H. Devoret, H. Dal, A. Thess, R.E. Smalley, L.J. Geerligs, and C. Dekker, Nature (London), 386, (6624), 1997.
76. P. M. Ajayan and O. Z. Zhou, Carbon Nanotubes, **80**, (391-425, 2001.
77. S. Reich, J. Maultzsch, C. Thomsen, P. Ordejon, Tight-binding description of graphene. Phys. Rev., B 2002, 66, 035412.
78. R. Saito, G. Dresselhaus, M.S. Dresselhaus, Electronic structure of chiral graphene tubules. Appl. Phys. Lett. 1992, 60, 2204.
79. S.J. Tans, M.H. Devoret, H. Dai, A. Thess, R.E. Smalley, L.J. Geerligs, C. Dekker, Individual single-wall carbon nanotubes as quantum wires. Nature 1997, 386, 474.
80. M. Bockrath, D.H. Cobden, P.L. McEuen, N.G. Chopra, A. Zettl, A. Thess, R.E. Smalley, Singleelectron transport in ropes of carbon nanotubes. Science 1997, 275, 1922.
81. M.R. Buitelaar, W. Belzig, T. Nussbaumer, B. Babic, C. Bruder, C. Scho¨nenberger, Multiple andreev reflection in a carbon nanotube quantum dot. Phys. Rev. Lett. 2003, 91, 057005.
82. M. Ouyang, J. L. Huang, and C. M. Lieber, Accounts of Chemical Research, **35**, (12)**,** 1018-1025, 2002.
83. A. Zuttel, C. Nutzenadel, P. Sudan, P. Mauron, C. Emmenegger, S. Rentsch, L. Schlapbach, A. Weidenkaff, and T. Kiyobayashi, Journal of Alloys and Compounds, **330**, (676-682, 2002.
84. 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, Synthetic Metals, **113**, (3)**,** 209-216, 2000.

85. A. Zuttel, P. Sudan, P. Mauron, T. Kiyobayashi, C. Emmenegger, and L. Schlapbach, International Journal of Hydrogen Energy, **27**, (2)**,** 203-212, 2002.
86. X. P. Gao, Y. Lan, G. L. Pan, F. Wu, J. Q. Qu, D. Y. Song, and P. W. Shen, Electrochemical and Solid State Letters, **4**, (10)**,** A173-A175, 2001.
87. X. Qin, X. P. Gao, H. Liu, H. T. Yuan, D. Y. Yan, W. L. Gong, and D. Y. Song, Electrochemical and Solid State Letters, **3**, (12)**,** 532-535, 2000.
88. N. Rajalakshmi, K. S. Dhathathreyan, A. Govindaraj, and B. C. Satishkumar, Electrochimica Acta, **45**, (27)**,** 4511-4515, 2000.
89. C. Nutzenadel, A. Zuttel, D. Chartouni, and L. Schlapbach, Electrochemical and Solid State Letters, **2**, (1)**,** 30-32, 1999.
90. G. P. Dai, C. Liu, M. Liu, M. Z. Wang, and H. M. Cheng, Nano Letters, **2**, (5)**,** 503-506, 2002.
91. A. K. M. F. Kibria, Y. H. Mo, K. S. Park, K. S. Nahm, and M. H. Yun, International Journal of Hydrogen Energy, **26**, (8)**,** 823-829, 2001.
92. E. Frackowiak and F. Beguin, Carbon, **40**, (10)**,** 1775-1787, 2002.
93. M. Hirscher, M. Becher, M. Haluska, A. Quintel, V. Skakalova, Y. M. Choi, U. Dettlaff-Weglikowska, S. Roth, I. Stepanek, P. Bernier, A. Leonhardt, and J. Fink, Journal of Alloys and Compounds, **330**, (654-658, 2002.
94. K. Atkinson, S. Roth, M. Hirsher, and W. Grunwald, Fuel Cells Bulletin, 38)**,** 9-12, 2002.
95. C. Liu, Y. Y. Fan, M. Liu, H. T. Cong, H. M. Cheng, and M. S. Dresselhaus, Science, **286**, (5442)**,** 1127-1129, 1999.
96. R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, Science, **297**, (5582)**,** 787-792, 2002
97. K. H. An, W. S. Kim, Y. S. Park, Y. C. Choi, S. M. Lee, D.C. Chung, D. J. Bae, S. C. Lim, and Y. H. Lee, Journal of Korean Physical Society, **39**, (511-517, 2001.
98. E. Frackowiak and F. Beguin, Carbon, **39**, (6)**,** 937-950, 2001.
99. E. Frackowiak, K. Jurewicz, K. Szostak, S. Delpeux, and F. Beguin, Fuel Processing Technology, **77**, (213-219, 2002.
100. E. Frackowiak, K. Jurewicz, S. Delpeux, and F. Beguin, Journal of Power Sources, **97-8**, (822-825, 2001.
101. K. Jurewicz, S. Delpeux, V. Bertagna, F. Beguin, and E. Frackowiak, Chemical Physics Letters, **347**, (1-3)**,** 36-40, 2001.
102. E. Frackowiak, S. Delpeux, K. Jurewicz, K. Szostak, D. Cazorla-Amoros, and F. Beguin, Chemical Physics Letters, **361**, (1-2)**,** 35-41, 2002.
103. Q. Jiang, M. Z. Qu, G. M. Zhou, B. L. Zhang, and Z. L. Yu, Materials Letters, **57**, (4)**,** 988-991, 2002.
104. J. M. Bonard, H. Kind, T. Stockli, and L. A. Nilsson, Solid-State Electronics, **45**, (6)**,** 893-914, 2001.
105. Y. Saito and S. Uemura, Carbon, **38**, (2)**,** 169-182, 2000.
106. D. Rotman, Natuur & Techniek, **70**, (6)**,** 30-37, 2002.
107. Avouris, Phaedon, Accounts of Chemical Research, 35, (12), 2002.
108. S. J. Tans, A. R. M. Verschueren, and C. Dekker, Nature, **393**, (6680)**,** 49-52, 1998.
109. R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and Ph. Avouris, Applied Physics Letters, **73**, (17)**,** 2447-2449, 1998.
110. S. J. Wind, J. Appenzeller, R. Martel, V. Derycke, and Ph. Avouris, Journal of Vacuum Science Technology B, **20**, (6)**,** 2798-2801, 2002.
111. P. L. Mceuen, M. Bockrath, D. H. Cobden, Y. G. Yoon, and S. G. Louie, Physical Review Letters, **83**, (24)**,** 5098-5101, 1999.
112. M. Bockrath and et.al., Physical Review B, R10606-R10608, 2000.
113. V. Derycke, R. Martel, J. Appenzeller, and Ph. Avouris, Applied Physics Letters, **80**, (15)**,** 2773-2775, 2002.
114. S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and Ph. Avouris, Physical Review Letters, **89**, (10)**,** 106801-1-106801-4, 2002.

115. J. R. Tucker, C. Wang, and P. S. Carney, Applied Physics Letters, **65**, (618)**,** 1994.
116. J. Kong, J. Cao, H. J. Dai, and E. Anderson, Applied Physics Letters, **80**, (1)**,** 73-75, 2002.
117. H. W. C. Postma, T. Teepen, Z. Yao, M. Grifoni, and C. Dekker, Science, **293**, (5527)**,** 76-79, 2001.
118. D. Kasuya, M. Yudasaka, K. Takahashi, F. Kokai, S. Iijima. Selective Production of Single-Wall Carbon Nanohorn Aggregates and Their Formation Mechanism. J. Phys. Chem. B **106** (19), 4947 (2002).
119. T. J. Bernatowicz; et al. Constraints on stellar grain formation from presolar graphite in the Murchison meteorite. Astrophysical Journal **472** (2), 760–782 (1996).
120. P. Fraundorf, M. Wackenhut. The core structure of presolar graphite onions. Astrophysical Journal Letters **578** (2), L153–156 (2002).
121. R. Zan, Q.M. Ramasse, U. Bangert, K.S. Novoselov. Graphene re-knits its holes. Mesoscale and Nanoscale Physics **12** (8), 3936 (2012).
122. J. Meyer, A. K. Geim, M. I. Katsnelson, K. S. Novoselov, T. J. Booth, S. Roth. The structure of suspended graphene sheets. Nature **446** (7131), 60–63 (2007).
123. J. M. Carlsson. Graphene: Buckle or break. Nature Materials **6** (11), 801–2 (2007).
124. A. Fasolino, J. H. Los, M. I. Katsnelson. Intrinsic ripples in graphene. Nature Materials **6** (11), 858–61 (2007).
125. M. Ishigami, et al. Atomic Structure of Graphene on SiO_2. Nano Lett **7**(6): 1643–1648 (2007).
126. E. Stolyarova, et al. High-resolution scanning tunneling microscopy imaging of mesoscopic graphene sheets on an insulating surface. Proceedings of the National Academy of Sciences **104** (22), 9209–9212 (2007).
127. P. A. Denis, F. Iribarne. Comparative Study of Defect Reactivity in Graphene. Journal of Physical Chemistry C **117** (37), 19048 (2013).
128. Y. Yamada, K. Murota, R. Fujita, J. Kim, et al. Subnanometer vacancy defects introduced on graphene by oxygen gas. Journal of American Chemical Society **136** (6), 2232 (2014).
129. A. Eftekhari, P. Jafarkhani. Curly Graphene with Specious Interlayers Displaying Superior Capacity for Hydrogen Storage. Journal of Physical Chemistry C **117** (48), 25845 (2013).
130. Y. Yamada, H. Yasuda, K. Murota, M. Nakamura, T. Sodesawa, S. Sato. Analysis of heat-treated graphite oxide by X-ray photoelectron spectroscopy. Journal of Material Science **48** (23), 8171 (2013).
131. Y. Yamada, J. Kim, K. Murota, S. Matsuo, S. Sato. Nitrogen-containing graphene analyzed by X-ray photoelectron spectroscopy. Carbon **70,** 59 (2014).
132. P. Blake, E. W. Hill, et al. Making graphene visible. Applied Physics Letters 91(6), 063124 (2007).
133. X. Liang, Z. Fu, et al. Graphene Transistors Fabricated via Transfer-Printing In Device Active-Areas on Large Wafer. Nano Lett. 7(12), 3840-3844 (2007).
134. A. N. Sidorov, M. M. Yazdanpanah, et al. Electrostatic deposition of graphene. Nanotechnology (13), 135301 (2007).
135. J. S. Bunch, Y. Yaish, et al. Coulomb oscillations and Hall effect in quasi-2D graphite quantum dots. Nano Letters 5(2), 287-290 (2005).
136. C. Berger, Z. M. Song, et al. Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics. Journal of Physical Chemistry B 108(52), 19912-19916 (2004).
137. Y. Zhang, J. Small, et al. Fabrication and electric-field-dependent transport measurements of mesoscopic graphite devices. Applied Physics Letters 86(7), 073104 (2005).
138. X. Li, X. Wang, et al. Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors. Science 319(5867), 1229-1232 (2008).
139. C. Berger, Z. M. Song, et al. Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics. Journal of Physical Chemistry B 108(52), 19912-19916 (2004).

140. C. Berger, Z. Song, et al. Electronic Confinement and Coherence in Patterned Epitaxial Graphene. Science 312(5777), 1191-1196 (2006).
141. Y. Han, B. Ozyilmaz, et al. Energy Band-Gap Engineering of Graphene Nanoribbons. Physical Review Letters 98(20): 206805 (2007).
142. A. C. Neto, N. M. R. Peres, K. S. Novoselov, A. K. Geim. The electronic properties of graphene. Rev Mod Phys **81,** 109 (2009).
143. J.-C. Charlier, P.C. Eklund, J. Zhu, A.C. Ferrari, A. Jorio, G. Dresselhaus and M.S. Dresselhaus, eds. *Electron and Phonon Properties of Graphene: Their Relationship with Carbon Nanotubes*. *Carbon Nanotubes: Advanced Topics in the Synthesis, Structure, Properties and Applications* (Berlin/Heidelberg: Springer-Verlag) (2008).
144. G. W. Semenoff. Condensed-Matter Simulation of a Three-Dimensional Anomaly. Physical Review Letters **53** (26): 2449 (1984).
145. P.R. Wallace. The Band Theory of Graphite. Physical Review **71** (9): 622 (1947).
146. P. Avouris, Z. Chen, V. Perebeinos. Carbon-based electronics. Nature Nanotechnology **2** (10), 605–15 (2007).
147. C.A. Lamas, D.C. Cabra, N. Grandi. Generalized Pomeranchuk instabilities in graphene. Physical Review B **80** (7), 75108 (2009).
148. K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, A. A. Firsov. Two-dimensional gas of massless Dirac fermions in graphene. Nature **438** (7065), 197–200 (2005).

149. S.V. Morozov, K. Novoselov, M. Katsnelson, F. Schedin, D. Elias, J. Jaszczak, A. Geim. Giant Intrinsic Carrier Mobilities in Graphene and Its Bilayer. Physical Review Letters **100** (1), 016602 (2008).
150. J. H. Chen, C. Jang, S. Xiao, M. Ishigami, M. S. Fuhrer. Intrinsic and Extrinsic Performance Limits of Graphene Devices on SiO_2. Nature Nanotechnology **3** (4), 206–9 (2008).
151. A. Akturk, N. Goldsman. Electron transport and full-band electron–phonon interactions in graphene. Journal of Applied Physics **103** (5), 053702 (2008).
152. J. Baringhaus, M. Ruan, F. Edler, A. Tejeda, M. Sicot, A. Taleb-Ibrahimi, A. P. Li, Z. Jiang, E. H. Conrad, C. Berger, C. Tegenkamp, W. A. De Heer. Exceptional ballistic transport in epitaxial graphene nanoribbons. Nature **506** (7488), 349 (2014).
153. J. H. Chen, C. Jang, S. Adam, M. S. Fuhrer, E. D. Williams, M. Ishigami. Charged Impurity Scattering in Graphene. Nature Physics **4** (5): 377–381 (2008).
154. F. Schedin, A. K. Geim, S. V. Morozov, E. W. Hill, P. Blake, M. I. Katsnelson, K. S. Novoselov. Detection of individual gas molecules adsorbed on graphene. Nature Materials **6** (9), 652–655 (2007).
155. S. Adam, E. H. Hwang, V. M. Galitski, S. Das Sarma. A self-consistent theory for graphene transport. Proc. Nat. Acad. Sci. USA **104** (47), 18392–7 (2007).
156. H. Steinberg, G. Barak, A. Yacoby, et al. Charge fractionalization in quantum wires (Letter). Nature Physics **4** (2): 116–119 (2008).
157. A. Trisetyarso. Dirac four-potential tunings-based quantum transistor utilizing the Lorentz force. Quantum Information & Computation **12** (11–12), 989 (2012).
158. J. K. Pachos. Manifestations of topological effects in graphene. Contemporary Physics **50** (2): 375 (2009).
159. Y. B. Zhang, Y. W. Tan, H. L. Stormer, and P. Kim, Nature **438**, 201 (2005).
160. V. P. Gusynin, S. G. Sharapov. Unconventional Integer Quantum Hall Effect in Graphene. Physical Review Letters **95** (14), 146801. (2005).
161. Y. Zhang, Y. W. Tan, H. L. Stormer, P. Kim. Experimental observation of the quantum Hall effect and Berry's phase in graphene. Nature **438** (7065), 201–204 (2005).
162. K. S. Kim, H. Jang, S. Y. Lee, J. M. Kim, K. S. Kim, J.-H. Ahn, P. Kim, J.-Y. Choi, B. H. Hong. Large-scale pattern growth of graphene films for stretchable transparent electrodes. Nature **457**(7230), 706–10 (2009).

163. J. Jobst, D. Waldmann, F. Speck, R. Hirner, D. K. Maude, T. Seyller, H. B. Weber. How Graphene-like is Epitaxial Graphene? Quantum Oscillations and Quantum Hall Effect. Physical Review B **81** (19), 195434 (2009).
164. T. Shen, J.J. Gu, M. Xu, Y.Q. Wu, M.L. Bolen, M.A. Capano, L.W. Engel, P.D. Ye. Observation of quantum-Hall effect in gated epitaxial graphene grown on SiC (0001). Applied Physics Letters **95** (17), 172105 (2009).
165. X. Wu, Y. Hu, M. Ruan, N. K. Madiomanana, J. Hankinson, M. Sprinkle, C. Berger, W.A. de Heer. Half integer quantum Hall effect in high mobility single layer epitaxial graphene. Applied Physics Letters **95** (22), 223108 (2009).
166. J.A. Alexander-Webber, A.M.R. Baker, T.J.B.M. Janssen, A. Tzalenchuk, S. Lara-Avila, S. Kubatkin, R. Yakimova, B. A. Piot, D. K. Maude, R.J. Nicholas. Phase Space for the Breakdown of the Quantum Hall Effect in Epitaxial Graphene. Physical Review Letters **111** (9): 096601 (2013).
167. M. S. Fuhrer. A physicist peels back the layers of excitement about graphene. Nature **459** (7250), 1037 (2009).
168. Y. Zhang, Z. Jiang, J. P. Small, M. S. Purewal, Y.-W. Tan, M. Fazlollahi, J. D. Chudow, J. A. Jaszczak, H. L. Stormer, P. Kim. Landau-Level Splitting in Graphene in High Magnetic Fields. Physical Review Letters **96** (13), 136806 (2006).
169. X. Du, I. Skachko, F. Duerr, A. Luican, E. Y. Andrei. Fractional quantum Hall effect and insulating phase of Dirac electrons in graphene. Nature **462** (7270), 192–195 (2009).
170. K. Bolotin, F. Ghahari, M. D. Shulman, H. L. Stormer, P. Kim. Observation of the fractional quantum Hall effect in graphene. Nature **462** (7270), 196–199 (2009).
171. S.-E. Zhu, S. Yuan, G. C. A. M. Janssen. Optical transmittance of multilayer graphene. EPL (Europhysics Letters) **108** (1), 17007 (2014).
172. R. R. Nair, P. Blake, A. N. Grigorenko, K. S. Novoselov, T. J. Booth, T. Stauber, N. M. R. Peres, A. K. Geim. Fine Structure Constant Defines Visual Transparency of Graphene. Science **320** (5881), 1308–1308 (2008).
173. A. B. Kuzmenko, E. Van Heumen, F. Carbone, D. Van Der Marel. Universal infrared conductance of graphite. Physical Review Letters **100** (11), 117401 (2008).
174. Y. Zhang, T.-T. Tang, C. Girit, Z. Hao, M. C. Martin, A. Zettl, M. F. Crommie, Y. R. Shen, F. Wang. Direct observation of a widely tunable bandgap in bilayer graphene. Nature **459** (7248), 820–823 (2009).
175. J. Liu, A. R. Wright, C. Zhang, Z. Ma. Strong terahertz conductance of graphene nanoribbons under a magnetic field. Appl Phys Lett **93**(4): 041106–041110 (2008).
176. U. Kurum, B. Liu, K. Zhang, Y. Liu, H. Zhang. Electrochemically tunable ultrafast optical response of graphene oxide. Applied Physics Letters **98** (2): 141103 (2011).
177. K.V. Sreekanth, S. Zeng, J. Shang, K.-T. Yong, T. Yu. Excitation of surface electromagnetic waves in a graphene-based Bragg grating. Scientific Reports **2,** 737 (2012).
178. Q. Bao, H. Zhang, Yu. Wang, Z. Ni, Y. Yan, Z. X. Shen, K. P. Loh, D. Y. Tang. Atomic-Layer Graphene as a Saturable Absorber for Ultrafast Pulsed Lasers. Advanced Functional Materials **19** (19), 3077 (2009).
179. Z. Zheng, C. Zhao, S. Lu, Y. Chen, Y. Li, H. Zhang, S. Wen. Microwave and optical saturable absorption in graphene. Optics Express **20** (21), 23201–23214 (2012).
180. H. Zhang, S. Virally, Q. Bao, L. Kian Ping, S. Massar, N. Godbout, P. Kockaert. Z-scan measurement of the nonlinear refractive index of graphene. Optics Letters **37** (11): 1856–1858 (2012).
181. H. Dong, C. Conti, A. Marini and F. Biancalana, Terahertz relativistic spatial solitons in doped graphene metamaterials, Journal of Physics B: Atomic, Molecular and Optical Physics 46, 15540 (2013).
182. C. Lee, Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene, Science **321** (385), 385–388 (2008).

183. C. Lee, X. Wei, J. W. Kysar,. J. Hone, Measurement of the elastic properties and intrinsic strength of monolayer graphene, Science **321** (5887), 385–8 (2008)..

184. I.W. Frank, D.M. Tanenbaum, A.M. Van Der Zande, P.L. McEuen, Mechanical properties of suspended graphene sheets, J. Vac. Sci. Technol. B **25**(6), 2558–2561 (2007).

185. S. Braga, V. R. Coluci, S. B. Legoas, R. Giro, D. S. Galvão, R. H. Baughman, Structure and dynamics of carbon nanoscrolls, Nano Letters **4**(5), 881 (2004).

186. D. Bolmatov, C.-Y. Mou, Graphene-based modulation-doped superlattice structures, Journal of Experimental and Theoretical Physics (JETP) **112**, 102 (2011). D. Bolmatov, Thermodynamic properties of tunneling quasiparticles in graphene-based structures, Physica C **471** (23–24), 1651 (2011).

187. J. Chen, M. Ishigami, C. Jang, D. R. Hines, M. S. Fuhrer, E. D. Williams, Printed graphene circuits, Advanced Materials **19** (21): 3623–3627 (2007).

188. L. A. Ponomarenko, F. Schedin, M. I. Katsnelson, R. Yang, E. W. Hill, K. S. Novoselov, A. K. Geim, Chaotic dirac billiard in graphene quantum dots, Science **320**(5874), 356–8 (2008).

189. X. Wang, X. Li, L. Zhang, Y. Yoon, P. K.Weber, H. Wang, J. Guo, H. Dai, N-doping of graphene through electrothermal reactions with ammonia, Science **324**(5928), 768–71 (2009).

190. F. Traversi, V. Russo, R. Sordan, Integrated complementary graphene inverter, Appl. Phys. Lett **94** (22): 223312 (2009).

191. Y.-M. Lin, C. Dimitrakopoulos, K. A. Jenkins, D. B. Farmer, H.-Y. Chiu, A. Grill, Ph. Avouris, 100-GHz transistors from wafer-scale epitaxial graphene, Science **327** (5966), 662 (2010).

192. Y.-M. Lin, A. Valdes-Garcia, S.-J. Han, D. B. Farmer, I. Meric, Y. Sun, Y. Wu, C. Dimitrakopoulos, A. Grill, P. Avouris, K. A. Jenkins, Wafer-scale graphene integrated circuit, Science **332** (6035), 1294–1297 (2011).

193. M. C. Lemme, T. J. Echtermeyer, et al. A graphene field-effect device, IEEE Electron Device Letters **28** (4): 282 (2007).

194. J. Kedzierski, P.-L. Hsu, P. Healey, P.W. 0Wyatt, C. L. Keast, M. Sprinkle, C. Berger, W. A. De Heer, Epitaxial graphene transistors on SiC substrates, IEEE Transactions on Electron Devices **55**(8), 2078–2085 (2008).

195. J.S. Moon, D. Curtis, M. Hu, D. Wong, C. McGuire, P.M. Campbell, G. Jernigan, J.L. Tedesco, B. Vanmil, R. Myers-Ward, C. Eddy, D.K. Gaskill, Epitaxial-graphene RF field-effect transistors on Si-face 6H-SiC substrates, IEEE Electron Device Letters **30** (6), 650–652 (2009).

196. T.J. Echtermeyer, M.C. Lemme, et al. Nonvolatile switching in graphene field-effect devices, IEEE Electron Device Letters **29**(8), 952 (2008).

197. R. Sordan, F. Traversi, V. Russo, Logic gates with a single graphene transistor, Appl. Phys. Lett. **94**(7), 073305 (2009).

198. G. Fiori, G. Iannaccone, On the possibility of tunable-gap bilayer graphene FET, IEEE Electr. Dev. Lett., 30, 261 (2009).

199. G. Fiori, G. Iannaccone, Ultralow-voltage bilayer graphene tunnel FET, IEEE Electr. Dev. Lett., 30, 1096 (2009)

200. M. Lin; C. Dimitrakopoulos, K.A. Jenkins, D.B Farmer, H.Y Chiu, A. Grill, P. Avouris, "100-GHz transistors from wafer-scale epitaxial graphene, *Science* (Science) **327**(5966), 662 (2010).

201. F. Torrisi, T. Hasan, W. Wu, Z. Sun, A. Lombardo, T. Kulmala, G. W. Hshieh, S. J. Jung, F. Bonaccorso, P. J. Paul, D. P. Chu, A. C. Ferrari, Ink-jet printed graphene electronics, ACS Nano , **6** (2992), 4970 (2011).

202. Y. Kawano, Wide-band frequency-tunable terahertz and infrared detection with graphene, Nanotechnology **24** (21), 214004 (2013).

203. L. Britnell, R. V. Gorbachev, A. K. Geim, L. A. Ponomarenko, A. Mishchenko, M. T. Greenaway, T. M. Fromhold, K. S. Novoselov, L. Eaves, Radical new graphene design operates at terahertz speed, Nature Communications **4**(4), 1794 (2013).

204. G. Liu, S. Ahsan, A.G. Khitun, R.K. Lake, A.A. Balandin, Graphene-based non-boolean logic circuits, J. Appl. Phys. **114**(10), 4310 2013).
205. G. Eda, G. Fanchini, M. Chhowalla, Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material, Nat Nanotechnol. **3**(5): 270 (2008).
206. Y. Wang, X. Chen, Y. Zhong, F. Zhu, K.P. Loh, Large area, continuous, few-layered graphene as anodes in organic photovoltaic devices, Applied Physics Letters **95** (6), 063302 (2009).
207. J.B. Wu, M. Agrawal, H.A. Becerril, Z. Bao, Z. Liu, Y. Chen, P. Peumans, Organic light-emitting diodes on solution-processed graphene transparent electrodes, *ACS Nano* **4,** 43(2010).
208. P. Matyba, H. Yamaguchi, et al. Graphene and mobile ions: The key to all-plastic, solution-processed light-emitting devices, ACS Nano **4** (2), 637–42 (2010).
209. R. R. Nair, H. A. Wu, P. N. Jayaram, I. V. Grigorieva, A. K. Geim, Unimpeded permeation of water through helium-leak-tight graphene-based membranes, *Science* **335**(6067), 442–4 (2012).
210. Zhu, Shou-En; Yuan, Shengjun; Janssen, G. C. A. M. (1 October 2014). "Optical transmittance of multilayer graphene". *EPL (Europhysics Letters)* **108** (1), 17007.
211. X. Li, D. Xie, H. Park, T.H. Zeng, K. Wang, J. Wei, M. Zhong, D. Wu, J. Kong, H. Zhu, Anomalous behaviors of graphene transparent conductors in graphene–silicon heterojunction solar cells, Advanced Energy Materials **3** (8), 1029 (2013).
212. J. T. W. Wang, J. M. Ball, E. M. Barea, A. Abate, J. A. Alexander-Webber, J. Huang, M. Saliba, I. N. Mora-Sero, J. Bisquert, H. J. Snaith, R. J. Nicolas, Low-temperature processed electron collection layers of graphene/TiO_2 nanocomposites in thin film perovskite solar cells, Nano Letters **14** (2), 724–730 (2013).
213. Y. Dan, Y. Lu, N.J. Kybert, Z. Luo, A. T. C. Johnson, Intrinsic response of graphene vapor sensors, Nano Letters **9** (4), 1472–1475 (2009).
214. F. Schedin, A.K. Geim, S.V. Morozov, E.W. Hill, P. Blake, M.I. Katsnelson, K.S. Novoselov, Detection of individual gas molecules adsorbed on graphene, Nature Materials **6** (9), 652–655 (2007).
215. M. Ong, E.J. Reed, Engineered piezoelectricity in graphene, ACS Nano **6** (2), 1387–94 (2012).
216. G. Lalwani, A.M. Henslee, B. Farshid, L. Lin, F.K. Kasper, Y.X. Qin, A.G. Mikos, B. Sitharaman, Two-dimensional nanostructure-reinforced biodegradable polymeric nanocomposites for bone tissue engineering, Biomacromolecules **14** (3), 900–9 (2013).
217. S. Kanakia, J.D. Toussaint, S.M. Chowdhury, G. Lalwani, T. Tembulkar, T. Button, K.R. Shroyer, W. Moore, B. Sitharaman, Physicochemical characterization of a novel graphene-based magnetic resonance imaging contrast agent, International Journal of Nanomedicine **8,** 2821–33 2013).
218. G. Lalwani, J.L. Sundararaj, K. Schaefer, T. Button, B. Sitharaman, Synthesis, characterization, *In Vitro* phantom imaging, and cytotoxicity of a novel graphene-based multimodal magnetic resonance imaging - X-ray computed tomography contrast agent, Journal of Materials Chemistry B **2** (22), 3519–3530 (2014).
219. G. Lalwani, X. Cai, L. Nie, L.V. Wang, B. Sitharaman, Graphene-based contrast agents for photoacoustic and thermoacoustic tomography, Photoacoustics **1** (3-4), 62–67 (2013).
220. C.S. Mullick G. Lalwani, K. Zhang, J.Y. Yang, K. Neville, B. Sitharaman, Cell specific cytotoxicity and uptake of graphene nanoribbons, Biomaterials **34** (1), 283–93 (2013).
221. Y. Talukdar, J.T. Rashkow, G. Lalwani, S. Kanakia, B. Sitharaman, The effects of graphene nanostructures on mesenchymal stem cells, Biomaterials **35** (18), 4863–77 (2014).
222. R. Abdul Khaliq, R. Kafafy, H.M. Salleh, W.F. Faris, Enhancing the efficiency of polymerase chain reaction using graphene nanoflakes, Nanotechnology **23**(45), 455106 (2012).
223. N. Mohanty, V. Berry, Graphene-based single-bacterium resolution biodevice and DNA-transistor – interfacing graphene-derivatives with nano and micro scale biocomponents, Nano Letters **8** (12), 4469–76 (2008).

224. L. Donaldson, Graphene: Invisible to water, Materials Today **15** (3), 82 (2012).

225. M.S. Xu, D. Fujita, N. Hanagata, Perspectives and challenges of emerging single-molecule DNA sequencing technologies, Small **5** (23): 2638–49 (2009).

226. M. Majumder, R. Tkacz, R. Oldenbourg, S. Mehta, M. Miansari, A. Verma, pH dependent isotropic to nematic phase transitions in graphene oxide dispersions reveal droplet liquid crystalline phases, Chemical Communications **50** (50), 6668–6671 (2014).

Chapter 15

1. W.U. Huynh, J.J. Dittmer, A.P. Alivisator. Hybrid nanorod-polymer solar cells. Science 2002, 295, 2425–2427.

2. M. Law, D.J. Sirbuly, J.C. Johnson, J. Goldberger, R.J.Saykally, P. Yang, Nanoribbon waveguides for subwavelength photonics integration. Science 2004, 305, 1269–1273.

3. A.D. Yoffe, Semiconductor quantum dots and related systems: electronic, optical, luminescence and related properties of low dimensional systems. Adv. Phys. 2001, 50, 1–208.

4. Y. Cui, Z. Zhong, D. Wang, W. Wang, C.M. Lieber, High performance silicon nanowire field effect transistors. Nano Lett. 2003, 3, 149–152.

5. G. Zheng, W. Lu, S. Jin, C.M. Lieber, Synthesis and fabrication of high-performance n-type silicon nanowire transistors. Adv. Mater. 2004, 16, 1890–1893.

6. C.M. Lieber, The incredible shrinking circuit. Sci. Am. 2001, 285 (3), 58.

7. R. Dagani, Building from the bottom up. Chem. Eng. News 2000, 78 (42), 27.

8. J.R. Heath, P.J. Kuekes, G.S. Snider, R.S. Williams, A defect-tolerant computer architecture:opportunities for nanotechnology. Science 1998, 280, 1716.

9. Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, H. Yan, Onedimensional nanostructures: synthesis, characterization, and applications. Adv. Mater. 2003, 15, 353–389.

10. C.M. Lieber, One-dimensional nanostructures: chemistry, physics & applications. Solid State Commun. 1998, **107**, 607–616.

11. M. Law, J. Goldberger, P. Yang, Semiconductor nanowires and nanotubes. Ann. Rev. Mater. Res. 2004, **34**, 83–122.

12. Z.L. Wang, Ed. Nanowires and Nanobelts: Materials, Properties and Devices; Kluwer Academic Publishers: Boston, 2003; Vols. I and II.

13. R.S. Wagner, W.C. Ellis, Vapor–liquid–solid mechanism of single crystal growth. Appl. Phys. Lett. 1964, **4**, 89–90.

14. Y. Lu,; J. Zhong, (2004). Todd Steiner, ed. *Semiconductor Nanostructures for Optoelectronic Applications*. Norwood, MA: Artech House, Inc. pp. 191–192.

15. R. S. Wagner, A.P. Levitt (1975). *Whisker Technology*. Wiley – Interscience – New York.

16. M. H. Huang, Y. Wu, H. Feick, N. Tran, E. Weber, P. Yang, (2001). "Catalytic Growth of Zinc Oxide Nanowires by Vapor Transport". *Adv. Mater.* **13** (2): 113–116.

17. J.-T. Wang, (2002). *Nonequilibrium Nondissipative Thermodynamics: With Application to Low-pressure Diamond Synthesis*. Berlin: Springer Verlag. p. 65.

18. B. Bhushan, *Springer Handbook of Nanotechnology*. Berlin: Spinger-Verlag. p. 105.

19. A.M. Morales, C.M. Lieber, A laser ablation method for the synthesis of crystalline semiconductor nanowires. Science 1998, **279**, 208–211.

20. K. Hiruma, M. Yazawa, T. Katsuyama, K. Ogawa, K. Haraguchi, M. Koguchi, H. Kaklbayashi, Growth and optical properties of nanometer-scale GaAs and InAs whiskers. J. Appl. Phys. 1995, **77**, 447–462.

21. X. Duan, C.M. Lieber, General synthesis of compound semiconductor nanowires. Adv. Mater. 2000, 12, 298–302.

22. M.S. Gudiksen, J. Wang, C.M. Lieber, Synthetic control of the diameter and length of single crystal semiconductor nanowires. J. Phys. Chem. B 2001, **105**, 4062–4064.

23. T.J. Trentler, K.M. Hickman, S.C. Goel, A.M. Viano, P.C. Gibbons, W.E. Buhro, Solution–liquid–solid growth of crystalline III–V semiconductors: an analogy to vapor–liquid–solid growth. Science 1995, **270**, 1791–1794.
24. H. Yu, W.E. Buhro, Solution–liquid–solid growth of soluble GaAs nanowires. Adv. Mater. 2003, **15**, 416–419.
25. Moore’s Law. http://www.intel.com/research/silicon/mooreslaw.htm.
26. The International Technology Roadmap for Semiconductors. http://www.intel.com/research/silicon/itroadmap.htm.
27. D.J. Meindl, Q. Chen, J.A. Davis, Limits on silicon nanoelectronics for terascale integration. Science 2001, **293**, 2044.
28. J.A. Rogers, K.E. Paul, R.J. Jackman, G.M. Whitesides, Using an elastomeric phase mask for sub-100 nm photolithography in the optical near field. Appl. Phys. Lett. 1997, **70**, 2658–2660.
29. A.C. Irvine, Z.A.K. Durrani, H. Ahmed, S. Biesemans, Single-electron effects in heavily doped polycrystalline silicon nanowires. Appl. Phys. Lett. 1998, **73**, 1113–1115.
30. F. Hua, Y. Sun, A. Gaur, M.A. Meitl, L. Bilhaut, L. Rotkina, J. Wang, P. Geil, M. Shim, J.A. Rogers, Polymer imprint lithography with molecular-scale resolution. Nano Lett. 2004, **4**, 2467–2471.
31. Y. Yin, B. Gates, Y. Xia, A soft lithography approach to the fabrication of nanostructures of single crystalline silicon with well-defined dimensions and shapes. Adv. Mater. 2000, **12**, 1426–1430.
32. S.C. Lee, S.R.J. Brueck, Nanoscale twodimensional patterning on Si (001) by large-area interferometric lithography and anisotropic wet etching. J. Vac. Sci. Technol. B 2004, **22**, 1949–1952.
33. Y. Xia, G.M. Whitesides, Shadowed sputtering of gold on V-shaped microtrenches etched in silicon and applications in microfabrication. Adv. Mater. 1996, **8**, 765–768.
34. S. Jeon, E. Menard, J.-U. Park, J. Maria, M. Meitl, J. Zaumseil, J.A. Rogers, Three-dimensional nanofabrication with rubber stamps and conformable photomasks. Adv. Mater. 2004, **16**, 1369–1373.
35. S.Y. Chou, C. Keimel, J. Gu, Ultrafast and direct imprint of nanostructures in silicon. Nature 2002, **417**, 835–837.
36. Z.L. Wang, Transmission electron microscopy of shape-controlled nanocrystals and their assemblies. J. Phys. Chem. B 2000, **104**, 1153–1175.
37. Y. Sun, J.A. Rogers, Fabricating semiconductor nano=microwires and transfer printing ordered arrays of them onto plastic substrates. Nano Lett. 2004, **4**, 1953–1959.
38. Y. Sun, D.-Y. Khang, F. Hua, K. Hurley, R.G. Nuzzo, J.A. Rogers, Photolithographic route to the fabrication of micro=nanowires of III–V semiconductors. Adv. Funct. Mater. 2005, **15**, 30–40.
39. E. Menard, K.J. Lee, D.-Y. Khang, R.Z. Nuzzo, J.A. Rogers, A printable form of silicon for high performance thin film transistors on plastic substrates. Appl. Phys. Lett. 2004, **84**, 5398–5400.
40. K.J. Lee, J.S. Lee, H. Hwang, R.F. Davis, J.A. Rogers, R.G. Nuzzo, A printable form of single crystal gallium nitride for flexible optoelectronic systems. 2005.
41. M.A. Reed, J.A. Tour, Computing with molecules. Sci. Am. 2000, **282** (6), 86.
42. C. Joachim, J.K. Gimzewski, A. Aviram, Electronics using hybrid-molecular and monomolecular devices. Nature 2000, 408, 541.
43. C.P. Collier, E.W. Wong, M. Belohradsky, F.M. Raymo, J.F. Stoddart, P.J. Kuekes, Semiconductor Nanowires: Nanoscale Electronics and Optoelectronics 29 Williams, R.S.; Heath, J.R. Electronically configurable molecular-based logic gates. Science 1999, **285**, 391.
44. C.P. Collier, G. Mattersteig, E.W. Wong, Y. Luo, K. Beverly, J.Sampaio, F.M. Raymo, J.F. Stoddart, J.R. Heath, A [2]catenane-based solid state electronically reconfigurable switch. Science 2000, 289, 1172.

45. M.A. Reed, J. Chen, A.M. Rawlett, D.W. Price, J.M. Tour, Molecular random access memory cell. Appl. Phys. Lett. 2001, 78, 3735.
46. A.P. Alivisatos, Semiconductor clusters, nanocrystals, and quantum dots. Science 1996, 271, 933.
47. D.L. Klein, R. Roth, A.K.L. Lim, A.P. Alivisatos, P.L. McEuen, A single-electron transistor made from a cadmium selenide nanocrystal. Nature 1997, 389, 699.
48. M.H. Devoret, R.J. Schoelkopf, Amplifying quantum signals with the single-electron transistor. Nature 2000, 406, 1039.
49. P.G. Collins, P. Avouris, Nanotubes for electronics. Sci. Am. 2000, 283 (6), 62.
50. T.W. Odom, J.-L. Huang, P. Kim, C.M. Lieber, Atomic structure and electronic properties of single-walled carbon nanotubes. Nature 1998, 391, 62.
51. T.W. Odom, J. Huang, P. Kim, C.M. Lieber, Structure and electronic properties of carbon nanotubes. J. Phys. Chem. B 2000, 104, 2794–2809.
52. C. Dekker, Carbon nanotubes as molecular quantum wires. Phys. Today 1999, 52 (5), 22.
53. H. Dai, J. Kong, C. Zhou, N. Franklin, T. Tmobler, A. Cassell, S. Fan, M. Chapline, Controlled chemical routes to nanotube architectures, physics and devices. J. Phys. Chem. B 1999, 103, 11,246.
54. S.J. Tans, R.M. Verschueren, C. Dekker, Room temperature transistor based on a single carbon nanotube. Nature 1998, 393, 49.
55. R. Martel, T. Schmidt, H.R. Shea, T. Hertel, P. Avouris, Single- and multi-wall carbon nanotube field effect transistors. Appl. Phys. Lett. 1998, 73, 2447.
56. Z. Yao, H.W.C. Postma, L. Balents, C. Dekker, Carbon nanotube intramolecular junctions. Nature 1999, 402, 273.
57. M.S. Fuhrer, J. Nygrad, L. Shih, M. Forero, Y.G. Yoon, M.S.C. Mazzoni, H.J. Choi, Ihm, J. S.G. Louie, A. Zettl, P.L. McEuen, Crossed nanotube junctions. Science 2000, 288, 494.
58. V. Derycke, R. Martel, J. Appenzeller, P. Avouris, Carbon nanotube inter- and intramolecular logic gates. Nano lett. 2001, 1, 453.
59. A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker, Logic circuits with carbon nanotube transistors. Science 2001, 294, 1317.
60. P.G. Collins, M.S. Arnold, P. Avouris, Engineering carbon nanotubes and nanotube circuits using electrical breakdown. Science 2001, 292, 706.
61. C.M. Lieber, One-dimensional nanostructures: chemistry, physics and applications. Solid State Commun. 1998, 107, 106.
62. J. Hu, T.W. Odom, C.M. Lieber, Chemistry and physics in one dimension: synthesis and properties of nanowires and nanotubes. Acc. Chem. Res. 1999, 32, 435.
63. X. Duan, C.M. Lieber, General synthesis of compound semiconductor nanowires. Adv. Mater. 2001, 12, 298.
64. M.S. Gudiksen, J. Wang, C.M. Lieber, Synthetic control of the diameter and length of single crystal semiconductor nanowires. J. Phys. Chem. B 2001, 105, 4062–4064.
65. Y. Cui, X. Duan, J. Hu, C.M. Lieber, Doping and electrical transport in silicon nanowires. J. Phys. Chem. B 2000, 104, 5213.
66. X. Duan, Y. Huang, Y. Cui, J. Wang, C.M. Lieber, Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices. Nature 2001, 409, 66.
67. Y. Cui, C.M. Lieber, Functional nanoscale electronic devices assembled using silicon nanowire building blocks. Science 2001, 291, 851.
68. Y. Huang, X. Duan, Y. Cui, L. Lauhon, K. Kim, C.M. Lieber, Logic gates and computation from assembled nanowire building blocks. Science 2001, 294, 1313.
69. Y. Huang, X. Duan, Q. Wei, C.M. Lieber, Directed assembly of one-dimensional nanostructures into functional networks. Science 2001, 291, 630.
70. F. Seker, K. Meeker, T.F. Kuech, A.B. Ellis, Surface chemistry of prototypical bulk II–VI and III–V semiconductors and implications for chemical sensing. Chem. Rev. 2000, 100, 2505.
71. R.K. Iler, The Chemistry of Silica; Wiley: New York, 1979.

72. Y. Huang, X. Duan, Q. Wei, C.M. Lieber, Directed assembly of one-dimensional nanostructures into functional networks. Science 2001, 291, 630–633.
73. P.A. Smith, C.D. Nordquist, T.N. Jackson, T.S. Mayer, B.R. Martin, J. Mbindyo, T.E. Mallouk, Electric-field assisted assembly and alignment of metallic nanowires. Appl. Phys. Lett. 2000, 77, 1399–1401.
74. X. Duan, Y. Huang, Y. Cui, J. Wang, C.M. Lieber, Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices. Nature 2001, 409, 66–69.
75. D. Whang, S. Jin, Y. Wu, C.M. Lieber, Largescale hierarchical organization of nanowire arrays for integrated nanosystems. Nano Lett. 2003, 3, 1255–1259.
76. A. Tao, F. Kim, C. Hess, J. Goldberger, R. He, Y. Sun, Y. Xia, P. Yang, Langmuir–Blodgett silver nanowire monolayers for molecular sensing using surface-enhanced Raman spectroscopy. Nano Lett. 2003, 3, 1229–1233.
77. J. Sun, J.A. Rogers. Semiconductor nanowires for application in macroelectronics, in Encyclopedia of nanoscience and nanotechnology, Marcel Dekker, Inc., New York, 2005.
78. Y. Wu, H. Yan, M. Huang, B. Messer, J. H. Song, P. Yang, Inorganic semiconductor nanowires: rational growth, assembly, and novel properties. Chem. Eur. J. 2002, 8, 1261–1268.
79. M.H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, P. Yang, Room-temperature ultraviolet n anowire nanolasers. Science 2001, 292, 1897–1899.
80. S. Bhunia, T. Kawamura, S. Fujikawa, K. Tokushima, Y. Watanabe, Free-standing Semiconductor Nanowires for Applications in Macroelectronics 13 S and vertically aligned InP nanowires grown by metalorganic vapor phase epitaxy. Physica E 2004, 21, 583–587.
81. Z.H. Wu, X.Y. Mei, D. Kim, M. Blumin, H.E. Ruda, Growth of Au-catalyzed ordered GaAs nanowire arrays by molecular-beam epitaxy. Appl. Phys. Lett. 2002, 81, 5177–5179.
82. B. Nikoobakht, C.A. Michael, S.J. Stranick, M.D. Vaudin, Horizontal growth and in situ assembly of oriented zinc oxide nanowires. Appl. Phys. Lett. 2004, 85, 3244–3246.
83. Y. Shan, K. Kalkan, C.-Y. Peng, S.J. Fonash, From Si source gas directly to positioned, electrically contacted Si nanowires: the self-assembling ‘‘grow-in-place’’ approach. Nano Lett. 2004, 4, 2085–2089.
84. D.C. Duffy, J. Cooper McDonald, O.J.A. Schueller, G.M. Whitesides, Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). Anal. Chem. 1998, 70, 4974.
85. C.A. Stover, D.L. Koch, C. Cohen, Observations of fiber orientation in simple shear-flow of semidilute suspensions. J. Fluid Mech. 1992, 238, 277.
86. D.L. Koch, E.S.G. Shaqfeh, The average rotation rate of a fiber in the linear flow of a semidilute suspension. Phys. Fluids A 1990, 2, 2093.
87. J. Liu, M.J. Casavant, M. Cox, D.A. Walters, P. Boul, W. Lu, A.J. Rimberg, K.A. Smith, D.T. Colbert, R.E. Smalley, Controlled deposition of individual single-walled carbon nanotubes on chemically functionalized templates. Chem. Phys. Lett. 1999, 303, 125.
88. M. Burghard, G. Dueberg, G. Philipp, J. Muster, S. Roth, Controlled adsorption of carbon nanotubes on chemically modified electrode arrays. Adv. Mater. 1998, 10, 584.
89. C. De Rosa, C. Park, B. Lotz, J.C. Wittmann, L.J. Fetters, L.E. Thomas, Control of molecular and microdomain orientation in a semicrystalline block copolymer thin film by epitaxy. Macromolecules 2000, 33, 4871.
90. M. Gleiche, L.F. Chi, H. Fuchs, Nanoscopic channel lattices with controlled anisotropic wetting. Nature 2000, 403, 173.
91. Y. Wu, J. Xinag, C. Yang, W. Lu, C.M. Lieber, Single-crystal metallic nanowires and metal= semiconductor nanowires heterostructures. Nature 2004, 430, 61–65.
92. E. Menard, R.G. Nuzzo, J.A. Rogers, Bendable single crystal silicon thin film transistors formed by printing on plastic substrates. Appl. Phys. Lett. 2005, 86 (093507), 1–3.
93. Y.-M. Lin, X. Sun, S. Cronin, Z. Zhang, J. Y. Ying, M. S. Dresselhaus: Fabrication and transport properties of Te-doped bismuth nanowire arrays, Molecular Electronics: MRS Symp. Proc., Boston 1999, 1 Nanowires References 139 ed. by S. T. Pantelides, M. A. Reed, J. Murday, A. Aviran (Materials Research Society Press, Pittsburgh 2000) **582** H10.3 (1–6).

94. C. R. Martin: Nanomaterials: A membrane-based synthetic approach, Science **266** (1994) 1961–1966.
95. M. S. Sander, A. L. Prieto, R. Gronsky, T. Sands, A. M. Stacy: Fabrication of high-density, high aspect ratio, large-area bismuth telluride nanowire arrays by electrodeposition into porous anodic alumina templates, Adv. Mater. **14** (2002) 665–667.
96. M. H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, P. Yang: Room-temperature ultraviolet nanowire nanolasers, Science **292** (2001) 1897–1899.
97. Y. Huang, X. Duan, Y. Cui, C. M. Lieber: Gallium nitride nanowire nanodevices, Nano. Lett. **2** (2002) 101–104.
98. Y. H. Tang, Y. F. Zhang, N. Wang, C. S. Lee, X. D. Han, I. Bello, S. T. Lee: Morphology of Si nanowires synthesized by high-temperature laser ablation, J. Appl. Phys. **85** (1999) 7981–7983.
99. Y. F. Zhang, Y. H. Tang, N. Wang, C. S. Lee, I. Bello, S. T. Lee: Germanium nanowires sheathed with an oxide layer, Phys. Rev. B **61** (2000) 4518–4521.
100. A. M. Morales, C. M. Lieber: A laser ablation method for the synthesis of crystalline semiconductor nanowires, Science **279** (1998) 208–211.
101. Y. Wu, R. Fan, P. Yang: Block-by-block growth of single-crystalline Si/SiGe superlattice nanowires, Nano. Lett. **2** (2002) 83–86.
102. S. B. Cronin, Y.-M. Lin, O. Rabin, M. R. Black, G. Dresselhaus, M. S. Dresselhaus, P. L. Gai: Bismuth nanowires for potential applications in nanoscale electronics technology, Microsc. Microanal. **8** (2002) 58–63.
103. M. S. Sander, R. Gronsky, Y.-M. Lin, M. S. Dresselhaus: Plasmon excitation modes in nanowire arrays, J. Appl. Phys. **89** (2001) 2733–2736.
104. L. J. Lauhon, M. S. Gudiksen, D. Wang, C. M. Lieber: Epitaxial core-shell and core-multishell nanowire heterostructures, Nature **420** (2002) 57–61.
105. L. Venkataraman, C. M. Lieber: Molybdenum selenide molecular wires as one-dimensional conductors, Phys. Rev. Lett. **83** (1999) 5334–5337.
106. Z. Zhang, D. Gekhtman, M. S. Dresselhaus, J. Y. Ying: Processing and characterization of single-crystalline ultrafine bismuth nanowires, Chem. Mater. **11** (1999) 1659–1665.
107. K. Nielsch, R. Wehrspohn, S. F. H. Kronmuller, J. Barthel, J. Kirschner, U. Gosele: Magnetic properties of 100 nm nickel nanowire arrays obtained from ordered porous alumina templates, MRS Symp. Proc. **636** (2001) D19 (1–6).
108. A. Majumdar: Scanning thermal microscopy, Annual Rev. Mater. Sci. **29** (1999) 505–585.
109. D. D. D. Ma, C. S. Lee, Y. Lifshitz, S. T. Lee: Periodic array of intramolecular junctions of silicon nanowires, Appl. Phys. Lett. **81** (2002) 3233–3235.
110. Y.-M. Lin, S. B. Cronin, J. Y. Ying, M. S. Dresselhaus, J. P. Heremans: Transport properties of Bi nanowire arrays, Appl. Phys. Lett. **76** (2000) 3944–3946.
111. 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, G. A. C. Jones: One-dimensional transport and the quantization of the ballistic resistance, J. Phys. C: Solid State Phys. **21** (1988) L209–L214.
112. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouvenhoven, D. van der Marel, C. T. Foxon: Quantized conductance of point contacts in a two-dimensional electron gas, Phys. Rev. Lett. **60** (1988) 848–850.
113. C. J. Muller, J. M. van Ruitenbeek, L. J. de Jongh: Conductance and supercurrent discontinuities in atomic-scale metallic constrictions of variable width, Phys. Rev. Lett. **69** (1992) 140–143.
114. C. J. Muller, J. M. Krans, T. N. Todorov, M. A. Reed: Quantization effects in the conductance of metallic contacts at room temperature, Phys. Rev. B **53** (1996) 1022–1025.
115. J. L. Costa-Krдmer, N. Garcia, H. Olin: Conductance quantization in bismuth nanowires at 4 K, Phys. Rev. Lett. **78** (1997) 4990–4993.
116. J. L. Costa-Krдmer, N. Garcia, H. Olin: Conductance quantization histograms of gold nanowires at 4 K, Phys. Rev. B **55** (1997) 12910–12913.

117. C. Z. Li, H. X. He, A. Bogozi, J. S. Bunch, N. J. Tao: Molecular detection based on conductance quantization of nanowires, Appl. Phys. Lett. **76** (2000) 1333–1335.
118. H. Masuda, M. Satoh: Fabrication of gold nanodot array using anodic porous alumina as an evaporation mask, Jpn. J. Appl. Phys. **35** (1996) L126–L129.
119. Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K.-H. Kim, C. Lieber: Logic gates and computation from assembled nanowire building blocks, Science **294** (2001) 1313–1317.
120 J.-R. Kim, H. Oh, H.M. So, J.-J. Kim, J. Kim, C. J. Lee, S. C. Lyu: Schottky diodes based on a single GaN nanowire, Nanotechnology **13** (2002) 701–704.
121. X. Duan, Y. Huang, C. M. Lieber: Nonvolatile memory and programmable logic from molecule-gated nanowires, Nano. Lett. **2** (2002) 487–490.
122. E. C. Walter, R. M. Penner, H. Liu, K. H. Ng, M. P. Zach, F. Favier: Sensors from electrodeposited metal nanowires, Surf. Inter. Anal. **34** (2002) 409–412.
123. E. C. Walter, K. H. Ng, M. P. Zach, R. M. Penner, F. Favier: Electronic devices from electrodeposited metal nanowires, Microelectron. Eng. **61-62** (2002) 555–561.
124. Y.-M. Lin, X. Sun, M. S. Dresselhaus: Theoretical investigation of thermoelectric transport properties of cylindrical Bi nanowires, Phys. Rev. B **62** (2000) 4610–4623.
125. J. Heremans, C. M. Thrush, Y.-M. Lin, S. Cronin, Z. Zhang, M. S. Dresselhaus, J. F. Mansfield: Bismuth nanowire arrays: synthesis and galvanomagnetic properties, Phys. Rev. B **61** (2000) 2921–2930.
126. K. Hong, F. Y. Yang, K. Liu, D. H. Reich, P. C. Searson, C. L. Chien, F. F. Balakirev, G. S. Boebinger: Giant positive magnetoresistance of Bi nanowire arrays in high magnetic fields, J. Appl. Phys. **85** (1999) 6184–6186.
127. K. Liu, C. L. Chien, P. C. Searson: Finite-size effects in bismuth nanowires, Phys. Rev. B **58** (1998) R14681–R14684.
128. Z. Zhang, X. Sun, M. S. Dresselhaus, J. Y. Ying, J. Heremans: Magnetotransport investigations of ultrafine single-crystalline bismuth nanowire arrays, Appl. Phys. Lett. **73** (1998) 1589–1591.
129. J. Heremans, C. M. Thrush, Z. Zhang, X. Sun, M. S. Dresselhaus, J. Y. Ying, D. T. Morelli: Magnetoresistance of bismuth nanowire arrays: A possible transition from one-dimensional to three dimensional localization, Phys. Rev. B **58** (1998) R10091–R10095.
130. L. Sun, P. C. Searson, C. L. Chien: Finite-size effects in nickel nanowire arrays, Phys. Rev. B **61** (2000) R6463–R6466.
131. Y.-M. Lin, O. Rabin, S. B. Cronin, J. Y. Ying, M. S. Dresselhaus: Semimetal-semiconductor transition in Bi1−*x*Sb*x* alloy nanowires and their thermoelectric properties, Appl. Phys. Lett. **81** (2002) 2403–2405.
132. Y.-M. Lin, S. B. Cronin, O. Rabin, J. Y. Ying, M. S. Dresselhaus: Transport properties and observation of semimetal-semiconductor transition in bi-based nanowires, Nanocrystalline Semiconductor Materials and Devices: MRS Symp. Proc., Boston 2002, ed. by J.M. Buriak, D. D. M. Wayner, L. Tsyeskov, F. Priolo, B. E. White, Boston 2003) **737** F3.14.
133. Y.-M. Lin, M. S. Dresselhaus: Transport properties of superlattice nanowires and their potential for thermoelectric applications, Nanocrystalline Semiconductor Materials and Devices: MRS Symp. Proc., Boston 2002, ed. by J.M. Buriak, D. D. M. Wayner, L. Tsyeskov, F. Priolo, B. E. White (Materials Research Society Press, Boston 2003) **737** F8.18.
134. J. Heremans, C. M. Thrush, Y.-M. Lin, S. B. Cronin, M. S. Dresselhaus: Transport properties of antimony nanowires, Phys. Rev. B **63** (2001) 085406(1–8).
135. Y.-M. Lin, S. B. Cronin, O. Rabin, J. Y. Ying, M. S. Dresselhaus: Transport properties of Bi1−*x*Sb*x* alloy nanowires synthesized by pressure injection, Appl. Phys. Lett. **79** (2001) 677–679.
136. D. E. Beutler, N. Giordano: Localization and electron–electron interaction effects in thin Bi wires and films, Phys. Rev. B **38** (1988) 8–19.
137. J. P. Heremans, C. M. Thrush, D. T. Morelli, M.-C. Wu: Thermoelectric power of bismuth nanocomposites, Phys. Rev. Lett. **88** (2002) 216801(1–4).

138. Z. Zhang, X. Sun, M. S. Dresselhaus, J. Y. Ying, J. Heremans: Electronic transport properties of single crystal bismuth nanowire arrays, Phys. Rev. B **61** (2000) 4850–4861.
139. K. Liu, C. L. Chien, P. C. Searson, Y. Z. Kui: Structural and magneto-transport properties of electrodeposited bismuth nanowires, Appl. Phys. Lett. **73** (1998) 1436–1438.
140. J. Heremans, C. M. Thrush: Thermoelectric power of bismuth nanowires, Phys. Rev. B **59** (1999) 12579–12583.
141. L. D. Hicks, M. S. Dresselhaus: Thermoelectric figure of merit of a one-dimensional conductor, Phys. Rev. B **47** (1993) 16631–16634.
142. Y.-M. Lin, S. B. Cronin, O. Rabin, J. Heremans, M. S. Dresselhaus, J. Y. Ying: Transport properties of Bi-related nanowire systems, Anisotropic Nanoparticles: Synthesis, Characterization and Applications: MRS Symp. Proc., Boston 2000, ed. by S. Stranick, P. C. Searson, L. A. Lyon, C. Keating (Materials Research Society Press, Pittsburgh 2001) **635** C4301–C4306.
143. Y.-M. Lin, M. S. Dresselhaus: Thermoelectric properties of superlattice nanowires, Phys. Rev. B **68** (2003) 075304 (1–14).
144. M. S. Gudiksen, L. J. Lauhon, J. Wang, D. C. Smith, C. M. Lieber: Growth of nanowire superlattice structures for nanoscale photonics and electronics, Nature **415** (2002) 617–620.
145. M. T. Bjцrk, B. J. Ohlsson, T. Sass, A. I. Persson, C. Thelander, M. H. Magnusson, K. Deppert, L. R. Wallenberg, L. Samuelson: One-dimensional steeplechase for electrons realized, Nano. Lett. **2** (2002) 87–89.
146. M. T. Bjцrk, B. J. Ohlsson, C. Thelander, A. I. Persson, K. Deppert, L. R. Wallenberg, L. Samuelson: Nanowire resonant tunneling diodes, Appl. Phys. Lett. **81** (2002) 4458–4460.
147. D. Li: Thermal transport in individual nanowires and nanotubes. Ph.D. Thesis (Univ. California, Berkeley 2002).
148. T. S. Tighe, J. M. Worlock, M. L. Roukes: Direct thermal conductance measurements on suspended monocrystalline nanostructures, Appl. Phys. Lett. **70** (1997) 2687–2689.
149. S. T. Huxtable, A. R. Abramson, C.-L. Tien, A. Majumdar, C. LaBounty, X. Fan, G. Zeng, J. E. Bowers, A. Shakouri, E. T. Croke: Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices, Appl. Phys. Lett. **80** (2002) 1737–1739.
150. R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn: Thin-film thermoelectric devices with high room-temperature figures of merit, Nature **413** (2001) 597–602.
151. C. Dames, G. Chen: Modeling the thermal conductivity of a SiGe segmented nanowire, 21st Int. Conf. Thermoelectrics: Proc. ICT Symposium, Long Beach 2002 (IEEE, Piscataway 2002) 317–320.
152. G. Chen, M. S. Dresselhaus, G. Dresselhaus, J.-P. Fluerial, T. Caillat: Recent developments in thermoelectric materials, Int. Mater. Rev. **48** (2003) 45–66.
153. K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes: Measurement of the quantum of thermal conductance, Nature **404** (2000) 974–977.
154. K. Schwab, J. L. Arlett, J. M. Worlock, M. L. Roukes: Thermal conductance through discrete quantum channels, Physica E **9** (2001) 60–68.
155. M. Cardona: *Light scattering in solids* (Springer, Berlin, Heidelberg 1982).
156. P. Y. Yu, M. Cardona: *Fundamentals of Semiconductors* (Springer, Berlin, Heidelberg 1995) Chap. 7.
157. J. C. M. Garnett: Colours in metal glasses, in metallic films, and in metallic solutions, Philos. Trans. Roy. Soc. London A **205** (1906) 237–288.
158. D. E. Aspnes: Optical properties of thin films, Thin Solid Films **89** (1982) 249–262.
159. U. Kreibig, L. Genzel: Optical absorption of small metallic particles, Surf. Sci. **156** (1985) 678–700.
160. M. R. Black, Y.-M. Lin, S. B. Cronin, O. Rabin, M. S. Dresselhaus: Infrared absorption in bismuth nanowires resulting from quantum confinement, Phys. Rev. B **65** (2002) 195417(1–9).
161. M. W. Lee, H. Z. Twu, C.-C. Chen, C.-H. Chen: Optical characterization of wurtzite gallium nitride nanowires, Appl. Phys. Lett. **79** (2001) 3693–3695.

162. M. J. Zheng, L. D. Zhang, G. H. Li, W. Z. Chen: Fabrication and optical properties of large-scale uniform zinc oxide nanowire arrays by one-step electrochemical deposition technique, Chem. Phys. Lett. **363** (2002) 123–128.
163. D. M. Lyons, K. M. Ryan, M. A. Morris, J. D. Holmes: Tailoring the optical properties of silicon nanowire arrays through strain, Nano. Lett. **2** (2002) 811–816.
164. J. C. Johnson, H. Yan, R. D. Schaller, L. H. Haber, R. J. Saykally, P. Yang: Single nanowire lasers, J. Phys. Chem. B **105** (2001) 11387–11390.
165. S. Blom, L. Y. Gorelik, M. Jonson, R. I. Shekhter, A. G. Scherbakov, E. N. Bogachek, U. Landman: Magneto-optics of electronic transport in nanowires, Phys. Rev. B **58** (1998) 16305–16314.
166. A. Sugawara, T. Coyle, G. G. Hembree, M. R. Scheinfein: Self-organized Fe nanowire arrays prepared by shadow deposition on NaCl(110) templates, Appl. Phys. Lett. **70** (1997) 1043–1045.
167. J. P. Pierce, E. W. Plummer, J. Shen: Ferromagnetism in cobalt-iron alloy nanowire arrays on W(110), Appl. Phys. Lett. **81** (2002) 1890–1892.
168. J. C. Johnson, H. Yan, R. D. Schaller, P. B. Petersen, P. Yang, R. J. Saykally: Near-field imaging of nonlinear optical mixing in single zinc oxide nanowires, Nano. Lett. **2** (2002) 279–283.
169. M. R. Black, Y.-M. Lin, S. B. Cronin, M. S. Dresselhaus: Using optical measurements to improve electronic models of bismuth nanowires, 21st Int. Conf. Thermoelectrics: Proc. ICT Symposium, Long Beach 2002 (IEEE, Piscataway 2002) 253–256.
170. H. Richter, Z. P. Wang, L. Ley: The one phonon Raman-spectrum in microcrystalline silicon, Solid State Commun. **39** (1981) 625–629.
171. I. H. Campbell, P. M. Fauchet: The effects of microcrystal size and shape on the one phonon
Raman-spectra of crystalline semiconductors, Solid State Commun. **58** (1986) 739–741.
172. H.-L. Liu, C.-C. Chen, C.-T. Chia, C.-C. Yeh, C.-H. Chen, M.-Y. Yu, S. Keller, S. P. DenBaars: Infrared and Raman-scattering studies in single-crystalline GaN nanowires, Chem. Phys. Letts. **345** (2001) 245–251.
173. R. Gupta, Q. Xiong, C. K. Adu, U. J. Kim, P. C. Eklund and: Laser-induced Fanoresonance scattering in silicon nanowires, Nano. Lett. **3** (2003) 627–631.
174. Y. Huang, X. Duan, C.M. Lieber, Semiconductor nanowires: nanoscale electronics and optoelectronics, in Encyclopedia of nanoscience and nanotechnology, Marcel Dekker, Inc., New York, 2005.
175. O. Madelung, Semiconductors: intrinsic properties of group IV elements and III–V and II–VI and I–VII compounds. In Landolt–Bornstein New Series; Madelung, O., Ed.; Springer: Berlin, Heidelberg, 1987; Vol. III=22a.
176. S.M. Sze, Semiconductor Devices, Physics and Technology; Wiley: New York, 1985.
177. Y. Huang, X. Duan, C.M. Lieber, Nanowires for multi-color photonics. Small 2005, 1, 142.
178. S. Datta, Electronics Transport in Mesoscopic Systems; Cambridge University Press: Cambridge, 1995.
179. F. Capasso, S. Datta, Quantum electron devices. Phys. Today 1990, 43 (2), 74.
180. A. Steane, Quantum computing. Rep. Prog. Phys. 1998, 61, 117.
181. W. Liang, M. Bockrath, D. Bozovic, H.H. Hafner, M. Tinkham, H. Park, Fabry–Perot interference in a nanotube electron waveguide.Nature 2001, 411, 665.
182. M. Buttiker, Y. Imry, R. Landauer, Generalized many-channel conductance formula with application to small rings. Phys. Rev. B 1985, 31, 6207.
183. K. Tsukagoshi, B.W. Alphenaar, H. Ago, Coherent transport of electron spin in a ferromagnetically contacted carbon nanotube. Nature 1999, 401, 572.

Ch16

[1]. J.D.Joannopoulos, Pierre R.Villeneuve and Shanhui Fan, Photonic crystals^putting a new twist on light, Nature, 386 (1997) 143-149

[2]. D. R. Smith, J. B. Pendry, M. C. K. Wiltshire, Metamaterials and Negative Refractive Index, Science, 305 (2004), 788-792

Ch18

1.Porous Silicon Membranes, Leigh Canham Ed. L.Canham, Springer, Switzerland, 2014, 163-170

2. MACE Silicon Nanostructures, Ciro Chiappini, Ed. L.Canham, Springer, Switzerland, 2014, 172-186

3.Porous Silicon Formation by Metal Nanoparticle-Assisted Etching, Claude Le´vy-Cle´ment Ed. L.Canham, Springer, Switzerland, 2014, 49-59

4.Polymer - Porous Silicon Composites, Ester Segal and Maksym A. Krepker Ed. L.Canham, Springer, Switzerland, 2014, 188-198

5.Imprinting Porous Silicon, Judson D. Ryckman and Sharon M. Weiss,Contents Ed. L.Canham, Springer, Switzerland, 2014, 552-560

6. Porous Silicon Application Survey, Leigh Canham Ed. L.Canham Springer, Switzerland, 2014, 734-740

7. Porous Silicon Micromachining Technology, Giuseppe Barillaro Ed. L.Canham, Springer, Switzerland, 2014, 779-786

8. Porous Silicon Functionalities for BioMEMS, Julien Schweicher and Tejal A. Desai, , Ed. L.Canham, Springer Switzerland, 2014, 788-796

9. Silicon for Microdevices, and Microsystems, Luca De Stefano and Ilaria Rea Ed. L.Canham, Springer Switzerland, 2014, 797-804

10. Porous Silicon Optical Waveguides, Sharon M. Weiss and Xing Wei. L.Canham, Springer Switzerland, 2014, 816-822

11. Michael J. Sailor, Porous Silicon in Practice,Preparation, Characterization and Applications, 2012 Wiley-VCH Verlag & Co. KGaA, Boschstr. 12, 69469 Weinheim, Germany

12.Porous silicon: a quantum sponge structure, for silicon based optoelectronics
O. Bisi, Stefano Ossicini, L. Pavesi Surface Science Reports 38 (2000) 1-126

13. Valeriy A. Skryshevsky, Tetiana Serdiuk, Yuriy E. Zakharko, Sergei A.Alekseev, Alain Géloën,Vladimir Lysenko. *Preparation, Luminescent Properties and Bioimaging Application of Quantum Dots Based on Si and SiC, In: Functional Nanomaterials and Devices for Electronics, Sensors and Energy Harvesting*. Ed. Alexei Nazarov, Francis Balestra, Valeriya Kilchytska, Denis Flandre, Springer, Switzerland, 2014, pp.323-348.

14.Valeriy A.Skryshevsky. *Porous Si Structures for Gas, Vapor and Liquid Sensing. In: Comprehensive Guide for Mesoporous Materials, Volume 3: Properties and Development* Ed. Mahmood Aliofkhazraei, Nova Science Publishers Inc.,US, 2015, pp. 123-146.

Ch 19

1. Brinker CJ, Scherer GW (1990) Sol-Gel Science. The Physics and Chemistry of Sol-Gel Processing. Academic Press, New-York

2. Pierre AC (1998) Introduction to Sol-Gel Processing Kluwer, Boston

3. Soleimani DA, Abbasi MH (2008) Silica aerogel; synthesis, properties and characterization. J Mater Proc Technol 199:10–26

4. Pierre AC, Pajonk GM (2002) Aerogels and their applications. Chem Rev 102:4243–4265

5. Ambekar AP, Bagade P (2006) A review on: ’aerogel - world’s lightest solid’. Popular Plastics & Packaging, 51:96–102

6. Kocon L, Phalippou J (2005) Aerogels. Material aspect Techniques de l’Ingenieur, Sciences Fondamentales,AF196:AF3610/1-AF3610/21
7. Carraher CE Jr (2005) General topics: silica aerogels-properties and uses. Polymer News, 30(12), 386–388
8. Carraher CE Jr (2005) Silica aerogels - synthesis and history. Polymer News, 30:62–64
9. Pajonk GM (2003) Some applications of silica aerogels. Colloid and Polymer Science, 281:637–651
10. Akimov YK (2003) Fields of Application of Aerogels (Review) Instruments and Experimental Techniques (Translation of Pribory i Tekhnika Eksperimenta), 46:287–299
11. Venkateswara Rao A, Parvathy Rao A, Kulkarni MM (2004) Influence of gel aging and Na2SiO3/H2O molar ratio on monolithicity and physical properties of water–glass-based aerogels dried at atmospheric pressure.J Non-Cryst Solids 350:224–229
12. Hwang S-W, Jung H-H, Hyun S-H, Ahn Y-S (2007) Effective preparation of crack-free silica aerogels via ambient drying. J Sol Gel Sci Technol 41, 139–146
13. Kistler SS (1932) Coherent expanded aerogels. J Phys Chem 36:52–64
14. Lee CJ, Kim GS, Hyun SH (2002) Synthesis of silica aerogels from waterglass via new modified ambient drying. J Mater Sci 37:2237–2241
15. Tang Q, Wang T (2005) Preparation of silica aerogel from rice hull ash by supercritical carbon dioxide drying. J Supercrit Fluids 35:91–94
16. Li T, Wang T (2008) Preparation of silica aerogel from rice hull ash by drying at atmospheric pressure. Materials Chemistry and Physics 112:398–401
17. Nakanishi K, Minakuchi H, Soga N, Tanaka N (1998) Structure Design of Double-Pore Silica and Its Application to HPLC. J Sol Gel Sci Technol 13:163–169
18. Wagh PB, Begag R, Pajonk GM, Venkasteswara Rao A., Haranath D (1999) Comparison of some physical properties of silica aerogel monoliths synthesized by different precursors. Mater Chem Phys 57:214–218
19. EbelmenM(1846) Recherches sur les combinaisons des acides borique et silicique avec les e´thers. Ann Chim Phys 16:129–166; (1847) Sur l’hyalite artificielle et l’hydrphane. C R Acad Sci Paris 25:854–856
20. Einarsrud MA, Nilsen E, Rigacci A, Pajonk GM, Buathier S, Valette D, Durant M, Chevalier P, Nitz P, Ehrburger-Dolle F (2001) Strengthening of silica gels and aerogels by washing and aging processes J. Non-Cryst. Solids 285:1–7
21. Deng Z, Wang J, Wei J, Shen J, Zhou B, Chen L (2000) Physical Properties of Silica Aerogels Prepared with Polyethoxydisiloxane. J Sol Gel Sci Technol 19:677–680
22. Venkastewara Rao A, Bhagat SD, Hirashima H, Pajonk GM (2006) Synthesis of flexible silica aerogels using methyltrimethoxysilane (MTMS) precursor. J Colloid Interface Sci 300:279–285
23. El Rassy H, Buisson P, Bouali B, Perrard A, Pierre AC (2003) Surface Characterization of Silica aerogels with Different Proportions of Hydrophobic Groups, dried by the CO2 Supercritical Method; Langmuir, 19:358–363
24. Harreld JH, Ebina T, Tsubo N, Stucky G (2002) Manipulation of pore size distributions in silica and ormosilgels dried under ambient pressure conditions. J Non-Cryst Solids 298:241–251
25. Brinker CJ, Keefer KD, Schaefer DW, Ashley CS (1982) Sol-gel transition in simple silicates. J Non-Cryst Solids 48:47–64
26. Tillotson TM, Hrubesh LW (1992) Transparent ultralow-density silica aerogels prepared by a two-step sol-gel process J Non-Cryst Solids 145:44–50
27. Tsou P (1995) Silica aerogel captures cosmic dust intact. J Non-Cryst Solids 186:415–427
28. De la Rosa-Fox N, Esquivias L, Craievich AF, Zarzycki J (1990) Structural study of silica sonogels. J Non- Cryst Solids 121: 211–15
29. Zarzycki J (1994) Sonogels. Heterog Chem Rev 1:243–253
30. Matson DW, Smith RD (1989) Supercritical fluid technologies for ceramic-processing applications. J AmCeram Soc 72:871–881

31. Pajonk GM (1994) A short history of the preparation of aerogels and carbogels. In: Attia YJ (ed) Sol-Gel Processing and Applications, Plenum Press, New-York, 201–209
32. Livage J, Henry M, Sanchez C (1988) Sol-Gel Chemistry of Transition Metal Oxides. Prog Solid State Chem 18:259–341
33. Land VD, Harris TM, Teeters DC (2001) Processing of low-density silica gel by critical point drying or ambient pressure drying. J Non-Cryst Solids 283:11–17
34. Zarzycki J, Prassas M, Phalippou J (1982) Synthesis of glasses from gels: the problem of monolithic gels. J Mater Sci 17:3371–3379
35. Hench, LL (1986) Use of dying control chemical additives (DCCAs). In controlling sol-gel processing. In: Hench LL, Ulrich DR (eds) Science of Ceramic Chemical Processing, Wiley, New-York, 52–64
36. Hæreid S, Einarsrud MA, Scherer GW (1994) Mechanical strengthening of TMOS-based alcogels by aging in silane solutions. J Sol-Gel Sci Tech 3:199–204
37. Schwertfeger F, Glaubitt W, Schubert U (1992) Hydrophobic aerogels from tetramethoxysilane/methyltrimethoxysilane mixtures J Non-Cryst Solids 145:85–89
38 Smith DM, Stein D, Anderson JM, Ackermann W (1995) Preparation of low-density xerogels at ambient pressure. J Non-Cryst Solids 186:104–112
39. Kocklenberg R, Mathieu B, Blacher S, Pirard R, Pirard JP, Sobry R, VandenBossche G (1998) Texture control
of freeze-dried resorcinol-formaldehyde gels. J Non-Cryst Solids 225:8–13
40. Tretyakov YD, Shlyakhtin OA (1999) Recent progress in cryochemical synthesis of oxide materials. J Mater Chem 9:19–24
41. Pajonk GM, Repellin-Lacroix M, Abouarnadasse S, Chaouki J, Klvana D (1990) From sol-gel to aerogels and cryogels. J. Non-Cryst. Solids. 121: 66–67
42. Fricke J (ed) (1986) Aerogels – Proceedings of the First International Symposium, Wurzburg, FRG, Sept. 23–25, 1985, Springer-Verlag, Berlin
43. Vacher R, Phalippou J, Pelous J, Woignier T (eds) (1989) Proceedings of the Second International Symposium on Aerogels (ISA2), Rev. Phys. Appl. Colloq, 24–C4
44. Fricke J (ed) (1992) Proceedings of the Third International Symposiumon Aerogels (ISA 3), JNon-Cryst Solids 145 45. Pekala RW, Hrubesh LW (eds) (1995) Proceedings of the Fourth International Symposium on Aerogels (ISA 4), J Non-Cryst Solids 186
46. Phalippou J, Vacher R (eds) (1998) Proceedings of the Fifth International Symposium on Aerogels (ISA 5), J Non-Cryst Solids 225
47. Ashley CS, Brinker CJ, SmithDM(eds) (2001) Aerogels 6. Proceedings of the Sixth International Symposium on Aerogels (ISA6), Albuquerque, NM, USA; 8–11 October 2000. J Non-Cryst Solids 285
48. Fricke J, Emmerling A (1992) Aerogels – preparation, properties, applications. Struct Bonding (Berlin) 77:37–87
49. Fricke J, Emmerling A (1998) Aerogels – Recent progress in production techniques and novel applications. J Sol-Gel Sci Technol 13:299–303
50. Burger T, Fricke J (1998) Aerogels: Production, modification and applications. Berichte der Bunsen Gesellschaft Phys Chemi Chem Phys 102:1523–1528
51. Hrubesh LW (1998) Aerogel applications. J Non-Cryst Solids 225:335–342
52. Schmidt M, Schwertfeger F (1998) Applications for silica aerogel products. J Non-Cryst Solids 225:364–368
53. Husing N, Schubert U (1998) Aerogels – Airy materials: Chemistry, structure, and properties. Angew Chem Int Ed 37:23–45
54. Caps R, Doell G, Fricke J, Heinemann E, Hetfleisch J (1989) Thermal transport in monolithic silica aeroge. In Vacher R, Phalippou J, Pelous J, Woignier T (eds) Proceedings of the Second International Symposium on Aerogels (ISA2), Rev. Phys. Appl. Colloq, 24-C4:113–118

55. Gronauer M, Kadur A, Fricke J (1986) Mechanical and acoustic properties of silica aeroge. In ISA1 Springer Proc Phys 6:167–173
56. Pajonk GM (2003) Some applications of silica aerogels. Colloid Polym Sci 38:4407–4413
57. Venkateswara Rao A, Pajonk GM (2001) Effect of methyltrimethoxysilane as a co-precursor on the optical properties of silica aerogels. J Non-Cryst Solids 285:202–209
58. Tscheuschner D, Ratke L (1999) Crystallization of InSb in aerogel crucibles. Cryst Res Technol 34:167–174
59. Kim KK, Jang KY (1991) Hollow silica spheres of controlled size and porosity by sol-gel processing. J Am Ceram Soc 74:1987–1992
60. Gerlach R, Kraus O, Fricke J, Eccardt PC, Kroemer N, Magori V (1992) Modified silica aerogels as acoustic impedance matching layers in ultrasonic devices. J Non-Cryst.Solids 145:227–232
61. Forest L, Gibiat V, Woignier T (1998) Biot’s theory of acoustic propagation in porous media applied to aerogels and alcogels. J Non-Cryst Solids 225:287–292
62. Woignier T, Phalippou J (1989) Scaling law variation of the mechanical properties of silica aerogels. In Vacher R, Phalippou J, Pelous J, Woignier T (eds) Proceedings of the Second International Symposium on Aerogels (ISA2), Rev. Phys. Appl. Colloq, 24-C4:179–184
63. LeMay JD, Tillotson TM, Hrubesh LW, Pekala RW (1990) Microstructural dependence of aerogel mechanical properties Mat Res Soc Sym Proc 180:321–324
64. Holmes NC, Radousky HB, Moss MJ, Nellis WJ, Henning S (1984) Silica at ultrahigh temperature and expanded volume. Appl Phys Lett 45:626–628
65. Guise MT, Hosticka B, Earp BC, Norris PM (1995) An experimental investigation of aerosol collection utilizing packed beds of silica aerogel microspheres. J Non-Cryst Solids 285:317–322
66.Advances in Sol-Gel Derived Materials and Technologies, Aerogels Handbook, ed.Michel A. Aegerter, Springer New York, 2011
67.. A.Yu.Karlash, G.V.Kuznetsov, Yu.S.Milovanov, V.A.Skryshevsky. Impedance spectroscopy of composites based on porous silicon and silica aerogel for sensor applications, *Functional Materials*, 2013, **20**, 68-74.
68. A.Yu.Karlash, V.A.Skryshevsky, G.V.Kuznetsov, V.P.Kladko, Evolution of visible photoluminescence of Si quantum dots embedded in silicon oxide matrix, *J.Alloys and Compounds*, 2013, **577,** 283-287.
69.A.Yu.Karlash, Yu.E.Zakharko, V.A.Skryshevsky, A.I.Tsiganova, G.V.Kuznetsov. Photoluminescence properties of silica aerogel/porous silicon nanocomposites. *J.Phys. D: Appl.Phys*, 2010, **43**, art.335405, 2010.